DESIGN, CONSTRUCTION, AND CHARACTERIZATION OF THE YSGR MINIMAL CODON FAB LIBRARY FOR CHAPERONE-ASSISTED RNA CRYSTALLOGRAPHY

by

SEAN HOLMES

A thesis submitted in partial fulfillment of the requirements
for the Honors in the Major Program in Microbiology and Molecular Biology
in the College of Medicine, and in the Burnett School of Biomedical Sciences
at the University of Central Florida.
Orlando, Florida

Spring term 2012

Thesis Chair: Jingdong Ye, Ph.D.

Abstract

Of the entire human genome, 90% of all genetic information is transcribed but only a fraction of that subsequent RNA is translated into proteins. RNAs which are not translated into proteins are deemed non-coding RNAs. Little is known about this large category of noncoding RNAs, although they perform a variety of functions within the cell. RNA crystallography is used to study RNA tertiary structure, which gives insight to the function of these non-coding RNAs. However, complications associated with RNA crystallography arise due to RNA's lack of surface functional group diversity, flexible tertiary structure, and conformational heterogeneity.

A novel technique, Chaperone-assisted RNA crystallography (CARC), can greatly improve the success in crystallization of RNA. With this technique, synthetic antibodies called Antigen Binding Fragments (Fabs) are employed as crystallization chaperones to promote the structure elucidation of certain target RNAs.

To identify Fabs that complex with RNA molecules of interest, we constructed a randomized library of synthetic antibodies enriched in ligand binding regions with tyrosine, serine, glycine, and arginine residues. This Fab protein library was constructed with a minimal codon design, and then screened against a variety of RNA targets. Several Fabs have been identified and isolated through phage display selection against three RNA targets. These Fabs were expressed and biochemically characterized for their binding affinities and specificities. To date, five Fabs have been identified and they have outstanding capabilities in binding their specific RNA antigen.

Table of Contents

List of Figures

List of Tables

A Background on Antibody Research and RNA Targeting

The development of antibody technologies has led to a wide array of biomedical applications including the use of antibodies as research tools, diagnostic agents, and therapeutic drugs. Antibodies have been widely employed for protein targeting in assays such as the Western Blot and ELISA. They have also been used for therapeutics, and they are currently being developed into cancer treatment agents. Other uses of antibodies which target proteins include detection of disease antigens and cell surface marker analysis.

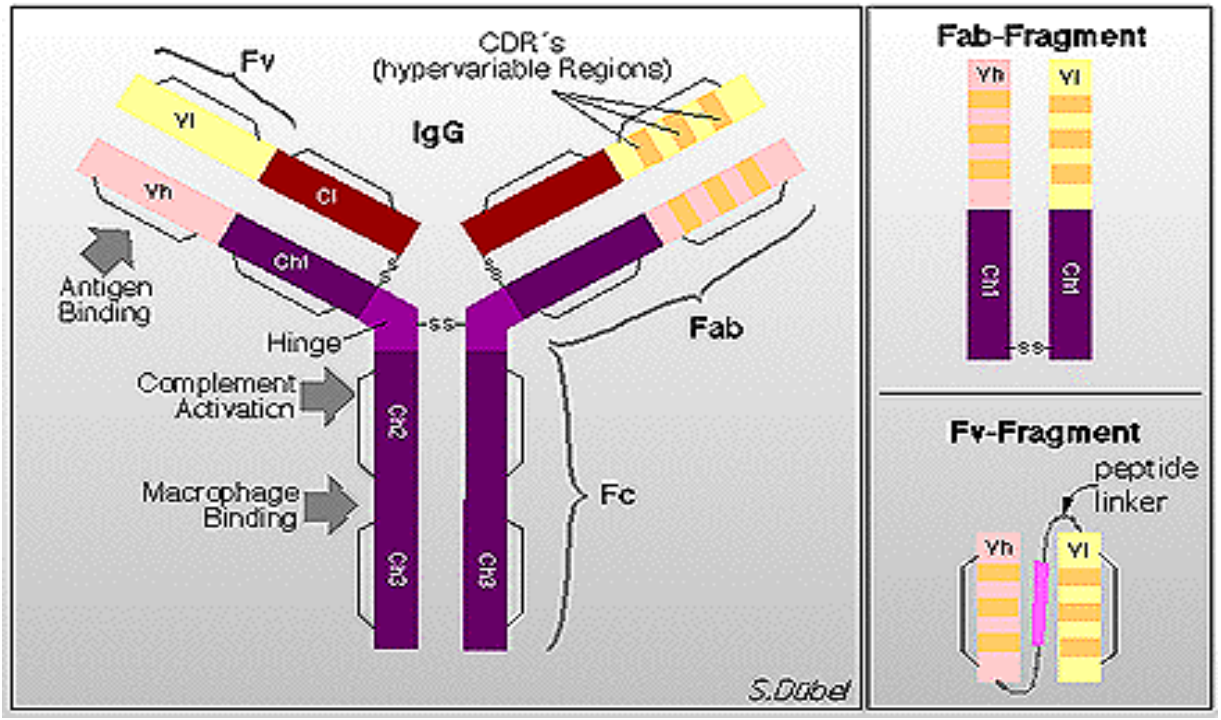
However a lack of endogenous nucleic acid-binding antibodies has largely restricted their use to only proteins. Nucleic acids lack immunogenicity because they get degraded too quickly upon entrance into a host by nuclease activity, and the host has no time to develop specific antibodies towards that antigen. For these reasons, there is no solid foundation for obtaining RNA-binding antibodies.

Current sources of antibodies include the derivation of monoclonal antibodies by fusing a B cell to an immortal hybriboma cell and producing clones of that hybridoma. These clones all secrete identical antibodies to the original fused B cell. Another method for obtaining antibodies is harvesting the serum of a host or mammal after injection of the antigen into that host. Antibodies specific to the injected antigen can then be purified out of the serum, these are said to be a collection of polyclonal antibodies because they were generated from a pool of B cells. Both of these methods for obtaining antibodies are expensive and time consuming. Furthermore, these methods have not worked well for obtaining anti-RNA antibodies due to RNA's low immunogenicity.

Of the entire human genome, 90% of all genetic information is transcribed but only a fraction of that subsequent RNA is translated into proteins. RNAs which are not translated into proteins are deemed non-coding RNAs. Little is known about this large category of noncoding RNAs, although they perform a variety of functions within the cell. Examples of non-coding RNAs include rRNA, tRNA, siRNA, and many others. Very little is known about the structure and function of this vast array of non-coding RNAs, and better tools to study RAN structure must be developed to correct this issue.

A novel method has been designed for obtaining RNA-binding synthetic antibodies. This method involves designing a randomized antigen binding fragment (Fab) library, screening that library with phage display, expressing and purifying the selected Fabs, and characterizing each of these Fabs for their RNA-binding potentials (Fig 1). The development of nucleic acid-binding antibodies is very important to the study of RNA structure and function.

Figure 1 Antibody and Fab



In past work, Ye et al. designed a Fab library and also developed phage display methods to obtain specific RNA-binding Fabs from their Fab library (Fig 1) (1). With the produced specific Fabs, the crystal structures of the P4P6 domain as well as the class I ligase were elucidated by Chaperone-Assisted RNA crystallography (CARC). This crystallization method employs Fabs as crystal chaperones to aid in crystal packing, molecular stability, and obtaining initial phasing information from the target Fab-RNA complex.

However, this Fab library was not initially designed for the targeting of RNA. It was later found that the Fabs in this library had low RNA-binding affinities and many Fabs bound to their RNA targets via nonspecific interactions (1). Although their developed phage display methods were able to isolate the RNA-binding Fabs from their Fab library, it was realized that a Fab library specifically for RNA recognition was needed.

Designing a RNA-binding Fab library requires vital knowledge in RNA-Fab recognition. More specifically, one must know exactly where to install randomized sites on a Fab for the design of a Fab library. Additionally, it must be known which residues should be kept constant. One must also know what the amino acid composition of those randomized sites should be in order to yield a Fab library that specifically targets RNA structures (2,3). The study and analysis of Fab-RNA interfaces can lead to the foundational knowledge which is required for the design of RNA-specific Fab libraries. By studying the complementarity determining region (CDR) of an RNA-specific binding Fab, vital knowledge can be gained. The CDR of each Fab is the region which determines complementarity between that Fab and its RNA target. Thus by studying the amino acid content of the CDRs of selected Fabs, the active residues that are making RNA-

specific contacts can be identified. This information can be applied for initial design of other Fab libraries that target other tertiary RNA structures.

For my project, a minimal amino acid code was used to design a Fab library specific for Fab-RNA binding (4). The YSGR minimal codon library of Fab fragments, as it has been called, was screened by phage display selection in order to select for Fabs that bind tightly and specifically to their RNA targets of interest. Fabs that were isolated from the YSGR minimal codon library via initial phage display screening were sequenced and further biochemically characterized. Specific RNA-binding Fabs which were selected from the YSGR minimal codon Fab library were characterized for their RNA-binding capabilities.

Design and Construction of the YSGR Minimal Codon Fab library

Theory Behind the Minimal Codon Design: Fab structure and the YSGR Minimal Codon Fab library

The YSGR minimal codon Fab library was designed by engineering randomized point mutations into each of the CDRs of the Fab 2 scaffold. Fabs within the YSGR minimal codon library were randomized in all three light chain CDRs, as well as all three heavy chain CDRs. As a consequence of this minimal amino acid code, variable sites were not randomized with just any of the 20 amino acids from the genetic code; they were randomized using only a four amino acid genetic code (5).

One of four amino acids was inserted into each position of the CDR that was engineered to be randomized. This means that each randomized residue on each Fab within the library could only be coded into one of four amino acids rather than into one of the usual twenty. This minimal codon design yielded a randomized Fab library actual size of 3.1×10^9 sequence variants. This minimal codon design also reduced the gap between the theoretical library size and practical library size (6).

The four amino acids which were chosen consisted of Tyrosine, Serine, Glycine, and Arginine. Tyrosine and serine were chosen because previous crystal structure analysis has shown that they both make key RNA-binding interactions (3). While Arginine was chosen for its potential to form stabilizing electrostatic interactions with RNA, and Glycine was chosen to promote flexibility of the Fab within the CDR (2,3).

Each randomized position on the Fab structure had equal potential to be one of any four of the designated amino acids. This is a minimal codon Fab library design, as compared to a

normal library design where the randomized positions on each Fab would have their normal potential to encode for one of any of the 20 amino acids.

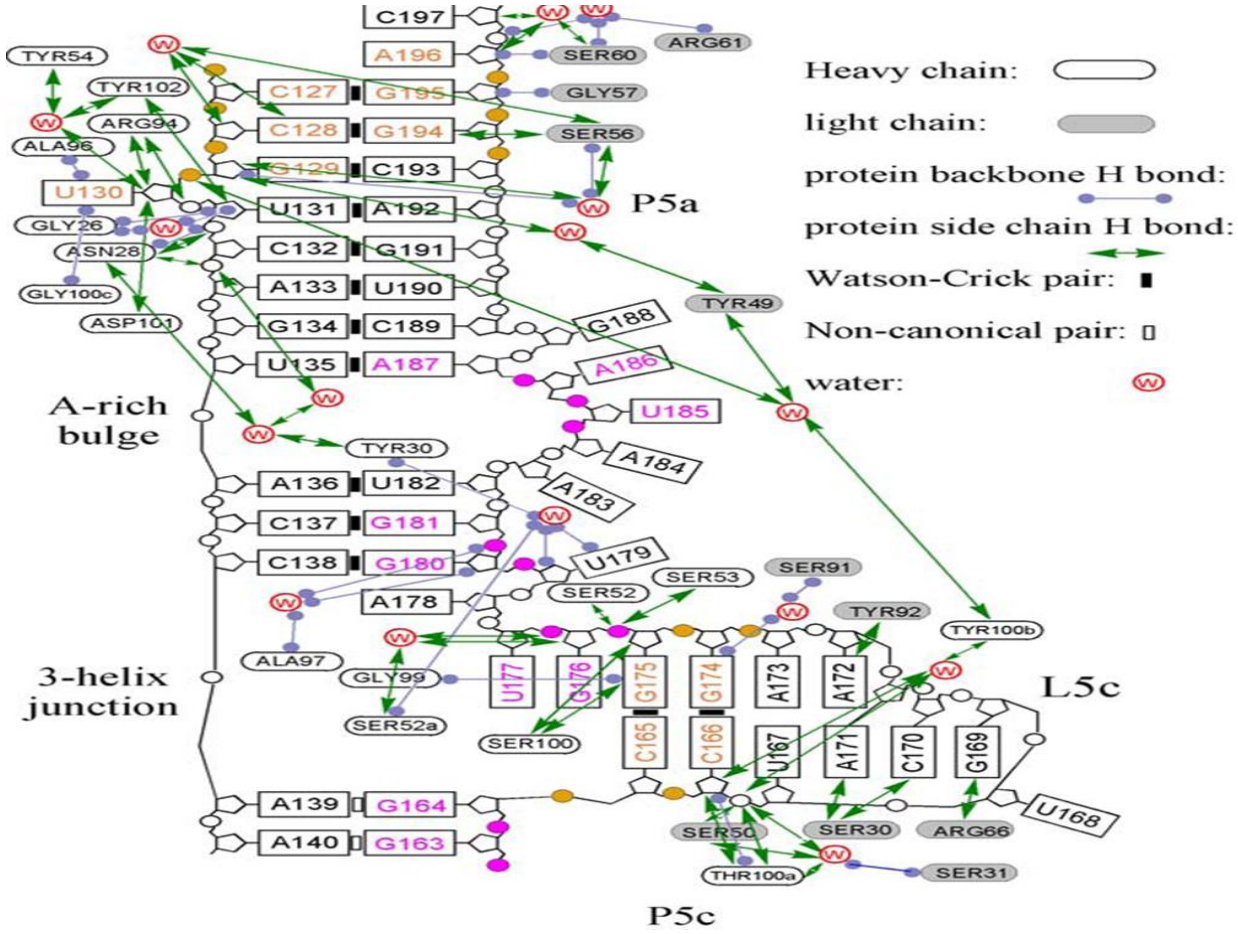
This project will serve as a challenge to the Fab-RNA interface. It will be tested whether or not a minimal code design within a Fab library will be sufficient to obtain tight and specific RNA-binding Fabs. It will thus be determined whether or not this minimal codon design will yield enough CDR variability to produce specific RNA-binding Fabs from a minimal codon Fab library design such as this one.

Experimental Design and Methods for Library Construction

The positions of the Fab to be randomized were selected based on their solvent exposure and their potential to interact with the RNA antigen as visualized and analyzed in the 3D crystal structure viewer program called PyMol. More specifically, a crystal structure of the complex of Fab2 and the P4-P6 RNA Ribozyme Domain was viewed and analyzed in PyMol to identify CDR residues that were making specific points of contact with the P4P6 RNA antigen (1). These identified interactions were either by direct contact or by water mediated Hydrogen bonds. Library construction was also based on analysis of the Fab2 -P4-P6 RNA binding interface (Fig 2).

Each amino acid residue that was chosen to be randomized was then correlated to the location of its codon within the corresponding DNA sequence. This correlation was made by the Kabat numbering system which was designed for identification of residues within immunologic proteins (7). Scaffold residues were chosen to be conserved in all Fabs to maintain each protein's native structural integrity. A Kunkel mutagenesis reaction was employed to effectively mutate the CDR sequences of the template into desired randomized sequences (8).

Figure 2 Image of the specific synthetic Fab2 bound to the P4P6 Ribozyme



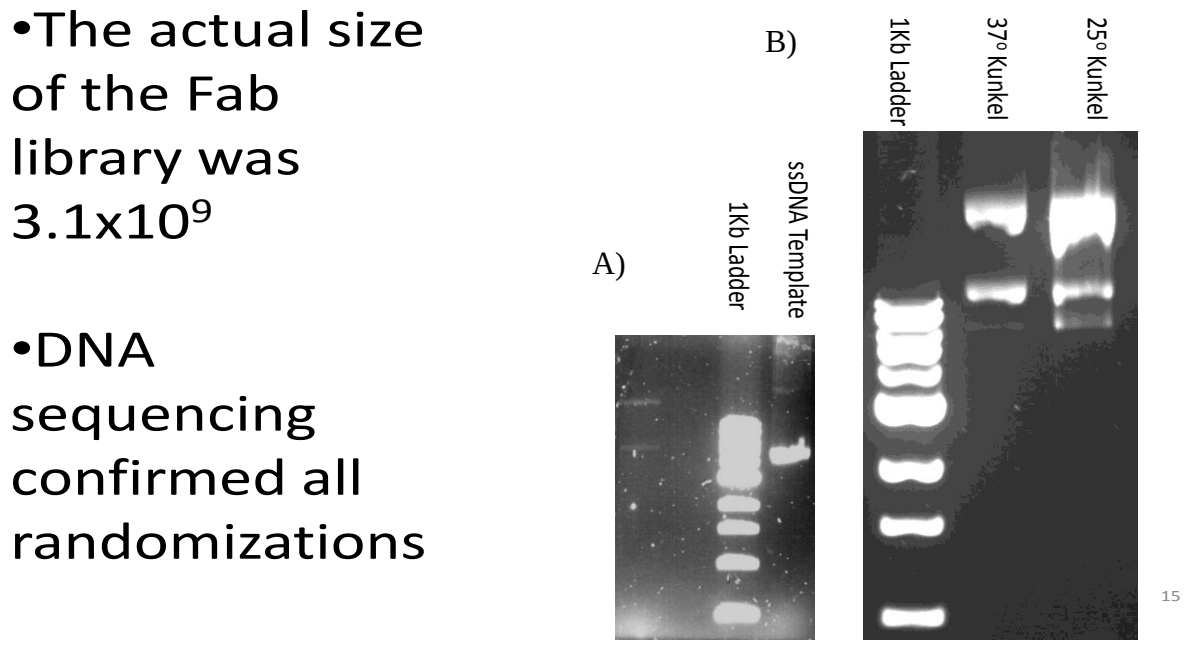
Shown here is an image of Fab2 bound to the P4P6 Ribozyme domain. Various amino acid residues (circles) of the Fab 2 are depicted here as making RNA-specific contacts. Randomized residues were chosen based on their potential to interact with the RNA antigen and also the level of their solvent exposure. For exmaple, shown at the bottom of this image is a group of serine residues that are making RNA-specific contacts. These are all examples of residues that should be randomized for the design of an anti-RNA Fab library.

To generate the YSGR minimal codon library, a template DNA strand was first made as a starting point for construction. The template strand was made by introducing stop codons into the six CDR regions of the P4-P6 Fab2 DNA scaffold through a Kunkel mutagenesis site-specific reaction (1). This design prevented the translation of unmutated DNA into Fab proteins (8). Then, sixteen library mutagenic primers were designed to engineer the amino acid residues Y, S, G, and R into all six CDRs of the heavy chain and light chain with position and length variability. Then, oligonucleotides containing these randomizations were ordered and synthesized by Integrated DNA Technologies and used as primers in the Kunkel mutagenesis reaction.

The Kunkel mutagenesis reaction was carried out with the synthesized template DNA and the sixteen library mutagenic primers which were ordered. This required a Kunkel mutagenesis protocol with an unprecedented number of primers. Thus, the reaction required experimental optimization to ensure proper annealing of all 16 primers to the template DNA plasmid. Reaction optimizations for temperature, primer-to-template ratio, enzyme concentration, and incubation time were all carried out to maximize the reaction yield.

The optimized conditions were shown to be: primers phosphorylated at 37°C with polynucleotide Kinase 4 for 1 hour, then annealed to the template DNA for 1.5 hours in 37°C. 25mM dNTP's, 10mM ATP. 100mM DTT were present in both phosphorylation and annealing (Fig 3).

Figure 3 Library construction



A). This agarose gel confirmed the optimized reaction conditions for the Kunkel reaction. The middle lane shows the Kunkel product band in highest yield (middle band). B) This agarose gel confirms the ssDNA template which was synthesized as a starting point. A Kunkel mutagenesis reaction is a site-specific mutagenesis which was employed here for CDR randomizations.

RNA Targets for Screening Characterization of the YSGR Minimal Codon Fab Library

To demonstrate the diverse functionality of the YSGR minimal codon Fab library, three RNA targets were chosen for Fab library screening. Two of these targets were riboswitches, and the third RNA target was a Ribozyme domain.

Description of Riboswitches and Why They Should be Studied

Riboswitches are cis-acting RNA elements that are present in the 5' untranslated region of mRNA where they act to control gene expression posttranscriptionally(9). Abundant in many pathogenic organisms, these regulatory RNAs mediate genes essential for life and virulence (10). Riboswitches fold to complex tertiary structures by binding to small organic metabolites and control gene expression through transcription termination, translation initiation, or alternative splicing (11,12). By targeting these riboswitches with novel antibiotics, the pathogen's ability to control gene expression can be disrupted and its virulence can be inactivated (13). Targeting novel aspects of bacterial physiology such as the control of gene expression by riboswitches will evade the current drug resistance mechanisms used by many human pathogens.

The level of antibiotic resistance possessed by pathogenic bacteria and fungi is constantly increasing throughout society. This issue presents dangerous implications for the worldwide control of pathogenesis, and it must be rectified (14). Antibiotic resistance is increasing by the day in our society, and current therapeutics which target very few cellular processes are becoming increasingly ineffective. Through the widespread use of these antibiotics and by natural selection, antibiotic resistance genes have allowed many pathogenic organisms to

overcome these current antibiotic inhibition mechanisms and proliferate in the presence of the antibiotics which were originally designed to kill them.

My project aims to alleviate the problem of increasing antibiotic resistance through contributing to the design of novel drugs that target the glycine riboswitch present in two pathogenic species (15) (Fig 4). With the crystal structure of this riboswitch, the detailed structural information of the metabolite binding sites can be employed to assist in the structure-based antibiotic drug design (13). However, riboswitches and RNA in general have a very low level of success in crystallization. This is due to a variety of factors such as RNA's lack of surface functional group diversity, flexible tertiary structure, and conformational heterogeneity (16). These issues present major road blocks in the process of crystallization because crystal lattice formation is dependent on tight packing of molecules and high stability (16). The unique approach employed by my project aims to overcome this issue by forming crystal contacts through Fab chaperones.

The YSGR Minimal Codon library of Fab Fragments has been synthesized specifically for selection against RNA targets. Fabs which were selected via initial screening were then expressed as soluble proteins. Biochemical analysis was then carried out on each Fab. This was completed to characterize the capability of each Fab to stabilize their RNA target and thus promote the crystallization of a Fab-RNA complex via CARC.

The Three Specific RNA Targets Which Were Chosen for Library Screening

Once the YSGR minimal codon Fab library was designed, it was then screened against three RNA targets: the *Vibrio cholerae* glycine riboswitch (VCLDI), the *Fusobacterium nucleatum* glycine riboswitch (FNIII), and the P4P6 ribozyme domain (P4P6). Each of these

RNA targets had to be synthesized in order to be used as targets in the screening of the YSGR minimal codon Fab library. The P4P6 domain was chosen as a positive target control and allows comparison to the selection results with previous previous P4P6 library targeting.

Two of these targets are highly conserved glycine riboswitches that are present in the vibrio cholerae and fusobacterium nucleatum pathogenic species: the VCLD1 riboswitch, and the FNIII riboswitch (Fig 4). In order to design antibiotics which will target the glycine riboswitch in these species and others, we must know all possible detailed structural information. The metabolite binding sites of the glycine riboswitch must be deduced as this will facilitate in-depth structural analysis. Crystallization of the glycine riboswitch will also provide insight on how to use novel antibiotics in the targeting of many different riboswitches that are present in several bacterial and fungal species.

Methods and Materials for Construction of the Three RNA Targets for Library Screening

The three RNA targets were synthesized by the following experimental methods (Fig 4):

Polymerase Chain Reaction: ddH20, 10x pfu buffer , 10 mM dNTPs , template plasmid DNA 6ng/ul, 6 uM FWD primer, 6 uM RVS primer, 2.5U/ul pfu. Regular PCR cycle was ran.

EcoRI and HindIII double digestion reaction : 10ug PCR product, NEB buffer 2, 15 u/uL Eco RI and HindIII. Purified this reaction with Qiagen PCR cleanup kit, measured the DNA concentration by UV spectrophotometer.

Ligation into digested pUC19 vector: 10ng/ul pUC19 vector, 1 ul double digested gene insert, 10x T4 DNA ligase buffer, ddH2O, T4 DNA ligase, 16^{o} C overnight.

50uL of this ligation reaction was then used to do a heat shock transformation into competent JM109 strain Escherichia coli cells, then the cells were plated on a LB/Ampicillin agar plate.

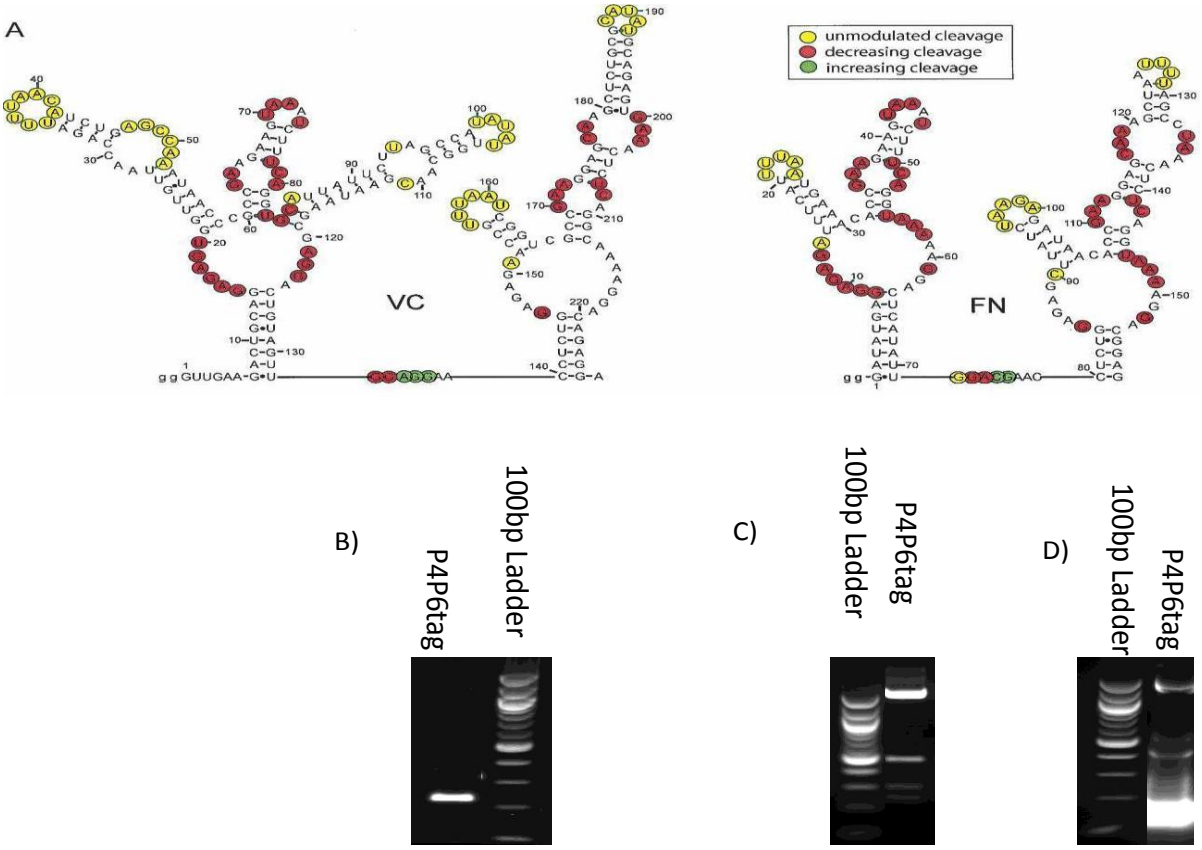
After transformation into JM109 cells, two colonies were picked from the LB/Amp plate and inoculated into 4 ml liquid autoclaved LB/Amp media in cell culture tubes; shaken for 12-16 hrs at 37 ^{o}C. Combined 750 ul overnight cell culture with 150 ul 80% glycerol, stored at -80 ^{o}C. Spun down the remaining cell cultures in the Allegra X-12R refrigerated tabletop centrifuge for 10 min at 4 ^{o}C, ~3500 rpm. Decanted the media into sink and follow the Qiagen miniprep kit to extract plasmid. Checked the concentration using 2 ul plasmid (dilute 35x with TE) on UV spectrometer.

A small scale EarI digestion of harvested plasmid DNA was then completed with NEB buffer 4. A 2% agarose gel was run to check the digested product. 500ng of DNA product was then sent for sequencing at ETON bio.

The correctly sequenced clones were chosen for amplification and harvest of DNA with Qiafilter plasmid Midi Kits. The DNA was extracted with PCA solution twice to get rid of RNase A in the late steps of the Midiprep Kit . Then, the larger scale EarI digestion of plasimd in NEB buffer 4 was completed, followed by in-vitro transcription.

RNA purification was carried out as follows: Buffer exchanged into 5 mM Sodium Citrate, completed PAGE gel purification, Electroeluted via dialysis in TE Buffer pH 8, PCA extracted the RNA, and completed the Ethanol precipitation. Then RNA was redissolved in TE Buffer and stored in -20°C.

Figure 4 VC and FN Riboswitches and synthesis of RNA targets



A) This figure shows a secondary structure prediction for the glycine riboswitch as it exists is the FN and VC species. This figure demonstrates the highly conserved genetic nature of the glycine riboswitch. B) This agarose gel confirmed the PCR reaction during cloning. C) This agaorse gel confirmed the Ear1 digestion reaction after cloning into the expression vector. D) This gel confirmed the final cloning step- the in-vitro transcription reaction which produced the RNA structures of interest.

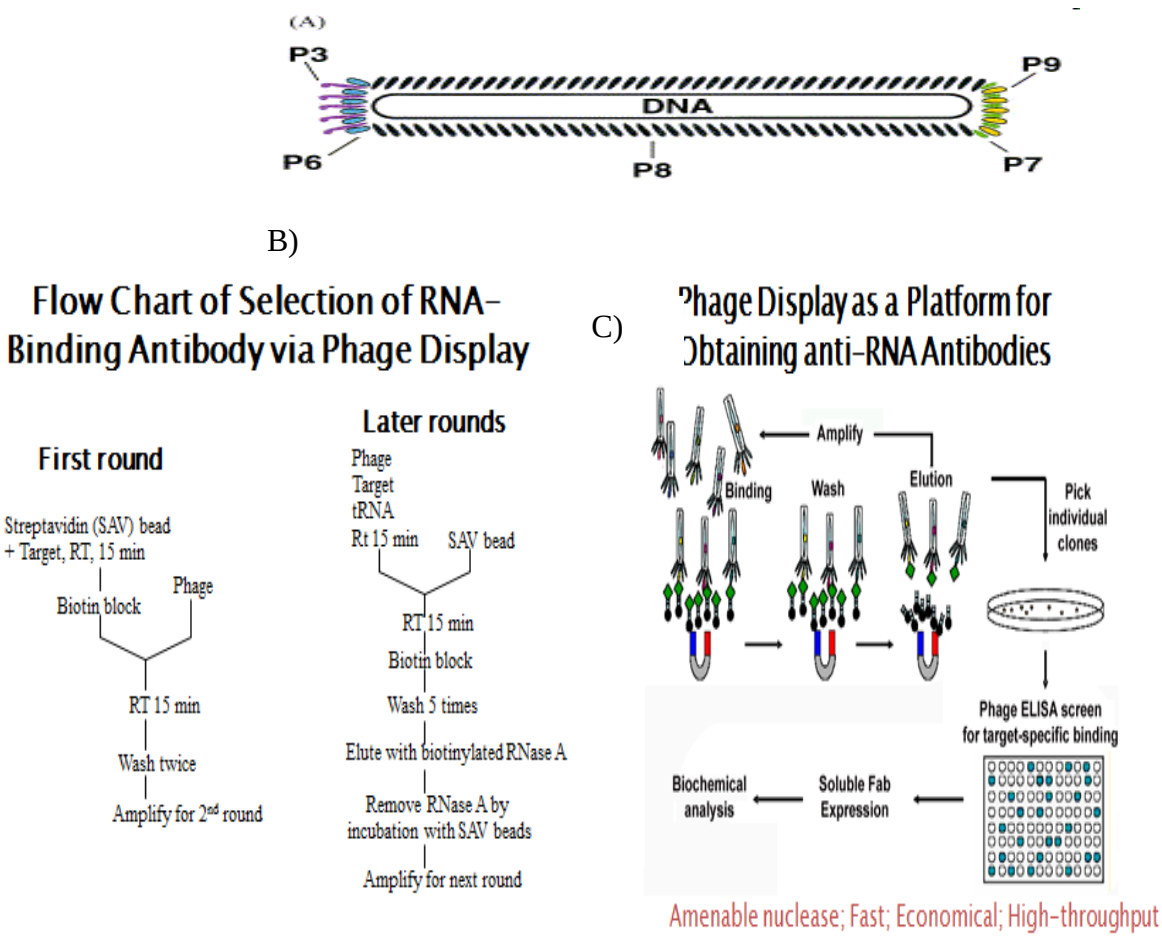
Initial Screening of the YSGR Minimal Codon Fab Library

General Process of Initial Screening for the YSGR Minimal Codon Fab Library

The YSGR minimal codon Fab library was selected against both the VC and FN riboswitches to isolate RNA-binding Fabs, as well as the P4P6 ribozyme domain for comparison of library functionality. The method employed to screen this library was called phage display selection, and it was followed by phage ELISA (17). This selection process used phage particles which displayed each of the Fab proteins containing the randomized CDRs on their surface (Fig 5) (18).

Phage ELISA was subsequently completed to isolate Fabs that proved to bind their RNA target (Fig 5). Positive clones were then sent for DNA sequencing to confirm and identify the amino acid sequences of each unique Fab clone. The harvested and sequenced DNA for each Fab was then used as template DNA for Fab protein expression and further characterization (19,20) .

Figure 5 Schematic of Phage Display Selection and Phage ELISA



A) This schematic shows a phage particle that is currently expressing a Fab as a surface protein on its coat (Sachdev Sidhu, Biomolecular Engineering vol. 18 issue 2 Sept. 2001, pages 57-63). The Fab is denoted in purple, note that the Fab is open for interactions with other Macromolecules. B) This flowchart demonstrates the process used for Phage Display Selection. C) This flowchart demonstrates the entire initial screening process in general and what it leads into.

Materials and methods for Initial Library Screening: Phage Display Selection, ELISA

Phage selection: First, the biotinylated RNA target was incubated with washed magnetic streptavidin beads for 15 min at 25°C. Phage particles with Fab proteins expressed on their coats were added, and 15 min incubation at 25°C was carried out. The washing was then completed in two steps to eliminate the Fabs that did not bind their RNA targets specifically. Fabs which were present after washes were then eluted out and amplified via XL1-blue cell culture. This was completed by infecting the E.coli cells with the DNA from within phage particles that expressed RNA-binding Fabs. Further rounds of selection were then completed with a similar protocol (Fig 5) (18).

Phage display selection has been completed for both the VC and FN riboswitches. Additionally, the P4P6 RNA domain was targeted with the YSGR minimal codon Fab library through Phage Display Selection to confirm library functionality. The results of Phage Selection for the YSGR minimal codon Fab library against all three RNA targets are described under the results of initial screening (Table 1).

Phage ELISA: 48 colonies were grown from the phage titer. Each colony was inoculated into 400 μL 2 YT, Amp, and 10^{10}/mL M13KO7 helper phage in 96-well deep-well plate at grown 37ºC for 16 hours. All 48 colonies were duplicated on to a LB/ Ampicillin Agar plate with a numbered grid for colony retrieval, and the plates were incubated at 37ºC for 16 hours.

Maxisorp plate was then coated with nuetravidin buffer, then blocked with BSA in 1x PBS. Columns: 1, 3, 5, 7, 9, 11 were loaded with 100 μL/well 25 nM RNA-ext/biotin (prefolded) in RNA binding buffer, and wells: 2,4,6,8,10,12 were loaded with only RNA binding buffer as a control. (RNA Folding reaction: 2.5 ul 50 uM RNA-ext/biotin + 97.5 ul RNA binding buffer, 50ºC for 30mins; room temperature for 10 mins; Add to 4.9 ml RNA binding buffer at RT in a

new falcon tube and mix). Then, the 96 deep well plate with cell cultures in each well was centrifuged at 3500rpm for 25 mins to pellet.

70 uL of supernatant was then transferred from the deep well plate to normal ELISA plate. 140 uL of phage dilution solution was then added and mixed with 1.5x RNA binding buffer (0.75% BSA, 30 ug/ml sheared salmon DNA, 1.5 mM vanadyl ribocomplexes or 0.15 U/uL RNaseIN, 0.66 uL for every 140 uL solution, 66 uL for 96-well plate). (Phage dilution solution 7ml total: 1050 ul 10x PBS, 52.5 ul 10% Tween20, 52.5 ul 500 mM EDTA, 131 ul 1M MgCl2, 10.5 ul 1M glycine, 52.5 mg BSA, 21 ul 10 mg/ml sheared salmon DNA, 52.5 ul 200 mM Vanadyl ribocomplex, 3ul 4 U/ul RNase IN, ddH2O to 7 ml, do not filter).

Maxisorp plate was then washed six times with RNA binding buffer. 100 μL of phage supernatant was then transferred from the diluted side of the ELISA plate into both RNA target containing wells and the corresponding control wells of the Maxisorp plate. The plate was then shaken for 30 minutes at room temperature. 6 more washes were then completed with RNA binding buffer.

100 uL diluted Anti-M13 antibody/HRP conjugate was then added to each well (diluted 5000x in RNA binding buffer, 0.5% BSA, 1.5 mM vanadyl ribocomplexes). Plate was shaken for 30 minutes at room temperature. Wells were then washed six times with RNA binding buffer, then twice more with RNA binding buffer minus Tween20.

Then, 100 uL of the TMB substrate was added and 96 well plate was incubated for 10 min at room temperature. Positive wells turned blue within a few minutes. The reaction was quenched with 100 μL of 1 M phosphoric acid into each well, at which point the positive wells

turned yellow immediately. The absorbance of each well on the 96 well plate was read at 450 nm in a plate reader. Positive clones were then sent for DNA sequencing at Genewiz.

Results of Initial Screening

Enrichment values that were obtained from Phage selection for both the VC and FN glycine riboswitches were proven to be comparable to enrichment values for P4P6 domain Phage Selection (Table 1). This confirmed the functionality and robustness of the YSGR minimal codon Fab library for all RNA targets being studied at the initial screening level. DNA sequencing of positive clones revealed 17 unique Fabs, all with varied heavy chain and light chain CDR regions (Table 2). These unique Fabs were to be expressed and further biochemically analyzed.

Once Phage ELISA and DNA sequencing were completed, each desired XL-1 blue cell colony from the numbered grid was amplified in LB/Amp liquid media. Then, Qiagen miniprep DNA harvests were carried out on each culture. This was completed to obtain the plasmid DNA sequence for each Fab that was isolated as a specific binder via Phage selection and Phage ELISA.

Table 1 Results of Phage selection

RNA Target in Phage Selection	Selection round (S = stringent conditions)	Enrichment	Reduction Rate
FNIIItag	2	.77	$2.6x10^7$
FNIIItag	3,S	8.12	$3.54x10^7$
FNIIItag	4,S	34.42	$9.06x10^4$
FNIIItag	5	75.6	$2.62x10^4$
VCLD1Tag	2	1.61	$8.89x10^5$
VCLD1Tag	3,S	6.84	$3.11x10^7$
VCLD1Tag	4	176.97	$9.27x10^4$
P4P6 domain	2	11.87	$2.92\ x10^6$
P4P6 domain	3	107.69	$1.31\ x10^6$
P4P6 domain	3, S	369.67	$4.46\ x10^5$

This table shows the results obtained from Phage Display selection for all three RNA targets that were screened against the YSGR minimal codon Fab library. An enrichment value of 10 showed that nonspecific binders were getting successfully screened out of this Fab library after each selection round. Selection was not completed until the necessary enrichment values could be achieved. "S" denoted stringent conditions.

Table 2 Sequencing results from Phage ELISA

RNA Target	Fab Name	Heavy Chain 1	Heavy Chain 2	Heavy Chain 3	Light Chain 1	Light Chain 2	Light Chain 3
P4P6	8	YGYRS	YSASGLYRGVPSR	GYRSPV	NLGSGYI	SYRPSSGSTR	SYSSRYSYAM
	11	SVSSA	YSASSLYSGVPSR	SYSSPI	NLYSSSI	SRSPRSGGTS	RAAGMSTYGF
VCLD1	24	YRYRS	SGAYGLYRGVPSR	YRSGLV	NVSSSSI	GSRSSSGSTG	SSYGSRSSAM
	25	YRYRS	SGAYSLYRGVPYR	YSSGLI	NFSSGSI	GSGPSRGSTS	SRSRSRSRAM
	26	YRYRS	SGAYSLYRGVPSG	YRYGLI	NFGYSSI	SYRSSSGSTG	SSSSSRSGAM
	27	SGSRS	YGASSLYGGVPSR	YSYGLL	NFRSRYI	GSGPSRGYTG	GSRYRYGSRGM
	29	YSYRS	YRASRLYGGVPSR	RSSYPV	NFSGSSI	YGSPGYGRTS	SSYGSRSGYAM
	30	SGSSY	YRASSLSSGVPSR	YSSRLL	NIYRYGI	YGSPGSGRTY	RYRYRYRSGL
	31	RYRSG	SGAYGLYSGVPSR	YGSGLI	NFRYRSI	SRSSRSGRTS	SRSGYSSGAL
FNIII	10	GYSYS	YGAYRLSRGVPYG	YSSGPV	NISYGSI	SSYSRYGRTY	SRSRSSYAM
	12	SRYRY	RSASSLSSGVPSY	YSYRLI	NIGSRSI	SGSSRYGGTY	SGSGSSSRYGM
	14	YGSGY	SGASGLSGGVPSG	SYRYPF	NVSGSRI	SGSSGYGGTS	SRSRYSYRYGL
	15	GYSYS	YGAYRLYGGVPSR	SSRSLV	NISSYSI	SRSPRSGRTS	YRYRSGSGL
	16	YSSRY	YGAYGLYRGVPYR	SYSYPI	NLGYRYI	SGSSGSGSTY	GSSYSYRYAF
	17	SSRSR	YSASGLSRGVPSR	YRYRPF	NFRSSYI	YRSPRSGRTY	GSGSRYRGM
	18	GYGSR	YSASRLSRGVPSS	YSYSLL	NFSSGYI	YGSPSSGRTS	SSYRSRYGAM

This table shows the amino acid sequences specific to each Fab. DNA sequencing was completed to detect if any duplicates of Fabs were positively isolated from this Fab library, and identify the true number of unique Fab clones. Information such as this gave insight on CDR region amino acid specificity for the glycine riboswitch. Each column on this table corresponds to a CDR region: three light chain CDRs and three heavy chain CDRs.

Expression and Biochemical Analysis of Fabs of Interest

Each sequenced and unique Fab was expressed and purified as a soluble protein in large quantities. The binding affinity and specificity of each Fab for their respective RNA target was then characterized by native gel shift assay and dot blot assay. Biochemical assays such as these were helpful in determining the potential for a Fab to be a crystallization chaperone, and allowed for direct comparison of Fabs to other Fabs yielded from this library (21).

Materials and Methods for Expression and Purification of Fabs of Interest

The following experimental procedures were carried out to express each selected Fab as a soluble protein. First, the single stranded DNA template was obtained from the harvested and sequenced double stranded DNA that was specific to each Fab. Then, a primer with a stop codon was annealed to the template DNA strand which was specific for each Fab. With the subsequent annealed DNA product, a heat-shock transformation into 34-B8 strain e.coli was then completed, and the resultant cells were grown on an LB/AMP agar plate. The DNA of interest which was contained in the 34-B8 cells was amplified via Colony Polymerase Chain Reaction, and confirmed by an agarose gel ran at 100V. After further culture amplification, the individual Fab proteins were then harvested from culture and purified via immunoaffinity, and ion affinity chromatography.

Biochemical Assays for Characterization of Purified Fabs as Soluble Proteins

Once the Fabs were expressed and purified as soluble proteins, then they were analyzed for their specificity in binding to their RNA target. This was completed by electrophoretic mobility shift assays (EMSA) of the Fab-RNA samples. For the Fabs that proved to be specific-

RNA target binders as soluble proteins, their binding affinity was then quantified. This was completed by dot blot assay.

EMSA materials and methods

First, the RNA target was folded to its native tertiary structure under the following conditions: 0.5 ul 40 uM RNA was added, 6.625 ul 10 mM Tris was added, 7.5 min incubation at 95ºC, then 2 mins at 50ºC, 1 ul 100 mM MgCl2 was then added, 0.375ul 4M NaCl was added, 1 ul 0.1M glycine was added, 50ºC for 20 mins, then room temperature for 10 mins. Then the Fab protein was added.

The last 30min room temperature incubation was carried out just after 0.5 ul of 40 uM of Fab was added, this was to allow time for the Fab proteins to complex with their RNA targets properly before the gel was ran. The assays were carried out immediately after this Fab-RNA complexation time was complete.

To run each sample in the native gel, 2 ul of 5x native gel loading dye was added to each sample (625 ul 80% glycerol, 1 mg xylene, cyanol to final concentration of 0.1%, 375 ul ddH2O). Samples were then ran on a 6% native gel (10 ml 10x TB, 29:1 acrylamide:bisacrylamide, 5 mM $MgCl_2$, 5 mM glycine, 74.5 ml ddH2O, filter through 0.22 um filter). The running buffer was for this assay was 100 ml 10x TB(890mM Tris, 890mM Boric Acid, H_2O), 5 ml 1 M $MgCl_2$, and the gel was run at a constant 120 V for 40 min. The gel was then stained with Ethidium bromide solution to stain the RNA and visualized with the UV illuminator.

Dot blot assay materials and methods

Dot blot assays were completed on each of the soluble Fabs, which demonstrated specificity for their RNA target. This assay was completed to establish the binding affinity between each ^{32}P labeled RNA target and its specific Fab protein.

For the dot blot assay, 50 ul of each of the following concentrations of each Fab was made: 20 uM, 10 uM, 5 uM, 2.5 uM, 1.25 uM, 0.625 uM, 0.3125 uM, 0.156 uM, 78 nM, 39 nM, and the dilutions were done in 1x PBS in a 96 well plate. The nitrocellulose and hybond membranes were equilibrated in PEM/glycine1 buffer (1x PBS, 0.1 mM EDTA, 10mM $MgCl_2$, 10 mM glycine).

Each radiolabeled RNA target was folded to its native teriary structure with the following conditions just before the dot blot assays were carried out. For one 96 well plate, ~650kcpm worth of radiolabeled RNA target was used, and it was diluted with 4546 uL PEM/glycine1 buffer in a 15 ml falcon tube and incubated for 10 min at 50°C. Then the falcon tube with diluted RNA was incubated at room temperature for 10 min. Then, 2 uL of RNase IN was added, and 130 μL 40 mg/mL Heparin was added.

Then, Fab and RNA were incubated for comlpexation: 4 ul of each Fab dilution from above was aliquoted into a new 96 well plate. The first and last columns were left empty. 36 ul of the folded RNA sample was then transferred to each well. The 96 well plate was then covered with sealing tape and shaken gently for 30 min at room temperature.

Then, 100 μL of cold PEM/glycine1 was added to five wells on the dot-blot apparatus. As soon as all five solutions were filtered through the membrane, 40 μL of the Fab-RNA samples

from above were transferred to the corresponding wells on the dot-blot apparatus from the 96 well plate. Pipette tips were then changed and the samples were allowed to fully drain through both membranes in each of the wells by vacuum. Then, 100 μL of cold PEM/glycine1 was added to the same five wells on the dot-blot apparatus. These steps were repeated for the rest of the samples. The membranes were developed in a Phosphor Imager cassette. Then, the Phosphor Image as scanned the next day by imageQuant software and the resultant data was analyzed in Kaliedograph. Each Fab was assayed in duplicate. Results of Biochemical Fab characterization are discussed in the following section.

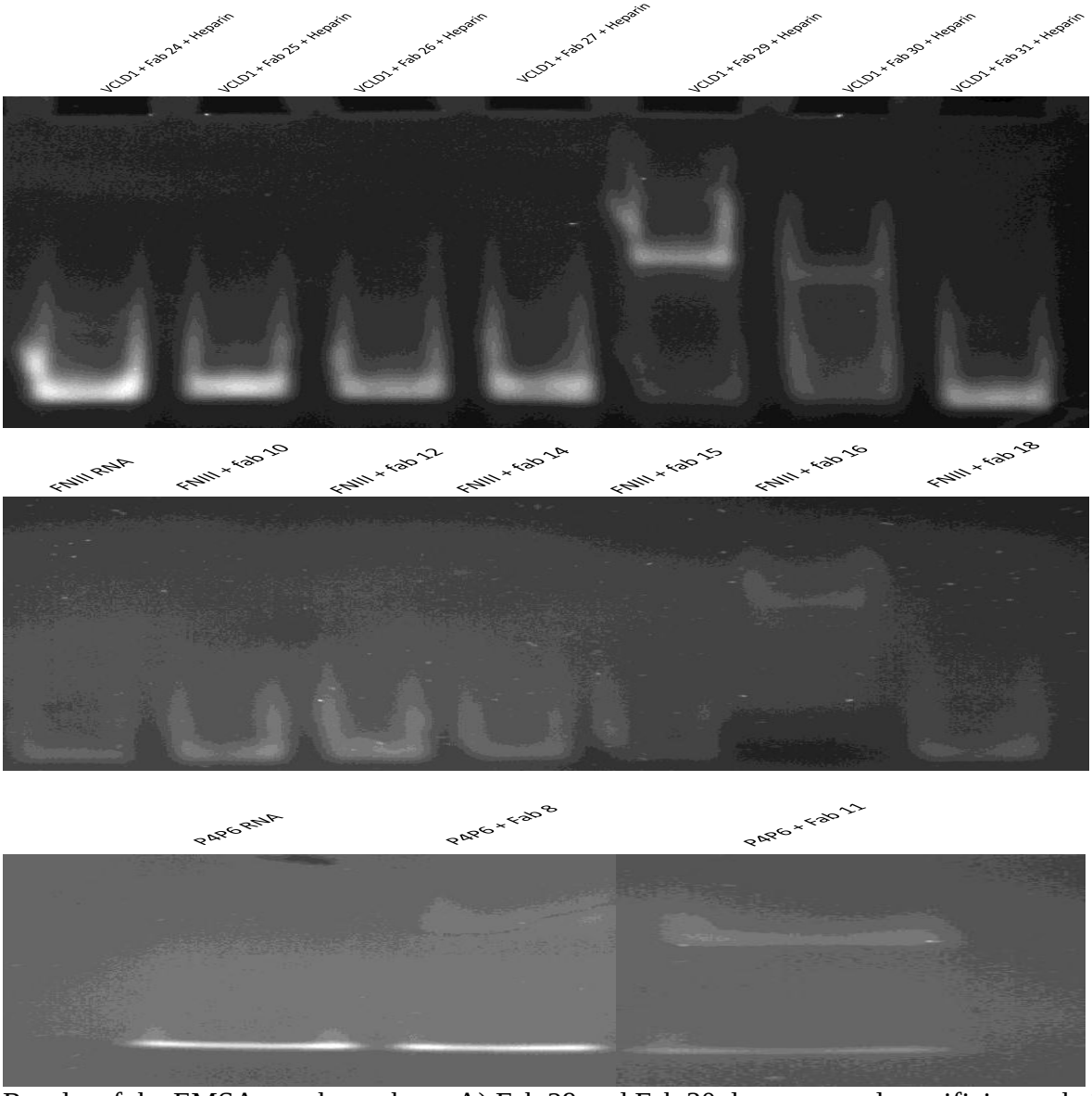
Results of Biochemical Characterization of Each Selected Fab

Results of the EMSA

Native gel shift assays were completed for the unique Fabs identified by phage ELISA and sequencing. Each selected Fab was assayed in complex with its cognate RNA target that it bound during the initial screening. Several Fabs showed RNA target specificity, although many Fabs did not. During the initial screening process each Fab was expressed on the coat of a phage particle, and this was in contrast to each Fab existing as a soluble entity after the expression. Thus this assay was used to reestablish the functionality of each Fab after expression to a soluble protein.

Of all Fabs that were expressed, the Fabs which demonstrated RNA target specificity for the P4P6 RNA target were Fabs 8 and 11. The Fabs which proved RNA target specificity for the VCLD1 RNA target were Fabs 29 and 30. Finally, the only Fab which proved RNA target specificity for the FNIII target was Fab 16. Target specificity was demonstrated by an upwards shift of the RNA in the native gel as it was bound to its Fab (Fig 6).

Figure 6 EMSA results



Results of the EMSA are shown here. A) Fab 29 and Fab 30 demonstrated specificity or their VCLD1 RNA target. B) Fab 16 demonstrated specificity or its FNIII RNA target. C) Fab 8 and Fab 11 demonstrated specificity or their P4P6 RNA target.

Results of the Dot Blot Assays

Dot Blot Assays were carried out on each Fab that demonstrated its functionality. Of the selected Fabs that were assayed via Dot Blot, nanomolar K_d values were achieved for all Fabs with their substituent RNA targets (Table 3). Fabs 8 and 11 for the P4P6 RNA target bound the tightest, and this was expected as P4P6 was used because it has already demonstrated that it can be effectively targeted by Fab proteins (1). Additionally, selected Fabs that were specific for the VCLD1 and FNIII RNA targets have demonstrated comparable K_d values to that of the P4P6 Fabs. This has indicated the true robustness of the YSGR minimal codon Fab library (fig 7).

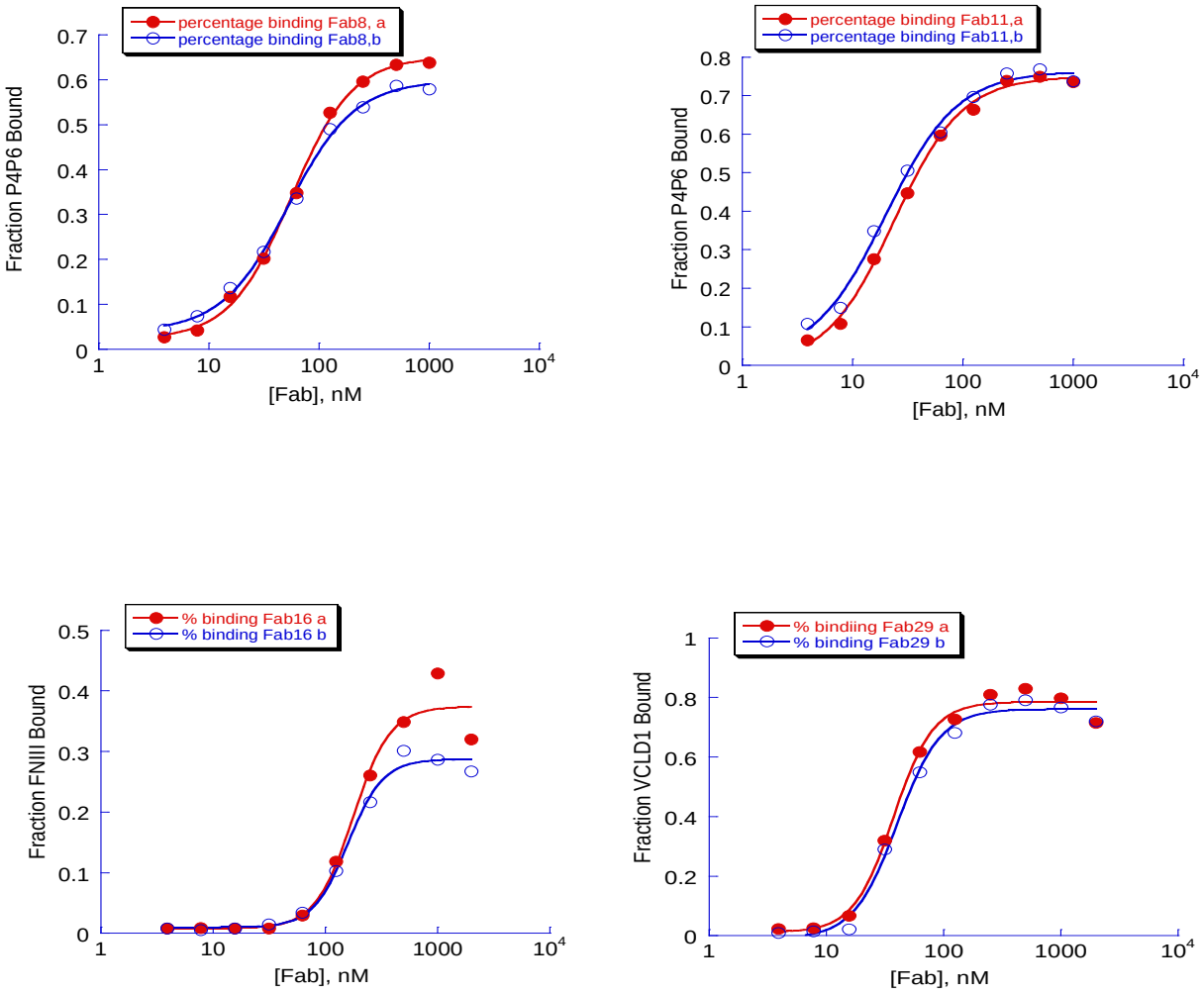
Since Fab 29 and Fab 30 demonstrated satisfactory binding affinities, these Fabs were taken a step further. These Fabs were raised against the VCLD1 riboswitch, and when a mutation was engineered into the VCLD1, it was hypothesized that this mutation should have decreased the binding affinity of Fab 29 and Fab 30 for their target. The results of this experiment are shown in Table 3.

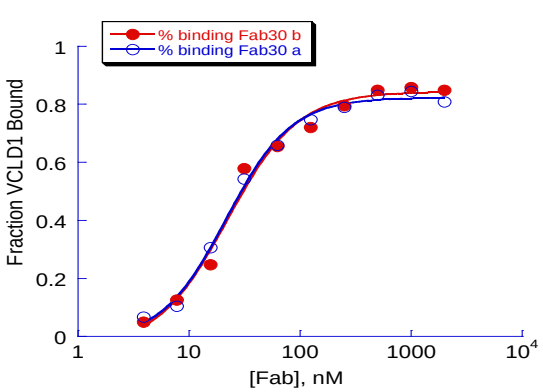
Table 3 K_d Values for each Fab

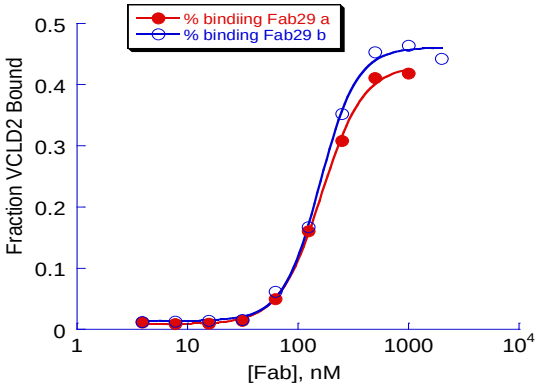
Fab Name	**RNA Target**	**K_d (nM)**
2	P4P6	56.51
8	P4P6	52.56
11	P4P6	20.13
16	FNIII	159.92
29	VCLD1	37.54
30	VCLD1	21.69
29	VCLD2	158.34
30	VCLD2	101.18

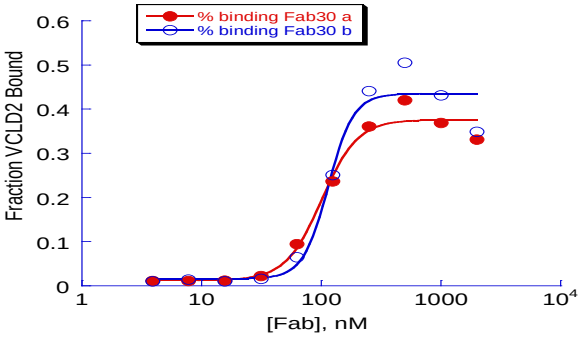
This table shows the nanomolar K_d values that were obtained for each Fab when binding to their cognate RNA. By this data, the functionality of our YSGR minimal codon Fab library was proven to be robust and functional. For the VC glycine riboswitch, a mutation was made to the leader sequence that increased the K_d. This demonstrated that Fab 29 and Fab 30 make specific points of contact with the RNA tertiary structure.

Figure 7 The dot blot assay binding curves









These were the binding curves which were quantified in kaliedagraph and used to extrapolate the K_d values that were shown in table 3.

Discussion

The minimal codon design of this Fab library proved to be sufficient for RNA binding. This was confirmed by a library performance against all three RNA targets that was of equal magnitude. A minimal codon design within a Fab library has been proven here to be successful for enriching Fabs with RNA-specific residues in CDRs. This minimal codon design paired with the phage display initial screening process can be viewed as a model system for obtaining anti-RNA antibodies which can then be used for a variety of clinical and biomedical applications.

After biochemical analysis, five Fab proteins have been identified as possible crystal chaperones. This biochemical analysis has indicated that their binding affinities and specificities are sufficient for potential downstream applications. This analysis has also further confirmed that a minimally designed Fab library can yield functional Fab proteins.

This project has demonstrated that it is possible to obtain specific RNA-binding antibodies by anti-RNA library designs and phage display screening. Anti-RNA antibodies have many applications including their use to identify novel RNA disease markers within the cell, capabilities in RNA immunoprecipitations, and capabilities in visualization of RNA. Furthermore, the success of this Fab library with only four amino acid codons at the RNA binding interface greatly improved our understanding of the minimal requirements needed for RNA-Fab interactions with high affinity and specificity.

References

1. Ye, J. et al. Synthetic antibodies for Specific Recognition and Crystallization of Structured RNA. *PNAS* **105**, 82-87 (2008).

2. Fellouse, F.A., Barthelemy, P.A., Kelley, R.F. & Sidhu, S.S. Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. *J Mol Biol* **357**, 100-114 (2006).

3. Birtalan, S. et al. The intrinsic contributions of tyrosine, serine, glycine and arginine to the affinity and specificity of antibodies. *J Mol Biol* **377**, 1518-1528 (2008).

4. Sidhu, S.S. & Kossiakoff, A.A. Exploring and designing protein function with restricted diversity. *Curr Opin Chemical Biology* **11**, 347-354 (2007).

5. Fellouse, F.A. et al. Molecular recognition by a binary code. *J Mol Biol* **348**, 1153-1162 (2005).

6. Fellouse, F.A. et al. High-Throughput Generation of Synthetic Antibodies from Highly Functional Minimalist Phage-Displayed Libraries. *J. Mol. Biol.* **373**, 924-940 (2007).

7. Kabat, E.A. & Wu, T.T. Attempts to Locate Complementariry Determining Residues in the Variable Positions of Light and Heavy Chains. *Annals New York Academy of Sciences* **190**, 382-393 (2006).

8. Kunkel, Roberts, & Zakour. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods in Enzymology* **154**, 368-382 (1987).

9. Mandal, M. et al. A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression. *Science* **306,** 275-279 (2004).

10. Winkler, W.C. & Breaker, R.R. Regulation of bacterial gene expression by riboswitches. *Annual Review Microbiology* **59**, 487-517 (2005).

11. R. & Breaker, R.R. Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis. *RNA* **13**, 573-82 (2007).

12. Edwards, T.E., Klein, D.J. & Ferre-D'Amare, A.R. Riboswitches: small-molecule recognition by gene regulatory RNAs. *Curr Opin Structural Biology* **17**, 273-279 (2007).

13. Blount, K.F. & Breaker, R.R. Riboswitches as antibacterial drug targets. *Nature Biotechnology* **24**, 1558-1564 (2006).

14. Center for Disease Control and Prevention: Antibiotic/ Antimicrobial Resistance. [http://www.cdc.gov/drugresistance/index.html]

15. Kwon, M. & Strobel, S.A. Chemical basis of glycine riboswitch cooperativity. *RNA* **14**, 25-34 (2007).

16. Ke, A. & Doudna, J.A. Crystallization of RNA and RNA-protein complexes. *Methods* **34**, 408-414 (2004).

17. Sidhu, S.S., Lowman, H.B., Cunningham, B.C. & Wells, J.A. Phage display for selection of novel binding peptides. *Methods Enzymol* **328**, 333-363 (2000).

18. Laird-Offringa, I.A. & Belasco, J.G. In vitro genetic analysis of RNA-binding proteins using phage display libraries. *Methods Enzymol* **267**, 149-168 (1996).

19. Griffiths, A.D. et al. Isolation of high affinity human antibodies directly from large synthetic repertoires. *Embo J* **13**, 3245-3260 (1994).

20. Agris, P.F. et al. Experimental models of protein-RNA interaction: isolation and analyses of tRNA(Phe) and U1 snRNA-binding peptides from bacteriophage display libraries. *J Protein Chem* **18**, 425-435 (1999).

21. Katsamba, P.S., Park, S. & Laird-Offringa, I.A. Kinetic studies of RNA-protein interactions using surface plasmon resonance. *Methods* **26**, 95-104 (2002).