ABSTRACT

DEAN, TIMOTHY RICHARD. Characterization of *In Vitro* Selected RNA Aptamers for the RB69 RegA Translational Repressor Protein and Optimization of *in vitro* Selection Utilizing Complex Targets. (Under the direction of Eric S. Miller.)

The purpose of the research has been to characterize the nucleic acid sequence requirements necessary for recognition by the translational repressor protein RegA. RegA, from bacteriophage RB69 and T4 related bacteriophages, is a translational repressor protein that is highly conserved and capable of repressing translation of numerous transcripts. The mRNA binding sites for RegA are devoid of structure and are AU rich. *in vitro* selection was carried out for 5 rounds generating specific sequences capable of binding to RB69 RegA. Characterization of those sequences included sequencing, dissociation constant (K_d*app*) determination, and use of the selected sequences to search the RB69 genome for potential binding sites.

The second aspect of the research focused on optimization of conditions for completing successful *in vitro* selection experiments against complex biological targets. Bacteriophage R17 is an F^+ coliphage infecting *E. coli* found in the intestinal tracts of swine. Development of a rapid diagnostic technique capable of identifying R17 would enable for detection of faeces in ground water. Generation of RNA aptamers against R17 was undertaken with the goal of optimizing selection conditions against a viable viral particle. Two selections were carried out one utilizing nitrocellulose to partition bound RNA from unbound RNA and the second using R17 immobilized on a agarose bead matrix. The results initiated the generation of specific considerations for initiating a selection experiment against a protein target.

Characterization of *In Vitro* Selected RNA Aptamers for the RB69 RegA Translational Repressor Protein and Optimization of *in vitro* Selection Utilizing Complex Targets

by

Timothy R. Dean

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Microbiology

Raleigh

2002

Approved By:

_____	_____
_____	_____
	Chair of Advisory Committee

DEDICATION

To my wife Cindy and my two lovely children Kaylie and Teun. Without your support I would never have been able to complete my degree. I thank you from the bottom of my heart and hope you all know that this work is as much yours as it is mine. Your loving husband and father,

Tim

PERSONAL BIOGRAPHY

TIMOTHY RICHARD DEAN

I was born October 21, 1968 in a suburb of Washington D.C. to Kerry R. Dean and Rebecca A. Dean. I am the oldest child of two; my younger sister Holly following me three years later. In 1986 I graduated from Downingtown High School. I attended West Chester University for 2 years when I put my education on hold to pursue a career as a bicycle racer in Europe. It was in Europe that I lived with 1998 summer Olympic silver medallist Leo Peelen and his wonderful family. Ties have remained close and Leo and his wife Madeleine are my daughter Kaylie's godparents. Following a succession of bad crashes I hung up cycling and returned to West Chester University, where in 1991 I received a Bachelor of Science in Economics. Following graduation I worked as a project accountant. Fortunes enabled me to change careers and become an operations manager at a printing facility in LaGrange, GA. I then decided to leave industry and pursue my advanced degree in Microbiology. I was accepted into the Department of Microbiology to receive the Doctor of Philosophy degree. I officially became a Doctoral candidate upon passing my preliminary examination in August 2001. Following the successful completion of my degree I will be taking a position at the United States Environmental Protection Agency.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my family for all their love and support. I wish to express my sincere gratitude to my parents for their financial support as well as for their words of encouragement. I would like to thank Dr. Miller for allowing me the opportunity to study and work in his laboratory, and to my committee members Dr. Dennis Brown, Dr. Mike Hyman, and Dr. James Brown for all their support and suggestions. My committee has enabled me to reach the position I am at today. I would like to thank all the staff and students in the Microbiology department. It has been with everyone's input that my work has progressed. In particular I want to thank Sherrice Allen who showed me the ropes and got me off to a great start in my research. Tom Hall, with all his helpful discussions and calm demeanor kept me on the strait and narrow. Andy Andrews for numerous suggestions and for his ability to listen to my problems when I'm sure he would rather have been elsewhere. A special thanks goes out to the best technician that I will ever work with, Bridget Szczypinski. Her friendship and support will never be forgotten. Thanks to Frances Trouth for being such a wonderful labmate and to Jong Lee for being a lab lightening rod and a good friend. Thanks to Dawn, Annette, and Yvonne for keeping me in line, and last but not least, my thanks goes out to Jimmy Cooper for keeping the department and our lab in tip top shape. It is with everyone's support that I am where I am today!! Thank you all.

Tim

TABLE OF CONTENTS

Page

LIST OF TABLES ..... **viii**

LIST OF FIGURES ..... **ix**

IN VITRO SELECTION, FROM BIOLOGY TO BIOTECHNOLOGY **1**

I. INTRODUCTION **2**

II. HISTORY **3**

III. *IN VITRO* SELECTION - THE PROCESS **5**

A. Nucleic acid library **6**

B. Incubation with the target **8**

Binding considerations 9

C. Partitioning bound nucleic acid from unbound **10**

D. RT-PCR of binding molecules **12**

E. Periodic assay for binding during the course of SELEX **13**

F. Post SELEX activities **14**

IV. INCREASING NUCLEIC ACID STABILITY IN *IN VITRO* SELECTION **15**

Incorporating modified nucleotides during RNA SELEX **16**

V. SELEX: EXAMPLE EXPERIMENTS **19**

Small molecule targets and the RNA World **19**

Protein targets **22**

***in vitro* selection of catalytic nucleotides 24**

VI. COMPARISON TO ANTIBODY TECHNOLOGY 26

VII. CONCLUSION: THE FUTURE OF SELEX. 28

REFERENCES:..... 29

CHARACTERIZATION OF *IN VITRO* SELECTED RB69 REGA RNA BINDING SITES.. 38

I. BACTERIOPHAGE T4 TRANSLATIONAL REPRESSORS 39

T4 translational repressor Gp32 39

T4 translational repressor Gp43 40

T4 translational repressor RegA 40

II. USING SELEX TO EXPLORE REGA-RNA INTERACTIONS..... 42

T4 RegA:RNA analysis using SELEX 42

RB69 RegA:RNA analysis using SELEX 45

REFERENCES:..... 46

***IN VITRO* SELECTED PHAGE RB69 REGA RNA BINDING SITES YIELDS UAA TRIPLETS AND A POTENTIAL ROLE IN TRANSLATIONAL REINITIATION 48**

ABSTRACT 49

INTRODUCTION..... 50

MATERIALS AND METHODS..... 53

RESULTS 59

DISCUSSION..... 67

CONCLUSIONS 72

REFERENCES..... 74

FIGURE LEGENDS 88

OPTIMIZATION OF SELEX USING BEAD IMMOBILIZED BACTERIOPHAGE R17 ... 90

INTRODUCTION..... 91

MATERIALS AND METHODS..... 95

RESULTS 104

DISCUSSION.....116

CONCLUSIONS 118

REFERENCES.....120

LIST OF TABLES

Table 1.1 Successful DNA SELEX experiments.....15

Table 1.2 Selections carried out using modified ribonucleotides.....16

Table 1.3 Nucleic acid aptamers for small molecules.....21

Table 1.4 Protein targets utilized in SELEX experiments.....24

Table 2.1 RegA sensitive mRNAs.....41

Table 2.2 Hierarchy of biological RNA affinities for T4 and RB69 Reg A proteins....42

Table 2.3 Binding affinity of selected RNA sequences for T4 RegA.....44

Table 2.4 Biological T4 mRNAs that are regulated by RegA.....44

Table 2.5 SELEX conditions using immobilized RB69 RegA-His_6.....80

Table 2.6 Cloned sequences after five rounds of SELEX.....81

Table 2.7 K_d*app* values for RB69 RegA on defined RNAs.....82

Table 2.8 Potential RB69 RegA binding sites in RB69 genome.....83

Table 3.1 Optimized reaction conditions.....99

Table 3.2 Selection of RNAs that bind to Phage R17 using N40 Nitrocellulose Partitioning.....107

Table 3.3 Selection of RNAs that bind to Phage R17 using bead partitioning.....114

LIST OF FIGURES

Figure 1.1 Typical selection scheme for RNA SELEX.....5

Figure 1.2 Template: Primer arrangement with subsequent RNA transcript.....7

Figure 2.1 Sequence of SELEX N14 DNA pools.....84

Figure 2.2 Gel shift analysis of RB69 RegA binding to selected and variant RNA 85

Figure 2.3 K_d*app* analysis of RB69 RegA-RNA interactions.....86

Figure 2.4 SELEX derived sequence logo.....87

Figure 3.1 Alignment and comparison of R17 and Qβ coat proteins.....94

Figure 3.2 N_{40} template design.....96

Figure 3.3 Biotinylated phage bound to nitrocellulose membranes.....101

Figure 3.5 SDS-PAGE of purified phage.....104

Figure 3.6 PCR and *in vitro* transcription optimization.....106

Figure 3.7 Modified western blot.....111

Figure 3.8 SDS-PAGE of R17 coat protein on the streptavidin matrix.....112

Chapter 1

Literature Review

In vitro Selection, From Biology to Biotechnology

I. Introduction

In vitro selection, *in vitro* evolution, or SELEX (Systematic Evolution of Ligands by Exponential enrichment) is a combinatorial technique wherein random pools of nucleic acids (RNA or DNA) are used as the starting materials to isolate rare functional molecules (1, 2, 3). Originally, *in vitro* selection was used to generate functional molecules that could bind to large or small ligands. Over the years numerous improvements to the procedure have greatly increased its potential while at the same time increasing the potential functionality of the nucleic acid aptamers. Functionality is determined by the conditions of the selection and include, but are not limited to, protein binding, small molecule binding, enzyme inhibition, and catalysis (4,5,6,7). *In vitro* selection straddles the biology/biotechnology interface being used to examine the interactions between proteins and their nucleic acid targets, as well as provide reagents for environmental, therapeutic, and diagnostic applications (8,9,10). The results of *in vitro* selection experimentation have caused a rethinking of the capabilities of nucleic acids, broadening their already diverse functionality. Selection results have also given plausibility to the "RNA World" hypothesis, where RNA stored information within its code while also carrying-out the reactions essential for "life" at the time.

Over the last dozen years numerous improvements have been made increasing the potential of SELEX. Improvements to the procedure include using partially randomized pools as the starting point when trying to elucidate the sequence or structure present in naturally occurring nucleic acids (11). Other improvements include the incorporation of

modified nucleotides, which can greatly increase the stability and half-life of potential therapeutic aptamers (12,13). Additional modification, by appending polyethylene glycol to an aptamer, have greatly increased the time it takes a therapeutic nucleic acid to be cleared from the human body (14). To fully understand the current research being performed using *in vitro* selection a review of its history is useful.

II. History

It is generally assumed that SELEX was initiated around 1990. However, there was quite a bit of groundwork laid out before the modern methods of selection came into being. When it was learned that nucleic acids could carry-out both informational and functional roles it became evident to many researchers that evolution could be carried out at the molecular level. The first to experiment with these characteristics was Sol Spiegelman who worked with bacteriophage Qβ (15). Due to the lack of pure enzymes and amplification techniques such as the polymerase chain reaction Spiegelman showed that Darwinian selection could operate in a cell free system (Reviewed in 16). The bacteriophage Qβ genome is replicated by the Qβ replicase protein. Using serial dilutions numerous generations of replication could be carried out quickly. The phenotype under the selective pressure was the replication speed itself. Due to the high inherent error rate of the replicase the Qβ genome deleted segments that were not necessary for recognition by the replicase. Thus replication speed increased as the genome

size decreased. Even though these experiments were carried-out years ago they formed the framework for much of the current work using *in vitro* selection.

With the development of automated methods for chemically synthesizing highly diverse nucleic acid pools, and amplification methods such as the polymerase chain reaction came the ability to create increasingly complex libraries of nucleic acids. These developments also allowed for selection experiments to be carried out in the absence of bacteriophage Qβ. These advances as well as the previous work by Spiegelman show that SELEX was not an idea generated by a single mind at a specific time, but instead a growing body of knowledge being put to use in numerous laboratories.

Loeb and Struhl were some of the first pioneers to take advantage of these new advancements as they probed for oligonucleotide function. They selected functional promoter sequences both *in vitro* and *in vivo* (17,18). The use of random sequences continued to grow with many researchers identifying binding sites for regulatory proteins (19,20).

The ability to use *in vitro* selection to probe for oligonucleotide function was becoming more obvious. Continued advancements moved beyond the realm of probing for regulatory binding sites to other small molecules and a variety of other targets. In 1990 two seminal papers were published simultaneously which brought selections to their current status. Ellington and Szostak published work whereby they generated RNA aptamers to a variety of small organic dyes (1). At the same time Tuerk and Gold were completing their first work with SELEX. Their work generated RNA aptamers to bacteriophage T4 DNA polymerase

(3). Both of these papers showed that nucleic acid aptamers could be developed to bind to numerous targets including basically any conceivable molecule. Since the early days of *in vitro* selection, experimentation has moved to include applications that have medical diagnostic, environmental diagnostic, and therapeutic capabilities.

III. *in vitro* Selection - The Process

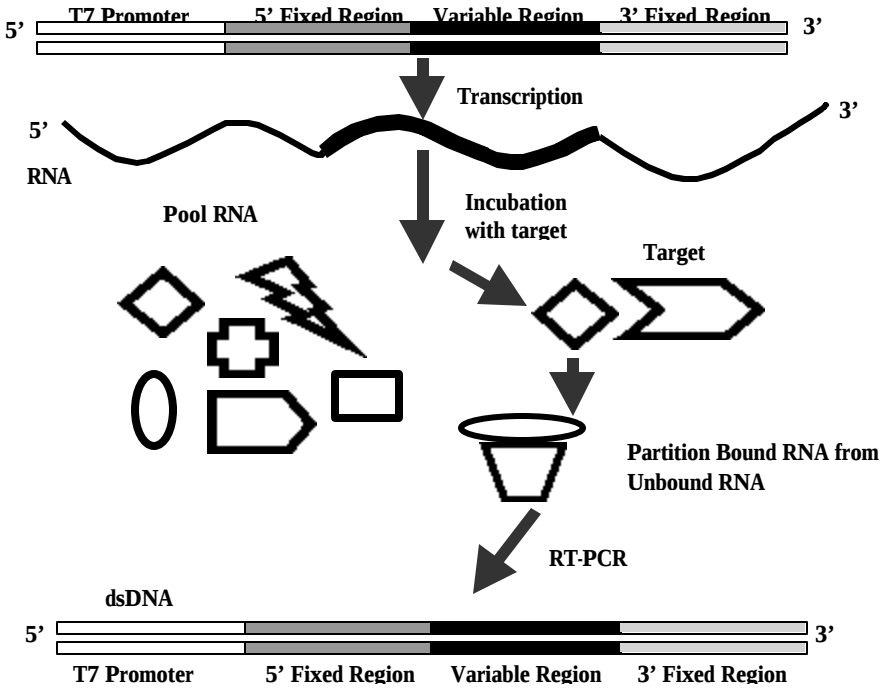


Figure 1.1 Typical selection scheme for RNA SELEX

A complete review of the methods of *in vitro* selection will allow for the comparison to other combinatorial methods, highlighting the capabilities that are possible with the procedure. The following description will outline a scheme geared toward generating nucleic acids that bind a target molecule (protein). However the same considerations would apply when performing a selection with the goal of generating a catalytic nucleic acid. The ligands that are generated using *in vitro* selection are typically called aptamers. This word is derived from the Greek word *aptus*, which means, "to fit" and describes early selection results with a lock and key mode of interaction between the nucleic acid and the target (21).

A. Nucleic acid library

Typically a selection experiment is initiated by generating a nucleic acid pool of high sequence and structural diversity. This is accomplished by synthetically synthesizing a DNA oligonucleotide using a DNA synthesizer (22). The single stranded DNA contains a randomized region flanked 5' and 3' by constant regions. The fixed regions contain known sequences for primer binding, enzymatic digestion, and possibly a T7 promoter region for transcription into RNA (see Figures 1 & 2). There are numerous considerations, however the success of a selection depends on the characteristics of the random sequence (23).

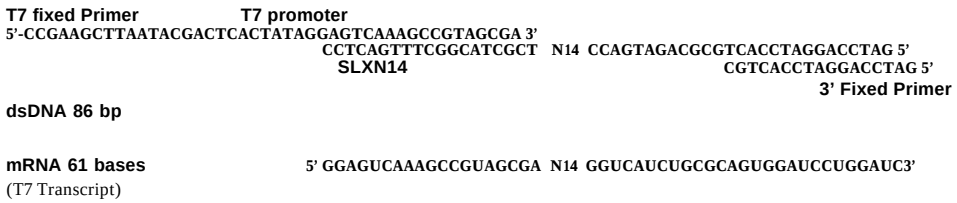


Figure 1.2 Template: Primer arrangement with subsequent RNA transcript.

The synthetic oligonucleotide is converted into double stranded DNA using the Polymerase Chain Reaction (PCR) with 5' T7 fixed and 3' fixed primers (24). At this point the strands of DNA can be separated and single stranded DNA (ssDNA) can be used in the selection experiment or the DNA can be transcribed into RNA which can then be used in the selection experiment. The use of DNA vs. RNA is a case of fitting the right nucleic acid with the desired terminal function. Due to the 2' OH group present in RNA it is inherently less stable than ssDNA of the same basic sequence (25). Therefore, in applications, which may require a stable aptamer it may be preferable use ssDNA.

Some selection experiments wish to better define a known binding site while some wish to identify a novel binding site. Each of the aforementioned experiments requires a random region with different characteristics. When doing an experiment to better identify a novel binding site it is appropriate for each base position of the random region to be synthesized with an equal probability of an A, C, G, or T. These types of selections can be done with different length randomized regions. A random region containing 40 nucleotides has the potential to have 4^{40} or $\sim 10^{24}$ possible sequences. Selections have been carried out

using as few as 8 random nucleotides to as many as 220 nucleotides (26). Typically when doing a binding experiment, randomized regions are kept below 70 nucleotides (23).

When doing a selection to better define a known binding site, the degree of randomization can be adjusted. For example, there may only be 10% "doping" at each base position using the known binding site as the original sequence. In this case if the wild type base is a C it would be present in the pool 70% of the time, while A, G, and T would all be at the same position 10% of the time. An excellent example of this type of experiment using a partially randomized pool was a selection to define the important residues of the Rev-response element (RRE) of HIV-I Rev (27). In these experiments the known binding site is the basis for the SELEX generated binding site.

B. Incubation with the target

Due to the unique chemical characteristics of each base the folding properties of the different members in the nucleic acid pool are different, allowing them to adopt different secondary structures (28). Each of these secondary structures makes up part of a “shape” library, with each molecule having different characteristics. The different properties of each molecule allow the pool to be placed under a specific pressure, usually a unique or novel function (binding to a target). Prior to incubation with the target the nucleic acid is heated sufficiently (approximately 70° C) to allow for it to be denatured completely. Following denaturation the pool is allowed to cool to room temperature over several minutes. The purpose of denaturing and renaturation is to allow the maximum amount of the nucleic acid

pool to adopt their proper “native” conformation. Once the pool of nucleic acids is prepared it is then incubated with the target protein. Incubation time should be long enough for the reaction to reach a binding equilibrium, usually between 30-60 minutes.

Binding considerations

Important considerations for a successful binding reaction include the content of the binding buffer and the temperature of the reaction. Many binding buffers have been used successfully in SELEX experiments (29,30). A major consideration when selecting a binding buffer should be the concentration of divalent metal cations, specifically $MgCl_2$. Divalent cations allow for the stable formation of RNA or ssDNA secondary structure by shielding the negative charges on the phosphate backbone of nucleic acids (31). The more stable the nucleic acid secondary structure the greater the percentage of nucleic acid molecules within the pool that will be in the correct conformation. The temperature of the binding reaction is also an important consideration. Many successful selections have been carried-out with binding taking place at room temperature (23° C) (32). Temperatures that are too high contain too much energy and fail to allow for stable complex formation, while temperatures that are too low fail to have sufficient energy to allow for binding to occur at all. Both of the above considerations should be taken into account when setting up a binding reaction in a SELEX protocol.

C. Partitioning bound nucleic acid from unbound

The next step in a SELEX protocol involves separating the nucleic acid bound to the target from all the non-binding species present in the reaction. This step is referred to as "partitioning" and can be done numerous ways. However, the goal is always to separate and collect the nucleic acid molecules that perform the desired function. Numerous methods of partitioning have been used and include separation on nitrocellulose membranes (33), separation via use of a solid support matrix (34), and separation based on differences in electrophoretic gel mobility (35). All three of the above methods have been used successfully.

One successful method of protein target binding is to allow the nucleic acid pool to incubate with the protein of interest in solution, followed by separation of the binding species from non-binders by filtration through a nitrocellulose membrane (36). Nitrocellulose binds proteins very well, while having a relatively low affinity for RNA (36). Thus, any RNA that is bound to the protein target will also be immobilized on the membrane while those that do not bind can be washed away. This low affinity also allows for the detection of target binding species as the selection continues. However, if one considers the surface area of a 13 mm nitrocellulose disk it is much greater than the potential binding area of the target protein. Couple the surface area with the ~ 1% of a RNA pool which will bind to the nitrocellulose membrane, and unwanted background binding will quickly overtake the selection experiment, making it impossible to pull-out protein binding species. The use of negative selection allows for the removal of nonspecific binding nucleic acid molecules.

Negative selection entails passing the pool of nucleic acids through the membrane, while collecting the flow through. The end result is that any RNA or DNA that does not bind to the nitrocellulose membrane will be caught in the flow-through and subsequently incubated with the protein target. Individual nucleic acid species that have an affinity for nitrocellulose will be removed from the pool allowing the selection to continue with little or no background binding. It is usually sufficient to perform negative selections when background binding reaches levels where it is difficult to discern a protein-dependent binding pattern.

Following binding of the target:nucleic acid to the membrane the bound nucleic acid must be removed from the protein. This elution step is carried out using a strong denaturant. 7 M urea and phenol/chloroform extraction have all been used successfully to separate bound nucleic acid molecules from the target (37,38).

In the above case the nucleic acid pool is incubated with the target molecule and binding allowed to occur in solution. However, this is not the only method of partitioning bound from unbound nucleic acid. The target may be immobilized on some solid matrix such as nickel agarose or streptavidin agarose (38). The pool of nucleic acids is then run over the column matrix, subsequently washed, leaving only the binding species attached to the immobilized target protein. In order to remove the binding species from the target the column can be washed with a highly denaturing solution such as a phenol:chloroform mixture or an elution buffer containing a high concentration of urea (~7 M). A second successful method of removing the bound nucleic acid molecules is to run a very high concentration of “free”

target molecule over the column transferring the binding species to the free target (39). The main consideration using this method is that the concentration of free target must be significantly higher than the concentration of the immobilized target. One other method of partitioning involves separation of nucleic acids that bind to the target from molecules that do not bind via gel shift mobility.

The use of gel shift mobility has proven to be a successful method of separating bound nucleic acids from unbound molecules (40). In this case the target is incubated with the pool of RNA or ssDNA in solution. The binding reaction is then run on a non-denaturing polyacrylamide gel. Separation is based upon the mobility difference between "free" unbound nucleic acid vs. bound molecules. Bound molecules will be much greater in size and their mobility will be greatly retarded with respect to non-binding species. This difference in mobility allows for the extraction and purification of binding molecules over non-binding molecules. Partitioning and separating the aptamers from the target allows for continued enzymatic manipulation.

D. RT-PCR of binding molecules

The binding nucleic acid molecules are then subjected to Reverse Transcription - Polymerase Chain Reaction (RT-PCR) in the case of RNA (41) or simply PCR in the case of ssDNA (24). In RNA SELEX, the purified binding molecules are subjected to reverse transcription generating a single stranded DNA pool. This pool of ssDNA is then transferred to a PCR reaction whereby the pool is converted into double stranded DNA. If the

experiment is a DNA SELEX, the purified bound DNAs are simply subjected to PCR. Both cases result in the generation of a dsDNA molecule which is used as the input for the next round of selection. These final enzymatic steps can be used to monitor the progress of a SELEX experiment.

The use of the final enzymatic steps (RT-PCR) in each round of selection to monitor the progress of the experiment can prove very beneficial. By tracking the number of cycles of PCR needed to generate sufficient double stranded product from round to round the significance of binding molecules from round to round can be analyzed. Analysis should focus on the difference in the number of cycles of PCR needed to generate dsDNA in the actual round of selection versus the number of cycles needed to amplify dsDNA using a pool of nucleic acids that is not selected with target. As the selection progresses it should take fewer cycles of PCR to amplify dsDNA than the number of cycles needed to amplify dsDNA in the "negative" control. Hence, as a greater percentage of the pool is enriched for binding species the number of PCR cycles will decrease.

E. Periodic assay for binding during the course of SELEX

It is important to periodically monitor the progress of a SELEX experiment by quantitatively assaying for binding (42). Filter binding is a highly successful method for determining the extent of binding in a pool of nucleic acid molecules (37). The initial pool of random sequence RNA or ssDNA is assayed prior to selection to set a baseline for improvement in binding. Following several rounds of selection the pool is assayed again. The

nucleic acid is kept constant while the concentration of target protein is increased in each assay. As the pool of nucleic acids is enriched for binding species there should be a dose-dependant curve generated in the binding assay (43). This dose- dependant curve is evidence that binding species are being selected. Rounds of selection are carried-out iterativly until the nucleic acid pool shows no improvement in binding over successive rounds. An excellent example using the binding assay to monitor the progress of a selection can be seen in a selection using the Ricin A-chain as the nucleic acid target (4). The importance of monitoring the progress of the selection is clearly seen by round 7 where ~ 20% of the input pool bound to the target. However, there was 7.8% background binding. In the following 2 rounds there was heavy negative selection incorporated into the procedure and at round 9 target binding was increased to 51% while background binding was reduced to ~2.5% (4). This example clearly demonstrates the usefulness of monitoring the selection by assaying for target binding, while at the same time using the assay results to incorporate negative selection into the scheme to reduce potential non-specific binding nucleic acid molecules.

F. Post SELEX activities

When the pool of nucleic acids has been enriched for binding species, individual molecules are cloned (44), sequenced (45), and characterized (46). The final PCR product DNA is typically cloned into a convenient plasmid, and sequences of 15-30 independent clones defined. Characterization involves aligning representative sequences and placing them into groups with similar sequence content. Secondary structure predictions can also be made

based on the sequence data. Using lowest energy of formation algorithims, programs such as mfold (47) are capable of predicting secondary structure. In addition, individual molecules can also be assayed for their ability to bind the target and this affinity is usually expressed as a dissociation constant (K_d). Individual K_d values can range from μM to low nM values depending on the target and the experiment.

IV. Increasing nucleic acid stability in *in vitro* selection

Nucleic acid selections are highly successful inside the laboratory, however, this success quickly coorelates into failure if the efficacy of the selected molecules is poor due to quick degradation, in real world applications. This degradation is more acute when using RNA over ssDNA because of the ribonucleases present in cells and blood serum. The half life of many RNAs are in the range of minutes or less (48). This quick degradation does not allow for successful target identification and binding. One method of circumventing the relative instability of RNA is to use ssDNA. The half life of many ssDNA molecules is increased to 1-2 hours (48). Conducting selection with ssDNA may allow for affective therapeutic or diagnostic use (Table 1).

Table 1.1 Successful DNA SELEX experiments.

Target	Pool	K_d (nM) of selected aptamer	Ref.
thrombin	ssDNA	25	49
IgE	ssDNA	10	50
Lrp	dsDNA	2	51
Hiv-1 RT	ssDNA	2	52

A highly successful method of increasing the stability of RNA aptamers is the use of modified nucleotides. Increased stability of ribonucleotides can be accomplished by the chemical modification at specific positions (Table 2). Modifications to ribonucleotides are at the 2' position of the ribonucleotide or at the phosphodiester backbone because modifications at these positions greatly interfere with nucleases that degrade RNA (48). The two main strategies that have been developed for using modified nucleotides are use of modified nucleotides during the selection, incorporating them in each round (48). The second method involves completing the SELEX experiment using unmodified nucleotides and then substituting modified nucleotides at the conclusion of the experiment (48). Each of these strategies will be reviewed.

Table 1.2 Selections carried out using modified ribonucleotides.

Target	**K_d (nM)**	**Ref.**
thrombin	400	53
HIV-1 Rev	1	55
elastase	15	56
bFGF	0.3	57

Incorporating modified nucleotides during RNA SELEX

The use of modified nucleotides during the selection is sometimes called presubstitution (48). This method of substitution relies on the ability of DNA and RNA

polymerases to incorporate the modified nucleotide into the synthesized RNA or DNA molecule (53). There have now been many experiments that have used incorporation of modified nucleotide incorporation by polymerases. The first successful experiment to use presubstitution was a selection using thrombin as a target. The anti-thrombin aptamer contained 5-(1-pentynyl)-2'-deoxyuridine in the molecules in the DNA pool (53). Interestingly, the modified thrombin aptamers showed decreased affinity for thrombin when compared to a selection carried out using natural unmodified nucleotides. Numerous modified nucleotides have been used successfully in selections. These include 2'-aminouridine, and 2'-aminocytidine (54). These two modified nucleotides have been used because they inhibit ribonuclease activity, while at the same time are still recognized by T7 RNA polymerase and AMV reverse transcriptase, allowing for their incorporation into successive rounds of selection.

Aptamers that bind to bFGF and VerF were selected using presubstitution with modified pyrimidines (57). Using these modified bases allowed for a successful selection with molecules having K_ds in the picomolar range. In addition to having such a strong affinity for bFGF, these aptamers where shown to have around a 1000-fold increased stability in serum (59). In yet another SELEX experiment, human neutrophil elastase was the target and modified nucleotides (pyrimidines) were used throughout the selection (54). The selected aptamers were reliant on the modified nucleotides and their half-lives were increased from minutes to hours in serum. These examples clearly show the success of using presubstitution

during the process of SELEX. Equally as successful has been the incorporation of modified nucleotides into aptamers at the post-SELEX stage.

Post-SELEX incorporation of modified ribonucleotides

Addition of modified nucleotides into selected aptamers has been shown to increase both affinity for the target as well as increase the stability of the aptamer. A key example of such a success was seen in a selection using Rous sarcoma virus (RSV) and a random pool of unmodified RNA molecules (60). Following selection, binding aptamers were completely substituted with 2'-fluoropyrimidines. With complete substitution there was a total loss of affinity by the aptamers for RSV. Three more rounds of selection were carried out using modified nucleotides, resulting in modified aptamers that could bind RSV and interfere with viral replication as well as the unmodified pool. The result of the final three rounds of selection was to remove all unsubstituted pyrimidines that were not critical for binding, while increasing the stability of the aptamer.

Additional SELEX experimentation using VegF aptamers and post modification proved to be successful as well. The selected anti-VegF aptamers contained pre-substituted 2'-aminopyrimidines and following selection were further substituted with 2'-O-methylpurines (58). When all purines were substituted there was a large decrease in the binding affinity of the aptamers, however, limited substitution led to an increase in affinity and a drastic increase in half-life in urine of 131 hours.

These examples clearly show that using modified nucleotides incorporated into the selection experiment, or incorporated following the generation of binding aptamers, leads not only to more stable molecules, but also molecules with increased binding affinity. Future advancements such as interference analysis (61) should lead to the generation of aptamers that are highly nuclease resistant as well as maintaining high binding affinity.

V. SELEX: Example Experiments

Small molecule targets and the RNA World

SELEX has been used extensively to develop aptamers to many small molecules. Aptamers directed at small molecules have helped delve into the "RNA World" hypothesis (62). In the RNA World RNA, is proposed to have stored the genetic code, as well as performed the role of chemical catalyst by functioning as an enzyme. For this to happen, RNA must have been able to recognize small molecules and substrates with high affinity and in a highly specific manner. *In vitro* selection is an excellent experimental procedure capable of showing the capabilities RNA has for recognizing small organic molecules.

The first experiments with the “modern” *in vitro* selection procedures utilized small organic dyes as the targets for RNA aptamers (1). These molecules are polycyclic, aromatic planar molecules, they have an overall negative charge, and numerous potential hydrogen bond donors and acceptors. Large random pools of RNA containing a random sequence

length of 100 nucleotides were used to select for molecules that could bind to the dyes. These experiments were very successful, and specific aptamers were derived for each of the six organic dyes used. The surprising result of these experiments was that the chemical makeup of the four bases was sufficient to enable stable interactions between the RNA and the organic dyes. Other small molecules such as the carbohydrate cellobiose (34) have generated successful aptamers, clearly showing that RNA has the capability to recognize and bind to molecules that have little diversity in chemical makeup.

In addition to cellobiose, other biological molecules have been used as targets for SELEX experiments. The first to be used was ATP (39). ATP is directly involved in the of energy of living systems (63). This selection was carried out by using ATP immobilized on a column matrix. RNA was passed over the column and binding species eluted by passing an excess of "free" ATP over the column. Following the selection, analysis of the cloned sequences showed that while highly divergent, all contained an 11 nucleotide stretch that folded into the same secondary structure. This structure was deduced to be a hairpin loop with an internal, asymmetric purine rich bulge. Once again SELEX was used to show that the four bases present in RNA were capable of recognizing small organic molecules, relatively simple chemical structures in biologically relevant molecules (Table 1.3).

RNA is not the only molecule that can recognize small organic molecules. Single stranded DNA is also capable of recognizing numerous small, simple targets including arginine (64), and porphyrins (65,66). In most cases aptamers that are developed as ssDNA

do not function when they are converted into RNA, which is not surprising. The 2'-OH group on RNA not only makes it more nuclease sensitive but also gives it added chemical makeup causing it to fold into secondary structures differing from ssDNA which lacks the 2'-OH.

Table 1.3 Nucleic acid aptamers for small molecules.

Target	Aptamer	K_d (mM)	Ref.
ATP	DNA	6	39
Cyanocobalamin	RNA	0.09	67
Valine	RNA	12,000	68
NAD	RNA	2.5	69
Tryptophan	RNA	18	70
Guanosine	RNA	32	71
Arginine	RNA	0.33	72
Dopamine	RNA	2.8	73
Streptomycin	RNA	1	75

What is clearly evident from the above examples is that RNA does indeed recognize simple chemical molecules. It is capable of doing this with only the four nucleotides folding into the proper secondary structure. This evidence lends credence to the RNA World hypothesis when RNA not only carried the genetic code but also acted as a catalyst in chemical reactions (61).

Protein targets

In addition to small organic targets the next logical approach for SELEX is the generation of aptamers against proteins. It would make sense that RNA or ssDNA would be able to form tight interactions with protein targets, which possess twenty amino acids with differing reactive groups. Proteins fold into complex three-dimensional structures enabling nucleic acid aptamers to bind in clefts or other surface structures. Numerous selections have been carried out using proteins that bind nucleic acids as their normal biological activity, as well as selections against proteins that do not function as nucleic acid binding proteins.

The first protein SELEX experiment was for bacteriophage T4 DNA polymerase (Gp43) (3). The binding of Gp43 to a hairpin structure located in the translation initiation region of its own mRNA allows for autogenous regulation. In this experiment the SELEX procedure was utilized to elucidate what is responsible for the loop nucleotide bias in this protein:RNA interaction. The RNA pool consisted of a random region of 8 nucleotides. Following 5 rounds of selection on nitrocellulose the pool of RNA exhibited the same affinity for Gp43 as the wild type binding sequence (K_d approximately 5 nM) (3). The end results of the experiment showed that there were two different loop sequences with very similar binding constants. In addition, the results showed that there were very few loop nucleotides that could tolerate substitution without loss of binding affinity.

A second SELEX experiment was carried out to identify RNA ligands to the bacteriophage R17 coat protein (74). This experiment was undertaken to determine if the

SELEX protocol could be used to predict a high affinity RNA binding site. The R17 coat protein is well characterized and naturally binds an RNA hairpin repressing expression of the replicase gene whose translation initiation region is contained in the hairpin (74). An RNA pool with a random region 32 nucleotides long was subjected to 11 rounds of selection. 47 individual sequences were cloned and sequenced, 38 were unique while 36 contained the hairpin structure necessary for high affinity to the R17 coat protein. These results showed that the SELEX procedure can accurately predict an RNA binding site for the R17 coat protein, and that this success could be expanded to other targets in the same class.

The first nucleic acid selection to utilize a protein target that does not biologically bind to RNA or DNA was the generation of a ssDNA aptamer against thrombin (49). The SELEX experiment was initiated with a starting pool consisting of a random region 60 nucleotides in length. The experiment resulted in numerous aptamers with nanomolar affinity for thrombin. This experiment was also important because the consensus sequence and structure was capable of inhibiting thrombin and blood clotting ability (76). Subsequent analysis showed the structure to be a G-quartet (77). There have been many reports of such experiments using non-nucleic acid binding proteins as targets with great success. It appears that many proteins are amenable to the SELEX procedure, which can be used to examine biological questions to potential therapeutic applications (Table 1.4).

Table 1.4 Protein targets utilized in SELEX experiments.

Target	K_d nM)	Ref.
Pepocin	20	78
Ricin A-chain	7.3	4
Integration host factor	0.8	79
T4 RegA	20	40

in vitro selection of catalytic nucleotides

Discussion thus far has focused on the ability of RNA or ssDNA to recognize and bind small organic molecules and proteins, both with high specificity and high affinity. The main goal of many SELEX experiments is to force polynucleotides to carry out some unique function. In addition to binding to targets, SELEX experiments have also been carried-out with the goal of finding RNA or ssDNA that can display enzyme-like functions. This section will deal with some of the experiments and advances that have been made on the road to polynucleotide catalysis.

The first successful SELEX protocol that isolated catalytic ribozymes was an experiment utilizing a pool of RNA containing a random region of 220 ribonucleotides (26). The selection was for catalytic formation of a phosphoester link to another RNA in a reaction that is very similar to the activity of RNA polymerase. The selected ribozymes form structures that bring the 5' terminal tri-phosphate into proximity to the 3' end of the target RNA

molecule. The experiment required 10 rounds to isolate ribozymes capable of catalyzing the reaction at a rate of 0.06 min^{-1}

Researchers also quickly realized that RNAs of known function could be used as the starting point in SELEX experiments designed to create molecules capable of completing different tasks (80). The first successful attempt at re-designing a known functional RNA was undertaken to create an RNA capable of catalyzing self phosphorylation (81). An initial pool of ~ 10^{15} different molecules that were already biased for binding to adenosine 5'-triphosphate was subjected to 13 rounds of selection for RNAs that acquire a thiophosphate group when incubated with γ-thioATP. The end result was a pool consisting of RNAs that were capable of self-phosphorylating at a rate of ~0.05 min^{-1}.

In yet other studies protein catalysis has been used as the initial starting point. Transition state theory has been used extensively in the development of protein catalysts (82). Many researchers felt that the same theory could be incorporated into the development of nucleic acid catalysts (83,84). The first experiments to use transition state theory with nucleic acid catalysts were carried-out by selecting for RNAs that could bind to the transition state analog of the interconversion between two diasteriomeric biphenyl compounds (85). The pool of RNA contained a random sequence 128 nucleotides long. Following 7 rounds of selection there were numerous individual RNAs that could bind to the transition state analog. However, only one of the selected molecules was capable of catalyzing the noncovalent

isomerization of the biphenyl substrate. The observed K_{cat} was 2.8 X 10^{-5} min^{-1} and the observed K_M was 542 μM.

The above examples are by no means exhaustive; catalytic RNA and ssDNA have been developed to perform numerous chemical reactions including RNA alkylation (86), transacylation and amide bond formation (87), porphyrin metalation (88), and DNA ligation (89). The results of these experiments show the diverse functions that RNA and ssDNA can possess, lending credence to their potential diagnostic and therapeutic abilities as well as strengthening the RNA World hypothesis.

VI. Comparison to antibody technology

Antibodies have been used for decades and are the most widely used class of molecules providing molecular recognition. The use of antibodies in detecting molecules predates the 1950's (90). During the 1970's, monoclonal antibody technology became available allowing for the production of large amounts of specific antibodies to meet the increasing demands in diagnostics (91). The use of antibodies has continued to grow over the years, however, there are certain limitations to antibodies (92), which bear mention.

Possibly the most obvious aspect of antibody preparation is that it takes place with animals. This may become a problem if the animal does not tolerate the antibody target, as could be the case with many toxins. In addition, the prevailing sentiment is against use of animals in science, making the use of antibody generation in animals an important ethical

issue. A second major issue with antibody generation is the potential time and cost expenditure. There may be a considerable time and money expenditure searching for a rare antibody, screening it, and purifying it for laboratory use. The performance of multiple antibody preparations is frequently different from preparation to preparation, most likely due to the differing immune systems of different animals. Finally, antibody preparations are subject to irreversible denaturation due to temperature sensitivity and have a definite “shelf” life (92). Many of these issues can be overcome by the use of oligonucleotides generated by the SELEX procedure.

Aptamer technology is completely *in vitro*, obviating the use of animals in their generation. Because there is no use of animals in aptamer generation, there is no concern for toxicity, or molecules that do not generate a strong immune response. In addition, the conditions of the experiment can be changed when needed, allowing for subtle changes in stringency, binding or catalytic activity. Aptamers for environmental diagnostics can be developed in buffers and at temperatures that are not physiological. Nucleic acid aptamers are sensitive to temperature as well, however, their denaturation is reversible and they can be refolded into the proper secondary structure. Finally, aptamers, once developed to a target, can be produced by chemical synthesis. This enables extreme precision in their production from batch to batch.

VII. Conclusion: The future of SELEX.

Significant advancements have been made over the last decade in the generation of aptamers, and in the variety of aptamers that have been developed. Their use in molecular recognition has made their use in therapeutics and diagnostics commonplace. Only eight years after the inception of SELEX, aptamers were already in clinical trials (92). Other areas that are beginning to use nucleic acid aptamers are *in vivo* diagnostics, and *in vivo* imaging (93). Due to the highly structured conformations (favorable binding energy for Watson-Crick base pairs) (94) RNA and ssDNA molecules should continue to grow in popularity and usage. The future of the technology is wide-open due to the simplicity of the enzymatic manipulation and the proven chemical functionalities of RNA and ssDNA aptamers.

REFERENCES:

1. Ellington, A.D. and Szostak, J.W. (1990) *Nature* 346: 818-822.

2. Joyce, G.F. (1989) *Gene* 82: 83-87.

3. Tuerk, C. and Gold, L. (1990) *Science* 249: 505-510.

4. Hesselberth, J.R., Miller, D., Robertus, J., and Ellington, A.D. (2000) *J Biol Chem* 275: 4937-4942.

5. Huizenga, D.E. and Szostak, J.W. (1995) *Biochemistry* 34: 656-665.

6. Rusconi, C.P., Yeh, A., Lyerly, H.K., Lawson, J.H. and Sullenger, B.A. (2000) *Thromb Haemost* 84: 841-848.

7. Joyce, G.F. (2001) *Methods Enzymol* 341: 503-517.

8. Bruno, J.G. and Kiel, J.L. (1999) *Biosens Bioelectron* 14: 457-464.

9. Brody, E.N. and Gold, L. (2000) *J Biotechnol* 74: 5-13.

10. Davis, K.A., Abrams, B., Lin, Y. and Jayasena, S.D. (1996) *Nucleic Acids Res* 24: 702-706.

11. Giver, L., Bartel, D.P., Zapp, M.L., Green, M.R., Ellington, A.D. (1993) *Gene* 137: 19-24.

12. Verma, S., and Eckstein, F. (1998) *Annual Rev. Biochem.* 67: 99-134.

13. Andreola, M., Calmels, C., Michel, J., Toulme, J., and Litvak, S. (2000) *Eur. J. Biochem.* 267: 5032-5040.

14. White, R., Sullenger, B., and Rusconi, C. (2000) *Journal of Clinical Inv.* 106: 929-934.

15. Mills, D., Peterson, R., and Spiegelman, S. (1967) *Proc. Natl. Acad. Sci.* 58: 217.

16. Spiegelman, S. (1971) *Q. Rev. Biophys.* 4: 213-53.

17. Oliphant, A., and Struhl, K. (1987) *Methods Enzymologyl.* 155: 558-568.

18. Horwitz, M., and Loeb, L. (1986) *Proc. Natl. Acad. Sci.* 83: 7405-7409.

19. Thiesen, H., and Bach, C. (1990) *Nucleic Acids Res.* 18: 3203-3209.

20. Sun, X., and Baltimore, D. (1991) *Cell* 64: 459-470.

21. Klug, S., and Famulok, M. (1994) *Mol. Bio. Rep.* 20: 97-107.

22. Pfleiderer, W., Matysiak, S., Bergmann, F., and Schnell, R. (1996) *Acta Biochimica Polonica* 43: 37-44.

23. Pollard, J., Bell, S., and Ellington, A. (1990) *Design, Synthesis, and Amplification of DNA Pools for In Vitro Selection* (personal communication).

24. Erlich, H.A. (1989) *in PCR Technology: Principles and Applications for DNA Amplification*, pp. 1-7, Stockton Press, New York, NY.

25. Moran, L., Scrimgeour, K., Horton, R., Ochs, R., and Rawn, J. (1994) *in Biochemistry*, pp. 24.1-24.33, Neil Patterson Publishers/Prentice-Hall, Inc., Englewood Cliffs, NJ.

26. Bartel, D., and Szostak, J. (1993) *Science* 261: 1411-1418.

27. Bartel, D., Zapp, M., Green, M., and Szostak, J. (1991) *Cell* 67: 529-539.

28. Woodson, S. (2000) *Cell Mol. Life Sci.* 57: 796-808.

29. Seiwert, S., Nahreini, T., Aigner, S., Ahn, N., and Uhlenbeck, O. (2000) *Chemistry and Biology* 7: 833-843.

30. Geiger, A., Burgstaller, P., Eltz, H., Roeder, A., and Famulok, M. (1996) *Nucleic Acids Research* 24: 1029-1036.

31. Saenger, W. (1984) *in Principles of Nucleic Acid Structure* (Cantor, C. Ed.) pp. 201-219, Springer - Verlag, New York, NY.

32. Rusconi, C., Yeh, A., Lyerly, H., Lawson, J., and Sullenger, B. (2000) *Thromb Haemost.* 84: 841-848.

33. Jellinek, D., Lynott, C., Riata, D., and Janjic, N. (1993) *Proc Natl Acad Sci USA.* 90: 11227-11231.

34. Yang, Q., Goldstein, I., Mei, H., and Engelke, D. (1998) *Proc Natl Acad Sci USA.* 95: 5462-5467.

35. Conrad, R., Giver, L., Tian, Y., and Ellington, A. (1996) *Methods Enzymology* 267: 336-367.

36. Jhaveri, S., and Ellington, A. (1990) *In Vitro Selection of RNA Aptamers to a Protein Target by Filter Immobilization* (personal communication).

37. Baskerville, S., Zapp, M., and Ellington, A. (1995) *Journal of Virology* 69: 7559-7569.

38. Allen, S. (1999) *Thesis* pp. 94-100.

39. Sassanfar, M., and Szostak, J. (1993) *Nature* 364: 550-553.

40. Brown, D., Brown, J., Kang, C., Gold, L., and Allen, P. (1997) *Journal Biological Chemistry* 272: 14969-14974.

41. Tuerk, C. (1995) *in Methods in Molecular Biology PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering* (White, B. Ed.) pp. 219-230, Humana Press Inc., Totowa, NJ.

42. Stone, S., Hughes, M., and Jost, J. (1991) *in A Laboratory Guide to In Vitro Studies of Protein-DNA Interactions* (Saluz, H., and Becker, M. Eds.) pp. 163-176, Birkhauser-Verlag, Boston, MA.

43. Bell, S., Denu, J., Dixon, J., and Ellington, A. (1998) *Journal of Biological Chemistry* 273: 14309-14314.

44. Validis, B. (1998) *in Basic Cloning Procedures* (Berzins, V. Ed.) pp. 1-169, Springer, New York, NY.

45. Graham, C., and Hill, A. (2001) *in Methods in Molecular Biology: DNA Sequencing Protocols* (Vol 167) pp. 1-244, Humana Press Inc., Totowa, NJ.

46. Patel, D., and Suri, A. (2000) *Journal of Biotechnology* 74: 39-60.

47. Zuker, M., Mathews, D., and Turner, D. (1999) *RNA Biochemistry and Biotechnology* 288: 911-940.

48. Osborne, S., and Ellington, A. (1997) *Chemical Rev.* 97: 349-370.

49. Bock, L., Griffin, L., Latham, J., Vermaas, E., and Toole, J. (1992) *Nature* 355: 564-566.

50. Wiegand, T., Williams, P., Dreskin, S., Jouvin, M., Kinet, J., and Tasset, D. (1996) *Journal Immunology* 157: 221-230.

51. Cui, Y., Wang, Q., Stormo, G., and Calvo, J. (1995) *Journal Bacteriology* 177: 4872-4880.

52. Schneider, D., Feigon, J., Hostomsky, Z., and Gold, L. (1995) *Biochemistry* 34: 9599-9610.

53. Latham, J., Johnson, R., and Toole, J. (1994) *Nucleic Acids Research* 22: 2817-2822.

54. Lin, Y., Qiu, Q., Gill, S., and Jayasena, S. (1994) *Nucleic Acids Research* 22: 5229-

55. Jensen, K., Atkinson, B., Willis, M., Koch, T., and Gold, L. (1995) *Proc Natl Acad Sci USA* 92: 12220-12224.

56. Lin, Y., Padmapriya, A., Morden, K., and Jayasena, S. (1995) *Proc Natl Acad Sci USA* 92: 11044-11048.

57. Jellinek, D., Green, L., Bell, C., Lynott, C., Gill, N., Vargeese, C., Kirschenheuter, G., McGee, D., Abesinghe, P., Pieken, W., Shapiro, R., Rifkin, D., Moscatelli, D., and Janjic, N. (1995) *Biochemistry* 34: 14413-14424.

58. Green, L., Jellinek, D., Bell, C., Beebe, L., Feistner, B., Gill, S., Jucker, F., and Janjic, N. (1995) *Chemical Biology* 10: 683-695.

59. Pieken, W., Olsen, D., Benseler, F., Aurup, H., and Eckstein, F. (1991) *Science* 253: 314-317.

60. Pan, W., Craven, R., Qiu, Q., Wilson, C., Wills, J., Golovine, S., and Wang, J. (1995) *Proc Natl Acad Sci USA* 92: 11509-11513.

61. Green, L., Waugh, S., Binkley, J., Hostomska, Z., Hostomsky, Z., and Tuerk, C. 1995) *Journal of Molecular Biology* 247: 60-68.

62. Atkins, J., and Gesteland, R. (1993) *in The RNA world : the nature of modern RNA suggests a prebiotic RNA world,* pp. 1-630, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

63. Moran, L., Scrimgeour, K., Horton, R., Ochs, R., and Rawn, J. (1994) *in Biochemistry*, pp. 14.1-14.20, Neil Patterson Publishers/Prentice-Hall, Inc., Englewood Cliffs, NJ.

64. Harada, K., and Frankel, A. (1995) *EMBO* 14: 5798-5811.

65. Conn, M., Prudent, J., and Schultz, P. (1996) *Journal American Chemical Society* 118: 7012-7013.

66. Li, Y., and Sen, D. (1996) *Nat Struct. Biology* 3: 743-747.

67. Lorsch, J., and Szostak, J. (1994) *Biochemistry* 33: 973-982.

68. Majerfeld, I., and Yarus, M. (1994) *Nat Struct. Biology* 1: 287-292.

69. Lauhon, C., and Szostak, J. (1995) *Journal American Chemical Society* 117: 246-1257.

70. Famulok, M., and Szostak, J. (1992) *Journal American Chemical Society* 114: 3990-3991.

71. Connell, G., and Yarus, M. (1994) *Science* 264: 1137-1141.

72. Geiger, A., Burgstaller, P., Eltz, H., Roeder, A., and Famulok, M. (1996) *Nucleic Acids Research* 24: 1029-1036.

73. Mannironi, C., Di Nardo, A., Fruscoloni, P., and Tocchini-Valentini, G. (1997) *Biochemistry* 36: 9726-9734.

74. Schneider, D., Tuerk, C., and Gold, L. (1992) *Journal of Molecular Biology* 228: 862-869.

75. Wallace, S., and Schroeder, R. (1998) *RNA* 4: 112-123.

76. Griffin, L., Tidmarsh, G., Bock, L., Toole, J., and Leung, L. (1993) *Blood* 81: 3271-3276.

77. Kelly, J., Feigon, J., and Yeates, T. (1996) *Journal of Molecular Biology* 256: 417-422.

78. Hirao, I., Madin, K., Endo, Y., Yokoyama, S., and Ellington, A. (2000) *Journal of Biological Chemistry* 275: 4943-4948.

79. Goodman, S., Velten, N., Gao, Q., Robinson, S., Segall, A. (1999) *Journal of Bacteriology* 181: 3246-3255.

80. Connell, G., and Christian, E. (1993) *Origins of Life* 23: 291-297.

81. Lorsch, J., and Szostak, J. (1994) *Nature* 371: 31-36.

82. Schultz, P., and Lerner, R. (1995) *Science* 269: 185-1842.

83. Abelson, J. (1990) *Science* 249: 488-489.

84. Ellington, A., and Szostak, J. (1992) *Nature* 355: 850-852.

85. Prudent, J., Uno, T., and Schultz, P. (1994) *Science* 264: 1924-1927.

86. Wilson, C., and Szostak, J. (1995) *Nature* 374: 777-782.

87. Illangasekare, M., Sanchez, G., Nickles, T., and Yarus, M. (1995) *Science* 267: 643-647.

88. Li, Y., Geyer, R., and Sen, D. (1996) *Biochemistry* 35: 6911-6922.

89. Cuenoud, B., and Szostak, J. (1995) *Nature* 375: 611-614.

90. Yallow, R., and Berson, S. (1959) *Nature* 185: 1648-1649.

91. Kohler, G., and Milstein, C. (1975) *Nature* 256: 496-497.

92. Jayasena, S. (1999) *Clinical Chemistry* 45: 1628-1650.

93. Charlton, J., Sennello, J., and Smith, D. (1997) *Chemical Biology* 4: 809-816.

94. Fontana, W., Konings, D., Stadler, P., and Schuster, P. (1993) *Biopolymers* 33: 1389-1404.

Chapter 2

Characterization of *in vitro* selected RB69 RegA RNA Binding Sites

I. Bacteriophage T4 Translational Repressors

Bacteriophage T4 utilizes translational repression as a method of gene regulation (1). The T4 proteins Gp43, Gp32, and RegA have all been studied extensively and much is known about their roles in gene regulation. While all three proteins repress translation, Gp32 and Gp43 carry out their regulation in a manner differing from that of RegA. Both Gp32 and Gp43 regulate their own synthesis by preventing translation of their cognate mRNAs, while RegA has been shown to regulate numerous genes, in addition to its own transcript (1).

T4 translational repressor Gp32

Gp32 is a single stranded DNA binding protein. Its primary function is to bind and stabilize single stranded DNA allowing DNA polymerase to replicate the opposite strand (2). While Gp32 is primarily a ssDNA binding protein, it will bind to RNA when the concentration of protein reaches an elevated level. When the ssDNA is saturated with Gp32 the excess protein then binds to its own mRNA repressing further translation (1). Gp32 recognizes an RNA pseudoknot structure at its own translation initiation region (TIR), then fills an adjacent unstructured region, thereby reducing binding and translating the mRNA (3). Gp32 is a Zn^{2+} protein, incorporating 1 mole of Zn^{2+} per mole of protein (4). In addition, three of the four cysteines present in Gp32 are found in a tyrosine-rich sequence (residues 72-116) and have been implicated in DNA binding by 1H NMR investigations (4).

T4 translational repressor Gp43

Gp43 is a DNA replication accessory protein with 5'→3' polymerase activity, as well as having 3'→5' exonuclease activity (1). Gp43 operates somewhat analogous to Gp32 function in translational repression. Gp43 binds to an RNA hairpin in the translation initiation region of the gene 43 mRNA. The minimum operator sequence to which Gp43 binds consists of 36 nucleotides that include a hairpin (containing a 5 base-pair helix and an 8 nucleotide loop) and a single-stranded segment that contains the Shine-Dalgarno sequence of the ribosome binding site (5). Binding by Gp43 at this site occludes the TIR preventing translation of the gene (5,6). Gp32 and Gp43 function in translational repression by recognizing RNA secondary structure and binding at that site. The third T4 translational repressor, RegA, functions much differently in translational repression than the other two repressors.

T4 translational repressor RegA

T4 RegA is a protein that, unlike Gp32 and Gp43, is only known to function as a translational repressor. Bacteriophage T4 is an unusual translational repressor in that it is capable of recognizing multiple mRNAs of different sequence, with different affinities (see Table 2.1 & 2.2), including its own transcript (1). A major feature of the RegA binding sites is their apparent lack of secondary structure (7). T4 RegA, like Gp32 and Gp43, binds at the TIR of an mRNA preventing ribosome binding and translation initiation.

The T4 translational repressor RegA protein folds into two structural domains (8). Domain I of the RegA protein contains a four-stranded beta-sheet and two alpha-helices. Domain II contains a four-stranded beta-sheet and an unusual 3/10 helix (8). Mutagenesis study has shown the RNA binding site to be a surface pocket formed by residues on two loops and an alpha-helix (8).

Table 2.1 RegA sensitive mRNAs[a]

T4 Gene	Sequence[b]	50% Repression (mM RegA)[c]
44	GAAUGAGGAAAUU**AUG**AUUACUGUAAAUG	0.6
rIIA	UAAUGAGGAAAUU**AUG**AUUAUCACCACUG	n.d.
αgt	AAAGAGGAAAACU**AUG**CGUAUUUGCAUUU	1
rpbA	UACUGACUUAUUAUGACUAAAGUGUAU**AUG**ACUAAAAUUACUGUGA	1.5
dexA	CGAGGAAAAUUUA**AUG**UUUGAUUUUAUUA	1.5
rIIB	AAUAAGGAAAAUU**AUG**UACAAUAUUAAAU	3.2
dexA.1	GAGGCGAUUAUUA**AUG**AUUGAAUUAAGUU	3.7
45	GAAUUGAAGGAAAUUAC**AUG**AAACUGUCUAAAG	3.8
βgt	AUAAAAGGAAUAU**AUG**AAAAUUGCUAUAA	4.5
42	CAUUGAGGUUUAA**AUG**AUUAGUGACUCUA	n.d.
regA	CAUUGGAAUGGUAAA**AUG**AUUGAAAUUACUC	10
1	UGAGGAGAAACAC**AUG**AAACUAAUCUUUU	11.5
62	CAGUGGAAGUGAU**AUG**AGCUUAUUUAAAG	n.d.

[a] Adapted from Miller et al., 1994.

[b] RegA sensitive TIRs aligned at the AUG initiation codon. Underlined nucleotides are protected from RNase digestion by bound RegA protein.

[c] The concentration of RegA causing 50% reduction in the yield of *in vitro* translated product from each message. n.d., not determined. Data compiled from (9,10,11,12,13,14).

Table 2.2 Hierarchy of biological RNA affinities for T4 and RB69 Reg A proteins[a]

T4 RegA Protein	Sequence[b]	$K_{app}(10^7\ M^{-1})$
RB *G45*	AAGGAAAUAAA	19.9
T4 *G44*	GAGGAAAUUAU	4.9
T4 *G45*	AAGGAAAUUAC	3.2
RB *G44*	GAGGAAAAUUA	2.5
RB *regA*	AUGGUAAAAAU	1.3
T4 *regA*	AUGGUAAAAUG	0.8
RB69 RegA Protein		**$K_{app}(10^7\ M^{-1})$**
T4 *G44*	GAGGAAAUUAU	34.0
RB *G44*	GAGGAAAAUUA	9.5
T4 *G45*	AAGGAAAUUAC	2.0
RB *G45*	AAGGAAAUAAA	1.8
RB *regA*	AUGGUAAAAAU	0.6
T4 *regA*	AUGGUAAAAUG	0.16

[a] Adapted from Sengupta et al., 2001.
[b] Underlining indicates RNase protected nucleotides.

II. Using SELEX to explore RegA-RNA interactions

T4 RegA:RNA analysis using SELEX

RegA is a protein, which binds to the translation initiation regions of upwards of 20 different mRNAs (15). Due to the lack of apparent secondary structure and high sequence conservation in the RegA regulated genes, SELEX was used to determine the RNA sequence requirements for binding (16). Starting with an initial RNA pool consisting of 30 randomized positions, fifteen rounds of selection were needed to generate aptamers that were similar to biologically relevant binding sites (16). Dissociation constants ranged from 5 to 200 nM as

determined using an electrophoretic mobility shift assay (Table 2.3) (16). An interesting aspect of this T4 SELEX experiment was the generation of multiple binding sites within the random region that was subjected to selection. One site consisted of the first thirteen 5' nucleotides with the second consisting of four 3' nucleotides. The 3' sequence of AAAA is relevant when the 3' fixed region, used for primer binding, (underlined) is taken into account (AAAAUUCGACAUG) with the whole sequence resembling the 5' sequence (AAAAUUGUUAUGUAA). These SELEX derived sequences compare favorably with T4 binding sites, with the up-stream poly A tract and UU repeats being notable similarities (Table 2.4). These experiments clearly show the efficacy of the SELEX procedure for use in elucidating biological protein:nucleic acid sequence issues.

Table 2.3 Binding affinity of selected RNA sequences for T4 RegA[a]

Putatuve binding site	Dissociation constant (nM)
AAAAUUGUUAUGUAA	5
AAAAUUAUUAUGUAA	5
AAAAUUGUUAUGGAA	5
AAAAUUGUUAUGAAA	6
AAAAUUGUGAUGUAA	8
AAAAUUGGUAUGUAA	12
AAAAUGGUGAUGAAA	11
AAAAUUGUUAUGAAC	24
AAAAUGGUGAUGAGA	24
AAGAUUGCAAUGGAA	40
AAAAAUAAAAUGAAA	200

[a] Adapted from Brown et al. (1997)

Table 2.4 Biological T4 mRNAs that are regulated by RegA[a]

T4 Gene	RegA binding site
rIIB	AAAAUU**AUG**UAC
44	AAAUU**AUG**AUU
DexA	AAAAUUUA**AUG**UUU
RpbA	AAAGGUGUAUU**AUG**ACU
45	AAAUUAC**AUG**AAA
agt	AAAACU**AUG**CGU
bgt	AAAAGGAAUAU**AUG**AAA
1	AAACAC**AUG**AAA
regA	AAUGGUAAA**AUG**AUU

[a] Adapted from Brown et al. (1997)

RB69 RegA:RNA analysis using SELEX

RB69 RegA translational repressor protein is 78% identical in amino acid sequence to the well-characterized T4 protein (15), binds both RB69 and T4 mRNAs, and displays a hierarchy of binding to RNAs that do not have a conserved sequence or structure (17). These differences in the sequences of T4 RegA and RB69 RegA, and their different affinities for the same mRNAs suggested RB69 RegA as a useful homolog of the T4 protein for studying RNA-binding and translational regulation.

RNAs randomized at 14 positions (the SLXN14 template) were carried-through rounds of SELEX in binding reactions with immobilized His-tagged RB69 RegA (RegA-His_6). After only five rounds of selection, RNAs that preferentially bound RB69 RegA-His_6 showed extensive sequence conservation among the 18 individual RNAs that were cloned and sequenced (18).

The following manuscript details work completed in this laboratory (T.R. Dean & S.V. Allen) to elucidate the sequence requirements for RB69 RegA binding to RNA in the absence of ribosome binding competition. Prior to publication of the T4 RegA SELEX analysis, our laboratory had initiated this selection approach with the bacteriophage RB69 RegA protein.

REFERENCES:

1. Miller, E., Karam, J., and Spicer, E. (1994) *in Molecular Biology of Bacteriophage T4* (Karam, J., and et al., Eds.) pp. 193-205, ASM Press, Washington, D.C.

2. Villemain, J., and Giedroc, D. (1996) *Biochemistry* **35**: 14395-14404.

3. McPheeters, D., Stormo, G., and Gold, L. (1988) *Journal of Molecular Biology* **201**: 517-535.

4. Giedroc, D., Keating, K., Williams, K., Konigsberg, W.,and Coleman, J.(1986) Proc Natl Acad Sci U S A **83**: 8452-6

5. Tuerk, C., Eddy, S., Parma, D., and Gold, L. (1990) *Journal of Molecular Biology* **213**: 749-761.

6. Andrake, M., Guild, N., Hsu, T., Gold, L., Tuerk, C., and Karam, J. (1988) *Proc Natl Acad Sci U S A* **85**: 7942-7946.

7. Szewczak, A., Webster, K., Spicer, E., and Moore, P., (1991) *Journal of Biological Chemistry* **266**: 17832-17837.

8. Kang, C., Chan, R., Berger, I., Lockshin, C., Green, L., Gold, L., and Rich, A. (1995) Science **268**: 1170-1173.

9. Winter, R., Morrissey, L., Gauss, P., Gold, L., Hsu, T., and Karam, J. (1987) *Proc Natl Acad Sci U S A* **84**: 7822-7826.

10. Miller, E., Karam, J., Dawson, M., Trojanowska, M., Gauss, P., and Gold, L. (1985)) *Journal of Biological Chemistry* **194**: 397-410.

11. Webster, K., Adari, H., and Spicer, E. (1989) *Nucleic Acids Research* **17**: 10047-10068.

12. Daegelen, P., and Brody, E. (1990) *Genetics* **125**: 237-248.

13. Karam, J., Gold, L., Singer, B., and Dawson, M. (1981) *Proc Natl Acad Sci U S A* **78**: 4669-4673.

14. Unnithan, S., Green, L., Morrissey, L., Binkley, J., Singer, B., Karam, J., and Gold, L. (1990) *Nucleic Acids Research* **18**: 7083-7092.

15. Jozwik, C., and Miller, S. (1992) *Proc Natl Acad Sci U S A* **89**: 5053-5057.

16. Brown, D., Brown, J., Kang, C., Gold, L., and Allen, P. (1997) *Journal of Biological Chemistry* **272**: 14969-14974.

17. Sengupta, T., Gordon, J., and Spicer, E. (2001) *Nucleic Acids Research* **29**: 1175-1184.

18. Allen, S. (1999) Ph.D. thesis (North Carolina State University, Raleigh, NC).

In vitro Selected Phage RB69 RegA RNA Binding Sites Yields UAA Triplets and a Potential Role in Translational Reinitiation

Timothy R. Dean, Sherrice V. Allen[+] and Eric S. Miller[*]

Department of Microbiology

North Carolina State University

Raleigh, NC 27695-7615

[*]Corresponding author:

Phone: 919-515-7922

Fax: 919-515-7867

email: eric_miller@ncsu.edu

[+]Current address:

Department of Natural Sciences

Fayetteville State University

1200 Murchison Road

Fayetteville, NC 28301

For submission to Nucl. Acids Res.

Abstract

The SELEX method of *in vitro* selection was used to isolate RNAs that bind the RB69 RegA translational repressor protein immobilized to Ni-NTA agarose. After five rounds of SELEX, the pool of selected RNA displayed striking sequence uniformity: UAAUAAUAAUAAUA was clearly enriched in the 14 nucleotides that underwent selection. Individual, cloned molecules displayed the $(UAA)_n$ sequence, with two RNAs having a 3' AUG. Removing the 3' AUG reduced binding in gel shift assays, moving the AUG 5' proximal of the $(UAA)_n$ improved binding, but $(UAA)_4$ alone still bound the purified protein. Dissociation constants showed that RNA shortened to $(UAA)_3$ and $(UAA)_2$ also retained binding, whereas cytosine clearly prevented binding by RB69 RegA. A sequence logo and the makewalker program were used to search the complete RB69 genome for UAA-rich sequences. Several potential RB69 RegA binding sites were identified within the RB69 genome. It also appears that UAA stop codons juxtaposed to a translation initiation codon are an important component of RB69 RegA binding sites. mRNA for the pentameric (4:1) phage gp44/62 DNA polymerase clamp loader complex has the RB69 gene *44* stop codon and 3'-adjacent gene *62* initiation codon in a sequence (AAAUAAUAUG) that is similar to *in vitro* selected RNA and binds RB69 RegA. RegA repressor proteins appear to interact at UAA-AUG junctions of translationally coupled genes.

Introduction

Protein-mediated translational repression is an important regulatory mechanism in prokaryotes and eukaryotes (1,2). Bacteriophage T4 encodes three translational repressors: the single stranded DNA binding protein (gp32), the DNA polymerase (gp43), and RegA (3). Gp32 and gp43 regulate their own translation by binding to a pseudoknot structure or a simple hairpin, respectively, contained in the translation initiation region (TIR) of their mRNA (4-6). As for most translational repressor proteins, RNA structure is an important component of the target site for unambiguous binding and repression of their target mRNAs.

T4 RegA is an unusual translational repressor. It represses translation of several different phage mRNAs, including its own (7,8). Several studies have focused on the specificity determinants in the RegA:RNA interaction. A striking feature of RegA binding sites on RNA is the unusual absence of secondary structure. The TIRs at RegA binding sites are AU-rich. An RNA structure study revealed that one of the binding sites (T4 *44* RNA) is single-stranded, with a 2'-endo conformation of the ribose that is more characteristic of B-form DNA than of A-form RNA (9). Comparing the TIR regions recognized by RegA with all T4 transcripts fails to show a marked difference between the two sets of RNA.

SELEX (Systematic Evolution of Ligands by Exponential Enrichment) is an *in vitro* selection method used to obtain RNA or DNA that has interesting binding or catalytic properties (6,10); reviewed in (11). T4 RegA protein was previously used as an RNA SELEX target to reveal bases required for RegA binding in the absence of sequence requirements for

ribosome binding (12). The SELEX experiments confirmed that T4 RegA recognizes RNA in a sequence-dependent manner and does not rely on RNA structure. The reported "consensus" sequence, 5'-AAAAUUGUUAUGUAA-3'' (12), again showed that T4 RegA binds AUG flanked by AU-rich bases. The crystal structure of the T4 protein has been solved (13), and mutation analyses (14) reveal that specific RNA site recognition is mediated through a unique RNA-binding domain defined by an N-terminal helix region (15,16) in a structural binding "cleft" (15).

Bacteriophage RB69 is a T4-related phage whose genes complement defects in some T4 genes (17). RB69 encodes a RegA protein that is composed of 122 amino acids and is 78% identical to T4 RegA (16). RB69 RegA can bind and translationally repress T4 mRNAs, including gene *44*, *45*, and *rpbA*, but does so with slightly different affinities for the respective RNAs (16,18,19). Thus, RB69 RegA provides a useful homolog of the T4 protein for studying RNA-binding and translational regulation.

RB69 RegA SELEX experiments were carried-out using immobilized RB69 RegA protein and a pool of RNA randomized at 14 positions. After 5 rounds of selection, all of the selected sequences were AU-rich. The preponderance of $(UAA)_n$ in the selected RNAs is consistent with naturally occurring sites, and distinguishes the SELEX binding sites of T4 and RB69 RegA. The results show that RB69 RegA binds $(UAA)_n$ RNA that lacks an AUG, and that an AUG slightly enhances the protein's affinity. SELEX generated RNAs were used to create a sequence logo which when "walked" through the RB69 genome located other

potential RB69 RegA binding sites. A role for RB69 RegA is proposed in which the protein binds RNA at stop (UAA) and start (AUG) codons of adjacent genes (e.g., the DNA polymerase 44/62 clamp loader complex), affecting their coordinate translational regulation.

Materials and Methods

Strains, plasmids, and media. *E. coli* JM109 and pGEM-T vector were from Promega Corp. (Madison, WI). pSA1 contains RB69 *regA* cloned into the polyhistidine fusion vector pET21 for expression of RegA69-His_6 (20). *regA* on pSA1 was sequenced throughout to confirm the absence of mutations in the gene. Procedures for growth of bacteria in LB media have been described (21). Ni-NTA agarose, Ni-NTA spin columns, and Qiaprep Miniprep were obtained from Qiagen (Valencia, CA). Ribomax T7 *in vitro* transcription system, T4 DNA ligase, *Taq* DNA polymerase, and restriction enzymes were obtained from Promega Corp. and used as recommended. Retrotherm RT was from Epicentre Technologies (Madison, WI). Ribogreen RNA Quantitation kit was from Molecular Probes (Eugene, OR).

RegA proteins. RegA69-His_6 C-terminal fusion protein was synthesized *in vitro* from pSA1 and immobilized on agarose Ni-NTA Spin Columns (Qiagen) as described (20). RB69 RegA was induced in *E. coli* B containing a heat-inducible λ expression vector (pEC69) with the wildtype gene, and the protein purified as described (18).

SELEX template. A fourteen nucleotide (nt) variable region DNA library, containing 3×10^8 unique sequences, was constructed by the polymerase chain reaction (PCR) using a 5' T7 fixed primer and a 3' fixed primer that anneal to the SLXN14 oligonucleotide (11). The randomized region was synthesized over argon using proportions of phosphoramidites to yield

equal amounts of each base at the 14 nt positions (IDT, Coralville, IA). The 61 nt SLXN14 oligo was PAGE purified.

SELEX parameters with RegA69-His$_6$. Four initial 50 µl PCR reactions were performed to generate the initial dsDNA pool (22). Each reaction contained: 0.5 mM each dNTP, 7.5 mM $MgCl_2$, 1.5 U Taq DNA polymerase, Buffer (50 mM KCl, 10 mM Tris-HCl, pH 9.0 at 25°C, 0.1% Triton X-100), 6 µM each T7 Fixed and 3' Fixed primers, 0.3 µM SLXN14 template. PCR was performed for 10 cycles at 93°C 30 sec; 55°C 10 sec; and 72°C 1 min. PCR products were separated by nondenaturing 8% PAGE, visualized by ethidium bromide - UV illumination, eluted in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.2% SDS, and ethanol precipitated.

SLXN14 RNA used for selection was generated by *in vitro* transcription with T7 RNA polymerase (Ribomax, Promega Corp.). Each 25 µl reaction contained 12 pmoles of the initial PCR product or 1/2 of the subsequent RT-PCR product (6 - 40 pmoles) in 80 mM HEPES-KOH, pH 7.5, 24 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, 7.5 mM each rNTP and T7 enzyme mix (T7 RNA polymerase, RNase inhibitor and yeast inorganic pyrophosphatase). Reactions were incubated at 37°C for 2 hrs. Following DNase treatment, RNA was phenol-extracted, precipitated and suspended in 30 µl RNase-free dH_2O. Transcripts were analyzed using denaturing (8.3M urea) 8% PAGE and concentrations determined by Ribogreen RNA quantitation (Molecular Probes Inc.).

In vitro translation mixtures containing 1 or 2 pmoles of RegA69-His_6 in 200 µl Ni-NTA buffer I (50 mM NaH_2PO_4, pH 8.0; 300 mM NaCl; 10 mM imidazole) were applied to 40 µl of Ni-NTA slurry, incubated at 4°C for 30 min, followed by centrifugation at 14,000 rpm for 1 min. To remove other extract proteins, the slurry was washed twice with 200 µl Ni-NTA buffer II (50 mM NaH_2PO_4, pH 8.0; 300 mM NaCl; 20 mM imidazole). Immobilized protein was then equilibrated 3 times with 200 µl of RNA binding buffer (20 mM Tris, pH 7.5; 250 mM NaCl; 5 mM $MgCl_2$; 1 mM β mercaptoethanol). The SLXN14 random RNA pool (100 pmoles) in 200 µl RNA binding buffer was denatured at 65°C for 5 min and cooled on ice. In negative pre-selections, the RNA was applied to 40 µl of Ni-NTA without RegA69-His_6, incubated for 10 min at 25°C, centrifuged at 7,000 rpm for 2 min, and the supernatant collected. RNA (40 pmoles) was then applied to 40 µl RegA69-His_6 Ni-NTA, incubated and centrifuged as above. The slurry was washed 3 times with 200 µl of RNA binding buffer to remove unbound RNAs. The flow-through and washes for each round were used in the Ribogreen RNA assay to measure the amount of random RNA that did not bind to the immobilized RegA69-His_6 protein. Selected RNAs were extracted from the slurry with 400 µl 1:1 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.2% SDS and phenol/$CHCl_3$ followed by centrifugation at 14,000 rpm for 1 min. Extracted RNA was ethanol precipitated and suspended in dH_2O.

Each 50 µl RT-PCR reaction contained 20 µl of RNA (10 to 20 pmoles) from the previous round of selection, 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.75 mM $MnSO_4$, 0.2 mM each dNTP, 1 µM 3' Fixed and T7 Fixed primers, and 5 U Retrotherm RT (EpicentreTechnologies). Primer annealing was done at 50°C for 5 min, first strand synthesis at 70°C for 10 min, and PCR for 25 cycles at 93°C 30 sec, 55°C 10 sec and 72°C 1 min. Amplified DNA was visualized by 8% non-denaturing PAGE to ensure correct product size. After rounds 1, 2, and 5, RT-PCR products were gel-purified. DNA was transcribed as above to create the RNA pool for the next round of selection. Round 5 RT-PCR product was cloned into pGEM-T. The ligation mixtures were transformed into JM109 as described (21) and insert-containing clones sequenced.

RNA binding assays and K_d*app* determinations. RNA was transcribed from dsDNA made using a sequence-specific oligo template, 5' T7 promoter primer, and a 3' primer. PCR conditions to create dsDNA for transcription were identical to those used in SELEX (see above). RNA was labeled with [α^{32}P-UTP] via a T7 *in vitro* RNA transcription system (Ribomax, Promega Corp). Each 25 µl reaction contained 10% of PCR, 80 mM HEPES-KOH, pH 7.5, 24 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, 2.5 mM each rATP, rGTP, and rCTP, 1 mM rUTP, 50 µCi α^{32}P-UTP, and T7 enzyme mix (T7 RNA polymerase, RNase inhibitor and yeast inorganic pyrophosphasase). Reactions were incubated at 37°C for 2 hr. Entire transcription reactions were run on 8% denaturing PAGE, appropriate size bands were

excised, and RNA extracted overnight into ribonuclease-free water. Following ethanol precipitation, purified RNA was suspended in 15 µl ribonuclease free water. RNA (1.5 µl) was mixed into 13 µl of RNA binding buffer (150 mM NaCl), heated to 70°C for 3 min and then cooled to room temperature over 10 min. Single-point gel shift assays used 1.5 µM purified RB69 RegA protein incubated with 115 nM of α^{32}P-UTP-labeled RNA for 20 min at room temperature. For K_d*app* determinations, RB69 RegA from 18 nM to 739 nM was used in a final binding volume of 15 µl. Binding reactions were placed at -20°C for 5 min and then 5 µl of loading dye added. Samples were run on 12% non-denaturing PAGE in 0.5X TBE for 3 hr at 4°C. Gels were placed on a phosphor screen, exposed for 30 min, and viewed using a Storm 840 phosphorimager (Molecular Dynamics). ImageQuant software was used to obtain the volume report of labeled bands and to derive the fraction of RNA shifted. K_d*app* values were obtained from Bjerrum plots of the fraction of free RNA vs. the log concentration of RB69 RegA (23). K_d*app* was calculated where the fraction of free RNA = 0.5. Values are the average of two or more separate experiments, with R ranging from 0.95 to 0.99.

Logo and walker analysis. RB69 RegA-selected RNA sequences (eleven UAA containing sequences) were aligned and used to generate a sequence logo (24). The total information content of the selected sites (R_{seq}) was calculated, as was R_{Freq} of the RB69 complete genome (Karam, J. 2002 at http://www.tmc.tulane.edu/sif/t4genome.html). To search for potential RegA binding sites within the RB69 genome, a "minimal binding site"

sequence logo (8 bases) was scanned (25) against the complete RB69 genome. The stringency of the search was adjusted to only include matches with an information content between 10 and 13 bits. Sites were then evaluated as to whether they were likely located on RNA transcripts.

Results

Conditions for selecting RB69 RegA RNA binding sites. *In vitro* selection (SELEX) was undertaken to identify high-affinity RNA binding sites of the phage RB69 RegA protein. Because many of the shorter, biologically relevant RNA sequences (e.g., the gene *44* RNA) bound by RegA are 9 – 12 nucleotides, a randomized region of 14 consecutive bases was used. At 4^{14}, approximately 3 X 10^{8} different RNAs are theoretically possible. Since 15 pmoles of the initial SLXN14 template were converted to dsDNA for transcription, at least 10^{4} over-representation of each sequence was achieved in the dsDNA. Using 40 pmoles of RNA transcript, 10^{3}–fold excess of the different RNAs was present during the first binding reaction with RegA. The relatively short randomized region limits the structural diversity in the pool. However, structure has not appeared relevant for RegA (9,12) and the shorter random region should reduce the number of rounds needed to select high affinity RNAs.

To partition complexes, RegA-His_6, previously shown to bind T4 *44* RNA (20), was immobilized on agarose. In the first two rounds, negative pre-selection was performed with Ni-NTA agarose alone to remove potential matrix-binding RNAs. Following round 2, the amount of immobilized RB69 RegA-His_6 was reduced to one pmole (RNA:RegA ratio of 40:1) to increase the stringency in the binding reaction. To further increase the selection stringency and reduce the number of rounds, 250 mM NaCl was used in the column wash buffer, which is the threshold ionic condition at which RegA remains bound to polyU (18,26). Overall,

SELEX parameters (**Table 1**) were used to enhance selection efficiency, while reducing the number of cycles needed to obtain high-affinity RB69 RegA binding sites.

The selected bulk RNA pool displays a repeating UAA sequence. The progress of selection, and the diversity of the RNA pool, were evaluated by sequencing a portion of the dsDNA (RT-PCR product) following the fifth round of SELEX. **Figure 1** shows the sequence of the template pool prior to selection (which showed a slight over-representation of Cs), and that of the pool selected for binding to RegA-His_6 after five rounds. A reduction in the randomization of the N_{14} region was seen, with an apparent selection for alternating Us and As. A clear pattern of $(UAA)_n$ is evident in the selected RNA pool. Therefore, the dsDNA at the end of round five was cloned and the nucleotide sequence of independent inserts determined.

Sequences of individual selected RNAs. Within the 14-base variable region of the selected RNAs, a highly similar sequence was identified in each of the eighteen clones analyzed. As shown in **Table 2**, the sequences can be divided into two separate groups, those that consist of UAA and those that contain UAA + C. The sequence 5'-UAAUAAUAA-3' was present in eleven of the eighteen RNAs. The related sequences 5'-UAAUAAUUAUAA-3' (SA5-7) and 5'-ACAAUAACAAUAA-3' (SA5-3) each appeared once, varying only at the underlined positions. Variation was also observed in a few 3' fixed positions where, during SELEX, PCR appeared to have caused base substitutions. A 3'-terminal AUG only occurred

twice (clones SA5-1 and SA5-2). Like the TIRs repressed by RegA, and the RB69 genome overall, the selected RNAs are AU-rich. $(UAA)_n$ is present in each selected RNA, yet an AUG is not a prominent feature of the pool or of the individual RNAs sequenced. Indeed, a "consensus" sequence (RB69-Con; 5'-UAAUAAUAAUAAUA-3'; Table 2), as well as the SELEX sequence logo derived from these RNAs, is variable in the 3'-termnal bases and lacks an AUG. Although an AUG in the recognition site has been shown to be important for RNA binding by T4 RegA (8,27-30), in the context of *in vitro* selection, $(UAA)_n$ emerged as the prominent feature for RB69 RegA.

Purified RB69 RegA binds phage and SELEX RNAs in the gel shift assay. Gel mobility shift assays were standardized on two naturally occurring RNA sequences and one mutant derivative RNA. RB69 *44*, T4 *44*, and T4 *44*-C (G→C substitutions eliminating the two AUGs in T4 *44*) RNAs were used with purified RB69 RegA protein lacking the His_6 terminal extension. To survey binding, excess protein (1.5 μM) was used in the gel shift assays, with an RNA concentration of 115 nM. Under the conditions used, purified RB69 RegA protein bound and shifted both RB69 and T4 *44* RNA, whereas the mutant *44*-C RNA showed no shift.

Three sequences were initially chosen as representative for analysis of SELEX RNAs. Each of these contained the $(UAA)_n$ sequence, with one having a 3' AUG (SA5-1), one having a 3' AGA (SA5-7), while the third was a modified "consensus" sequence (RB69-Con)

consisting of AAAA$(UAA)_3$UAG and flanking nucleotides. When transcribed, the selected RNAs were 61 bases long, including 5' and 3' fixed regions used during the SELEX procedure. Purified RB69 RegA generated a shift with SA5-1 RNA that was comparable to the shift with biological RB69 *44* RNA. RB69-Con RNA gave a similar shift. There was no apparent shift of the SA5-7 RNA. SA5-7 contains a "UUA" within the selected $(UAA)_n$ region, and lacks an AUG at any position in the RNA. Compared to the RB69 and the T4 *44* RNAs, SA5-1 and RB69-Con shifted equally well with excess purified RB69 RegA (**Table 3**).

To determine whether the flanking fixed regions contributed to RegA binding, the selected RNAs were truncated 5' and 3' to yield approximately 20-base RNAs. SA5-1 RNA was truncated to the 14 base random region plus 2 bases from the 3' fixed region that also displayed sequence variation. Because DNA templates and T7 RNA polymerase were used to produce the labeled RNA, three Gs were included at the 5' end to enhance transcription efficiency. The length of the truncated SA5-1 [$(UAA)_4$-3'AUG] RNA was 19 bases. Thus, $(UAA)_4$-3'AUG RNA showed a significant band shift with RB69 RegA **(Table 3)**.

Variants of the $(UAA)_4$-3'AUG RNA were constructed to evaluate the relevance of specific bases for RegA binding. To determine whether the position of an AUG was important, RNA was synthesized to position the AUG 5' of the $(UAA)_4$ region**.** Under the conditions used, $(UAA)_3$-5'AUG showed the maximum gel shift of selection-derived RNAs. $(UAA)_4$-3'AUG RNA was also changed at the terminal AUG to AUC [$(UAA)_4$-3'AUC] and

AAG [$(UAA)_4$-3'AAG]. These two RNAs produced gel shifts approximately half that of RB69 *44* RNA **(Table 3).** To evaluate requirements for the length of $(UAA)_n$ on RegA binding, RNAs were prepared with shorter intervals, flanked 5' and 3' by A residues. RB69 RegA bound RNAs $(UAA)_3$ and $(UAA)_2$ with shift values approximately half that of RB69 *44* RNA **(Table 3)**, further demonstrating that an AUG is not explicitly required for RB69 RegA to bind RNA.

RB69 RegA was assayed for binding to its own RNA. Two different length RegA sequences were used, with the opportunity for more than one RegA protein to bind the longer sequence. The longer sequence, RB69 *regA*-78, consisted of 78 bases previously shown to be protected by RB69 RegA protein in nuclease protection assays (18), although it contains only one 5' proximal AUG. RB69 RegA bound and shifted RB69 *regA*-78 RNA. The second, shorter RNA RB69 *regA*-28, consisted of 28 bases from the 5' end of *regA*-78, and included the initiator AUG. However, under these conditions, there was no observable shift with RB69 *regA*-28.

RB69 RegA binds the consensus T4 RegA-selected RNA. RNA was previously selected that binds T4 RegA (12); in that approach, nitrocellulose filters were used to partition bound from unbound RNA during SELEX. The T4-selected consensus RNA is AU-rich, contains only one UAA sequence, and has an AUG proximal to the 3' end (**Table 3**). The T4 RegA SELEX consensus RNA, with 5'-terminal G residues (T4-Con; 5'-

GGGAAAAUUGUUAUGUAA-3') was evaluated for binding RB69 RegA. The T4 SELEX consensus sequence RNA produced a shift with RB69 RegA that was slightly less than RB69-Con RNA.

C residues prevent RB69 RegA binding. The affect of Cs in the selected RNAs on binding RB69 RegA was evaluated using the selected SA5-3 and SA5-10 RNA sequences in truncated form (Figure 2). SA5-3 RNA contains two Cs where U is found in the other sequences and SA5-10 contains 5'-AACAAUA-3' which was found 4 times in the selected aptamers (SA5-3, SA5-8, SA5-10, and SA5-11). However, neither of these RNAs displayed any affinity for RB69 RegA in the gel shift assay (Table 3). Additional rounds of SELEX would likely have eliminated these RNAs from the selected pool. Together with the data on mutant T4 *44-C* RNA (see above) it appears that C's, whether they are substituted in the AUG or in the selected AU rich region, inhibit RNA binding by RegA.

Apparent K_d values for RB69 RegA-selected RNAs. Apparent K_d values (K_d *app*) were determined for three biological RNAs using varying concentrations of purified RB69 RegA, ^{32}P-labeled RNA and Bjerrum plots of gel mobility shift assay (**Figure 3** and **Table 3**). RB69 *44* RNA generated a K_d *app* of 122 nM, while T4 *44* RNA produced a higher affinity at 16 nM. A similar higher affinity of RB69 RegA protein for the T4 *44* RNA over the RB69 *44* RNA was previously reported (19). The RB69 *regA*-78 site produced a K_d *app* of 165 nM. The binding curve of RB69 RegA with $(UAA)_4$-3'AUG RNA yielded a K_d *app* of 295 nM. Variants

changing the 3' AUG to AAG [$(UAA)_4$-3'AAG] gave a protein-dependent binding curve with a K_d *app* of 337 nM and $(UAA)_3$ RNA generated a K_d *app* of 281 nM. Moving the AUG to the 5' end of the RNA [$(UAA)_3$-5'AUG] resulted in a K_d *app* of 350 nM. These latter values are approximately three-fold higher than the K_d *app* for RB69 *44* RNA.

RB69 RegA SELEX sites occur at gene junctions in the RB69 genome. The RB69 genome was searched with the query sequences $(TAA)_2$ and $(ATA)_2$ (GCG and BioEdit programs (31,32). 5'-AUAAUAUG-3' was found at the TIRs of genes *62* and *rpbA*, two T4 genes known to be translationally repressed by RegA (3). These RB69 mRNA sequences are highly similar to the *in vitro* selected RNAs. The RB69 *62* RNA was synthesized with A residues flanking the identified sequence. Purified RB69 RegA bound the specific ^{32}P-labeled RB69 *62* RNA in the gel shift assay, displaying a concentration-dependent mobility shift with an apparent K_d *app* of 437 nM (**Figure 3**).

Refining the search parameters, a sequence logo (24) was generated using the aligned (selected) UAA sequences, (Figure 5). Sequence walker (25) was then used to scan the complete RB69 genome for potential RB69 RegA binding sites. The 14 nucleotide logo had a information content (R_{seq}) of 21.27. The same logo generated an R_{freq} of 13.34. Gel-shift binding assays showed that $(UAA)_2$ bound RegA with affinity approximately 50% that of T4 *44*. A sequence logo (8 nucleotides) consisting of $(UAA)_2$ flanked by a single position both 5'

and 3' (position -4 to +3 in Figure 5) had a R_{seq} of 12.90 and an R_{freq} of 12.60, indicating the minimal site size necessary for recognition by RB69 RegA (where $R_{seq} \approx R_{freq}$). Therefore, the $(UAA)_2$ sequence logo was used with the RB69 genome, to identify sites with information contents between 12 and 13 bits. Initially 26 sites were found. Several of these sites are not transcribed into RNA, leaving 10 potential sites (Table 4). The site at position 6151 is exactly between *comC*-α and the adjacent hypothetical T4-like gene orf *39.2*, and includes the start codon of *39.2*. Two sites are found at the ends of genes (*rIIB* and T4 hypothetical orf *e.8*). Their placement at the end of the genes makes both of these sites possible targets for regulation by RB69 RegA. The remaining 6 sites fall in the middle of various genes (**Table 4**), four of which encode structural proteins (gp50 head completion protein, major capsid protein, gp24 head vertex protein, and gp29 tail tube assembly protein).

Discussion

T4 and RB69 RegA repressor proteins have closely related primary sequences, with molecular modeling predicting a common structure and conserved RNA-binding domain that is unique among RNA binding proteins (19). Differences between the two proteins include: RB69 RegA, when not bound to RNA, associates as a tetramer (18), whereas T4 RegA is a dimer (33); RB69 RegA displays obvious cooperativity when binding certain RNAs (i.e., *rpbA*) (18); and there are subtle differences in the selection of target RNAs by the two repressors (16,18,19). To help address differences in RNA site specificity by the two proteins, the SELEX method of *in vitro* RNA selection was undertaken with RB69 RegA.

After five rounds of SELEX, the bulk RNA pool clearly displayed a prominent sequence (Figure 1). In the very first RNA SELEX experiment (6), RNA ligands for T4 DNA polymerase were selected with four rounds using RNA randomized at eight hairpin loop positions. T4 RegA required 15 rounds of SELEX to derive a consensus sequence, and other selections (particularly for targets that do not normally bind nucleic acids) have required 20 rounds or more (34). However, for T4 RegA and many other SELEX protocols, a highly diverse 30+ base random region was used (11). By limiting the randomized region to 14 positions, we achieved selection efficiency approaching that for T4 DNA polymerase. A shorter randomized region was justified for the RB69 RegA selection by the established observation that RNA structure is not a component of RegA binding sites (9,12,18). Further, RB69 RegA-His_6 (20) immobilized on Ni-NTA agarose was an effective matrix for partitioning

RegA-RNA complexes from unbound RNA. Nitrocellulose, although widely used as a partitioning matrix in SELEX (11), has sufficient RNA-binding capacity that numerous negative selections and additional rounds of SELEX are usually required (35).

T4 RegA clearly binds the T4 *44* RNA in the absence of discernable RNA structure. The program MFold (36) predicts no stable secondary structure in any of the full-length RB69 RegA-selected, consensus or truncated RNAs. RB69 RegA, like the T4 protein, recognizes RNA in a structure-independent manner, and primary sequence is substantially responsible for RNA recognition and binding site selection. However, this does not preclude the possibility that RegA proteins bind RNA sequence with a propensity for the 2' *endo* conformation (9), or that upon binding the RNAs assume a specific, protein-induced conformation (37).

Biological T4 RNAs that bind RegA were identified in translational repressor assays and naturally contain an AUG. Frequently, the AUG is adjacent to 5' U residues, or is adjacent to AUUA (3,12,28). SELEX experiments with T4 RegA generated a consensus sequence 5'-AAAAUUGUUAUGUAA-3' that is primarily AU-rich RNA, with an AUG adjacent to UAA near the 3' end of the site (12).

RNA protection data indicated that RB69 RegA bound RB69 RNAs at sites immediately upstream of, but not including, the AUG (19). Independent RNase protection experiments with RB69 RegA (18) identified RegA-bound sequences that do not contain an AUG over 28 (T4 *rpbA* RNA) and 41 (RB69 *regA* RNA) nucleotides, although the minimal binding sequence and site size for RB69 RegA are only nine nucleotides. Therefore, through

differential site selection or RegA protein-protein interactions (18), RB69 RegA does bind RNA that lacks an AUG.

Our RB69 RegA SELEX experiments revealed a selected, bulk pool RNA sequence of $(UAA)_n$. All but five of the individual clones sequenced contained the recurring UAA sequence. Those that had an interrupted UAA pattern did not bind the purified repressor and apparently survived partitioning. Two identical RNAs that bound RB69 RegA contained the 3' terminal AUG, but other binding species lacked an AUG. All sequences showed a propensity for an A-rich tract at the 5' end of the selected interval (**Table 2**). The selection and gel shift results (Figures 2 & 3) show that RB69 RegA binds $(UAA)_n$ alone, with an affinity that is comparable to both $(UAA)_4$-3'AUG and $(UAA)_3$-5'AUG. Together, the T4 (12) and our RB69 RegA SELEX results suggest a site comprised of a 5' A-rich tract, followed by an AU-rich region, and then the variably occurring 3' AUG. RB69 RegA SELEX RNAs display UAA triplets in the AU-rich region. As noted above, the consensus T4 SELEX RNA displayed a UAA triplet immediately 3' of the selected AUG (12).

The gel shift results using the two RB69 *regA* RNAs also support the selection results. In the context of the biological RB69 *regA* site it appears that the RegA protein requires a site greater than the initial 5' 28 nucleotides protected in nuclease protection assays (18). Five rounds of selection produced RNAs that bound RB69 RegA with K_d*app* values similar to the larger *regA*-78 RNA, while the shorter *regA*-28 RNA showed no affinity for RegA. The

selection of short AU-rich RNAs with affinities similar to the 78 nucleotide biological *regA* site (containing only one 5' AUG) further suggests that an AUG is not a prerequisite for RB69 RegA binding to RNA.

Densely coding genomes of T-even type phages have a high occurrence of juxtaposed stop and start codons of adjacent genes. Genes on the same transcript are often translationally coupled; ribosomes translating the 5' gene reinitiate at the 3' gene (38) (2). In T4, and presumably in RB69, UAA is preferentially used as the translational stop codon over UGA and UAG (frequencies of 0.31, 0.16 and 0.03, respectively; (39)). The SELEX RNAs for T4 and RB69 RegA show a striking similarity to the start sites of translationally coupled genes. There is a suboptimal, A-rich Shine-Dalgarno sequence, a UAA stop codon (sometimes occurring in tandem, sometimes a few nucleotides 5' of the AUG, and at times just 3' of the AUG), and the AUG. RB69 RegA SELEX results indicate that the initiator codon AUG is less important for binding than is the UAA stop codon.

One outcome of SELEX experiments is the ability to use selected sequences to scan genomes for similar patterns. In pattern searches of the RB69 genome, 5'-AUAAUAUG-3', which is closely related to the SELEX sequence, was found in the mRNAs for genes *62* and *rpbA*. Gel shift analysis confirmed that RB69 RegA binds this sequence (Figure 3). For gene *62*, the AUAAUAUG includes the upstream *44* translation stop codon and the downstream *62* start codon (Figure 4; (40)). Although *rpbA* was inferred not to be coding in phage RB69 due to sequence divergence and stop codons occurring in all reading frames, the RegA binding

site on this RNA overlaps a TIR-like sequence that shares many features with translationally coupled genes and is just 3' of the expressed Orf *45.2*. The TIR and RegA site of RB69 "*rpbA*" may indeed be a genome relict, of a once functional gene.

Following our initial pattern searches, the complete RB69 genome became available. This allowed the use of the RB69 RegA SELEX sequence logo to examine the RB69 genome with the makewalker program (25). Previous work has shown that if the goal of a SELEX experiment is to learn about biological binding sites, sequence logos can be used to guide the experiment (41). Of the 26 sites that were initially uncovered, only 10 were inferred to be transcribed into RNA. Position 6151 in the RB69 genome lies directly between *comC-**a*** and the hypothetical T4 orf *39.2*. The proximity of these two genes (separated by 6 nucleotides) indicates they may be translationally coupled. Table 4 shows the sequence at this site which includes the stop codon of *comC-**a*** and the start codon of *39.2*. Future study to examine the relationship of expression between *comC-**a*** and the hypothetical protein could provide further insight into the action of RB69 RegA. Additional sites uncovered are located in phage structural proteins (gp50 head completion protein, major capsid protein, gp24 head vertex protein, and gp29 tail tube assembly protein). A function for RB69 RegA binding at these sites could prove informative.

Conclusions

We suggest the following for the nature of RegA binding sites to accommodate their lack of structure, low sequence conservation, and broad distribution in phage genomes. First, the absence of RNA structure in RegA sites reflects the trend for prokaryotic TIRs to be devoid of structure for efficient ribosome recognition (2). A low occurrence of C residues (which if present inhibit RegA binding) contributes to the absence of structure. RB69 RegA binds to poly-A as well as poly-U, however, the ability to bind to poly-C is not present. Second, the occurrence of RegA sites at adjacent genes requires a binding strategy that neither perturbs the end of the 5' gene nor the start of the 3' gene. These sites contain UAA (as the stop codon) and predominate at genes that are translationally coupled. RegA bound at these sites would serve to reduce independent translation initiation of the 3' gene. Ribosomes traversing to the site from the 5' gene would be affected by RegA bound at the distal junction. By binding at the 3' ends of RNA coding regions that contain UAA sequences, RegA may affect the kinetics of ribosome elongation and release, in addition to its recognized affects on translational repression. In these instances, RegA can be viewed as a translational repressor. However, if RegA bound at UAAs prevents release factor (RF) from binding the stop codon, diminished ribosome release may promote reinitiation and translation of the distal gene for a translationally coupled gene pair. Enhanced protein synthesis by RegA, from both phage and host mRNAs, has been noted (3,28). It appears that RB69 RegA mediates a form of protein-dependent "recoding" (42), as exemplified by its affect at the translational coupling site of the

44/*62* clamp loader subunit genes that are located on the same mRNA and immediately upstream of *regA.*

References

1. Okabe, M., Imai, T., Kurusu, M., Hiromi, Y., and Okano, H. 2001. Translational repression determines a neuronal potential in Drosophila asymmetric cell division. Nature. 411: 94-8.

2. Gold, L. 1988. Posttranscriptional regulatory mechanisms in Escherichia coli. Annu Rev Biochem. 57: 199-233.

3. Miller, E. S., Karam, J. D., and Spicer, E. K. 1994. Control of translation initiation: mRNA structure and protein repressors, vol. . ASM Press, Washington, DC.

4. McPheeters, D. S., Stormo, G. D., and Gold, L. 1988. Autogenous regulatory site on the bacteriophage T4 gene 32 messenger RNA. J Mol Biol. 201: 517-35.

5. Andrake, M. D., and Karam, J. D. 1991. Mutational analysis of the mRNA operator for T4 DNA polymerase. Genetics. 128: 203-13.

6. Tuerk, C., and Gold, L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science. 249: 505-10.

7. McPheeters, D. S. 1994. Rapid bacteriophage T4 mRNA preparations and the primer extension assay, vol. . ASM Press, Washington, DC.

8. Winter, R. B., Morrissey, L., Gauss, P., Gold, L., Hsu, T., and Karam, J. 1987. Bacteriophage T4 regA protein binds to mRNAs and prevents translation initiation. Proc Natl Acad Sci U S A. 84: 7822-6.

9. Szewczak, A. A., Webster, K. R., Spicer, E. K., and Moore, P. B. 1991. An NMr characterization of the regA protein-binding site of bacteriophage T4 gene 44 mRNA. Journal of Biological Chemistry. 266: 17832-17837.

10. Ellington, A. D., and Szostak, J. W. 1990. In vitro selection of RNA molecules that bind specific ligands. Nature. 346: 818-22.

11. Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M. 1995. Diversity of oligonucleotide functions. Annu Rev Biochem. 64: 763-97.

12. Brown, D., Brown, J., Kang, C., Gold, L., and Allen, P. 1997. Single-stranded RNA recognition by the bacteriophage T4 translational repressor, regA. Journal of Biological Chemistry. 272: 14969-14974.

13. Kang, C.-H., Chan, R. K., Berger, I., Green, L., Lockshin, C., Gold, L., and Rich, A. 1995. Crystal structure of the T4 regA translational regulator protein at 1.9 A resolution. Science. 268: 1170-1173.

14. Webster, K. R., Keill, S., Konigsberg, W., Williams, K. R., and Spicer, E. K. 1992. Identification of amino acid residues at the interface of a bacteriophage T4 regA protein-nucleic acid complex. J Biol Chem. 267: 26097-103.

15. Gordon, J., Sengupta, T. K., Phillips, C. A., O'Malley, S. M., Williams, K. R., and Spicer, E. K. 1999. Identification of the RNA binding domain of T4 RegA protein by structure-based mutagenesis. J Biol Chem. 274: 32265-73.

16. Jozwik, C. E., and Miller, E. S. 1992. Regions of bacteriophage T4 and RB69 RegA translational repressor proteins that determine RNA-binding specificity. Proceedings of the National Academy of Sciences. 89: 5053-5057.

17. Russell, R. L., and Huskey, R. J. 1974. Partial exclusion between T-even bacteriophages: an incipient genetic isolation mechanism. Genetics. 78: 989-1014.

18. Jozwik, C. E., and Miller, E. S. 2001. RNA-binding properties of the bacteriophage RB69 RegA translational repressor protein. Submitted. .

19. Sengupta, T. K., Gordon, J., and Spicer, E. K. 2001. RegA proteins from phage T4 and RB69 have conserved helix-loop groove RNA binding motifs but different RNA binding specificities. Nucleic Acids Res. 29: 1175-84.

20. Allen, S. V., and Miller, E. S. 1999. RNA-binding properties of in vitro expressed histidine-tagged RB69 RegA translational repressor protein. Anal Biochem. 269: 32-7.

21. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. 1994. Current Protocols in Molecular Biology, . John Wiley & Sons, Inc., New York.

22. Tuerk, C. 1994. In vitro selection and amplification of protein-bound RNA ligands, vol. . ASM Press, Washington, DC.

23. Carey, J. 1991. Gel Retardation. Methods Enzymol. 208: 103-117.

24. Schneider, T. D. 1996. Reading of DNA sequence logos: prediction of major groove binding by information theory. Methods Enzymol. 274: 445-55.

25. Schneider, T. D. 1997. Sequence walkers: a graphical method to display how binding proteins interact with DNA or RNA sequences. Nucleic Acids Res. 25: 4408-4415.

26. Miller, E. S., Winter, R., Campbell, K. M., Power, S., and Gold, L. 1985. Bacteriophage T4 *regA* protein: purification of a translational repressor. Journal of Biological Chemistry. 260: 13053-13059.

27. Karam, J., Gold, L., Singer, B. S., and Dawson, M. 1981. Translational regulation: identification of the site on bacteriophage T4 rIIB mRNA recognized by the regA gene function. Proc Natl Acad Sci U S A. 78: 4669-73.

28. Miller, E. S., Karam, J., Dawson, M., Trojanowska, M., Gauss, P., and Gold, L. 1987. Translational repression: biological activity of plasmid-encoded bacteriophage T4 RegA protein. Journal of Molecular Biology. 194: 397-410.

29. Unnithan, S., Green, L., Morrissey, L., Binkley, J., Singer, B., Karam, J., and Gold, L. 1990. Binding of the bacteriophage T4 regA protein to mRNA targets: an initiator AUG is required. Nucleic Acids Research. 18: 7083-7092.

30. Webster, K. R., Adari, H. Y., and Spicer, E. K. 1989. Bacteriophage T4 regA protein binds to the Shine-Dalgarno region of gene 44 mRNA. Nucleic Acids Res. 17: 10047-68.

31. Devereux, J., Haeberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12: 387-95.

32. Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symposium Series. 41: 95-98.

33. Phillips, C. A., Gordon, J., and Spicer, E. K. 1996. Bacteriophage T4 regA protein binds RNA as a monomer, overcoming dimer interactions. Nucleic Acids Research. 24: 4319-4326.

34. Morris, K. N., Jensen, K. B., Julin, C. M., Weil, M., and Gold, L. 1998. High affinity ligands from in vitro selection: complex targets. Proc Natl Acad Sci U S A. 95: 2902-7.

35. Hirao, I., Madin, K., Endo, Y., Yokoyama, S., and Ellington, A. D. 2000. RNA aptamers that bind to and inhibit the ribosome-inactivating protein, pepocin. J Biol Chem. 275: 4943-8.

36. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol. 288: 911-40.

37. Weeks, K. M. 1997. Protein-facilitated RNA folding. Curr Opin Struct Biol. 7: 336-

38. Napoli, C., Gold, L., and Singer, B. S. 1981. Translational reinitiation in the rIIB cistron of bacteriophage T4. J Mol Biol. 149: 433-49.

39. Nakamura, Y., Gojobori, T., and Ikemura, T. 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res. 28: 292.

40. Yeh, L.-S., Hsu, T., and Karam, J. D. 1998. Divergence of a DNA replication gene cluster in the T4-related bacteriophage RB69. Journal of Bacteriology. 180: 2005-2013.

41. Shultzaberger, R. K., Bucheimer, R. E., Rudd, K. E., and Schneider, T. D. 2001. Anatomy of Escherichia coli ribosome binding sites. J Mol Biol. 313: 215-28.

42. Baranov, P. V., Gesteland, R. F., and Atkins, J. F. 2002. Recoding: translational bifurcations in gene expression. Gene. 286: 187-201.

Table 2.5 SELEX conditions using immobilized RB69 RegA-His_6

Round[a]	pmoles RNA	pmoles RegA-His_6	[NaCl] wash	elution	preselection
1	40	2	250 mM	phenol/$CHCl_3$	+
2	40	2	250 mM	phenol/$CHCl_3$	+
3	40	1	250 mM	phenol/$CHCl_3$	-
4	40	1	250 mM	phenol/$CHCl_3$	-
5	40	1	250 mM	phenol/$CHCl_3$	-

a, SLXN14 RNA (40 pmoles; preselected over Ni-NTA in rounds 1 & 2) was applied to RB69 RegA-His_6 (1 or 2 pmoles) bound to Ni-NTA. Complexes were washed in binding buffer containing 250 mM NaCl to remove unbound RNA. Bound RNA was extracted from the matrix with phenol/chloroform followed by ethanol precipitation.

Table 2.6 Cloned sequences after five rounds of SELEX

Clone	UAA Containing
SA5-1	aAUAAUAAUAAUAAUg
SA5-2	aAUAAUAAUAAUAAUg
SA5-4	aAAAUAAUAAUAAAAg
SA5-5	aAAAUAAUAAUAAAAg
SA5-6	aAGAAUAAUAAUAAGa
SA5-7	aUAAUAAUUAUAAGAa
SA5-9	aAAGAAUAAUAGUAGg
SA5-13	aUAAUAAUAAUAAUAc
SA5-17	aUAAUAAUAAGAAUAa
SA5-19	aGAAUAAUAAUAAGAg
SA5-20	aUAAUAAUAAUAAGAg
	UAA + C Containing
SA5-3	aAAACAAUAACAAUAa
SA5-8	aUGAUAAUAACAAUAg
SA5-10	aUAAUAAUAACAAUAg
SA5-11	aUAAUAAGAACAAUAg
SA5-14	aCAAUAAUAAUAAUAg
SA5-16	aGAAUAAGAAUAACAa
SA5-21	aAUAAGAACAAUAGGc
RB69-Con	aUAAUAAUAAUAAUAg

The 14 nucleotide randomized region plus one nucleotide from the 5' and 3' fixed regions (lower case) were aligned. Sequences have been divided into two distinct groups (UAA containing and UAA + C sequences). The $(UAA)_n$ repeat present in most RNAs emerges in the RB69 "consensus". UAA containing sequences were used to generate the sequence logo. Tested UAA + C sequences do not bind to RB69 RegA.

Table 2.7 K_d*app* values for RB69 RegA on defined RNAs

Name	RNA Sequence	K_d*app*
T4 *44*	gggAAUGAGGAAAUUAUGAUUAC	1.60 X 10^{-8}
RB69 *44*	gggAAUGAGGAAAAUUACAUGAUUAC	1.22 X 10^{-7}
RB69 *regA*-78	GCAUUGGAAUGGUAAAAAUGAUUGAAAUUAAAUUGAAGAACCCUGAAGAUUUUCUGAAA	1.65 X 10^{-7}
UAA_3	gggAAAAUAAUAAUAAA	2.81 X 10^{-7}
$(UAA)_4$-3'AUG	gggAAUAAUAAUAAUAAUG	2.95 X10^{-7}
$(UAA)_4$-3'AAG	gggAAUAAUAAUAAUAAAG	3.37 X 10^{-7}
$(UAA)_3$-5'AUG	gggAUGAAUAAUAAUAAUA	3.50 X 10^{-7}
RB69 *62*	gggAAAUAAUAUGAAA	4.37 X10^{-7}
SA5-1	...guagcgaAUAAUAAUAAUAAUGucauc...	+ + +
RB69-Con	...guagcgaAAAUAAUAAUAAUAGucauc...	+ + +
T4-Con	gggAAAAUUGUUAUGUAA	+ +
$(UAA)_4$-3'AUC	gggAAUAAUAAUAAUAAUC	+ +
UAA_2	gggAAAAAUAAUAAAAA	+ +
T4 *44*-C	gggAAUCAGGAAAUUAUCAUUAC	-
SA5-7	...guagcgaUAAUAAUUAUAAGAAucauc...	-
SA5-3	gggAAACAAUAACAAUAAA	-
SA5-10	gggUAAUAAUAACAAUA	--
RB69 *regA*-28	gggCAUUGGAAUGGUAAAAAUGAUUGAA	--

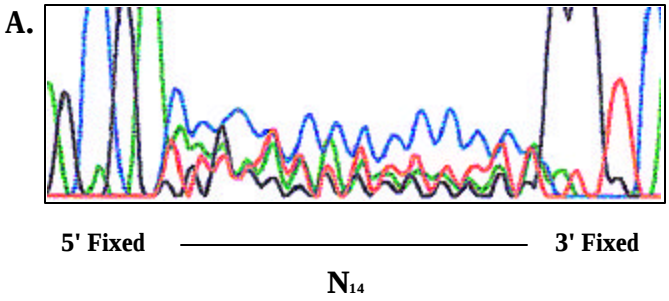
Apparent K_d values (K_d*app*), defined as [RegA] necessary to produce 1/2 maximal shift, were determined as described in *Material and Methods*. Complete RNA sequences are shown, except for SA5-1, SA5-7 and RB69-Con, which are 61 bases long, with additional bases originating from the SLXN14 template. For RNAs with no K_d*app* value gel shifts are indicated using the + notation, + + + indicating a greater percentage shift than + + > - (no shift).

Table 2.8 Potential RB69 RegA binding sites in RB69 genome

Coordinate[a]	Strand[b]	RNA Sequence (5' to 3')[c]	Postion[d]	Protein
6151	-	aaggaa**uaauaa**ugccguu	S,E	ComC-alpha & T4 orf 39.2
47424	-	ucuga**uaauaau**auccauc	M	Recombinant Endonuclease VII
72965	-	aguuuacca**uaauaa**uguc	E	Hypothetical T4 Protein orf e.8
80649	-	guauaaa**uaauaa**uacuau	M	Head Completion protein gp50
106784	+	auuauaaa**uaauaa**uguaa	S	Prohead Core Protein Protease Precursor gp21
109946	+	caaaua**uaauaa**uaucguc	M	Major Capsid protein gp23
111970	+	uauaaa**uaauaa**uaucaaa	M	Head vertex protein gp24
124320	+	uggcaua**uaauaa**uacagg	M	Tail Tube Assembly Protein gp29
147309	-	gauua**uaauaa**uaaaguuc	M	DNA helicase loader protein gp59
166614	-	gcagaa**uaauaa**uuggga	E	rIIB

a) Coordinate in RB69 genomic sequence (Karam, J. 2002 at http://www.tmc.tulane.edu/sif/t4genome.html). b) DNA strand, related to deposited sequence. c) Sequence and surrounding nucleotides. Underlined region indicates start codon for T4 hypothetical orf. 39.2 d) Location of identified sequence at start (S), middle (M), or end (E) of gene.

Figure 2.1 Sequence of SELEX N14 DNA pools.



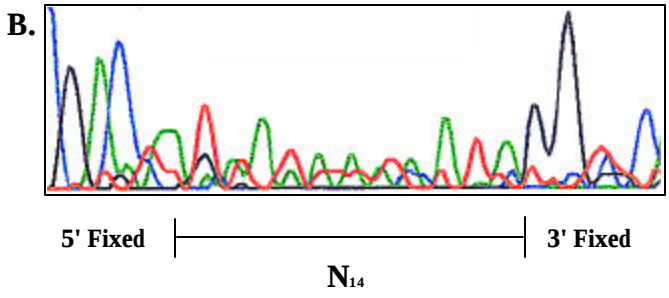


Figure 2.2 Gel shift analysis of RB69 RegA binding to selected and variant RNA.

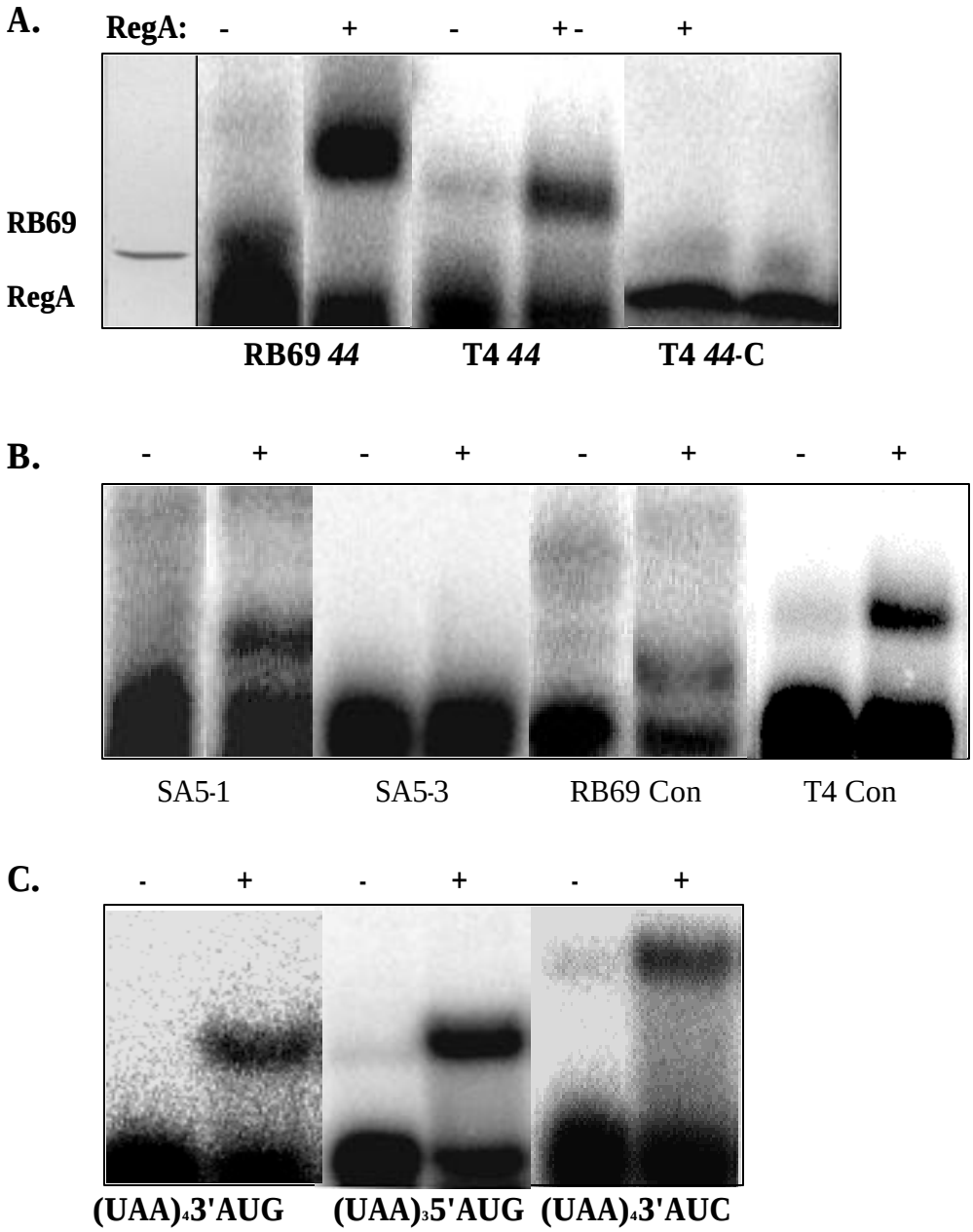
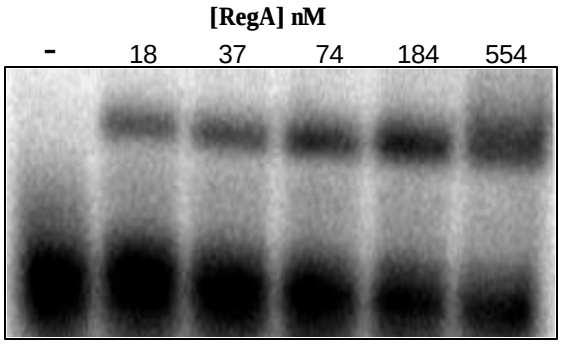
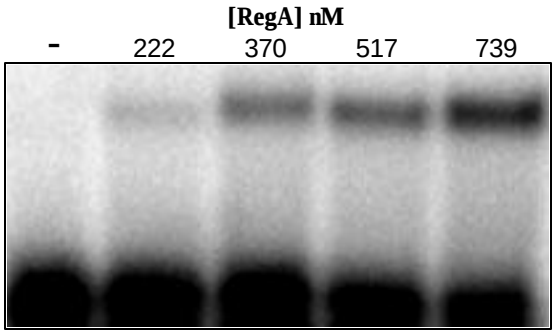


Figure 2.3 K_a*app* analysis of RB69 RegA-RNA interactions.

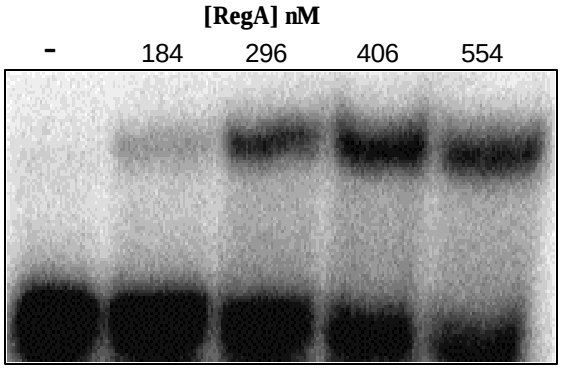
A. T4 *44*



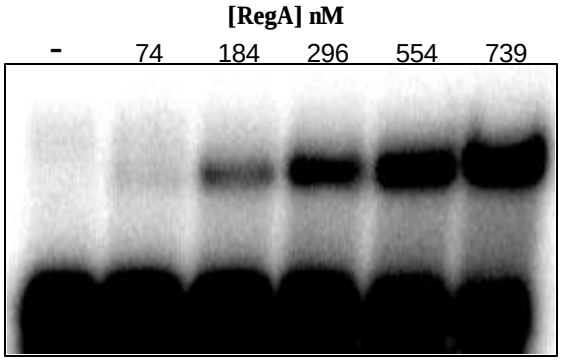
B. RB69 *62*



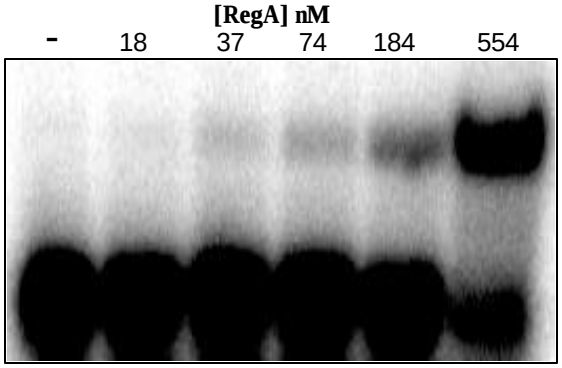
C. (UAA)$_3$



D. (UAA)$_3$-5'AUG



E. (UAA)$_4$-3'AUG



F. Bjerrum Plot

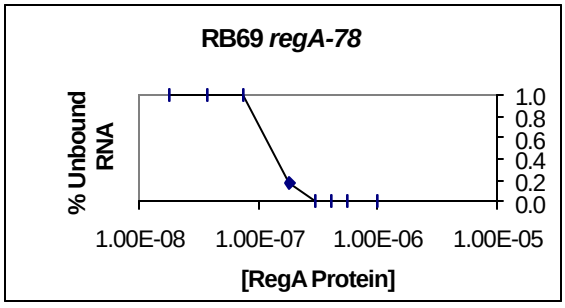


Figure 2.4 Sequence logo.

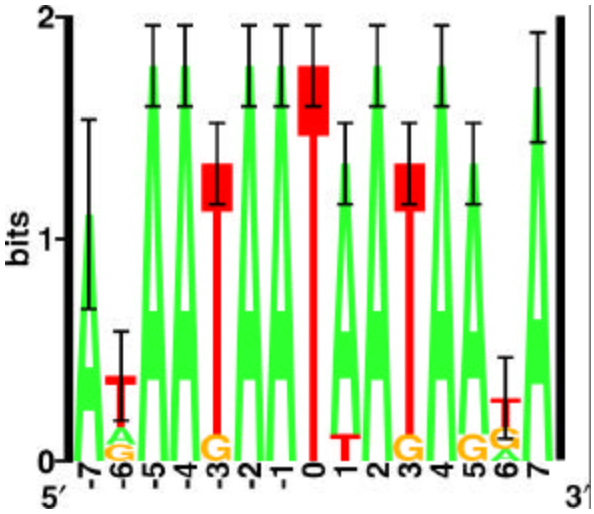


Figure Legends

Figure 1. Sequence of SELEX N14 DNA pools. Sequencing reactions on SLXN14 DNA at the start of round one (A), and following round five (B). The blue trace in panel A indicates that C residues had a slightly higher prevalence than A (green), G (black), or T (red). In panel B, the absence of C residues and the prevalence of alternating A and T residues is apparent.

Figure 2. Gel shift analysis of RB69 RegA binding to selected and variant RNA. (A) Purified RB69 RegA protein used in the binding assays was run on 12.5% SDS-PAGE. All other lanes are RNAs in binding reactions with 1.5 μM of RB69 RegA protein, followed by electrophoresis on 12% nondenaturing gels (see *Materials and Methods*). RB69 *44* and T4 *44* RNAs are natural RegA binding sites; T4 *44*-C is the T4 *44* RNA with the Gs of two internal AUG triplets substituted with C. (B) SA5-1, SA5-3, and RB69-Con RNAs are selected full length, 62 base RNAs containing the transcribed 5' and 3' fixed regions. T4-Con. is a 18 base RNA selected in previous experiments**.** (C) All other RNAs (modified versions of SA5-1) are 19 bases long and contain the 14 base random region, two bases from the 3' fixed region, and three 5' Gs used in T7 transcription.

Figure 3. K_d*app* analysis of RB69 RegA-RNA interactions. A-E) The binding of RB69 RegA to each RNA was evaluated in gel shift assays over a range of protein concentrations. F)

Bjerrum plots (Carey, J. 1991) were produced and dissociation constants (K_d*app*) calculated as described in *Materials and Methods*. RNA sequences are shown in Table 3. RB69 *62* RNA is a near match to the $(UAA)_2AUG$, and was identified in sequence searches of RB69 GenBank entries.

Figure 4. **Sequence logo generated using the 11 UAA containing sequences (see Figure 3)**. The logo contains 14 bases from the randomized region and one base from the 5' fixed region (position -7 in the logo). R_{seq} = 21.27; (see Schneider et al., 1986).

Chapter 3

Optimization of SELEX Using Bead Immobilized Bacteriophage R17

Introduction

Diseases caused by waterborne microbial pathogens are increasing at a high rate and their significance can, at best, only be estimated (1). Waterborne transmission is common for viral, bacterial, and protozoal pathogens associated with drinking water (1). Public health requires that there be adequate indicators of fecal pollution. Fecal contamination has commonly been monitored by screening for the presence of *Escherichia coli* . *E. coli* is found in all mammal faeces, at concentrations up to 10^9 (2). Comparison with other fecal indicators shows that *E. coli* is a far superior organism for the detection of fecal contamination (2).

In addition to *E. coli*, fecal contamination of surface and ground water has also been indicated by the presence coliphages. Some of these coliphages are *E. coli* F^+ specific, and have RNA genomes (3). Examples of F^+ RNA coliphages include R17, Qβ, and MS2. These bacteriophage have been isolated from human feces, animal feces, and sewage and are serologically placed into four groups (I-IV) (3). It has been noted that group I only occurs in animal feces (e.g. cows and pigs) (3).

There continues to be increasing demand for quantitative detection methods for fecal contamination, with increased sensitivity and faster turn-around time. Molecular detection methods have the ability to meet each of these requirements. A successful method for the detection of viral particles in water samples has been reverse transcription-polymerase chain reaction (RT-PCR). Human caliciviruses cause waterborne outbreaks of gastroenteritis. RT-PCR was used to detect the presence and titre of calcivirus in water samples (4). The results

showed that the limits of detection for calcivirus using RT-PCR were 0.75 to 1.5 plaque forming units (PFU) in 40 liters of water (4). Clearly molecular methods have merit as highly sensitive detection devices.

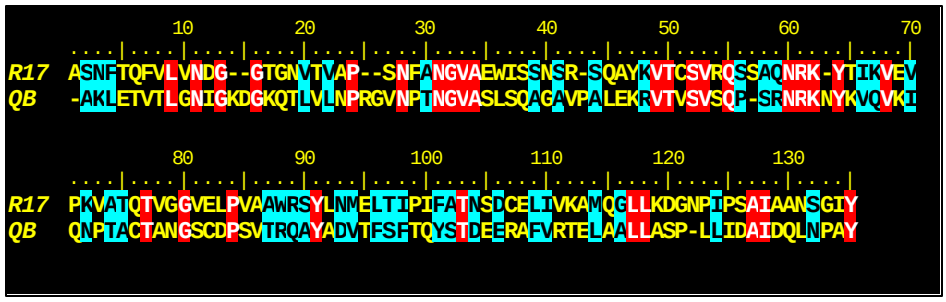
Systematic Evolution of Ligands by EXponential enrichment (SELEX), also called *in vitro* selection, is a procedure which uses "molecular diversity" to evolve nucleic acid molecules (RNA or ssDNA) with unique binding or catalytic capabilities. The SELEX procedure has been applied to numerous targets ranging from small molecules (ATP) (5), to proteins (pepocin) (6), to viral particles (Rous sarcoma virus) (7).

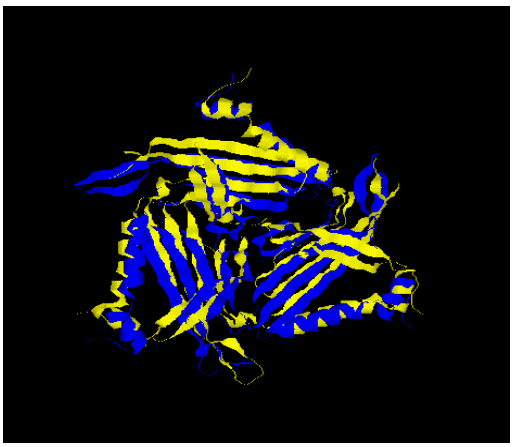
The Rous sarcoma virus (RSV) is an enveloped avian retrovirus displaying glycoproteins on the particle surface (7). SELEX was used to show that, lacking knowledge of the structures of the viral proteins, antiviral RNAs can be isolated in a systematic manner from a large pool of random sequences (7). The experiments confirmed that following 12 rounds of SELEX, the RNA pool was enriched for binding molecules, with a pool K_d of 3 μg of viral protein per ml (7). In addition, some of the selected sequences showed virus neutralizing capabilities (7).

Bacteriophage R17 and Qβ are classified group I coliphages. The virions are spherical, exhibit icosahedral symmetry (T=3), and have a diameter of 26 nm (8). The virions contain one molecule of positive sense RNA ~3,466 nucleotides long (8). The RNA comprises approximately 30% of the total molecular weight of the viral particle (4.2×10^6 Mr) (8). The capsids are composed of 180 units of the coat protein, arranged in 60 identical triangular units

and exhibit icosahedral symmetry (T=3). The capsids also contain a single copy of the A protein which is required for maturation of the virion, and for pilus attachment (8). Bacteriophage R17 infects *E. coli* that are found in the intestinal tract of swine, while bacteriophage Qβ infects *E. coli* found primarily in the intestinal tract of humans (3). R17 and Qβ coat protein sequences show 41% similarity (16) (**Figure 3.1**). The three dimensional structures of the coat proteins of both R17 and Qβ are very similar and can be superimposed upon each other (**Figure 3.1**). An environmental diagnostic capable of screening for R17, at the same time not recognizing Qβ, would be a useful indicator of swine fecal contamination of surface waters.

SELEX experiments were carried-out using bacteriophage R17 as the primary target and pools of RNA randomized at either 40 positions or at 25 positions. Experimentation followed standard selection protocols (9,10,11), with useful insights into the efficacy of different partitioning methods becoming a major outcome of the experiments.





R17-Blue
Qb-Yellow

Figure 3.1 Alignment and comparison of R17 and Qβ coat proteins. The top panel shows the alignment of the two proteins highlighting the similar and identical amino acids (16). The lower panel shows the secondary structures of each protein superimposed upon one another using RasMol (15). The domains clearly fold into similar structures making a selection an excellent method for distinguishing between the two bacteriophage.

Materials and Methods

Strains and Media. *E. coli* XL1Blue (EM86) (*recA1*, *lac-*, *endA1*, *gyrA96*, *thi*, *hsdR17*, *supE44*, *relA1/F' proAB*, *lacIQ*, *LacZΔM15*, *Tn10*) was obtained from Strategene (La Jolla, CA). Procedures for growth of bacteria in LB media have been described (12). Bacteriophage R17 and Qβ were obtained from the laboratory at a titre of 10^{12} plaque forming units (PFU) per milliliter. Ribomax T7 *in vitro* transcription system, *Taq* DNA polymerase, and AMV reverse transcriptase were obtained from Promega Corp. and used per the manufacturers recommendations.

Growth and purification of bacteriophage R17 and Qβ. One liter of LB media was inoculated with a fresh overnight culture of *E. coli* XL1Blue and grown to mid-log phase (10^8 cells or approximately 40 Klett units) at 37° Celsius. At mid-log phase the cells were inoculated with bacteriophage R17 or Qβ at a multiplicity of infection (MOI) of 1.0 either from phage lysate or from an individual plaque. Cells were then grown overnight at 37° C to allow for maximal phage infection. The cleared lysate was then treated with 1 ml of chloroform for 15 minutes to release any bacteriophage still within the cells. The chloroform was then removed and the lysate was spun down at 4,300 X g for 15 minutes. The lysate was then removed from the pellet and stored at 4° C. The lysate was then precipitated by adding NaCl to a concentration of 0.5 M followed by the addition of (polyethelene glycol) PEG 8000 to 10% by volume with stirring at 4° C overnight. Following the overnight precipitation the

lysate was centrifuged at 4,000 X g for 30 minutes, the supernatant poured off, and the pellet resuspended in 5 ml LB medium. This concentrated phage suspension was then layered over a sucrose step gradient consisting of a bottom layer that was 52.5% and a top layer that was 12.5% sucrose. The samples were then spun at 100,000 X g for 40 minutes. Following centrifugation top lighting showed a "bluish" phage band present at the gradient interface. This band was then removed using a 20 gauge needle and syringe. The phage sample was then placed in dialysis tubing with a molecular weight cutoff of 12,000 Da and dialyzed overnight into phage binding buffer (150 mM PBS containing 2.5 mM KCl, 8 mM Na_2HPO_4, 1 mM K_2HPO_4, 150 mM NaCl, and 2.5 mM $MgCl_2$). Following dialysis the phage samples were removed, titred, and stored at 4° C. Standard methods were used to determine the titre of the lysate on an overlay lawn of XL1-Blue (F^+).

Template construction. Two DNA libraries were constructed, with variable regions consisting of 25 or 40 nucleotides in length, using PCR with a 5' T7 fixed primer and a 3' fixed primer that annealed to the template as shown in **Figure 3.2**.

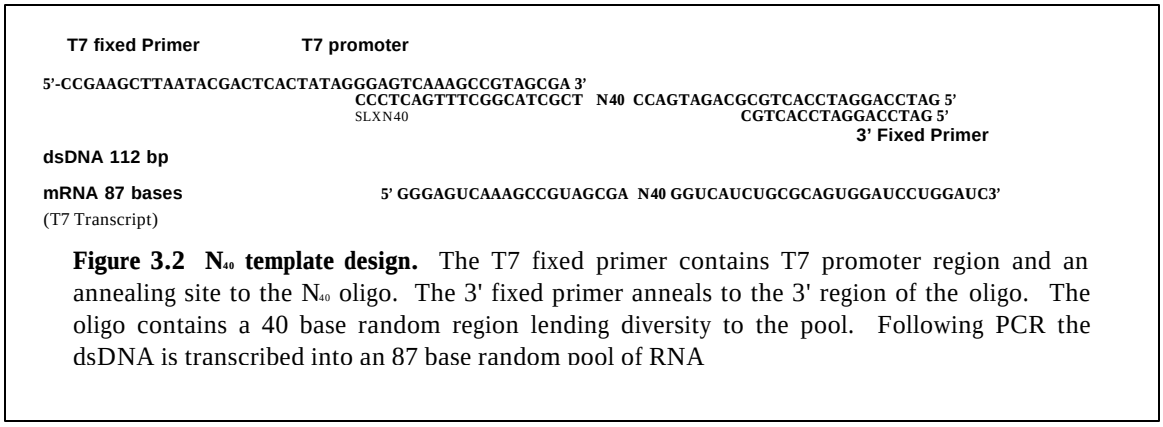


Figure 3.2 N_{40} template design. The T7 fixed primer contains T7 promoter region and an annealing site to the N_{40} oligo. The 3' fixed primer anneals to the 3' region of the oligo. The oligo contains a 40 base random region lending diversity to the pool. Following PCR the dsDNA is transcribed into an 87 base random pool of RNA

The random regions were created during synthesis of the ssDNA oligonucleotide by mixing the phosphoramidite A, T, G, and C reagent amounts to allow equal probability of each nucleotide for each position. The pool diversity is a function of the random region with 4^n (n = number of variable positions) equaling the number of potential sequences. The initial PCR reactions were carried out and optimized following standard protocols (13) (see **Table 3.1**) in 1X *Taq* buffer (10mM Tris-HCl (pH 8.0 at 25°C), 50mM KCl, and 0.1% Triton X-100). PCR was performed for 7 cycles at 93° C for 30 sec.; 55 ° C for 30 sec.; 72° C for 1 min. to generate the initial dsDNA pool. The PCR products were separated on a 8% non-denaturing polyacrylamide gel and visualized by ethidium bromide staining and UV illumination. The PCR products were gel purified for all selections and eluted into dH_2O and ethanol precipitated.

Generation of large pool of random RNA. The 5' fixed region of the dsDNA contains the recognition sequence used by T7 RNA polymerase. The pool of RNA was generated using the Ribomax T7 *in vitro* transcription kit (Promega Corp.) (see **Table 3. 1**). Each 25 µl reaction contained 10% of the previous PCR reaction in 80mM Hepes-KOH, pH 7.5, 24mM $MgCl_2$, 40mM DTT. The T7 enzyme mix provided in the kit consists of T7 RNA polymerase, RNasin ribonuclease inhibitor, and yeast inorganic pyrophosphatase. Reactions were incubated at 37°C for 2 hours. To remove any residual "contaminating" dsDNA, from the reaction the tube was heated to 93°C for 2 minutes, followed by the addition of 3 µl ribonuclease free-DNase (1 unit/µl). After mixing, the tube was placed at 37°C and allowed to slow cool for 30

minutes. The resulting transcripts were analyzed on an 8% denaturing polyacrylamide polyacrylamide gel (12) and quantified spectrophotometrically ($1A_{260}$ = 40 μg/ml; Concentration = A_{260} X Factor) .

Table 3.1 Optimized reaction conditions. Panels a, b, and c show the optimized conditions (initial/final concentrations and volumes) used in the SELEX experiments

a. PCR Reaction Conditions:

	Initial Conc.	Volume	Final Conc.
10X dNTP	5mM	5µl	0.5mM
10X Taq Buffer	10X	5µl	1X
3' Fixed Primer	50mM	6µl	300pmoles
T7 Fixed Primer	50mM	6µl	300pmoles
Template	15mM	3µl	15pmoles
MgCl	25mM	15µl	7.5mM
dH_2O	----	9.7µl	----
Taq Pol.	5U/µl	0.3µl	1.5U
	Total:	50µl	

b. Transcription Reaction Conditions:

	Initial Conc.	Volume	Final Conc.
T7 5X Buffer	5X	5µl	1X
rNTP's	25mM	7.5µl	7.5mM
linear DNA	≈4mM	5µl	≈20pmoles
RF-dH_2O	----	5µl	----
Enzyme Mix		2.5µl	
	Total:	25µl	

c. Reverse Transcription Conditions:

	Initial Conc.	Volume	Final Conc.
5X AMV Buffer	5X	4µl	1X
AMV RT	8U/µl	0.5µl	4U
3' Fixed Primer	50mM	1µl	50pmoles
dNTP's	5mM	2µl	0.5mM
Template		1µl	
RF-dH_2O	----	12.5µl	----
	Total:	20µl	

Removal of unincorporated ribonucleotides was accomplished by gel purification (12) or through size exclusion chromatography using MicroSpin G-25 spin columns (Amersham Pharmacia Biotech Inc.) following recommended protocols

Reverse transcription-polymerase chain reaction (RT-PCR). During each round of SELEX target-bound RNAs were subjected to RT-PCR to generate dsDNA for the next round of selection. Reverse transcription was performed in reaction buffer consisting of 250 mM Tris-HCl (pH 8.3 @ 25° C), 250 mM KCl, 50 mM $MgCl_2$, 2.5 mM spermidine, and 50 mM DTT. Optimized conditions are given in **Table 3.1**. Total reaction volume was 20 µl and was incubated at 42° C for 45 minutes. Following reverse transcription, 10 µl of the reaction (1/2) was transferred to the PCR. PCR was performed exactly as previously described with number of cycles variable depending on the selection conditions (**Table 3.1**).

Bacteriophage R17:streptavidin bead preparation. Bacteriophage R17 was initially biotinylated using EZ-Link photoactivatable biotin (Pierce Corp.) following the manufacturers protocol. 10^{11} PFU (90 µl) phage was incubated with 100 µl phage binding buffer and 10 µl photoactivatable biotin (100 µg/ml) in a darkroom. The reaction mixture was then exposed to full spectrum light (General Electric plant light) for 5 minutes on ice (distance 20 cm), cooled for 5 minutes and re-exposed for an additional 5 minutes. Following exposure, uncross-linked biotin was removed by size exclusion chromatography using MicroSpin G-25 spin columns (Amersham Pharmacia Biotech Inc.) following recommended protocols. The extent of biotinylation was assayed using a modified western blotting protocol (**Figure 3.4)**.

Biotinylated phage particles were then incubated in 100 µl pre-equilibrated ImmunoPure Immobilized streptavidin agarose (Pierce Corp.) for 1 hour at room temperature. The column was washed 3 times in 1ml phage binding buffer to remove unbound phage particles. The extent of column binding was assayed by titre test and SDS-PAGE to observe the presence of R17 coat protein on the column.

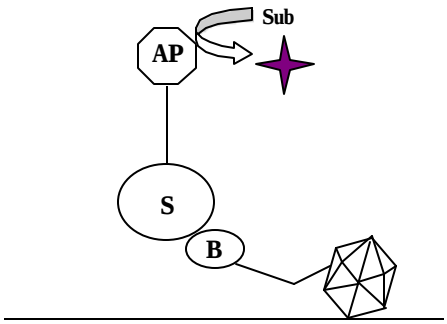


Figure 3.3 Biotinylated phage was bound to nitrocellulose membranes. The nitrocellulose was then blocked with milk (1 g/100 ml) for 30 minutes at room temperature. Nitrocellulose was then washed and incubated with a 1:10,000 dilution of streptavidin-alkaline phosphatase conjugate (Bio-Rad Corp.) at room temperature for 1 hour. Membrane was then washed with phage binding buffer. Membrane was then incubated in alkaline phosphatase conjugate substrate kit (Bio-Rad Corp.) for 30 minutes at room temperature or until color development was complete.

***In vitro* selection with bacteriophage R17.** The SELEX scheme used in the R17 experiments followed standard format and is summarized in (**Figure 1.1**). **Table 3.2 (discussed in results)** outlines the details and progress of the SELEX experiments using both nitrocellulose and immobilized streptavidin:biotinylated phage as different partitioning methods. Briefly, following transcription the random pool of RNA was incubated (in solution

or with beads) with bacteriophage R17 in phage binding buffer (2.5 mM KCl, 8 mM Na_2HPO_4, 1 mM K_2HPO_4, 150 mM NaCl, and 2.5 mM $MgCl_2$) for 30 minutes at room temperature. Bound RNA molecules were then partitioned from unbound molecules either via nitrocellulose filtration or bead centrifugation. Bound RNA was removed from the phage by boiling in elution buffer (7 M urea, 3 mM EDTA, and 100mM sodium citrate) for 5 minutes followed by phenol:chloroform extraction and ethanol precipitation. Concentrated RNA was then subjected to RT-PCR (previously described) and readied for the following round of selection. The stringency of selection was controlled by varying the concentration of bacteriophage R17 in the binding reaction, lowering the amount of R17 to increase the stringency.

RNA binding assay. The extent of RNA binding was determined using nitrocellulose binding assays or column assays. RNA was transcribed from dsDNA made using a random sequence oligo template, 5' primer, and a 3' primer. PCR conditions to create dsDNA for transcription were identical to those used in SELEX (see above). RNA was labeled with [α^{32}P-UTP] via a T7 *in vitro* RNA transcription system (Ribomax, Promega Corp). Each 25 μl reaction contained 10% of the PCR, 80 mM HEPES-KOH, pH 7.5, 24 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, 2.5 mM each rATP, rGTP, and rCTP, 1 mM rUTP, 50 μCi [α32P-UTP], and T7 enzyme mix (T7 RNA polymerase, RNase inhibitor and yeast inorganic pyrophosphatase). Reactions were incubated at 37°C for 2 hr. Entire transcription reactions were run on 8% denaturing PAGE, appropriate size bands were excised, and RNA extracted

overnight into ribonuclease-free water. Following ethanol precipitation, purified RNA was suspended in 15 μl ribonuclease free water. RNA (1.5 μl) was mixed into 188 μl of RNA binding buffer (150 mM PBS), heated to 70°C for 3 min and then cooled to room temperature over 10 min. Single-point assays used 10 nM purified R17 bacteriophage incubated with 115 nM of α^{32}P-UTP-labeled RNA for 20 min at room temperature.

Results

Purification of bacteriophage R17. One important aspect of a successful selection experiment is the purity of the target. Spurious proteins or molecules can cause unwanted parallel selections making it difficult to identify target specific aptamers. In a common bacteriophage lysate there is contamination in the form of cellular debris and growth media components. Increase in purity must be attained to use bacteriophage particles as a SELEX

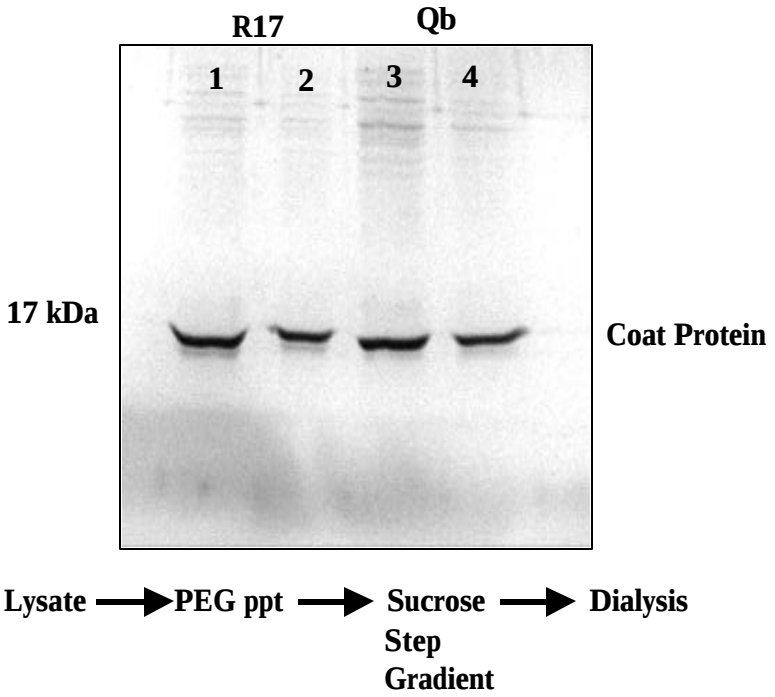


Figure 3.5 SDS-PAGE of purified phage. 12% polyacrylamide gel showing prominent phage coat protein band. 25 µl (2.5 X 10^{11} PFU/ml) loaded in lanes 1 and 3; 20 µl (2 X 10^{11} PFU/ml) loaded in lanes 2 and 4. Note the relative absence of spurious proteins in the purified lysate.

target. Subjecting the "crude" bacteriophage lysate to PEG precipitation, ultracentrifugation, and dialysis was done to help remove contaminants. Titre testing followed purification to quantify the number of viable phage particles present (plaque forming units or PFU/ml). Serial dilutions allowed countable plaques at the 10^{9} dilution for both R17 and Qβ. Plating 100 µl of the 10^{9} dilution resulted in 117 R17 plaques or 1.17 X 10^{12} infectious phage particles per milliliter pure lysate, while Qβ resulted in 141 plaques at the 10^{9} dilution or 1.41 X 10^{12} PFU/ml. The purity of the lysate was assayed via SDS-PAGE, assaying for the presence of phage coat protein. 12% SDS-PAGE gels showed the presence of the coat protein at approximately 14kDa with few other proteins present (**Figure 3.5**).

Optimization of the polymerase chain reaction. Optimization of PCR was completed to generate a "clean" single product band with as little incorrect sized DNA possible. Conditions initially chosen (see materials and methods) resulted in the formation of product, however, the number of cycles of PCR greatly affected the outcome of the reaction. Seven cycles of PCR (**Figure 3.6 a**) generated a suitable product of the appropriate size (112 bp for N_{40} and 94 bp for N_{25}). PCR reactions with cycles greater than 7 consistently generated spurious products that were larger than that desired. PCR reactions in SELEX have a large quantity of initial template to amplify (15 pmoles). Combined with the large random region it is quite possible that more PCR cycles allow for primer miss-annealing into the random region, or even template fixed regions annealing to the random region of other oligos in the

reaction. This would explain the larger product bands present when reactions were cycled more than 7 times.

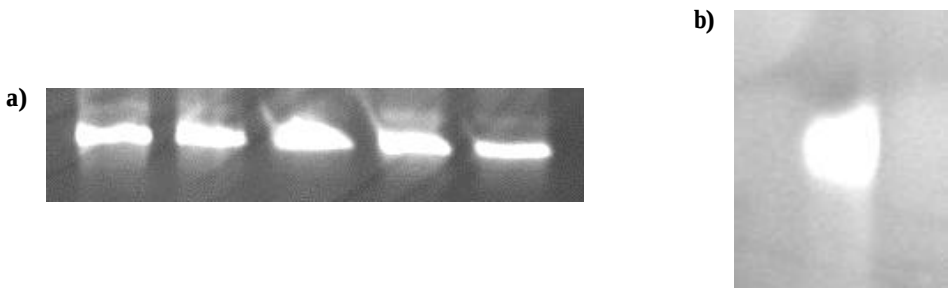


Figure 3.6 PCR and *in vitro* transcription optimization **a)** Optimization of N_{40} PCR. 8% nondenaturing gel showing generation of 112 bp product with limited background banding. Above reactions were pooled and used in first round of nitrocellulose selection. **b)** *in vitro* transcription product from optimized N_{40} PCR. Note large amount of RNA produced (1/25 of transcription reaction run on 8% denaturing polyacrylamide gel.

Optimization of *in vitro* transcription. Generation of large quantities of relatively pure RNA initially relied on the input of "clean" PCR product as template, and is an important factor in the success of a selection experiment. The RNA in a selection should be of uniform length. Competition for binding sites by short terminated transcription products coupled with possible negative interactions between nucleic acid molecules could prevent the generation of high-affinity aptamers. Therefore it is important to generate pure RNA in the transcription reaction. **Figure 3.6b** shows an example of an N_{40} transcription reaction. The Ribomax T7 *in vitro* transcription kit (Promega Corp.) generated large quantities of RNA, with reactions generally producing between 500 ng/µl to 1.0 µg/µl. Reaction times that proceeded longer than 2 hours failed to produce appreciably more product RNA.

R17 SELEX utilizing nitrocellulose partitioning. Initial selection experiments utilized nitrocellulose membranes (HAWP01300 Millipore Inc.) as the method of partitioning phage-bound RNAs from RNA molecules that did not bind phage. **Table 3.2** details the progress of the selection from round to round.

Table 3.2 Selection of RNAs that bind to Phage R17 using N40 Nitrocellulose Partitioning

										Filter Binding Assay	
Round	Pool RNA (M)	RNA ug	Phage R17 (M)	Phage ug	Phage R17 (PFU)	RNA:R17 Ratio	Volume	Filtration (Pre-Post)	PCR	R17 Binding	Filter Binding
0										5.50%	4.10%
1	870nM	5	10nM	5	1.00E+10	87:1	200ul	1\0	11		
2	870nM	5	10nM	5	1.00E+10	87:1	200ul	1\0	11		
3	870nM	5	2nM	1	2.00E+09	435:1	200ul	1\0	11		
4	870nM	5	2nM	1	2.00E+09	435:1	200ul	1\0	11	5.70%	5.50%
5	870nM	5	10nM	5	1.00E+10	87:1	200ul	3\2	22	32.0%	18.0%
6	870nM	5	10nM	5	1.00E+10	87:1	200ul	3\2	15	34.0%	34.0%

Prior to the first round of selection "Round 0" the first pool of RNA was assayed (see materials and methods) to determine the extent of phage R17 binding, as well as nitrocellulose background binding. Nitrocellulose is very good at binding protein (phage coat protein) and should have low affinity for RNA (11). The initial assay utilized 115 nM [^{32}P-UTP] labeled RNA transcripts incubated with 10 nM bacteriophage R17. This single point assay was used throughout the experiment to monitor the progress of the selection. Surprisingly, our membrane assay showed 4% background binding. Such a high initial level of background binding could quickly bias the RNA pool for those that bind nitrocellulose. Therefore, it was necessary to use an aggressive series of negative selections.

The first four rounds of SELEX were carried out with a single pre-nitrocellulose filtration to remove membrane binding RNAs. Prior to incubation with phage R17, RNA was passed through a pre-wetted nitrocellulose membrane and the flow-through collected for later incubation with the phage. The flow-through RNA should contain fewer molecules that have affinity for nitrocellulose.

Following SELEX round 4, a nitrocellulose binding assay was performed to monitor the progress of the experiment. The results showed (**Table 3.2**) that there was no appreciable increase in the amount of RNA that bound to phage compared to round 0. Nitrocellulose background binding continued to be substantial, increasing to 5.5%. Clearly, a single pre-selection negative filtration to remove membrane binding RNAs was not sufficient. In subsequent rounds, additional negative selection was used to remove nitrocellulose binding RNA molecules. To increase the pressure to remove membrane- binding RNAs in rounds 5 and 6, three sequential pre-selection passes were made through nitrocellulose. Following these negative selections the round 5 RNA was then incubated with phage R17 and partitioned. The recovered RNA was then subjected to 2 additional negative selections to remove any remaining RNAs with affinity for nitrocellulose. RT-PCR results showed an increase in the number of cycles needed to generate dsDNA increasing from 11 cycles to 22 cycles, consistent with a lower yield of RNA molecules. Following RT-PCR, RNA binding assay results were promising, showing 32% binding to phage with an 18% background binding to nitrocellulose. The difference between the phage and membrane alone was attributed to

RNAs that bound specifically to bacteriophage R17. The increase in background binding was large, indicating that the pool of RNA still contained a large percentage of molecules with high affinity for nitrocellulose. Nevertheless, RNA binding to Bacteriophage at 32% was sufficiently high enough for increased reduction in background binding to be the main goal in round 6.

Round 6 was carried out exactly like round 5, utilizing a 3/2 pre/post negative selection scheme. The ratio of RNA:R17 was also kept constant to monitor the progress from round 5. RT-PCR was conducted and a reduction to only 15 cycles of PCR was needed to produce the desired dsDNA product. Nitrocellulose binding assay results showed that R17 binding had increased to 34%, however, background binding had also increased to 34%. This result indicated that some characteristic of the RNA pool enabled it to continue to bind to nitrocellulose even under heavy pressure to reduce nitrocellulose binding affinity.

The above results are indicative of numerous attempts to isolate phage-binding RNAs using nitrocellulose partitioning. Different pools of RNA were generated with varying length random regions (N_{40}, N_{30}, N_{14}). Each pool of RNA produced similar results. Initial nitrocellulose binding was never below 3%. Intact bacteriophage R17 particles do not normally bind RNA. This characteristic makes it a difficult target in a SELEX experiment. High background binding to nitrocellulose apparently made it too difficult to effectively isolate the rare phage-binding RNA molecules. A different approach, one which keeps background

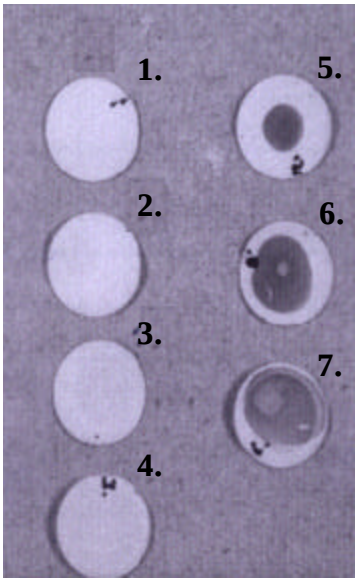
binding levels sufficiently low, was needed to successfully select for the rare phage-binding RNA.

Phage biotinylation and streptavidin column binding. The first step in generating a partitioning method that significantly reduces background binding was the biotinylation of the phage particles, without loosing phage viability. Previous work with adenovirus showed the ability of the virion to be photobiotinylated efficiently without loss of infectivity (14). EZ-Link photoactivatable biotin (Pierce Corp.) was chosen as the reagent for biotinylation due to the nonspecific nature (replacing C-H or C-C bonds), high reactivity, and ability of cross link under physiological conditions. This enabled completely random cross linking to each phage particle.

Cross-linking was initiated by placing the phage lysate with the photoactivatable biotin and irradiating with light between 340-360 nm. Numerous light sources were attempted including a UVGL-25 mineral light (366 nm) (UVP Corp.) and a TM-20 transilluminator (340 nm) (UVP Corp.). Following irradiation with each of these light sources, phage assays showed a significant loss in the amount of viable phage. This loss of phage viability is most likely due to the destructive nature of UV irradiation, causing mutations in nucleic acids, thereby rendering the RNA genomes of the bacteriophage inactive (17). To alleviate the loss of viable phage a "full spectrum" plant light (General Electric Corp.) was used in cross-linking trials. Following irradiation with the plant light, plaque assays showed no loss in infectious phage (10^{12} PFU/ml).

Following successful cross-linking, unbound photobiotin was removed by size exclusion column filtration. Microspin G25 columns (Amersham Pharmacia Biotech Inc.) were used to remove unreacted photobiotin. Qualitatively, the success of the filtration was monitored colorimetrically. The photobiotin is red, and following filtration, removal of unbound photobiotin caused the flow-through to loose the red coloration. Plaque assays showed that there was no loss in titre (10^{12} PFU/ml) following removal of unbound photobiotin.

The success of the cross-linking reaction was assayed using a modified western protocol (see materials and methods). The results of the assay (**Figure 3.7**) showed that there was extensive biotin cross-linking. Additionally, the assay also showed that the biotin was still capable of binding to streptavidin, indicating that the phage should be capable of binding to the streptavidin matrix.



1. No protein control
2. 10^{10} Non-biotinylated phage
3. 3 X 10^{11} Non-biotinylated phage
4. 6 X 10^{11} Non-biotinylated phage
5. 10^{10} Biotinylated phage
6. 3 X 10^{11} Biotinylated phage
7. 6 X 10^{11} Biotinylated phage

Results clearly show no interaction between non-biotinylated phage and the streptavidin alkaline phosphatase conjugate. When phage is biotinylated there is a colorimetric change indicating biotinylation.

Figure 3.7 Modified western blot showing extent of photobiotin-cross linking and the ability to bind to streptavidin.

Bacteriophage:streptavidin bead preparation was initiated by incubating approximately 10^{11} biotinylated phage with 100 µl prewashed streptavidin beads. The loading capacity for the beaded agarose is approximately 1-3 mg biotinylated BSA per ml of slurry, ensuring sufficient capacity to bind the biotinylated bacteriophage. Following incubation for 1 hour, the slurry was washed to remove all un-bound phage particles. To determine the extent of binding, two tests were performed. First, plaque assays of the washes showed a reduction from 10^{11} PFU/ml to 10^{10} PFU/ml. Roughly 9 X 10^{10} PFU were bound to the beads.

To visualize the phage coat protein on the beads SDS-PAGE was performed

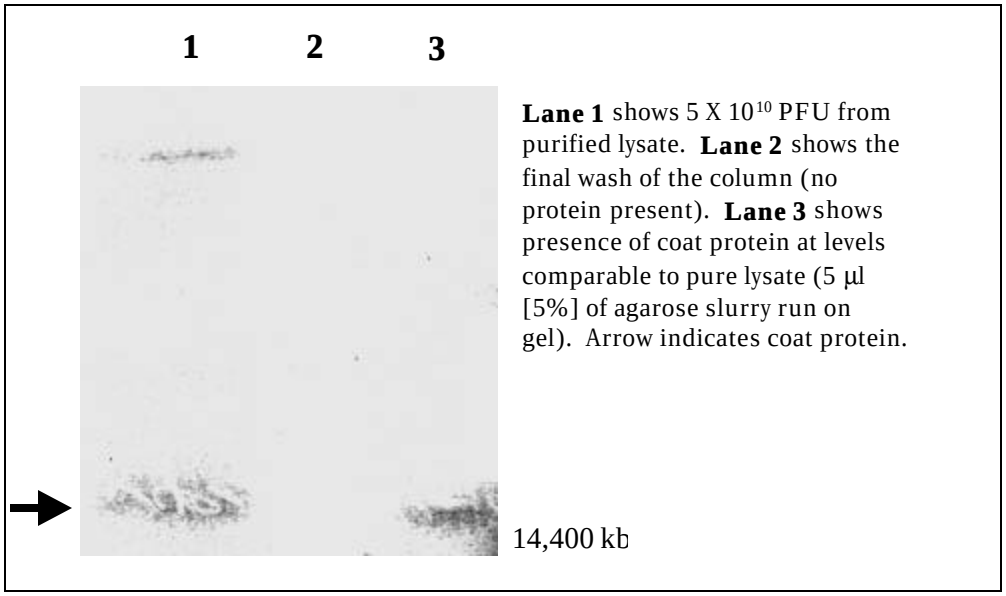


Figure 3.8 SDS-PAGE showing presence of R17 coat protein on the streptavidin matrix.

(Figure 3.8). The results confirmed the outcome of the plaque assay. Phage particles were on the beads at approximately 10^{10} PFU/ml. A second important result of the gel showed that

there were no phage removed from the beads during the final wash. Taken together, the plaque assay and SDS-PAGE indicated that bacteriophage R17 was on the beads at a titre of approximately 10^{10} PFU/ml, an amount of phage sufficient for SELEX experimentation.

R17 SELEX utilizing streptavidin agarose column partitioning. A second SELEX experiment was carried-out using bead immobilized bacteriophage R17 as the target. **Table 3.2** highlights the progress of the experiment. Selection was carried out using a pool of RNA randomized at 25 positions. An N_{25} pool was chosen because there are a possible 4^{25} or 10^{15} different molecules present in the pool. Using 5 µg of RNA (10^{14} molecules) in each round of selection covers most of the sequence diversity present in the pool. If different sequences fold into similar structural motifs it is probable that there is complete coverage of all possible RNA secondary structures in the initial N_{25} RNA pool used.

Initial binding assays showed that while there was no inherent pool affinity for bacteriophage R17, there was also no affinity for the streptavidin bead matrix. Initial rounds of selection had a 10,000:1 ratio of RNA:phage. This ratio should enable any RNA molecule with affinity for R17 the ability to bind. Following rounds 4 and 7 binding assays were performed to assess the progress of the selection.

Assays were performed in duplicate using two methods of binding. In the first method R17 was incubated with ^{32}P labeled RNA in solution and then passed through nitrocellulose membranes. Results were obtained using a Storm 840 (Molecular Dynamics Inc.) and densities measured using Imagequant software (Molecular Dynamics Inc., Sunnyvale, CA).

Table 3.3 Selection of RNAs that bind to Phage R17 using bead partitioning. Data resulting from R17 bead meethod. Initial background binding was below 1% indicating a low affinity for the bead matrix. Binding assays following rounds 4, 7, 9, and 14 all show similar low background affinity. While there was no increase in the affinity of the pool for bacteriophage R17, the biotinylation and bead immobilization could prove useful, keeping background binding sufficiently low to allow for selection of nucleic acid binding molecules to difficult targets. **Note: ****** indicates partitioning via gel-shift mobility, excision of the shifted band, and amplification of RNAs from gel slice.

								Bead Binding Assay		
Round	Pool RNA molecules	RNA ug	Phage R17 (PFU)	RNA:R17 Ratio	Volume	Neg. Selection	PCR	R17 Binding	Column Binding	
0								~1%	~1%	
1	9.7E+13	5	9.00E+09	10,777:1	200ul	Yes	25			
2	9.7E+13	5	9.00E+09	10,777:1	200ul	Yes	25			
3	9.7E+13	5	9.00E+09	10,777:1	200ul	No	25			
4	9.7E+13	5	9.00E+09	10,777:1	200ul	No	23	~1%	~1%	
5	9.7E+13	5	9.00E+09	10,777:1	200ul	No	20			
6	9.7E+13	5	9.00E+08	107,777:1	200ul	No	18			
7	9.7E+13	5	9.00E+08	107,777:1	200ul	Yes	15	~1%	1.5%	
8	9.7E+13	5	9.00E+08	107,777:1	20ul	No	18			*********
9	9.7E+13	5	9.00E+08	107,777:1	200ul	No	25	1%	1%	
10	9.7E+13	5	9.00E+09	10,777:1	200ul	No	20			
11	9.7E+13	5	9.00E+09	10,777:1	200ul	No	18			
12	9.7E+13	5	9.00E+09	10,777:1	200ul	No	18			
13	9.7E+13	5	9.00E+09	10,777:1	200ul	No	15			
14	9.7E+13	5	9.00E+09	10,777:1	200ul	No	18	~1%	~1%	

The second assay method resembled a round of selection. ^{32}P- labeled RNA was incubated with bead immobilized R17 and radioactive counts taken using a liquid scintillation counter (LS-6500) (Beckman Corp.) and percent binding calculated.

Results from each method showed similar results. After rounds 4 and 7 there was little background binding present. However, when R17 was present there was no appreciable binding over background, indicating that the pool still had no affinity for bacteriophage R17. An attempt was made to "jump start" the selection in round 8 by utilizing a different partitioning method.

A gel-shift mobility procedure was used to partition bound RNAs from unbound molecules. ^{32}P-labeled RNA was incubated with phage R17 in solution and then run on a 6% non-denaturing polyacrylamide gel containing 0.5X TBE. The appropriate band was excised and the RNA subjected to RT-PCR to start the following round of selection. A single round of selection was performed on bead-immobilized phage following the gel-shift and the pool assayed for the ability to bind R17. Once again, the pool showed no affinity for the column matrix, and still no affinity for R17. While the agarose column method works to reduce background levels, in the case of R17 (following 9 rounds) the target does not seem amenable to rapid selection experimentation.

The low amount of background binding present after 9 rounds of bead selection indicated that bacteriophage R17 was not amenable to rapid selection results. The low background, however, indicated that R17 SELEX had the potential to be successful because low background levels should allow for the amplification of specific R17 binding molecules, with vigorous rounds of selection. Five additional rounds of selection were carried out using bead-immobilized bacteriophage R17. The conditions were kept constant during each of these subsequent rounds (Table 3.2). Following 14 rounds of selection, once again, there was very little background binding present, even in the absence of additional negative selections. However, there was still a complete lack of R17 dependent binding. This result suggested that bacteriophage R17 is not tractable to positive selection results.

Discussion

Targets utilized in SELEX experimentation, that are not nucleic acid binding in nature, can prove difficult when attempting to generate RNA or ssDNA aptamers. One issue that frequently leads to difficulty is the high incidence of background binding that can occur when utilizing nitrocellulose as the method of partitioning. *In vitro* selection experimentation was undertaken in an attempt to generate nucleic acid aptamers against bacteriophage R17. Fast, reliable, and cost effective field diagnostic capabilities are becoming increasingly important as the quality of surface and ground waters declines due to fecal contamination. Bacteriophage R17 is a suitable target for screening for fecal contamination in surface waters. R17 infects *E. coli* found in the intestinal tract of swine, hence, the presence of R17 in surface or ground waters indicates the presence of *E. coli* and fecal contamination.

Initial bacteriophage purification methods needed optimization and improvement to generate a phage lysate with no contaminating components. For successful selection, the target need to be as pure as possible to avoid selecting aptamers against a spurious component. The purification procedure generated a lysate that was highly uniform from batch to batch, as well as sufficiently pure. Of the minimal protein bands present (non-capsid) in the final lysate one is most likely the phage A protein.

Initial SELEX experimentation utilized nitrocellulose as the method of separating target binding RNAs from non-binding molecules. RNA has a low affinity for nitrocellulose, however, with SELEX targets that are not amenable to the procedure small increases in

background binding level can take over the nucleic acid pool. In the experiments performed using nitrocellulose, background binding grew rapidly making the detection of phage binding RNAs virtually impossible. In such a case it was decided to devise a different method of partitioning designed to keep background binding minimal, allowing for the selection of target binding molecules.

Utilizing photoactivatable biotin (Pierce Corp.) proved to be a simple procedure to cross-link a biotin moiety to a protein target without loss of function (R17 lost no ability to infect *E. coli*). Because of the random nature of the cross-linking, there is no specific amino acid or side chain targeted. There are no potential nucleic acid binding sites that are blocked on every molecule making this a good method for a SELEX experiment. These characteristics make cross-linking a protein target to biotin acceptable for successful selection.

Due to the strong nature of the biotin:streptavidin bond, appending the phage to an agarose matrix is easily accomplished with high efficiency. The very nature of the agarose:streptavidin column matrix is one of very low affinity for RNA. The low affinity for the matrix keeps the background binding low and allows for the selective amplification of target binding nucleic acid molecules only. Only minimal negative selection need be incorporated into a selection to keep background binding levels optimum.

Following 14 rounds of selection against column bound R17 there was no enrichment for phage binding RNAs. It is highly possible that bacteriophage R17 is a target that is unsuitable for SELEX experimentation.

Conclusions

As described earlier, a wide range of proteins have been successful targets for selection experiments. These targets show no common functional theme or structure. There are 3 classes that successful SELEX experiments fall under. The first class of proteins that enable successful SELEX experimentation are proteins that normally bind nucleic acids as their biological functional role. Proteins that normally bind nucleic acids have a higher probability to extract specific sequences from a random pool of RNA or ssDNA. This class of proteins can be expanded to include proteins that bind to nucleotides.

The second class of proteins that allow for successful selection results are those proteins that contain basic patches in their primary sequences or on their surfaces. Many cytokines or other signal transduction proteins bind heparin and can be used to select aptamers that bind the same sites as heparin (18). Also, proteins that bind to phosphate or phosphodiester bonds have proven to be successful generators of aptamers (19).

A third class of proteins that frequently enable successful SELEX results are those that have large clefts or pockets on their surface. These surface structures allow smaller nucleic acid molecules to bind in a manner where the nucleic acid is surrounded and not "holding on" to the surface of the protein. This enables for the selection of tighter binding aptamers. The most common aptamers are those that fit into surface crevices (11).

Scanning the primary sequence of the bacteriophage R17 coat protein shows that there are no histidine residues, only 4 arginine, and 6 lysine residues present in the protein.

There are no basic patches in the primary sequence or in the secondary structure of the protein. In addition, the surface of the bacteriophage R17 is devoid of surface structures such as crevices or clefts. These conditions make bacteriophage R17 possibly a difficult target for use in a SELEX experiment.

These conditions do not preclude the success of a target that contains any or all of the aforementioned characteristics. These are simply considerations that should be accounted for when setting up a SELEX experiment. A final consideration that should be taken into account are the initial binding results from the round 0 pool of RNA (or ssDNA). A slight indication of protein dependent binding prior to SELEX experimentation can be used as a positive indicator of potential success.

References

1. Ford, T. (1999) *Environ Health Perspective* **107**: 191-206.

2. Edberg, S., Rice, E., Karlin, R., and Allen, M. (2000) *Symp Ser Soc Appl Microbiol* **29**: 106S-116S.

3. Havelaar, A., Furuse, K., and Hogeboom, W. (1986) *Journal of Applied Microbiology* **60**: 255-262.

4. Huang, P., Laborde, D., Land, V., Matson, D., Smith, A., and Juang, X. (2000) *Applied and Environmental Microbiology* **66**: 4383-7388.

5. Sassanfar, M., and Szostak, J. (1993) *Nature* 364: 550-553.

6. Hirao, I., Madin, K., Endo, Y., Yokoyama, S., and Ellington, A. (2000) *Journal of Biological Chemistry* 275: 4943-4948.

7. Pan, W., Craven, R., Qiu, Q., Wilson, C., Wills, J., Golovine, S., and Wang, J. (1995) *Proc Natl Acad Sci USA* 92: 11509-11513.

8. Murphy, F., Fauquet, C., Bishop, D., Ghabrial, S., Jarvis, A., Martelli, G., Mayo, M., and Summers, M. (1995) *in Virus Taxonomy: Sixth Report of the International Committee on Taxonomy of Viruses* pp. 324-325, Springer-Verlag Wien., New York, NY.

9. Klug, S., and Famulok, M. (1994) *Mol. Bio. Rep.* 20: 97-107.

10. Osborne, S., and Ellington, A. (1997) *Chemical Rev.* 97: 349-370.

11. Jhaveri, S., and Ellington, A. (1990) *In Vitro Selection of RNA Aptamers to a Protein Target by Filter Immobilization* (personal communication).

12. Sambrook, J., Fritsch, E., and Maniatis,T. (1989) *in Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY.

13. Erlich, H.A. (1989) *in PCR Technology: Principles and Applications for DNA Amplification*, pp. 1-7, Stockton Press, New York, NY.

14. Smith, J., Keller, J., Lohrey, N., McCauslin, C., Ortiz, M., Cowan, K., and Spence, S. (1999) *Proc Natl Acad Sci U S A*. **96**: 8855-8860.

15. Sayle, R., and James Milner-White. "**RasMol: Biomolecular graphics for all**", *Trends in Biochemical Sciences (TIBS)*, September 1995, Vol. 20, No. 9, p. 374.

16. Hall, T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser*. (1999) 41:95-98.

17. Snyder, L., and Champness, W. (1997) *in Molecular Genetics of Bacteria*, pp. 246-247, ASM Press, Washington, D. C.

18. Jellinek, D., Lynott, C., Rifkin, D., and Janjic, N. (1993) *Proc Natl Acad Sci U S A*. 90: 11227-31.

19. Bell, S., Denu, J., Dixon, J., and Ellington, A. (1998) *Journal of Biological Chemistry.* 273: 14309-14.