STRUCTURAL AND FUNCTIONAL STUDIES OF GLYCINE RIBOSWITCHES AND DEVELOPMENT OF FAB CHAPERONE ASSISTED RNA CRYSTALLOGRAPHY

by

EILEEN SHERMAN
AA&AS Lorain County Community College 2001
BS Ohio University 2004

Dissertation submitted in partial fulfillment of requirements
for the degree of Doctor of Philosophy
in the Department of Chemistry
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2014

Major Professor: Jingdong Ye

ABSTRACT

The glycine riboswitch is a structured RNA found upstream of genes in mRNA transcripts in many bacteria, functioning as a biofeedback gene regulator. Upon binding glycine, a complete RNA transcript including gene sequences is transcribed, effectively turning on gene expression. In an effort to understand the intricacies of its functioning, many mutants of the riboswitch were made and characterized during Ph. D. work, resulting in discovery of a P0 duplex/kink-turn motif involving a few nucleotides upstream of the established glycine riboswitch sequence which changed its ligand binding characteristics (Chapter 1). Previously, the two aptamers of the riboswitch were thought to cooperatively bind glycine, but with the inclusion of this leader sequence which forms a kink turn motif with the linker between the two aptamers, glycine binding in one aptamer no longer requires glycine binding in the other. Furthermore, the K_d from three species tested are now a similar, lower value of about 5 µM, indicating authenticity of this new consensus sequence. Glycine binding and interaptamer interaction both enhanced one another in trans aptamer assays. Another discovery from this was a shortened construct including all of aptamer II but only part of aptamer I in which a few specific nucleotides prevented glycine binding in aptamer II (Chapter 2). This may provide insight into the nature of interaptamer interactions in the full switch; addition of an oligonucleotide complimentary to these nucleotides restored glycine binding ability to aptamer II. With future development, this could also be a useful molecular biology tool, using two signals, glycine and an oligonucleotide, to allow gene expression.

To precisely understand how any macromolecule functions, a 3D structure, obtainable by x-ray crystallography, is vital. A new technique to accomplish that for RNA, precedented in the protein world, is Fab chaperoned crystallography, which has advantages compared to RNA alone. A phage displayed library of Fabs with reduced codon diversity designed for RNA was created, the YSGR Min library (Chapter 3). Its Fabs had specificities and affinities equal to or

greater than previous libraries which were originally created for phage displayed selection against proteins. Fab chaperoned RNA crystallography is currently in progress for the glycine riboswitch; the best resolution thus far is 5.3 Å (Chapter 4). In addition to providing molecular insight into its gene regulation mechanism, a structure of the glycine riboswitch could be applied for use in structure based drug design of novel antibiotics targeting the riboswitch to disrupt important downstream carbon cycle genes in pathogenic bacteria.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 AN ENERGETICALLY BENEFICIAL LEADER-LINKER INTERACTION ABOLISHES LIGAND COOPERATIVITY IN GLYCINE RIBOSWITCHES

1.1 Introduction

Riboswitches are *cis*-acting mRNA elements that regulate gene expression by direct binding to small molecule metabolites (1). Upon binding of the biologically important metabolites, riboswitches undergo conformational changes which may affect the structure of the associated expression platforms and control gene expression through transcriptional termination, translational initiation, or alternative splicing. Many riboswitches have been computationally discovered and experimentally verified (1, 2). Most of the currently reported 24 classes of riboswitches are found in the 5' UTRs of bacterial genes, though TPP riboswitches are also present in plants and fungi (3, 4). They vary widely in their structures and functional mechanisms, discriminatively sensing a plethora of metabolites (1, 2, 5).

Glycine riboswitches are among the earliest discovered riboswitches (6). Using BLASTN searches followed by iterative secondary structure refinement and extension with the SEQUENCESNIFFER program, Breaker and coworkers identified a *gcvT* motif that resides upstream of bacterial genes expressing enzymes involved in the metabolism of glycine (6, 7). The *gcvT* motif, termed the glycine riboswitch, forms an on-switch when bound to glycine, controlling the expression of enzymes that catalyze the initial reactions for the use of glycine as an energy source. These enzymes are components of the glycine cleavage system which breaks down glycine and produces carbon dioxide, ammonia, and 5-10-methylenetetrahydrofolate in the presence of tetrahydrofolate (8). 5-10-methylenetetrahydrofolate serves as sources of the methyl groups for methionine and 2'-deoxythymidine biosynthesis (9). In addition, 5-10-methylenetetrahydrofolate can also serve as the substrate in the reversible conversion of glycine to serine. Combined together, glycine and serine account for up to 15% of carbon assimilated by *E. coli* growing on glucose (10). Besides

their direct roles in protein synthesis, glycine and serine serve as precursors for purines, heme groups, cysteine, phosphatidylserine, enterochelin and tryptophan (11, 12). Thus, the glycine riboswitch self-regulates by biofeedback an important metabolite that serves as precursor to one of the main biosynthetic trunks in bacteria (12).

Interestingly, glycine riboswitches contain two tandem glycine-sensing domains and purportedly employ, uniquely, ligand binding cooperativity. The most biochemically characterized construct, *Vibrio cholerae* (VC) glycine riboswitch was reported to bind glycine with a Hill coefficient of 1.5 to 1.6 (6, 13, 14), the largest amongst all reported glycine riboswitches. Deleterious mutations in each aptamer of the VC glycine riboswitch also greatly reduced the glycine binding affinity in the other aptamer, providing direct evidence of the ligand binding cooperativity of the tandem riboswitch (6, 14). As the first example of naturally occurring RNA elements reported to bind small organic molecules cooperatively (6, 15), glycine riboswitches have inspired many biochemical and biophysical studies, including nucleotide analog interference mapping (NAIM) to investigate the chemical basis of the cooperativity (13) and small-angle X-ray scattering (SAXS) to dissect the roles of metal ions and ligand binding in structural transitions (16, 17). Recent efforts have resulted in the proposal of the asymmetrical model of glycine binding cooperativity (14) as well as crystal structures of the single aptamer VCII with an extended P1 helix (18) and the tandem glycine riboswitch from *Fusobacterium nucleatum* (19).

In working with the tandem VC glycine riboswitch, we have found our minimal construct exhibited a distinctive protection pattern in the linker region compared to literature results. This led to the discovery of a surprising leader-linker interaction involving nucleotides upstream of the reported consensus glycine riboswitch sequences. We have defined the energetic benefit of the leader-linker interaction through in-line probing of the tandem riboswitches and further characterized its effect on glycine binding and inter-aptamer interaction using biochemical assays with two aptamers in *trans*. Our mutational analysis surprisingly showed inter-aptamer

inhibition does not exist in the full-length VC glycine riboswitch. Together, our data suggest glycine binding and inter-aptamer interaction synergize to fold the global structure of the glycine riboswitch in which its two aptamers do not bind glycine cooperatively.

1.2 Results

1.2.1 Discovery of the leader-linker interaction

As RNA tertiary structures are generally more flexible than proteins, to reduce terminal entropy and increase crystallizability of RNAs, crystallographers generally remove the single stranded flanking regions from their RNA constructs. In an attempt to crystallize the VC tandem glycine riboswitch, we have constructed the minimal sequence VCI-IIs (nts 7 to 225, Fig. 1.1A). The overall glycine binding characteristics of VCI-IIs (K_d = 47 ± 14 µM, Table 1.1, Fig. 1.2) are very similar to VCI-II (nts 1 to 225, K_d = 59 ± 12 µM, Table 1.1). In the linker region (nts 135 to 137), VCI-IIs remained unstructured independent of glycine concentration (Fig. 1.2) while VCI-II exhibited local structure flexibility with increasing glycine concentration (6). This suggested that its 5' leader sequence may influence the structure of the linker region. One simple scenario is direct duplex formation between the leader and the linker region, which proved correct upon sequence retrieval from the NCBI database and manual comparison of the leader with the linker sequences. Thus, a leader-linker duplex (termed P0) beyond the published consensus glycine sequence was unearthed. Supported by Mfold (20), the P0 helix contains 3-7 base pairs and is 96% conserved (53 out of 55) in the tandem glycine riboswitch sequences reported by Breaker and coworkers (6). In addition, the P0 helix is present in 91% of the 22 tandem glycine riboswitch seed sequences in Rfam (21). Further analysis showed that in 40% of these glycine riboswitches, P0 helices are connected to P1 through simple internal loops/bulges (Fig. 1.3). In the rest (over 50%) of the glycine riboswitch sequences (Fig. 1.1B), however, P0 helices are connected to P1 through kink-turn motifs (22-25). Three characteristics show that these motifs fit the pattern of the consensus kink-turns derived from the ribosome kink-turns: 1) The C stems

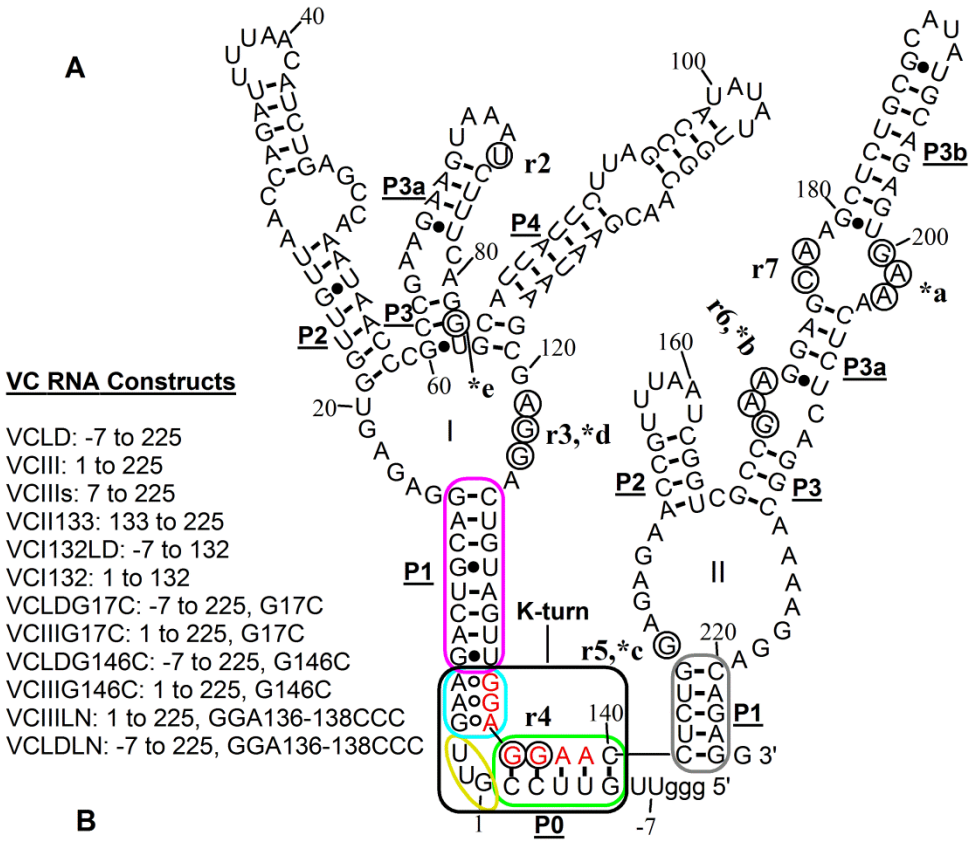
most often end with two C-G/G-C pairs; 2) The internal loop follows the RNN(N) consensus rule; 3) the NC stems usually start with two sheared G-A/A-G base pairs. Interestingly, the canonical base pair adjacent to the internal loop (-1b/-1n pair) has a strong preference for G-C, just the opposite to the C-G base pair preference in the ribosome, indicating a possible deviation from the consensus kink-turn motif, as the G-C pair may not support the type I A-minor interaction as well as C-G (22).

Table 1.1 Binding affinities and Hill coefficient values of glycine riboswitches and mutants[a]

RNA	K_d (μM)	Hill coefficient (n)	Description of the constructs
VCI-II	59 ± 12 64 ± 17[b,c]	1.3 ± 0.2	nts 1 to 225
VCI-IIs	47 ± 14	1.1 ± 0.2	nts 7 to 225
VCLD	5.0 ± 4.0	0.6 ± 0.2	nts -7 to 225
VCI-IILN	15 ± 3	1.4 ± 0.2	nts 1 to 225, GGA136-138CCC
VCLDLN	61 ± 11	1.3 ± 0.2	nts -7 to 225, GGA136-138CCC
VCII133[d]	(2.0 ± 0.1) × 10^3	≤ 1	nts 133 to 225
VCII133 (+VCI132)[d]	191 ± 44	≤ 1	VCII133 with 15 μM VCI132
VCII133 (+VCI132LD)[d]	7.4 ± 1.0	≤ 1	VCII133 with 15 μM VCI132LD
VCI132 (+VCII133)[e]	(7.4 ± 1.0) × 10^3	≤ 1	nts 1 to 132, with 15 μM VCII133
VCI132LD (+VCII133)[e]	190 ± 41	≤ 1	nts -7 to 132, with 15 μM VCII133
VCI-IIG17C[d]	(1.1 ± 0.1) × 10^3	≤ 1	nts 1 to 225, G17C
VCI-IIG146C	(1.9 ± 0.3) × 10^3	≤ 1	nts 1 to 225, G146C
VCLDG17C[d]	2.8 ± 0.4	≤ 1	nts -7 to 225, G17C
VCLDG146C[e]	12 ± 2	≤ 1	nts -7 to 225, G146C
FNI-II	14 ± 6 8.6 ± 0.4[b]	1.1 ± 0.1 0.9 ± 0.1	nts 1 to 158
FNLD	3.1 ± 2.5	0.7 ± 0.2	nts -11 to 158
BSI-II	311 ± 158[c]		nts 13 to 202
BSLD	3.6 ± 3.4	0.8 ± 0.2	nts 1 to 202

[a]Determined by in-line probing assay unless indicated otherwise. Binding affinities were calculated as average K_d values obtained from the individual glycine perturbed regions in at least two independent trials with curve fit coefficients better than 0.98. Hill coefficients were calculated in a similar way but only from those with curve fit coefficients better than 0.99.
[b]Determined by hydroxyl radical footprinting.
[c]Average Hill coefficient is not determined as the curve fit coefficients are less than 0.99.
[d]Binding parameters were calculated for aptamer II regions.
[e]Binding parameters were calculated for aptamer I regions.

A



B

```
Vch: ...UUGUUCCGUU-GAAGACUGCAG.....CUGUAGUUGGAGGAAC----CUCUG...
Fnu: ...AUAAUCGGAU-GAAGAUAUGAG.....CUCAUAUUGGACGAAC----CUCUG...
Bsu: ...AUGAGCGAAU-GACAGCAAGGG.....CUCUUGCUCGACGCAA----CUCUG...
Ban: ...UAAUUCGGAU-GAACCAUUCAG.....CUGCAUGGGGACGAAU----CUCUG...
Bce: ...UAAUUCGGAU-GAACCAUUCAG.....CUGCAUGGGGACGAAU----CUCUG...
Bha: ...AAAAGCGGAU-GAAAGCAAGGG.....CUCUUGCUCGACGCAGGCAGCUCUG...
Cac: ...UUGUUCAGAU-GAAGUUAGCGG.....CCGCUAUUGGAUGAAC----CCUUG...
Cpe: ...AUGUUCGGAU-GAAGGUAAUGG.....CCGUUACUGGACGAGC----CCUUG...
Cte: ...UAUUCUAGUU-GAAGAGUAUAA.....UUAUACUUUGAUAGA-----CUCUG...
Cte: ...AGGUUCGGAU-GAAGGUAAUAG.....CUGUUAUUGGACGAAA----CUCUG...
Lin: ...AUUAGCGGGU-GAAUGUAAACA.....UGUUUACGGGACGCAA----CUCUG...
Lmo: ...AUUAGCGGGU-GAAUGUAAGCA.....UGUUUACGGGACGCAA----CUCUG...
Oih: ...AUCAGCGGAA-GAUUACAAGGG.....CUCUUGUAUGACGCAA----CUCUG...
Sau: ...AAAAGCGAGU-GAUCAGUAUUA.....UAAUACUGUAACGCGU----UCCUG...
Sep: ...UUAAGCGAUU-GAUCAGUAUUA.....UAAUACUGUAACGCAU----UCCUG...
Sag: ....AUACCGAAU-GACGUCAUUCAG...CUGAUUGACAGACGGAC----UUCUG...
Spy: ....AAACCGAAU-GAUGUCAUGCAG...CUGUUUGAUGGACGGAC----UUCUG...
Sav: .....CGGCCGCUUGAAUCCGCGCGG...CCGCGCGGGCGAGGCAC----AUCUG...
Sav: ....AUGGCUGCU-GACCCCGUGCGG...CCGCACGGACGAGGUCA----CUCUG...
Tte: ...CUGUCUAGAU-GAAGAUAGCGA.....UCGCUAUCGGAUAGGC----CUCUG...
Tte: ...GAAAUGGGAU-GAAGAAUGCGG.....CCGCAUUCGGACCAUA----UCCCG...
Hdu: ...ACGUUCGGAU-GAAGGUGGCUG.....CAGUUACUGGACGAAC----CCUUG...
Hin: ...GAGUUCGGAU-GAAGGUAGCUG.....CUGUUACUGGACGAAC----CCUUG...
Pmu: ...AAGUUCGGAU-GAAGGUAGCUG.....CUGUUACUGGACGAAC----CCUUG...
Son: ...CCGUUCGGAU-GAAGGUAGCAG.....CUGUUACUGGACGAGC----CUCUG...
Vpa: ...AUGUUCCGAU-GAAGACUGCAG.....CUGUAGUUGGAGGAAC----CUCUG...
Vvu: ...UUGUUCCGAU-GAAGGCUACAG.....CUGUAGUUGGAGGAAC----CUCUG...
Mbo: ....ACAGCCGCU-GACGCGAUGUGG...CCACACCGCCGAGGCAA----CUCUG...
Mtu: ....ACAGCCGCU-GACGCGAUGUGG...CCACACCGCCGAGGCAA----CUCUG...
Sco: ....ACGGCUGCU-GACCCCGCGCGG...CCGCACGGACGAGGUCA----CUCUG...
Sco: .....CGGCCGUUUGAAUCCGCGCGG...CCGCGCGGGCGAGGCAC----AUCUG...
```

Figure 1.1 VC glycine riboswitch secondary structure and leader linker kink turn sequence alignment.

(*A*) Secondary structure of the full-length VC glycine riboswitch. r2-r7 and *a-*b indicate the representative glycine induced perturbation regions in the in-line probing assay and the hydroxyl radical footprinting assay of VCI-II, respectively; the corresponding nucleotides are circled. “I” and “II” refer to aptamer I and aptamer II, respectively. Green, yellow, cyan, purple, and grey regions indicate C stem (P0 helix), internal loop, NC stem, P1 helix of aptamer I, and P1 helix of aptamer II, respectively. The kink-turn motif is shown in a black box. Red nucleotides indicate previously defined linker sequence. (*B*) Alignment of the kink-turn motifs formed between leader and linker regions in 31 of 55 tandem glycine riboswitch sequences reported by Breaker and coworkers (Mandal et al. 2004). Abbreviations of the names of the organisms are taken from the same article.

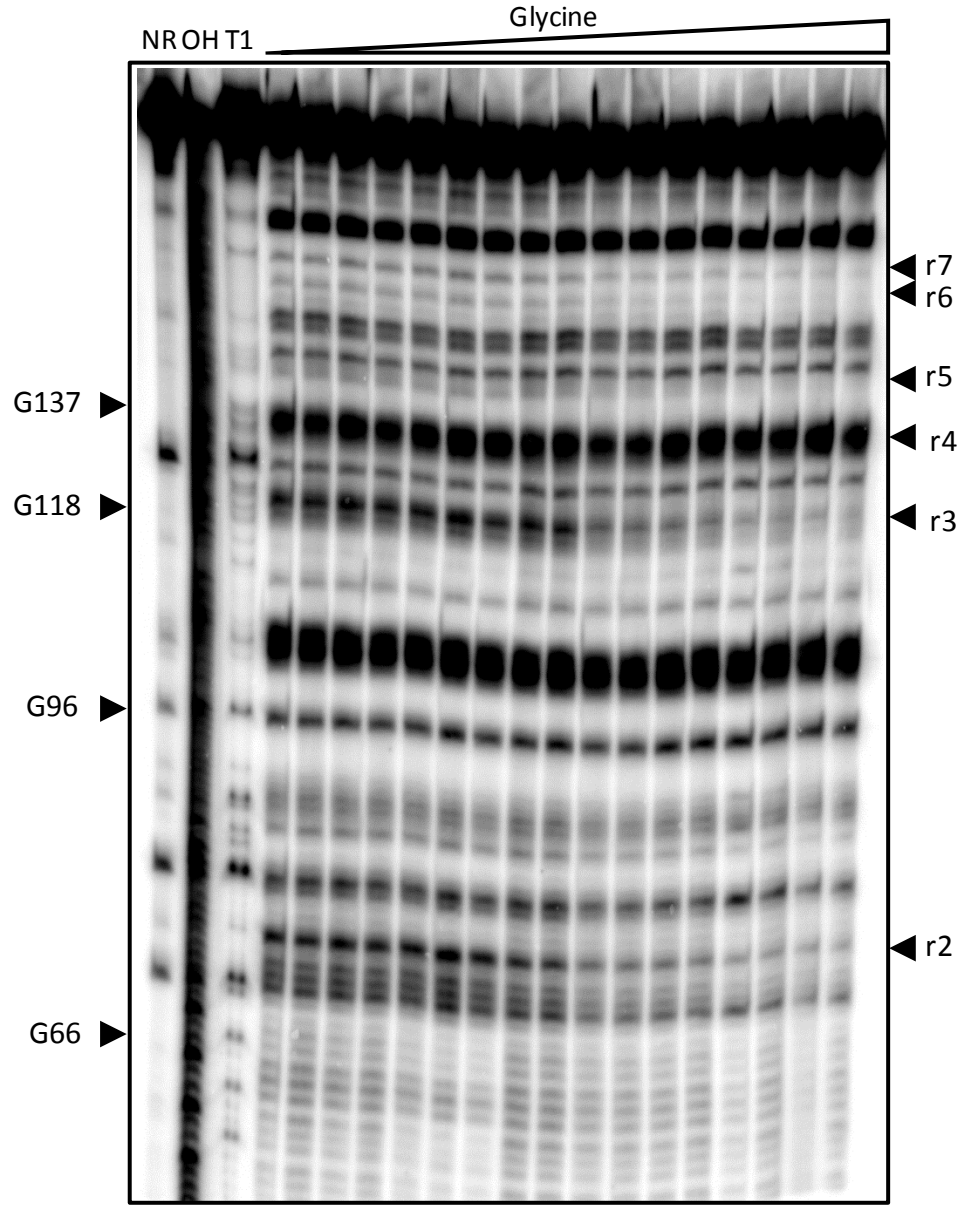


Figure 1.2 Representative in-line probing gel image of VCIIIs.

NR, T1, and OH represent no reaction, partial digestion with RNase T1, and partial digestion with alkali, respectively. r2-r7 indicate regions defined in Table 1.2.

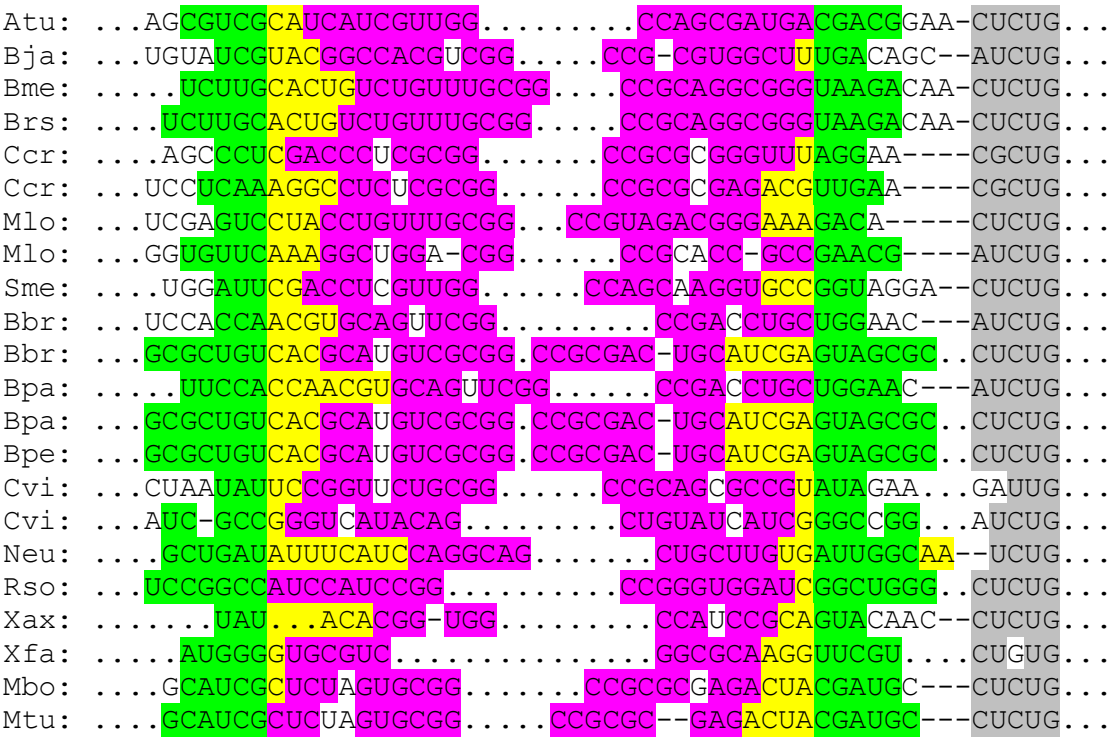


Figure 1.3 Leader linker sequence alignment for loops.

Alignment of leader-linker P0 duplex connected to P1 duplex with simple internal loops in 22 of 55 tandem glycine riboswitch sequences reported by Breaker and coworkers (Mandal et al. 2004). Abbreviations of the names of the organisms are taken from the same article. Green, yellow, purple, and grey regions indicate P0 duplex, internal loop, P1 helix of aptamer I, and P1 helix of aptamer II, respectively.

1.2.2 The global structures of the full-length glycine riboswitches are similar to those of the truncated constructs

To evaluate the global structures of the full-length glycine riboswitches, we have constructed both full-length and truncated versions of the glycine riboswitches from three organisms: *Vibrio cholerae* [VCLD (nts -7 to 225) and VCI-II], *Fusobacterium nucleatum* [FNLD (nts -11 to 158) and FNI-II (nts 1 to 158)] and *Bacillus subtilis* [BSLD (nts 1 to 202) and BSI-II (nts 13 to 202)]. These RNA sequences are downstream of the transcription start sites predicted by Prokaryotic Promoter Analysis Using Sequence Alignment Kernel software (26). Interestingly, the *Bacillus subtilis* glycine riboswitch previously reported (6) contains the leader sequence although the leader-linker interaction was not recognized. In-line probing assays map the local structure changes by measuring the flexibility of the phosphodiester bond to achieve in-line geometry for non-enzymatic cleavage through internal transphosphorylation (27, 28). Under standard in-line probing conditions (50 mM Tris-HCl, pH 8.3, 20 mM $MgCl_2$, 100 mM KCl, 25 ^{o}C) with saturating glycine concentrations, the full-length glycine riboswitches exhibit the same protection patterns as the truncated ones outside of the linker region in all three VC, FN and BS glycine riboswitches (Fig. 1.4A, 1.5, 1.6, 1.7). To compare the regions directly, we calculated the relative band intensities of the major cleavage bands and known glycine-perturbed regions. Results showed that these bands have similar relative intensities within any given construct in the in-line probing assay of glycine riboswitches with or without leader sequences (Table 1.2). In the linker region, the full-length glycine riboswitches are protected from spontaneous cleavage while the truncated glycine riboswitches show cleavage bands, consistent with a leader-linker interaction. Therefore, in-line probing assays showed that except for the linker regions, the full-length glycine riboswitch folds into the same structure as the truncated constructs at high magnesium and glycine concentrations.

Table 1.2 Comparison of glycine riboswitches' protection patterns with and without the leader[a]

Regions	VCI-II[b]	VCLD[b]	FNI-II[c]	FNLD[c]	BSI-II[d]	BSLD[d]
r7	0.14 ± 0.01	0.13 ± 0.05	0.10 ± 0.03	0.10 ± 0.04	0.78 ± 0.13	0.42 ± 0.03
r6	0.10 ± 0.02	0.13 ± 0.06	0.07 ± 0.01	0.08 ± 0.03	0.29 ± 0.02	0.11 ± 0.02
r5	0.12 ± 0.01	0.12 ± 0.07	0.04 ± 0.01	0.09 ± 0.08	0.21 ± 0.01	0.16 ± 0.01
r4[e]	1.90 ± 0.50	0.11 ± 0.07	11.4 ± 6.1	0.11 ± 0.05	2.2 ± 0.3	0.20 ± 0.02
r3	0.29 ± 0.09	0.16 ± 0.06	0.03 ± 0.01	0.02 ± 0.02	0.64 ± 0.21[f]	0.13 ± 0.02[f]
r2	0.18 ± 0.02	0.12 ± 0.02	0.18 ± 0.02	0.08 ± 0.07	0.23 ± 0.07[f]	0.05 ± 0.01[f]
ref1	1	1	1	1	1	1
ref2	0.72 ± 0.09	0.65 ± 0.06	5.1 ± 1.6	4.6 ± 0.1	1.2 ± 0.1	1.1 ± 0.1

[a]Determined as relative intensities of each region normalized to non-variable band ref1 at 20 mM magnesium and 3.16 mM glycine in in-line probing assay. Data are averaged from at least two independent trials.

[b]For VCI-II and VCLD, regions are as the following: C177-A178 (r7), G170-A172 (r6), G146 (r5), G136-G137 (r4), A121-G123 (r3), U74 (r2), U156-U161 (ref1), and U94 (ref2), which are indicated in Figure 1.1.

[c]For FNI-II and FNLD, regions are as the following: C117-A119 (r7), G110-A112 (r6), G85 (r5), G72-G76 (r4), G61 (r3), U46 (r2), C90 (ref1), and U96-A100 (ref2).

[d]For BSI-II and BSLD, regions are as the following: A158-G160 (r7), A151-A153 (r6), G121 (r5), G109-A115 (r4), A93-A96 (r3), U83 (r2), G136-U138 (ref1), and U126 (ref2).

[e]Data in r4, the linker regions, are shown in bold to show that the truncated glycine riboswitches are cleaved much more than the full-length constructs.

[f]In some regions, BSI-II showed 5 fold more cleavage than BSLD, indicating BSI-II may be more flexible and dynamic.

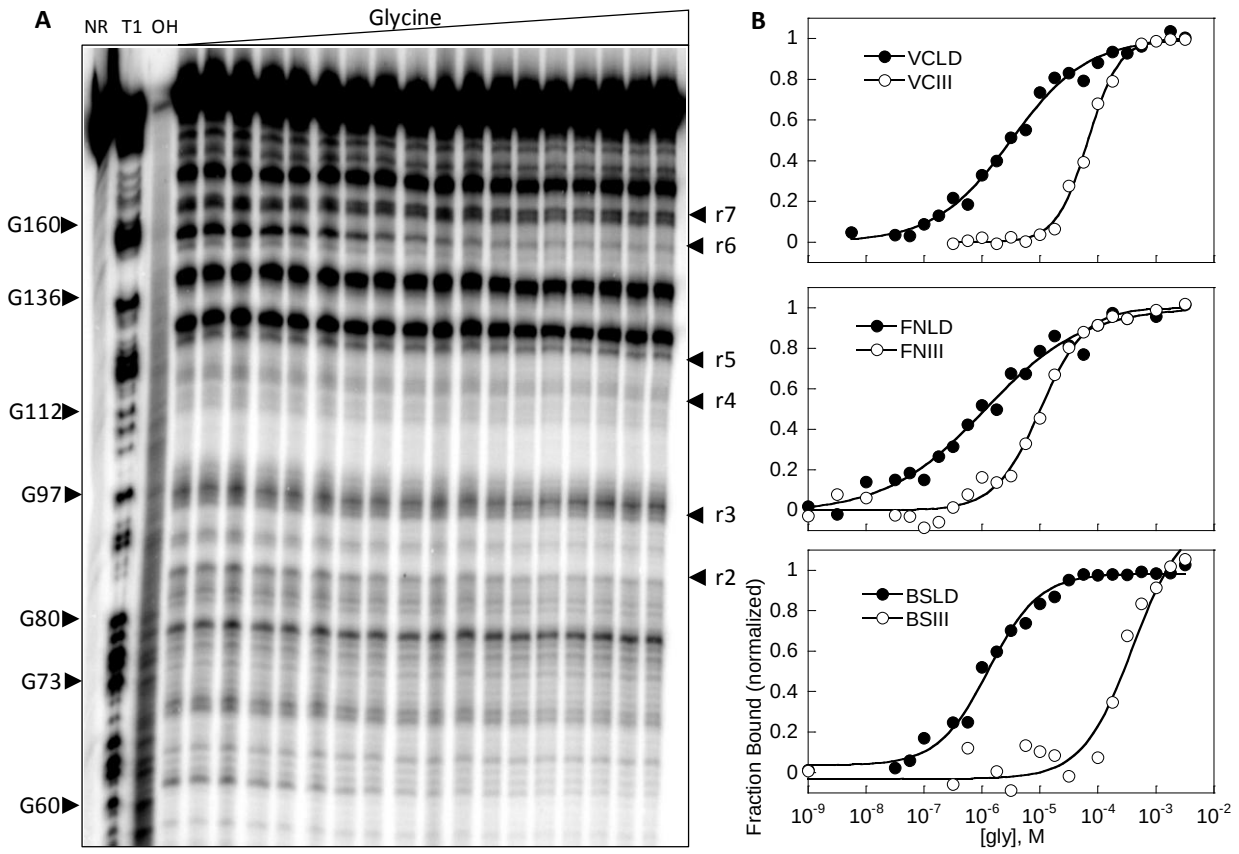


Figure 1.4. Semi-log K_d plots comparing in-line probing of full-length and truncated glycine riboswitches.

(A) Representative in-line probing gel image of BSLD. NR, T1, and OH represent no reaction, partial digestion with RNase T1, and partial digestion with alkali, respectively. r2-r7 indicate regions defined in Table 1.2. (B) Representative semi-log plots of the normalized fraction of RNA bound versus glycine concentration. Shown from top to bottom: VCLD/VCI-II (r3); FNLD/FNI-II (r6); BSLD/BSI-II (r6).

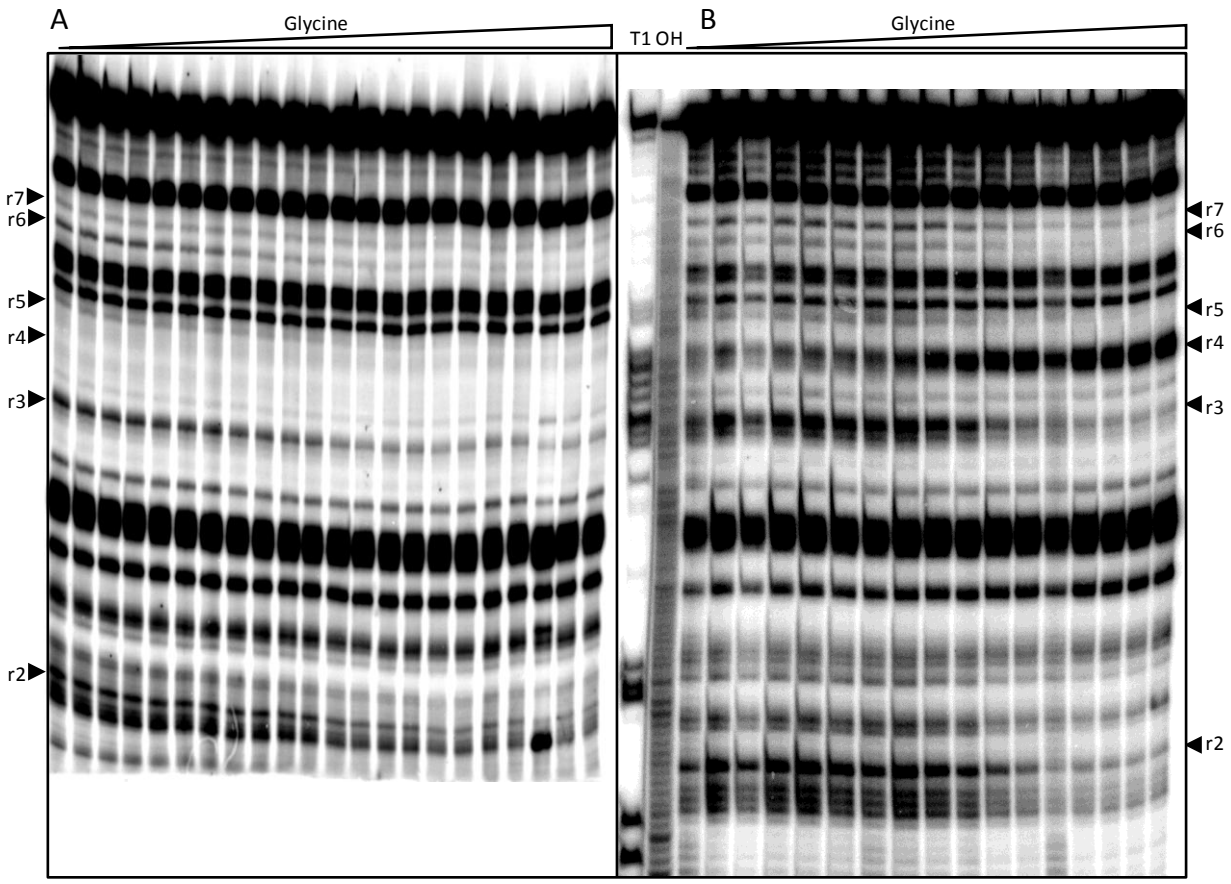


Figure 1.5 In-line probing gel images of VCLD (*A*) and VCI-II (*B*).

T1 and OH represent partial digestion with RNase T1 and partial digestion with alkali, respectively. r2-r7 indicate regions defined in Table 1.2.

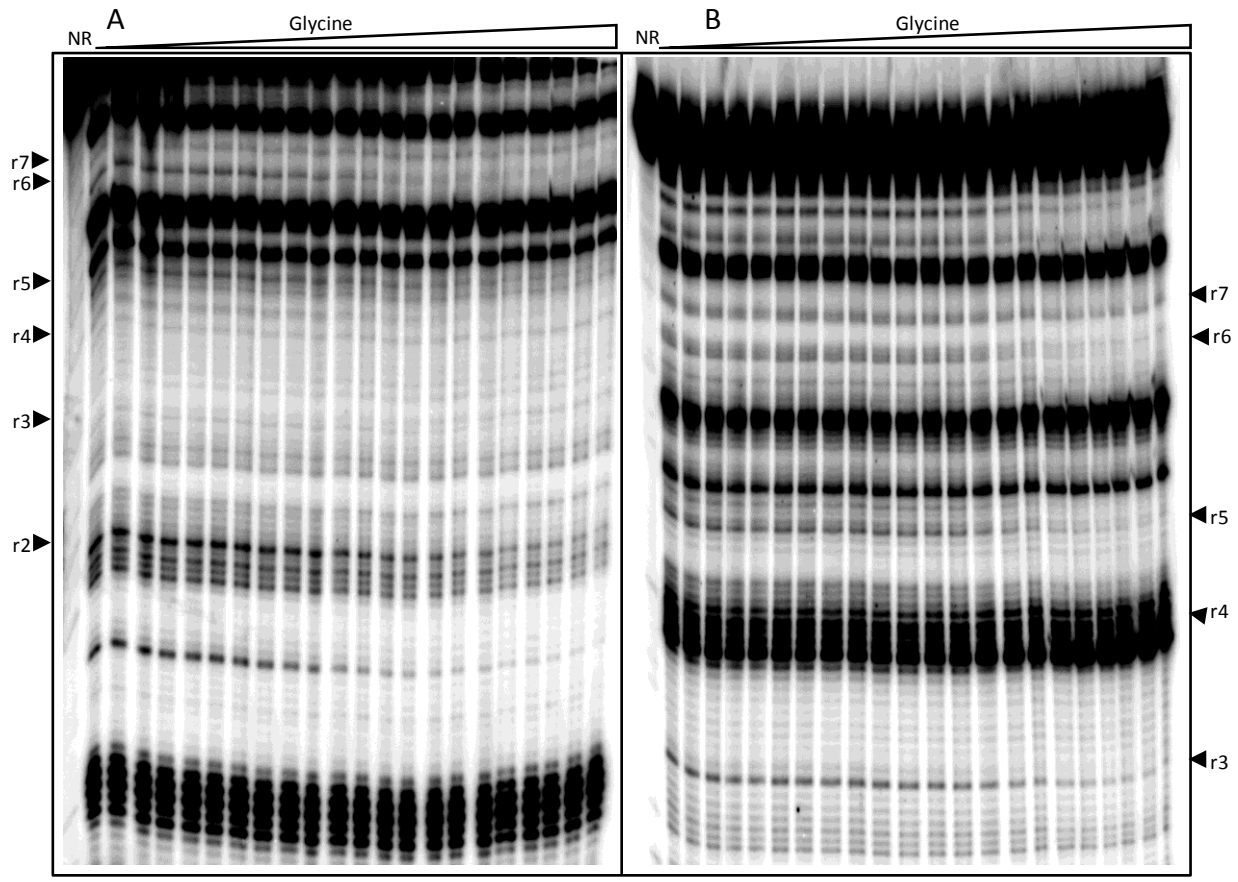


Figure 1.6 In-line probing gel images of FNLD (*A*) and FNI-II (*B*).

NR represents no reaction. r2-r7 indicate regions defined in Table 1.2.

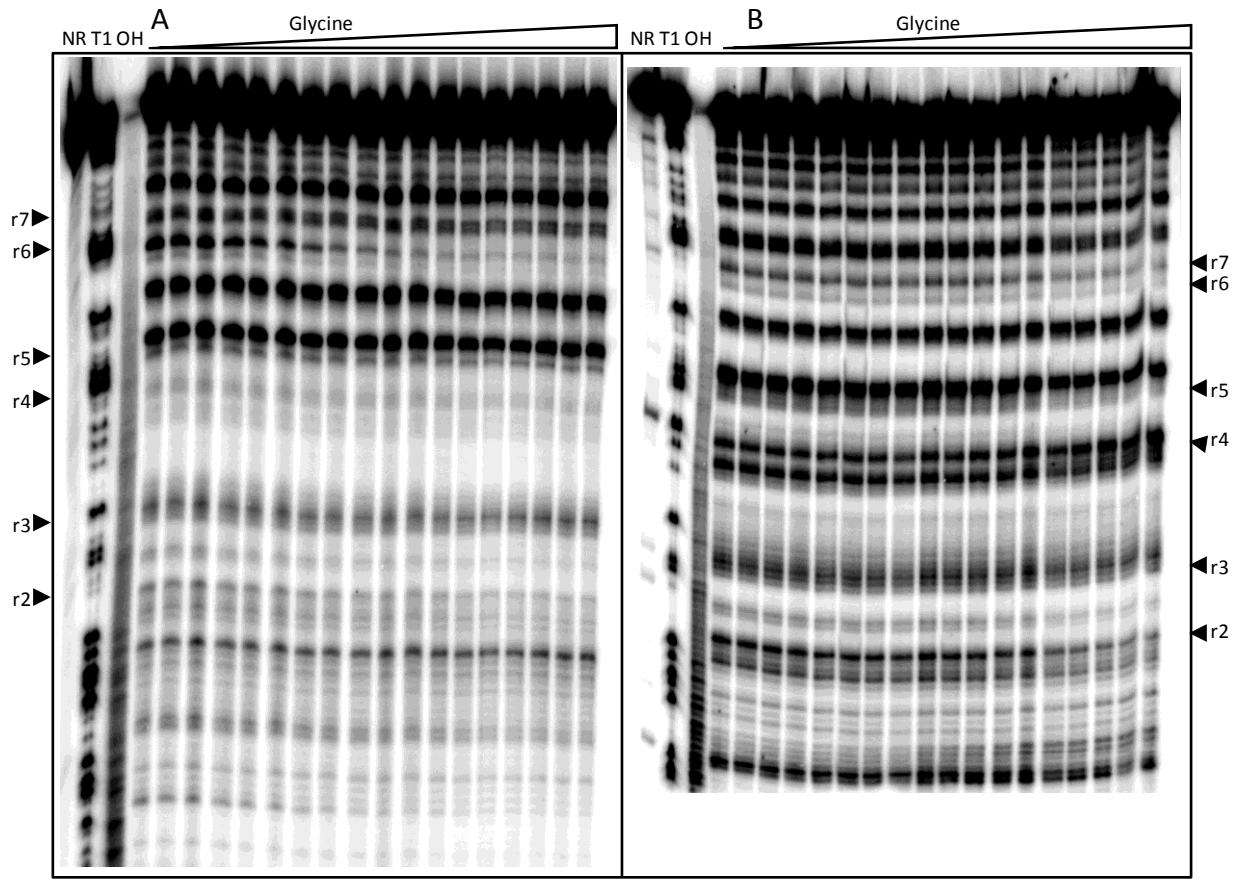


Figure 1.7 In-line probing gel images of BSLD (*A*) and BSI-II (*B*).

NR, T1, and OH represent no reaction, partial digestion with RNase T1, and partial digestion with alkali, respectively. r2-r7 indicate regions defined in Table 1.2.

As in-line probing assays may detect structure formation with heavy focus on secondary structures, to compare tertiary structures of the glycine riboswitches, we employed hydroxyl radical footprinting (HRFP), one of the most frequently used methods for studying RNA folding (29-32). HRFP is sequence independent and measures the perturbation of solvent accessibility of local tertiary structures during conformational change by hydroxyl radical triggered RNA backbone cleavage. Similar to the in-line probing results, HRFP experiments showed that, at saturating glycine concentrations, the full-length glycine riboswitches exhibit the same protection

patterns as the truncated ones outside of the linker region in all three VC, FN and BS glycine riboswitches (Fig. 1.8, 1.9, 1.10).

Combining in-line probing and HRFP experiments, results showed that full-length glycine riboswitches fold into the same global structure as the truncated version at saturating magnesium and glycine concentrations.

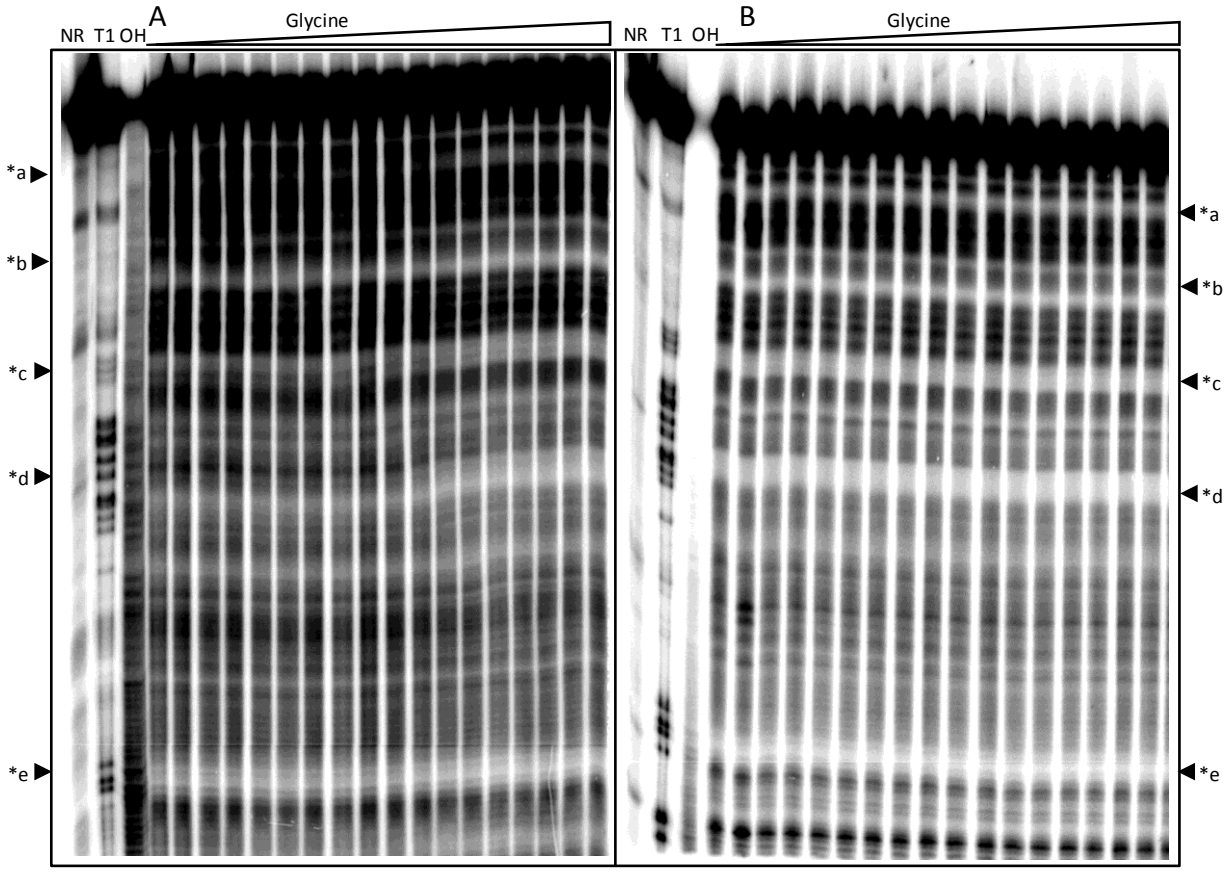


Figure 1.8 Hydroxyl radical footprinting gel images of VCI-II (*A*) and VCLD (*B*).

NR, T1, and OH represent no reaction, partial digestion with RNase T1, and partial digestion with alkali, respectively. *a (nts 200-202), *b (nts 170-172), *c (around nt 146), *d (around nt 123) and *e (around nt 82), as indicated in Figure 1.1, represent the glycine-induced protections in VCI-II, which correlate well with previous reports (Lipfert et al. J. Mol. Biol. 2007, 365, 1393; Erion and Strobel RNA 2011, 17, 74).

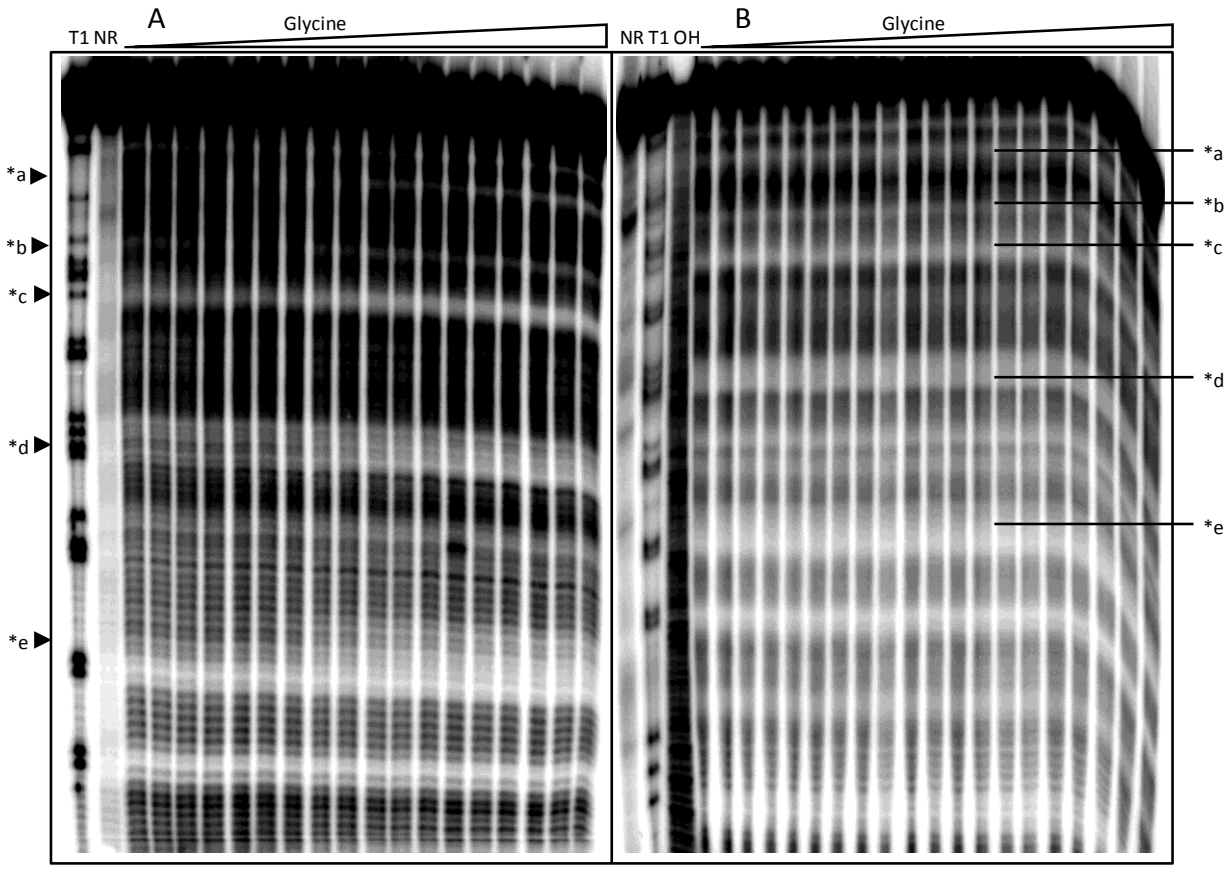


Figure 1.9 Hydroxyl radical footprinting gel images of FNI-II (*A*) and FNLD (*B*).

NR, T1, and OH represent no reaction, partial digestion with RNase T1, and partial digestion with alkali, respectively. *a (nts 134-136), *b (nts 117-119), *c (nts 110-112), *d (nts 85-86), and *e (nts 63-67) represent the glycine-induced protections in FNI-II.

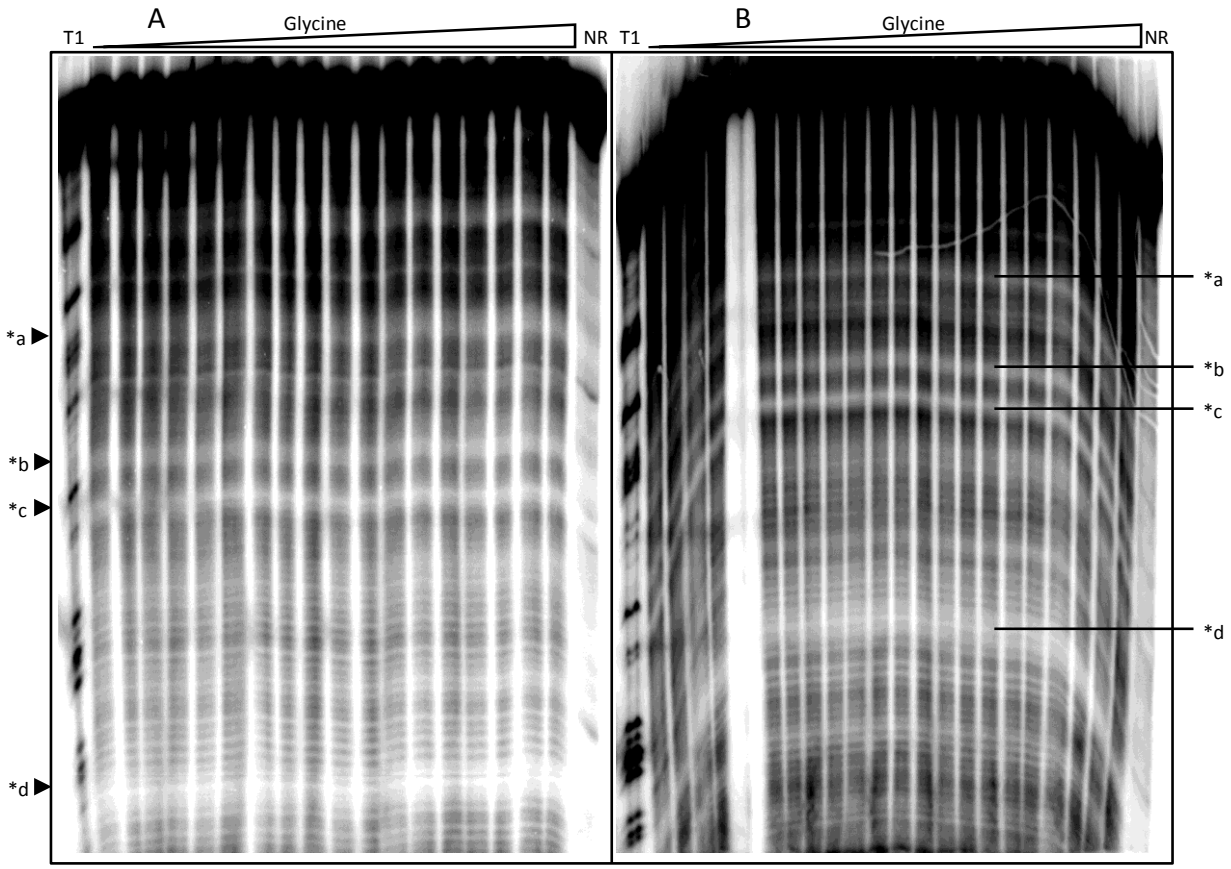


Figure 1.10 Hydroxyl radical footprinting gel images of BSI-II (*A*) and BSLD (*B*).

NR and T1 represent no reaction and partial digestion with RNase T1, respectively. X represents the lane below a damaged well without sample loading. *a (around G121), *b (around G97), *c (around C88), and *d (around G60) represent the glycine-induced protections in BSI-II.

1.2.3 Energetic benefit of the leader-linker interaction

Unexpected tertiary structures involving flanking sequences of functional RNAs have been reported before and are functionally beneficial. For instance, the loop/bulge interaction involving the 5' flanking sequence of the hammerhead ribozyme increases the catalytic rate substantially (33, 34). To understand the functional benefit of the leader-linker interaction, we

obtained the ligand binding affinities of the full-length glycine riboswitches through in-line probing assays at standard conditions similar to that reported previously for glycine riboswitches (6, 13, 14) (Fig. 1.4, 1.5, 1.6, 1.7). In the linker regions, the full-length glycine riboswitches showed constant protection with or without glycine, suggesting the leader-linker interaction is formed independent of glycine. Other glycine-perturbed regions, including both aptamer I and aptamer II regions, showed similar glycine binding affinities for the full-length constructs within experimental error. The affinities calculated from all the major glycine-perturbed regions over at least two duplicated experiments were averaged together and tabulated in Table 1.1. Results showed that glycine binding affinities of the full-length constructs have been increased by 4.5 to 86 fold compared to the truncated ones. This corresponds to a free energy benefit of 0.9-2.7 kcal/mol under standard in-line probing conditions. Interestingly, all three full-length riboswitches showed similar glycine binding affinities with K_ds of 5.0 ± 4.0 μM, 3.1 ± 2.5 μM, and 3.6 ± 3.4 μM for VCLD, FNLD and BSLD, respectively, while the truncated constructs showed much more varied glycine binding affinities. We have also obtained glycine binding affinities with HRFP, which gave similar values for VCI-II and FNI-II. However, we could not obtain any visible glycine perturbation for the full-length constructs using HRFP. At the corresponding regions where glycine induces protection in the truncated glycine riboswitches, full-length constructs exhibit protection with or without glycine, precluding visible perturbation. Typically, hydroxyl radical footprints are observed for the interior portion of closely packed helical segments for structured RNAs. We inspected the protected regions of FNI-II in the published crystal structure (19) in PyMol (www.pymol.org) and found these regions are mostly buried between the two aptamers. For instance, the sugar moieties (HRFP susceptible) of nts 63-67, nts 134-136 and nts 117-119 are buried by the α-α' inter-aptamer interaction; sugar moieties of nts 85-86 are partially buried by the β-β' inter-aptamer interaction and the packing against flipped base A111 which is very important for the formation of the glycine binding pocket; nts 110-112 directly participate in the glycine binding pocket. The HRFP protection in these regions of FNLD in the absence of glycine

suggests that the full-length constructs promote the inter-aptamer interaction, form more compact and organized pockets for glycine binding, and render any possible further compaction by the glycine binding undetectable through HRFP experiments.

1.2.4 Mutation in the linker region selectively reduces the ligand binding affinities of the full-length but not the truncated glycine riboswitches

The full-length glycine riboswitches exhibit significantly higher ligand binding affinities compared to the truncated ones. To link the energetic benefit directly to the leader-linker interaction, we focused on the VC glycine riboswitch, which was studied most extensively with biochemical analysis. Incorporating a GGA136-138CCC mutation to prevent the formation of the P0 helix, we prepared VCI-IILN and VCLDLN, the linker mutants in the truncated and full-length VC glycine riboswitches, respectively. Similar to the wild-type constructs, the linker mutants also showed similar glycine binding affinities in both aptamer I and aptamer II regions within experimental error. The glycine binding affinities were calculated as the average affinities of all the major regions over at least two duplicated experiments. Results showed that VCLDLN binds glycine with a K_d of 61 ± 11 μM (Fig. 1.11A; Table 1.1), very similar to that of VCI-II, 59 ± 12 μM, losing the entire energetic benefit (12 fold) of the leader-linker interaction observed in VCLD. This affinity loss could be potentially caused by a loss of an unidentified linker interaction with regions other than the 5' leader. To control for this possibility, we obtained the glycine binding affinity of VCI-IILN. Unlike VCLDLN, the linker mutation in VCI-IILN did not reduce the glycine binding affinity compared to VCI-II, consistent with the model that leader-linker interaction contributes solely to the energetic benefits in VCLD. In fact, there is about 4-fold increase in the glycine binding affinity of VCI-IILN compared with VCI-II (Fig. 1.11B; Table 1.1). Both VCI-II and VCI-IILN contain three guanosines at the 5' with two artificial guanosines added for improved transcription efficiency. Suggested by Mfold (20), this GGG sequence could potentially pair with CCC136-138 in VCI-IILN and form an imperfect kink-turn motif (only two nucleotides in the

internal loop). Thus, this 4-fold increase of glycine binding affinity of VCI-IILN versus VCI-II is also consistent with the functional role of the leader-linker interaction.

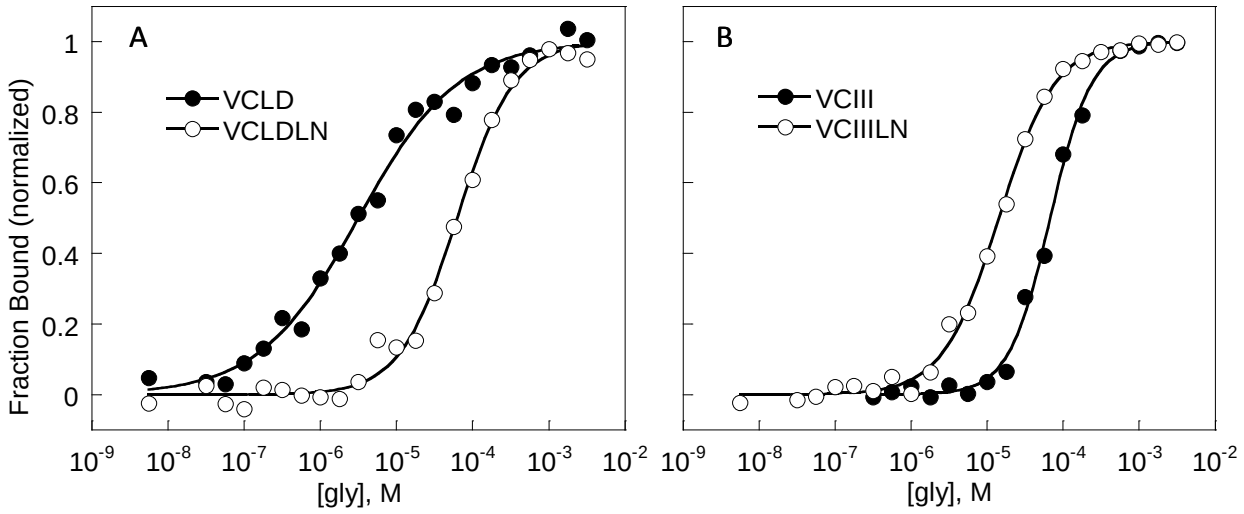


Figure 1.11 Comparison of the ligand binding of the linker mutants with wild-type VC glycine riboswitches.

(*A*) and (*B*) show the representative semi-log plots of the normalized fraction of RNA bound versus glycine concentration for r3 region (A121-G123) of VCLD with VCLDLN, and VCI-II with VCI-IILN, respectively.

1.2.5 Leader-linker interaction greatly enhances the inter-aptamer interaction

HRFP experiments showed that full-length glycine riboswitches promote the inter-aptamer interaction and form more compact structures in the absence of glycine. To observe the effect of the leader-linker interaction on inter-aptamer complexation directly, we designed the individual aptamers of the VC glycine riboswitch and analyzed the inter-aptamer binding with native gel assays. We have prepared two different aptamer I constructs, with (VCI132LD, nts -7 to 132) and without the 5' leader (VCI132, nts 1 to 132), and one aptamer II construct with the entire linker region (VCII133, nts 133 to 225). This design prevents self-association of aptamer I through leader-linker interaction by placing the entire linker region on aptamer II and precludes inter-aptamer duplex formation in P1 regions by avoiding any overlapping bases between

aptamers, thereby allowing direct measurement of the possible enhancement of the inter-aptamer complex formation through leader-linker interaction. We 5' ^{32}P labeled VCII133 and incubated it with unlabeled VCI132LD or VCI132 at concentrations ranging from 0 to ~100 µM under native conditions (10 mM $MgCl_2$, 5 mM glycine). The mixture was then resolved via native gel based on a polyacrylamide coelectrophoresis (PACE) method (13, 35) with radiolabeled VCLD as a size marker (Fig. 1.12A). VCII133 showed small fractions of probable self-complexation independent of aptamer I concentration (similar phenomenon noted previously in different experiments). With increasing amounts of VCI132, a new complex with increasing intensity was formed and co-migrated with VCLD, suggesting that it is a reconstituted complex between VCII133 and VCI132. The intensity of the reconstituted complex reached 16% with 25 µM VCI132 and dropped abruptly to 9% at ~100 µM VCI132, presumably due to possible dimerization of VCI132 at such a high concentration. With increasing amounts of VCI132LD, similar complexes with increasing intensity were formed and co-migrated with VCLD. Interestingly, on a well resolved PACE gel, we observed two closely migrating bands for the VCI132LD/VCII133 complex. We reasoned that without covalent linkage between two aptamers, the kink-turn motif formed between leader and linker may assume two different conformations or dock into the rest of the glycine riboswitch in two different geometries. The reconstituted complexes of VCI132LD/VCII133 were formed more readily and reached 52% with 50 µM VCI132LD before dropping abruptly to 44% at 100 µM VCI132LD similar to that observed for VCI132. Plotting complex formation vs. aptamer I concentration (excluding the data at ~100 µM aptamer I) gives K_d values of 0.66 ± 0.16 µM and ≥ 6.3 µM (Fig. 1.12B) for VCI132LD and VCI132, respectively (the K_d for VCI132 is estimated for lack of plateau at high aptamer I concentration). This shows that the leader-linker interaction enhanced the inter-aptamer interaction by ≥ 10 fold under native conditions, consistent with the energetic benefit obtained previously.

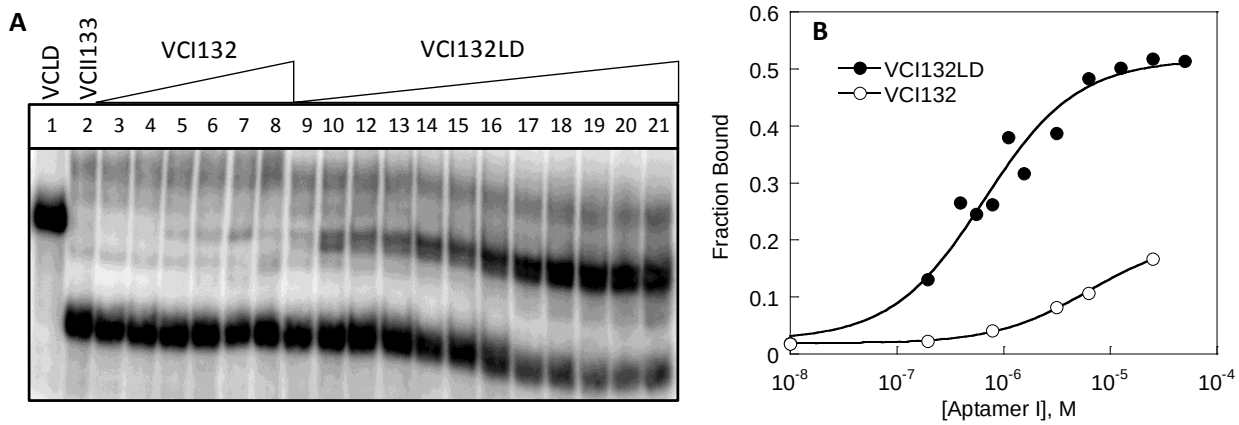


Figure 1.12 Native gel and graph of covalently disconnected aptamers I and II to reconstitute the switch from these *trans* aptamers by interaptamer interaction.

(*A*) Native gel image of ^{32}P labeled VCII133 titrated with aptamer I constructs in 10 mM $MgCl_2$ and 5 mM glycine. Lane 1, ^{32}P labeled VCLD; lane 2, ^{32}P labeled VCII133; lane 3-8, ^{32}P labeled VCII133 with 0.2 to 97 µM of VCI132; lane 9-21, ^{32}P labeled VCII133 with 0.2 to 100 µM of VCI132LD. (*B*) Semi-log plots of unnormalized fraction of VCII133 bound versus aptamer I concentration.

1.2.6 Glycine binding properties of individual aptamers

To better dissect the roles of inter-aptamer interaction and glycine binding on native glycine riboswitch folding in *cis* or in *trans*, we set out to analyze the glycine binding properties of the individual aptamers with in-line probing experiments. Neither VCI132 nor VCI132LD showed glycine-dependent modulation (data not shown), consistent with previous reports that aptamer I may not be capable of independent glycine binding, at least within the limits of the in-line probing method (14). In addition to the aforementioned two and two other aptamer I constructs reported previously (14),VCI139 (nts 1 to 139) and VCI150 (nts 1 to 150), we have also tested VCI145 (nts 1 to 145), VCI145LD (nts -7 to 145), VCI151 (nts 1 to 151), VCI151LD (nts -7 to 151), VCI166 (nts 1 to 166), VCI177 (nts 1 to 177) and VCI189 (nts 1 to 189), covering half of aptamer II. In-line probing experiments showed that none of these constructs bind glycine

(data not shown), not even with the 5' leader on some of the constructs (VCI145LD and VCI151LD). These experiments definitively showed that aptamer I does not bind glycine independently. Interestingly, aptamer II construct VCII133 does not bind glycine properly either (K_d of 2.0 ± 0.1 mM; Table 1.1), although VCII122 (nts 122 to 225) binds glycine with a K_d of 4.4 ± 1.9 µM, consistent with previous reports (6, 14). Further analysis showed that other VC aptamer II constructs shorter than VCII122 such as VCII127 (nts 127 to 225) and VCII130 (nts 130 to 225) also do not bind glycine (data not shown). An aptamer II construct with blunt end P1 helix VCII140m1 (nts 140 to 225, with CC140-141GG and GG224-225CC mutation) binds glycine with a K_d of 1.4 ± 0.2 mM, 300 fold weaker than VCII122, although it has been reported that a VC aptamer II construct with a wobble pair mutation and extended P1 helix binds glycine with a K_d of 3.5 ± 1.5 µM at 20 mM $MgCl_2$ analyzed by ITC experiment (18). These results strongly suggest that wild-type VC aptamer II sequences also need part of aptamer I structural elements to bind glycine properly. VCII122 contains 11 nucleotides from aptamer I, and it may happen to mimic the inter-aptamer interaction needed for aptamer II to bind glycine properly or assume an alternative conformation that allowed proper glycine binding. Further evidence supporting that individual aptamers need structural elements on the other aptamer comes from the FN glycine riboswitch. The corresponding constructs of VCI145, VCI151, and VCII122 on FN glycine riboswitch are FNI84 (nts 1 to 84), FNI90 (nts 1 to 90), and FNII61 (nts 61 to 158), none of which bind glycine in the in-line probing experiments (data not shown).

1.2.7 Leader-linker interaction greatly enhances the glycine binding affinities of the individual aptamers

Via native gel assays, we have also tried to analyze the complex formation between aptamer I and aptamer II without glycine. However, no complex was visible without glycine added in the gel when VCII133 was incubated with either VCI132 or VCI132LD (data not shown). The previous section showed that without proper inter-aptamer interaction, an individual aptamer does not bind glycine properly. These two observations suggest that glycine binding

and inter-aptamer interaction (assisted by leader-linker interaction) have to function synergistically in global glycine riboswitch folding. The PACE gel results showed that saturating glycine helps aptamers interact with each other in *trans* and form the inter-aptamer complex. We then sought to investigate whether a saturating concentration of an individual aptamer will enhance glycine binding on the other aptamer and what role the leader-linker interaction might play. We carried out the in-line probing experiment using radiolabeled VCII133 incubated with 15 µM VCI132 in the presence of a range of glycine concentrations. Results showed that VCI132 greatly enhances the glycine binding of VCII133 from 2 mM to a K_d of 191 ± 44 µM (Fig. 1.13A; Table 1.1). When 15 µM VCI132LD was added to the in-line probing assay, VCII133 showed glycine binding with a K_d of 7.4 ± 1.0 µM (Fig. 1.13A; Table 1.1), enhanced 26 fold further from that with VCI132, restoring the glycine affinity almost completely to that of the full-length tandem riboswitch. This enhancement is slightly better than the ~12 fold enhancement of VCLD versus VCI-II, suggesting that the leader-linker interaction benefits more in *trans*, which is consistent with its role in holding the two aptamers together. We also carried out the in-line probing experiment using radiolabeled VCI132 incubated with 15 µM VCII133 in the presence of a range of glycine concentrations. Results showed that VCII133 rescues the glycine binding of VCI132 from completely non-detectable to an apparent K_d of 8.1 ± 2.5 mM (Fig. 1.13B; Table 1.1). A similar in-line probing experiment was carried out with radiolabeled VCI132LD as well. Results showed that VCII133 also rescues the glycine binding of VCI132LD from completely non-detectable to a K_d of 190 ± 41 µM (Fig. 1.13B; Table 1.1), enhanced ~40 fold further from that with VCI132. This affinity is substantially smaller than that of VCLD, unlike the glycine binding affinity of VCII133 rescued by VCI132LD. A possible explanation is that aptamer I requires the interaction between the P3b hairpin of aptamer II and the P1 helix of aptamer I (13, 14, 18), which is right next to the break point in our in *trans* construct design. As evidenced by the two bands of the reconstituted complexes in native gel assays (Fig. 1.12A), in this non-covalently linked version, kink-turn leader-linker interaction in the VC glycine riboswitch may not

be formed perfectly, leading to imperfect interaction between the P3b hairpin of aptamer II and the P1 helix of aptamer I and exhibiting only partial rescue of the glycine binding affinity.

Combining the results of the *trans* in-line probing assays, we have demonstrated inter-aptamer interactions can enhance the glycine binding affinities of the individual aptamers while the leader-linker interaction provides an additional 26 to 40 fold enhancement.

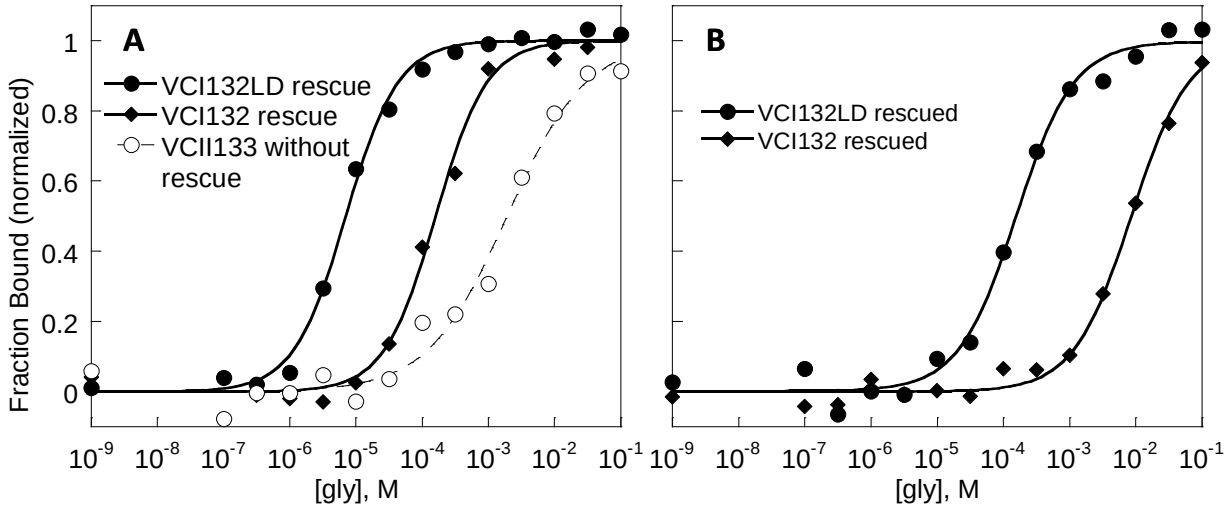


Figure 1.13 Representative semi-log plots of aptamers rescuing each other in *trans* in-line probing assays.

(*A*) r7 of VCII133 rescued by 15 µM of aptamer I constructs. (*B*) r2 of VCI132LD and VCI132 rescued by 15 µM of VCII133.

1.2.8 Full-length glycine riboswitch does not exhibit glycine binding cooperativity

We have shown that glycine binding and inter-aptamer interaction function synergistically in glycine riboswitch folding through *trans* assays with native gel and in-line probing experiments. There are two glycine binding sites and multiple inter-aptamer interactions, which may or may not take place concurrently. The leader-linker interaction forms in the absence of glycine, as indicated by the constant protection in the linker region in the in-line probing assay and the lack of glycine perturbation in the HRFP experiments of the full-

length glycine riboswitches. This seems to suggest that not all interactions are formed at the same time. One specific question is whether the two glycine molecules bind cooperatively. One crude measure is to use Hill coefficients of the glycine binding in the double aptamer riboswitches. Indeed, our results showed that the truncated VC and FN glycine riboswitches have Hill coefficient values larger than one, exhibiting reasonable apparent cooperativity, although smaller than reported (6, 13). Two linker mutants, VCLDLN and VCI-IILN which contain no leader-linker interaction, showed Hill coefficient values larger than one as well, exhibiting similar cooperativity to that of the truncated construct VCI-II. However, all three full-length glycine riboswitches showed Hill coefficient mean values less than or close to one and the upper limits of these values are all close to one, exhibiting no glycine binding cooperativity. We recognize that Hill coefficient values are extremely sensitive to the accidental error of the data points and therefore may not be the definitive measure for glycine binding cooperativity. The unequivocal evidence of the glycine binding cooperativity of the truncated construct VCI-II is from mutation analysis. Breaker and coworkers reported that deleterious mutations in each aptamer greatly reduced the glycine binding affinities of the other aptamer, demonstrating inter-aptamer inhibition and cooperativity between two glycine binding sites (6). Our results showed that VCI-IIG17C (nts 1 to 225, with G17C mutation) not only completely abolished glycine binding in the aptamer I regions but also reduced the aptamer II glycine binding affinity to 1.1 ± 0.1 mM while VCI-IIG146C binds glycine with a K_d of 1.9 ± 0.3 mM in both aptamers, comparable to the previous reports (6, 14). These experiments are consistent with the cooperative binding in the two glycine binding sites of the truncated glycine riboswitch VCI-II. To assess the glycine binding cooperativity of VCLD, we therefore constructed VCLDG17C (nts -7 to 225, with G17C mutation) and carried out in-line probing experiments. Results showed that most of the aptamer I regions modulated in VCI-II and VCLD do not show glycine modulation in VCLDG17C, such as nts 63-64, 70-72, 74 (r2), 78-80, and 83 (Fig. 1.14A). However, much to our surprise, unlike VCI-IIG17C, the aptamer II regions including nts 146 (r5), 170-172 (r6), 177-

178 (r7) and 207-208 showed strong glycine modulation with a K_d of 2.8 ± 0.4 μM, retaining full glycine binding capacity of VCLD in these regions (Fig. 1.14A,C; Table 1.1). Interestingly, A121-G123 (r3) is also perturbed with a similar glycine binding affinity, consistent with our finding that aptamer II needs structural elements from aptamer I to bind glycine. We have also mutated aptamer II and constructed VCLDG146C (nts -7 to 225, with G146C mutation). In-line probing experiments showed that most of the aptamer II regions modulated in VCI-II and VCLD do not show glycine modulation in VCLDG146C, such as nts 170-172 (r6), 177-178 (r7), and 207-208 (Fig. 1.14B). However, again unexpectedly, unlike VCI-IIG146C, the aptamer I regions including nts 70-72, 74 (r2), 78-80, and 121-123 (r3) showed great glycine modulation with a K_d of 12 ± 2 μM, retaining most of the glycine binding capability of VCLD in these regions (Fig. 1.14B,D; Table 1.1). Interestingly, nt 146 (r5) is also perturbed with a similar glycine binding affinity, consistent with our finding that aptamer I needs structural elements from aptamer II to bind glycine (though clearly, in this case, the type of nucleobase is flexible). The in-line probing results of VCLDG17C and VCLDG146C showed that an individual glycine binding site may retain its full or majority binding capacity even though the other site is not occupied, abolishing inter-aptamer inhibition observed in VCI-II, and strongly suggesting that VCLD does not bind two glycine molecules cooperatively.

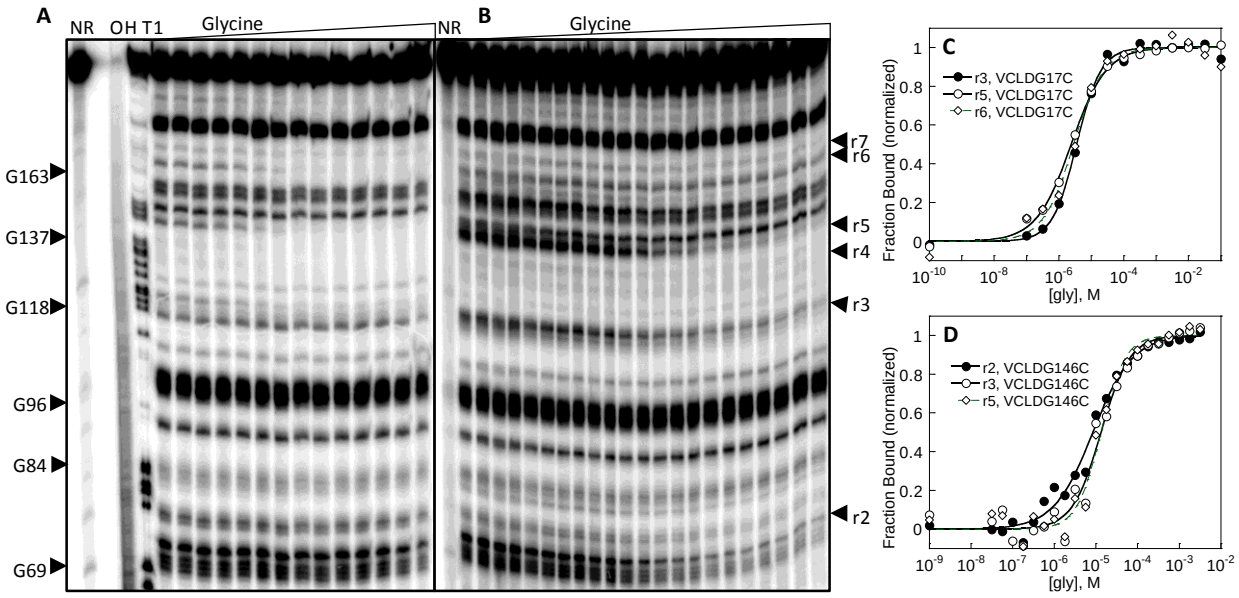


Figure 1.14 In-line probing gel images of VCLDG17C (*A*) and VCLDG146C (*B*).

NR, T1, and OH represent no reaction, partial digestion with RNase T1, and partial digestion with alkali, respectively. r2-r7 indicate regions defined in Table 1.2. (*C*) and (*D*) show the representative semi-log plots of the normalized fraction of RNA bound versus glycine concentration for VCLDG17C (r3, r5, and r6) and VCLDG146C (r2, r3, and r5), respectively.

1.3 Discussion

Prompted by the different protection patterns in the linker regions of two different truncated glycine riboswitches (VCI-II and VCI-IIs), we have identified a leader-linker interaction involving nucleotides upstream of the current published consensus glycine riboswitch sequences. This leader-linker interaction forms a P0 helix with 3-7 base pairs and is greatly conserved (>90%) in tandem glycine riboswitches. In more than half of the glycine riboswitches, the P0 helix is connected to the P1 helix through a kink-turn motif. These kink-turn motifs fit largely to the consensus kink-turns derived from the ribosome kink-turns, except that the base pair adjacent to the internal loop has a strong preference for G-C, just the opposite to the C-G

base pair preference in the ribosome kink-turns. The kink-turn motif is not unprecedented in riboswitches; it has been reported in SAM-I riboswitches connecting to the four-way junction motif (36) as well as in the class II c-di-GMP (37) and lysine riboswitches (38). This SAM-I kink-turn motif follows the exact consensus derived from the ribosome kink-turns and is stabilized by RNA tertiary structure (39), different from those stabilized by metal ions (25) or protein binding (40, 41). It will be interesting to investigate which strategy is used in the glycine riboswitch to stabilize the kink-turn motif located at the 5' boundary of this regulatory RNA.

We have shown that full-length glycine riboswitches bound glycine 4.5 to 86 fold tighter than the truncated versions under saturating magnesium concentrations (16). Interestingly, three full-length glycine riboswitches, VCLD, FNLD and BSLD showed glycine binding affinities very similar to each other, unlike their truncated versions, raising the question whether glycine riboswitches have convergently evolved to bind glycine at a cellularly specific value for optimum metabolism.

We have demonstrated direct binding between two aptamers in *trans* without any overlapping bases using native gel assays. The leader-linker interaction enhanced the inter-aptamer interaction by at least 10 fold with the riboswitches saturated with glycine. As inter-aptamer complexation cannot be detected by native gel assay in the absence of glycine, these results suggest that glycine binding enhances the inter-aptamer interaction. Conversely, aptamers also greatly enhanced the glycine binding ability of each other with the leader-linker interaction contributing an additional 26 to 40 fold improvement. These data suggest that inter-aptamer interaction and glycine binding complement each other. One prominent property of the truncated glycine riboswitches is their glycine binding cooperativity. Surprisingly, the Hill coefficients of the wild-type full-length glycine riboswitches and more definitively the mutational analysis with VCLDG17C and VCLDG146C showed that there is no ligand binding cooperativity between two glycine binding sites in the full-length glycine riboswitches.

Together, our results support a new global folding model of the tandem glycine riboswitches. Independent of glycine, the leader-linker interaction can be formed in the presence of magnesium ions and leads to a more compact structure which enhances glycine binding and other inter-aptamer interactions, such as the α-α', β-β' and γ-γ' inter-aptamer interactions reported previously (14, 18, 19). The glycine binding at either site acts synergistically with inter-aptamer interactions to achieve global folding of the tandem glycine riboswitch while the two glycine binding sites do not operate cooperatively.

Our biochemical results focused on the VC glycine riboswitch and strongly supported the lack of glycine binding cooperativity in the full-length VC glycine riboswitch. Although we cannot rule out possible ligand binding cooperativity in other full-length glycine riboswitches, since the truncated construct VCI-II is the most cooperative glycine riboswitch reported our finding is likely to be general. The lack of glycine binding cooperativity in the full-length glycine riboswitches may have been manifested in the great variability of the reported cooperativity values with Hill coefficients ranging from 1.1 to 1.6 (6, 12-14, 16).

One interesting question remains: without the advantage of cooperative and more "digital" response to the cellular glycine concentration, why do glycine-sensing RNAs evolve into the tandem arrangement? One possibility is that as our data pointed out, glycine-sensing may not be achieved efficiently with a single aptamer, given glycine is the smallest organic molecule sensed by RNAs. Individual aptamers need structural elements from other aptamers to achieve efficient glycine binding; the only reported functional wild-type single aptamer construct VCII122 uses part of the aptamer I sequence to mimic an efficient glycine binding strategy. The protection pattern changes shown by r3 in VCLDG17C and r5 in VCLD146C also support this idea. Another possibility is that through tandem arrangement, the glycine riboswitch employs the leader-linker interaction to promote interaptamer interactions and preorganize the glycine binding pockets. Three experimental results support the more organized and compact structure formation in the absence of glycine: a) protection of the regions important for inter-aptamer

interaction or the glycine binding pocket is independent of glycine in HRFP of the full-length glycine riboswitches (Fig. 1.8, 1.9., 1.10); b) the longer VCLD runs slightly faster than VCI-II in the absence of glycine on native gel while running comparable or slightly slower than VCI-II in the presence of glycine (Fig. 1.15); c) two antibody fragments (Fabs) selected against VCI-II (42) bind VCLD with similar affinities but higher percentage binding at saturating Fab concentrations presumably because VCLD is less flexible and retained better during the washing step in the filter-binding assay (data not shown). The more compact and organized structure of full-length glycine riboswitches may have advantages in glycine-sensing under physiological conditions in vivo if fast on rate is desired similar to other riboswitches under kinetic control (15, 43, 44)

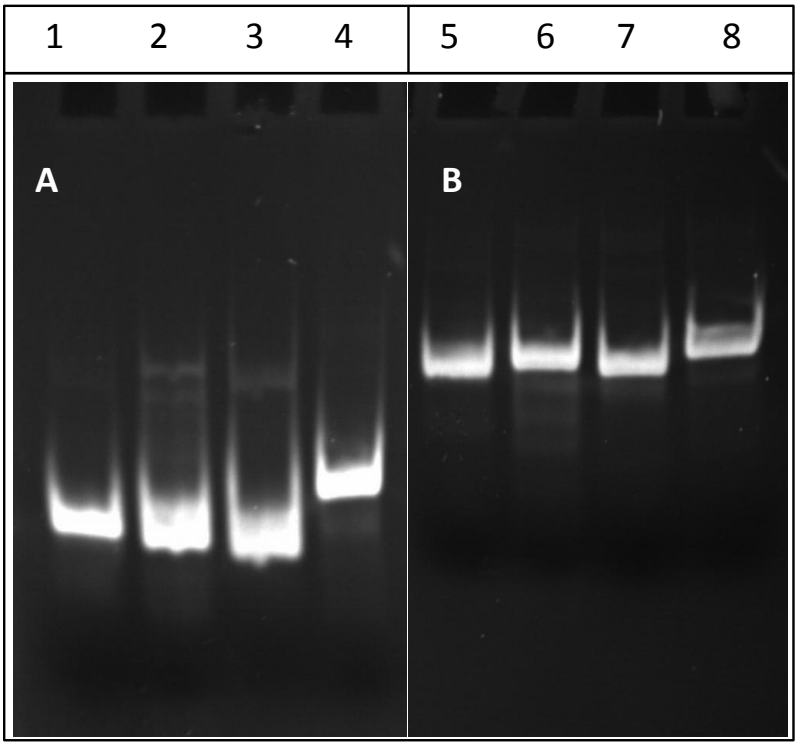


Figure 1.15 Native gel assay of glycine riboswitch compaction.

6% non-denaturing PAGE in 5 mM $MgCl_2$ and 5 mM glycine (*A*) and 5 mM $MgCl_2$ and no glycine (*B*). These two gels were prepared in two different halves of the same gel cassette separated by a spacer to ensure identical running conditions. Lane 1 and 5, VCLD; Lane 2 and 6, VCI-II; Lane 3 and 7, VCI-IIs; Lane 4 and 8, VCI-IIG146C.

1.4 Materials and Methods

1.4.1 Preparation of the RNAs

The glycine riboswitches VCI-II, FNI-II, BSI-II and single aptamers were generated by annealing DNA oligonucleotides (Integrated DNA Technologies) and amplified by recursive PCR with an EcoRI restriction site and T7 promoter sequence engineered at the 5' terminus, and EarI and HindIII sites engineered at the 3' terminus. Double digested riboswitch DNA inserts and pUC19 vector were ligated together to generate glycine riboswitch plasmids. The full-length glycine riboswitches and mutants were constructed by a similar method or QuikChange site-directed mutagenesis (Stratagene) from the corresponding wild type plasmids and proper DNA oligonucleotide primers. The plasmids were transformed into JM109 cells, amplified and purified with Qiagen plasmid preparation kit, and their sequences confirmed by sequencing (Genewiz Inc). A Qiagen Midiprep produced more plasmid which was then linearized with EarI and transcribed by T7 RNA polymerase. The RNA transcripts were purified by 6% to 8% denaturing PAGE, electroeluted, PCA extracted, ethanol precipitated, and redissolved in TE buffer.

1.4.2 In-line probing assay and quantification

RNA transcripts were dephosphorylated with Shrimp alkaline phosphatase (Fermentas Life Sciences) and 5' ^{32}P labeled with γ ^{32}P ATP (Perkin Elmer) and T4 polynucleotide kinase (New England Biolabs). For transcripts with the 5' end buried by or too close to a duplexed region, to improve the labeling yield, the transcript was first annealed to a short DNA oligonucleotide to free the 5' end before dephosphorylation and kinase reactions. In-line probing assays were carried out similar to the reported procedure (28). Briefly, ~50 kcpm ^{32}P labeled RNA transcript was incubated at 25 ^{o}C for 48 h in 50 mM Tris-HCl (pH 8.3), 20 mM $MgCl_2$, 100 mM KCl under various glycine concentrations. For the *trans* in-line probing assay, 15 µM unlabeled proper aptamer was also added to the reaction. Spontaneously cleaved RNA fragments were resolved by 8% denaturing PAGE, dried, exposed to PhosphorImager screens,

and quantified by ImageQuant (Molecular Dynamics). For VC glycine riboswitches and mutants, the following regions are quantified if contained on the constructs and well-resolved: U74 (r2), A121-G123 (r3), G136-G137 (r4), G146 (r5), G170-A172 (r6), C177-A178 (r7), and U207-C208. To control for loading differences, the band intensities were normalized to nonvariable bands (U94 and U156-U161). For FN glycine riboswitches, the following regions are quantified if contained on the constructs and well resolved: G35, U46 (r2), G61 (r3), G85 (r5), G110-A112 (r6), and C117-A119 (r7). Band intensities were normalized to nonvariable bands (U23 and C90). For BSLD and BSI-II, the following regions are quantified: G60, U83 (r2), C88-A89, A151-A153 (r6), and A158-G160 (r7). Band intensities were normalized to nonvariable bands (G136-U138). The Hill coefficients and dissociation constants were obtained by fitting each plot using the following equation in KaleidaGraph software: fraction bound = $m_1 \times [gly]^n/(K_d^n + [gly]^n) + m_2$.

1.4.3 Hydroxyl radical footprinting

For footprinting, 5' end labeled RNA was dissolved in 10 mM Na-MOPS (pH 7.0). 1 μl of this RNA solution (~100 kcpm) was incubated with 1 μl of 500 mM Na-MOPS pH 7, 1 μl of 100 mM $MgCl_2$, 1 μl of 10x stock glycine solution at various concentrations and 5 μl of water for 20 min at 50 ^oC. A 10x footprinting reagent was prepared separately that contained 1 mM $Fe(NH_4)_2(SO_4)_2$, 1.25 mM EDTA pH 8, and 60 mM sodium ascorbate. To each sample, 1 μl of this reagent was added to initiate the footprinting reaction. The reaction proceeded for 1 h at room temperature and was quenched by the addition of 5 μl of thiourea stop solution (9 M urea, 300 mM thiourea, 0.1% each xylene cyanol and bromophenol blue). The footprinting reactions were fractionated with 8% denaturing PAGE. The gels were dried, exposed to PhosphorImager screens, and quantified by ImageQuant (Molecular Dynamics). For VCI-II, the protections around G146 and G123 are quantified and normalized to nonvariable band C108. For FNI-II, the protections around G85-A86 and U65-C66 are quantified and normalized to nonvariable band G72. For BSI-II, the protection around C88 is quantified and normalized to nonvariable band

G75. The Hill coefficient and dissociation constant were obtained by fitting each plot using the following equation in KaleidaGraph software: fraction bound = $m_1 \times [gly]^n/(K_d^n + [gly]^n) + m_2$.

1.4.4 Native gel analysis of inter-aptamer complexation

Native gel electrophoresis was carried out similar to the procedure described previously (14, 35). 1 μl of 5' end labeled VCII133 (~5 kcpm) was incubated with 1 μl of 2x incubation buffer (180 mM Tris-borate, pH 8.3, 10 mM glycine, 20 mM $MgCl_2$) for 20 min at room temperature. In a separate Eppendorf tube, 1 μl of 4x stock unlabeled VCI132 (or VCI132LD) solution at various concentrations was incubated with 1 μl of 2x incubation buffer for 20 min at room temperature. The solution was combined together and incubated for an additional 20 min at room temperature. After homogenizing with 1 μl of 5x native loading dye (50% glycerol, 0.1% xylene cyanol), the reaction mixture was resolved via a 6% native PAGE in TB buffer containing 10 mM $MgCl_2$ and 5 mM glycine. Electrophoresis was carried out with TB buffer containing 10 mM $MgCl_2$ at 4 ^oC for 4 h under approximately 12.5 V per cm of gel height. The gels were dried, exposed to PhosphorImager screens, and quantified by ImageQuant (Molecular Dynamics). The dissociation constant was obtained by fitting each plot using the following equation in KaleidaGraph software: fraction bound = $m_1 \times [RNA]/(K_d + [RNA]) + m_2$.

CHAPTER 2 DNA-RESCUABLE ALLOSTERIC INHIBITION OF APTAMER II LIGAND AFFINITY BY APTAMER I ELEMENT IN THE SHORTENED *VIBRIO CHOLERAE* GLYCINE RIBOSWITCH

2.1 Introduction

Synthetic biology utilizes engineered genetic circuits to reprogram or rewire organisms to achieve diverse functions, such as synthesis of cheaper drugs, generation of renewable fuels, and identification of disease mechanisms and viable drug targets (45, 46). Novel, inexpensive, and versatile genetic control elements are thus essential to the success of synthetic biology (45, 46). Frequently, synthetic biologists turn to natural systems as inspiration for novel genetic control elements. Riboswitches are one class of naturally occurring genetic control elements which regulate gene expression by binding to small molecule metabolites independent of proteins (1, 47). Upon binding of the biologically important metabolites, riboswitches undergo conformational changes which may affect the structure of the associated expression platforms and control gene expression through transcriptional termination, translational initiation, or alternative splicing (1, 5, 47, 48). With increasing understanding of the gene control mechanism of riboswitches, researchers have designed and engineered several riboswitch-based genetic circuits (49, 50).

Glycine riboswitches are of special interest as potential genetic control elements. Containing two tandem ligand-sensing domains, glycine riboswitches turn on downstream gene expression upon binding glycine molecules (6, 13, 19, 51-53). The glycine riboswitch from *B. subtilis* has been employed to express recombinant proteins using inexpensive glycine as an inducer with yields comparable to that in xylose- and IPTG-driven expression systems (54). Glycine riboswitches may offer more than just a model of a simple metabolite-sensing genetic control gate. Two versions of the glycine riboswitches have been reported in the literature. The biologically relevant full-length glycine riboswitches were reported recently by our group and the Das group (51, 52), in which the interaptamer linker forms a P0 duplex with the previously

unrecognized 5' leader sequence. In over 50% of known tandem glycine riboswitches, P0 duplex complies with a kink-turn motif consensus which increases glycine affinity, allows further RNA compaction upon binding with kink-turn motif binding proteins, and abolishes glycine binding cooperativity (51-53). Interestingly, the previously reported glycine riboswitches lacking P0 duplex were reproducibly shown to bind two glycine molecules cooperatively (6, 12, 13, 16, 17, 51). Experiments showed that deleterious single-nucleotide mutation to individual aptamers not only disabled the mutated aptamer, but also greatly reduced the glycine binding affinity of the other, indicating structure elements in one aptamer may control glycine binding in the other through interaptamer interaction (6, 14, 18, 51). Further analysis of the P0-lacking glycine riboswitches may shed light on the molecular basis of the ligand-binding cooperativity and help design novel sophisticated biosensing genetic circuits for application in synthetic biology.

In the present work, we choose to study the P0-lacking *Vibrio cholerae* glycine riboswitch (nts 1-225, referred to as VCIII hereafter) as this construct exhibits the highest ligand binding cooperativity, with a Hill coefficient of 1.64 (6). Our goal is to analyze the ligand binding cooperativity in greater detail and identify structure elements that are not only responsible for allosteric control of glycine sensing but also amenable for future design of novel genetic circuits.

The Strobel group has proposed a sequential glycine binding model for the ligand binding cooperativity in VCIII glycine riboswitch (14). According to this model, one glycine molecule binds to the downstream aptamer II first and allows tertiary interactions to be made between the two aptamers, facilitating binding of a second glycine molecule to the upstream aptamer I, hence glycine-binding cooperativity. This model provides explanation for the inhibition of aptamer I's glycine affinity by a deleterious mutation in aptamer II. However, this model does not provide explanation for the inhibition of aptamer II's glycine affinity by a deleterious mutation in aptamer I, which has also been noted previously. Combining direct deletion studies, mutational analysis, and sequence examination, we have identified and verified an aptamer I element which effectively controls the aptamer II glycine affinity, especially in

the shortened VC glycine riboswitch. Interestingly, the inhibitive effect of this structural element can be rescued with a short nucleic acid strand, restoring aptamer II glycine binding. This provides the possibility of engineering a shortened glycine riboswitch into a genetic control element that senses both glycine metabolite and a DNA input signal.

2.2 Results

2.2.1 5'-extended aptamer II construct VCII66 lost most of its glycine-binding affinity

The sequential glycine binding model proposed by the Strobel group (14) provided explanation for the inhibition of aptamer I's glycine affinity by a deleterious mutation in aptamer II of the P0-lacking VC glycine riboswitch. However, this model does not readily explain vice versa, the inhibition of aptamer II's glycine affinity by a deleterious mutation in aptamer I. Breaker and coworkers showed that a deleterious mutation in aptamer I (VCIIIG17C, nts 1-225, G17C; Fig. 2.1, see figure legend for construct naming scheme) not only disabled the glycine binding affinity of aptamer I, but also greatly reduced the glycine binding affinity of aptamer II in the P0-lacking VC glycine riboswitch (6). This interaptamer inhibition, also reported by the Strobel group (14) and our group (51), is consistent with the ligand binding cooperativity in the P0-lacking VC glycine riboswitch (VCIII). Ligand binding cooperativity in a multi-subunit system typically suggests that binding to one site greatly enhances the binding at other sites (58). However, one version of the isolated aptamer II, VCII122, has been demonstrated to bind glycine even stronger than the double-aptamer riboswitch VCIII (59 ± 12 μM, Table 2.1) with a binding affinity of 4.4 ± 2.0 μM (Table 2.1) (51). The discrepancy between the glycine affinity of aptamer II in VCIIIG17C ([2.3 ± 0.5] $\times$ 10^3 μM, Table 2.1, Fig. 2.2) and that in VCII122 suggests that certain structural elements may exist in aptamer I and interfere with aptamer II function when aptamer I is disabled or not fully formed. To test this hypothesis, we extended the sequence of aptamer II at the 5' end to prepare the VCII66 construct (nucleotides, nts 66−225, Fig. 2.3), which contains the 3' half of aptamer I, the linker between two aptamers, and the fully

functional aptamer II construct VCII122. As Kwon and Strobel (13) have indicated that P3a of aptamer I might be important to the glycine binding cooperativity in VCIII, we included the 5' half of the P3a helix (nts 66–69) to accommodate the whole P3a helix in the construct. The glycine-binding affinity of VCII66 was then evaluated via in-line probing assays which map local structure changes by measuring the flexibility of the phosphodiester bond to achieve in-line geometry for non-enzymatic cleavage through internal transphosphorylation (27, 28). Results showed that the glycine affinity of VCII66 as calculated from aptamer II regions (Fig. 2.1) is much weaker than that of VCII122, with a K_d value of $(1.1 \pm 0.3) \times 10^3$ μM (Table 2.1, Fig. 2.3). The weaker binding affinity could also result from a mixture of multiple folding species in which the misfolded species may exhibit high cleavage background and mask the cleavage pattern of the correctly folded species. The overall protection pattern of VCII66 in 100 mM glycine is similar to that of VCII122 in the overlapping regions (Fig. 2.3), indicating that aptamer II folds into similar overall structures in these two constructs at saturating glycine concentrations. In addition, we have analyzed the riboswitch folding under in-line probing conditions with native gel electrophoresis. Results confirmed that VCII66 folds into a single species (Fig. 2.4 Lane 1). Combined together, these experiments suggest that there are critical structural elements between nt 66 and nt 122 interfering with glycine affinity of aptamer II in VCII66.

Table 2.1 Binding affinities of glycine riboswitches and mutants[a]

RNA	K_d (μM)	Description of the constructs
VCII122	4.4 ± 2.0[b]	nts 122 to 225
VCIII	59 ± 12[b]	nts 1 to 225
VCIIIG17C	$(2.3 \pm 0.5) \times 10^3$	nts 1 to 225, G17C
VCII66	$(1.1 \pm 0.3) \times 10^3$	nts 66 to 225
VCII81	6.1 ± 1.9	nts 81 to 225
VCII73	$(1.5 \pm 0.2) \times 10^3$	nts 73 to 225
VCIIIG17CM1	$(1.3 \pm 0.5) \times 10^3$	nts 1 to 225, G17C, CUUU75-78AAAA
VCII66P3a	Misfold: $(1.5 \pm 0.3) \times 10^3$ Correct fold: 15.5 ± 4.2	nts 66 to 225, GAAG66-69UUUU, CUUU75-78AAAA
VCII73M1	Misfold: $(2.5 \pm 1.1) \times 10^3$ Correct fold: 34.0 ± 31.6	nts 73 to 225, CUUU75-78AAAA
VCIIIG17CP3a	$(1.2 \pm 0.2) \times 10^3$	nts 1 to 225, G17C, GAAG66-69UUUU, CUUU75-78AAAA
VCII73 (+DNA)	12.5 ± 5.8	nts 73 to 225, with 12 μM complementary DNA[c]

[a] Determined by in-line probing assays. Binding affinities were calculated as average K_d values (and ± their standard deviation) obtained from the individual glycine perturbed regions with curve fit coefficients better than 0.97 in at least two independent trials.
[b] Data were taken from Sherman et al. 2012, *RNA*, 496-507.
[c] Complementary oligodeoxynucleotide sequence is ACCTGAAAGATCAGAT.

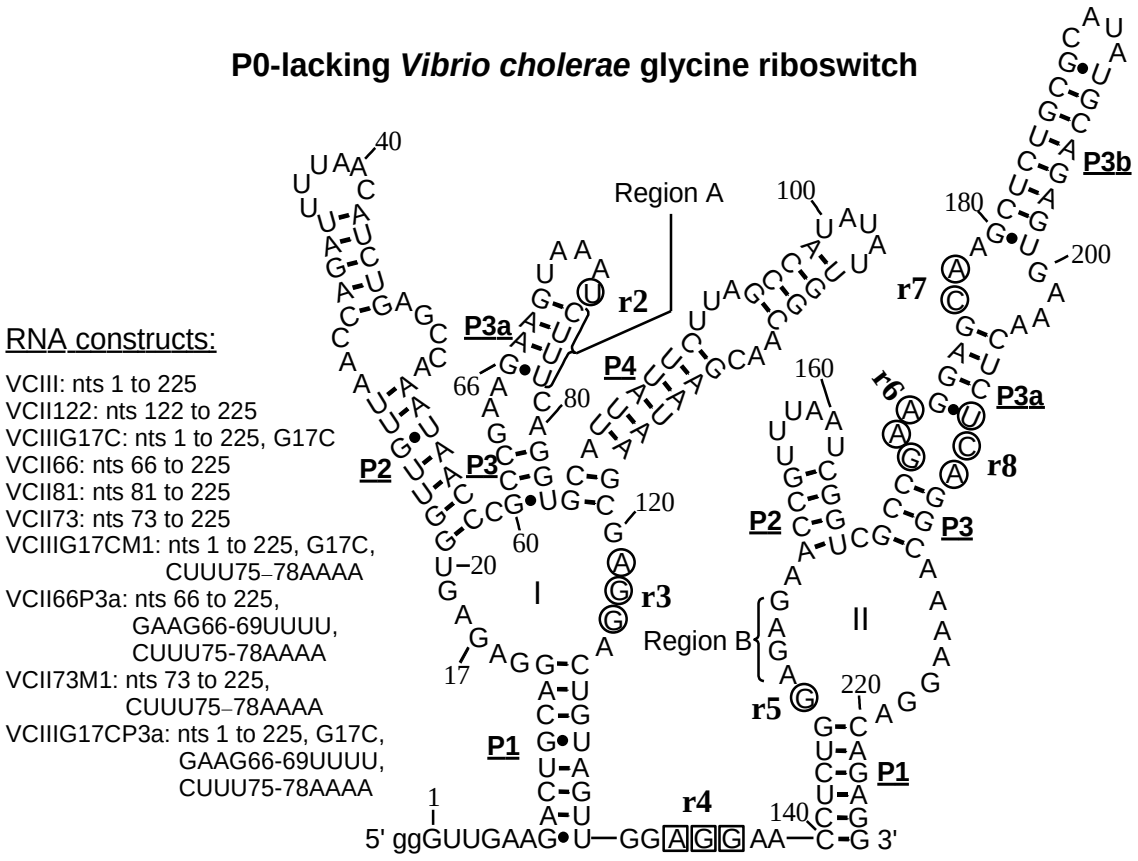


Figure 2.1 Secondary structure of the P0-lacking VC glycine riboswitch (VCIII).

r2–r8 indicate the representative glycine induced perturbation regions in the in-line probing assay of VCIII; the corresponding nucleotides are either circled or boxed with a square, which represents either decreased cleavage or increased cleavage with increasing glycine concentration, respectively. “I” and “II” refer to upstream aptamer I and downstream aptamer II, connected by a short linker. Definitions of construct lengths/mutations are given in the figure.

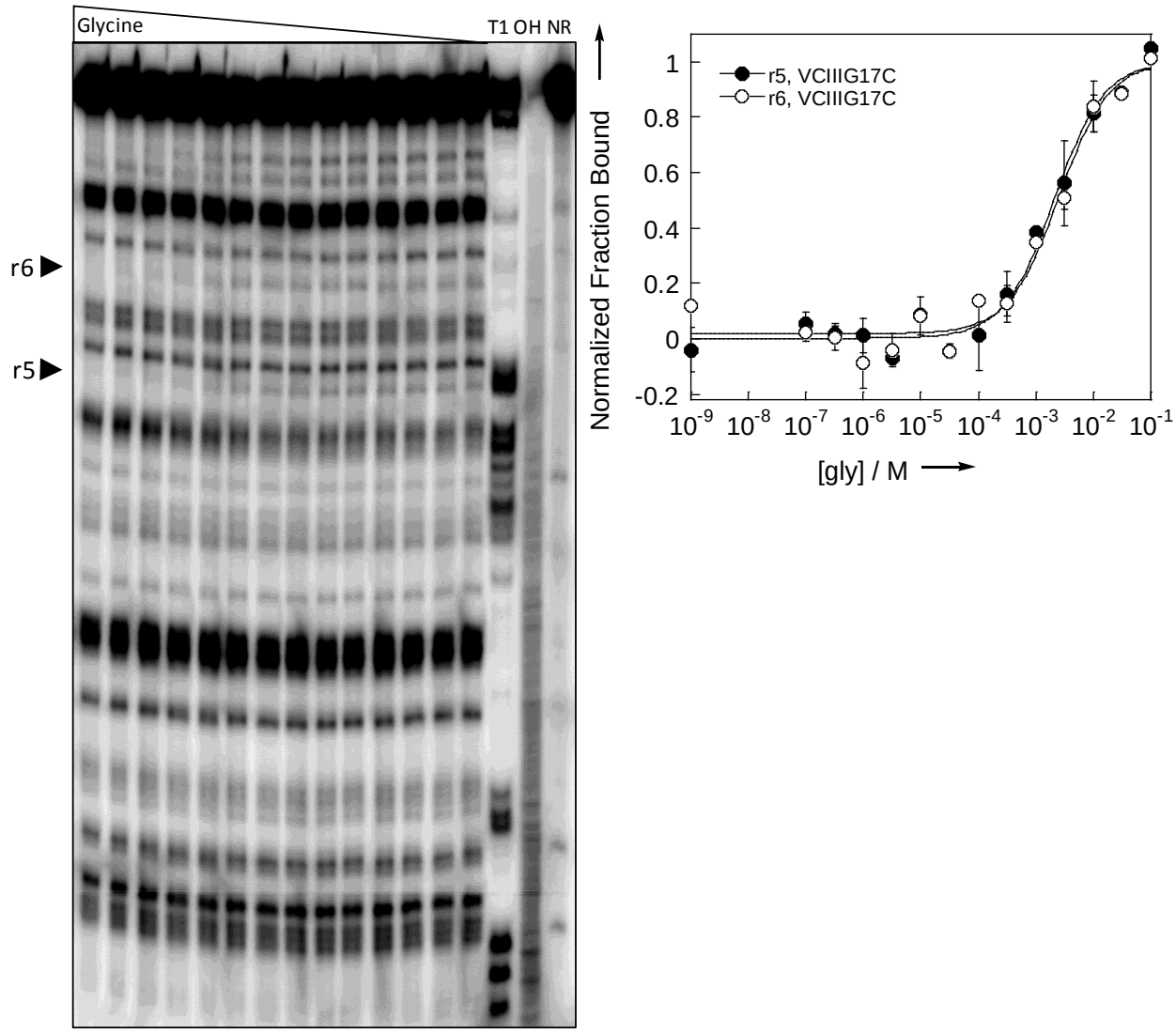


Figure 2.2 Representative in-line probing gel image and semi-log plots of the normalized fraction of RNA bound versus glycine concentration of VCIIIG17C.

NR, T1, and OH represent no reaction, partial digestion with RNase T1, and partial digestion with alkali, respectively. r5-r6 indicate regions defined in Fig. 2.1.

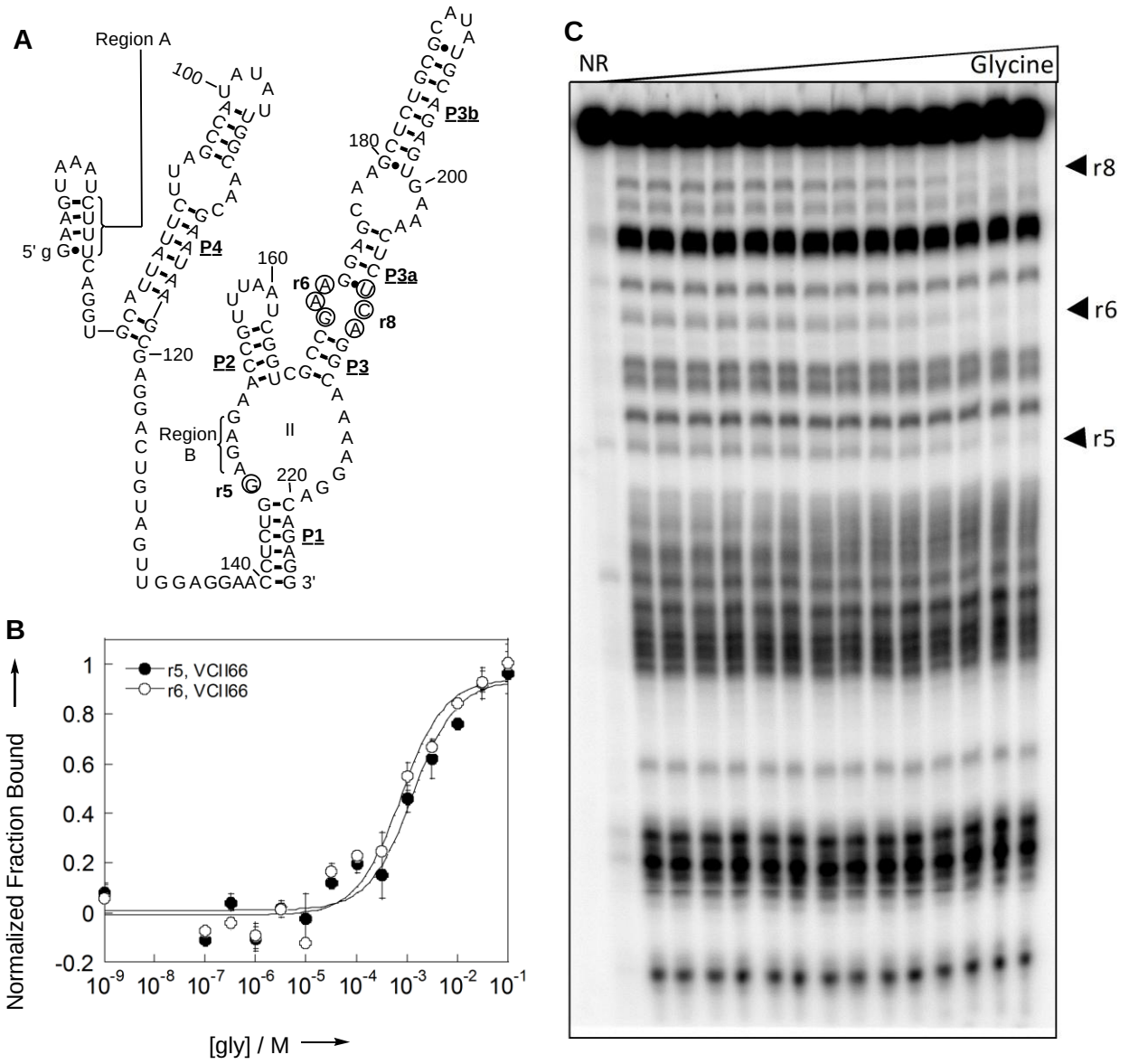


Figure 2.3 Glycine affinity of VCII66 is drastically reduced compared to that of VCII122.

A) Secondary structure of VCII66. r5–r8 and circled nucleotides are the same as defined in Fig. 2.1. B) Representative semi-log plots of the normalized fraction of RNA bound versus glycine concentration for r5 and r6 of VCII66. C) In-line probing gel image of VCII66. NR indicates no reaction.

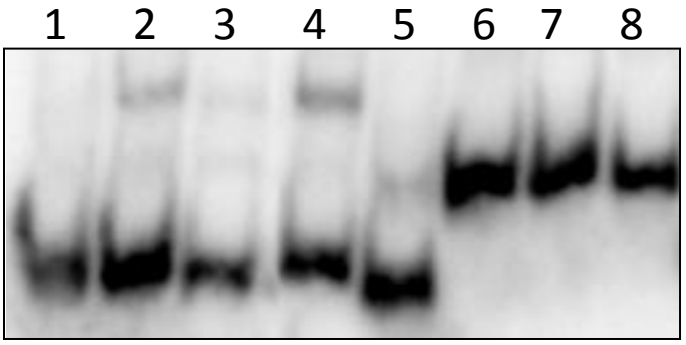


Figure 2.4 Native gel of 5' ^{32}P radiolabeled RNA constructs.

This shows the existence of a single species for each construct, with the exception of lane 2 and 4 showing an additional slower migrating band. Lane 1, VCII66; Lane 2, VCII66P3a; Lane 3, VCII73; Lane 4, VCII73M1; Lane 5, VCII81; Lane 6, VCIIIG17CP3a; Lane 7, VCIIIG17C; Lane 8, VCIIIG17CM1.

2.2.2 Identification of potential interaptamer interaction inhibiting the glycine affinity of VCII66

We denote the aptamer I structural element that interferes with aptamer II glycine binding as Region A, which is likely to interact with a specific region (or several regions) in aptamer II denoted Region B in the absence of glycine to reduce Region B's ability to participate in native structure formation and consequently decrease aptamer II's glycine affinity. Accordingly, in the in-line probing assays of VCIII or VCII66, Region B may interact with Region A in the absence of glycine while participating in the native structure formation in the presence of glycine. Thus, the in-line probing cleavage pattern of a potential Region B may appear to be unperturbed or only marginally perturbed at different glycine concentrations as either in the presence or in the absence of glycine it is structured. For VCII122, however, as Region A is not present in the construct, Region B may show structural perturbation from less structured in the absence of glycine to more structured in the presence of glycine. We compared the in-line probing patterns of the aptamer II regions in VCIII and VCII66 with that of VCII122 (Figures 1

and S2 in the report by Breaker and coworkers (6) and our in-line probing results) and found that nts 147–150 (AGAG) is the best region fit to this description of a candidate Region B. Examination of the corresponding regions in two other well-characterized glycine riboswitches from *Bacillus subtilis* (6) and *Fusobacterium nucleatum* (13) reveals the identical AGAG sequence; thusly nts 147-150 (AGAG) in VCIII is chosen as Region B. As long-range interactions in RNA structure involving a stretch of nucleotides often are through duplex formation with Watson-Crick or Wobble base pairs, such as kissing-loop interactions (33, 38, 59, 60) pseudoknots (60-63), etc., we postulated that Region A and Region B may interact through a simple duplex formation. Considering both Watson-Crick and Wobble pairing, Region B (AGAG) could pair with CUCU, CUUU, UUCU or UUUU. Searching the primary sequence of VCIII with these four short oligonucleotide sequences, we found three candidates in aptamer 1: nts 35–38 (UUUU), nts 75–78 (CUUU), nts 91–94 (UUCU). Comparing the corresponding sequences in the glycine riboswitches from *Bacillus subtilis* and *Fusobacterium nucleatum*, we found that only nts 75–78 is conserved to pair with Region B. Importantly, nts 75–78 locates between nt 66 and nt 122 as predicted in the previous section. Therefore, nts 75-78 (CUUU) is assigned as Region A in VC glycine riboswitch. Furthermore, nts 75–78 (CUUU) and nts 147–150 (AGAG) locate in the highly conserved regions of the 3' half of the P3a helix on aptamer I and J1/2 of aptamer II, respectively, strengthening their potential functional roles as Region A and Region B.

2.2.3 Verification and characterization of the inhibitory effect of Region A on aptamer II function in the shortened glycine riboswitch

First, we prepared a construct VCII81 (nts 81–225, Fig. 2.5A), removing the entire predicted Region A and taking advantage of the consecutive guanosines at position 81 and 82 for efficient *in vitro* transcription. If Region A is fully removed in this construct, we expect VCII81 to significantly rescue glycine binding affinity; in-line probing assays were used to determine the glycine binding K_d of VCII81. Results (Figures 2.5B and 2.5C) showed that VCII81 retains the

full ligand binding capacity of VCII122 with a K_d of 6.1 ± 1.9 μM (Table 2.1), consistent with the complete removal of Region A. Native gel electrophoresis results confirmed that VCII81 folds into a single species (Fig. 2.4 Lane 5). Interestingly, in the cleavage pattern of the in-line probing assay of VCII81 (Fig. 2.5C), in addition to the expected glycine perturbed regions r5–r8, we observed that nts 123–133 and nt 135 exhibit glycine-induced structural perturbation with similar glycine affinities. These observations rationalize the loss of glycine affinities in aptamer II for glycine riboswitches shorter than VCII122 reported previously (51). Particularly, nts 123–124 are located within region r3, a glycine-perturbed single stranded aptamer I region in the double-aptamer glycine riboswitches. Observation of glycine-induced structural perturbation in this region for VCII81 is reminiscent of the similar observation in the previously reported VCLDG17C construct (nts -7–225, G17C), in which region r3 exhibits similar glycine-induced structural perturbation as other aptamer II regions (r5–r7) although aptamer I is disabled (51). These results, combined together, confirm that aptamer II requires structural elements from aptamer I to bind glycine properly. In turn, this also rationalizes the potential existence of the aptamer I region (Region A) to modulate the glycine affinity of aptamer II.

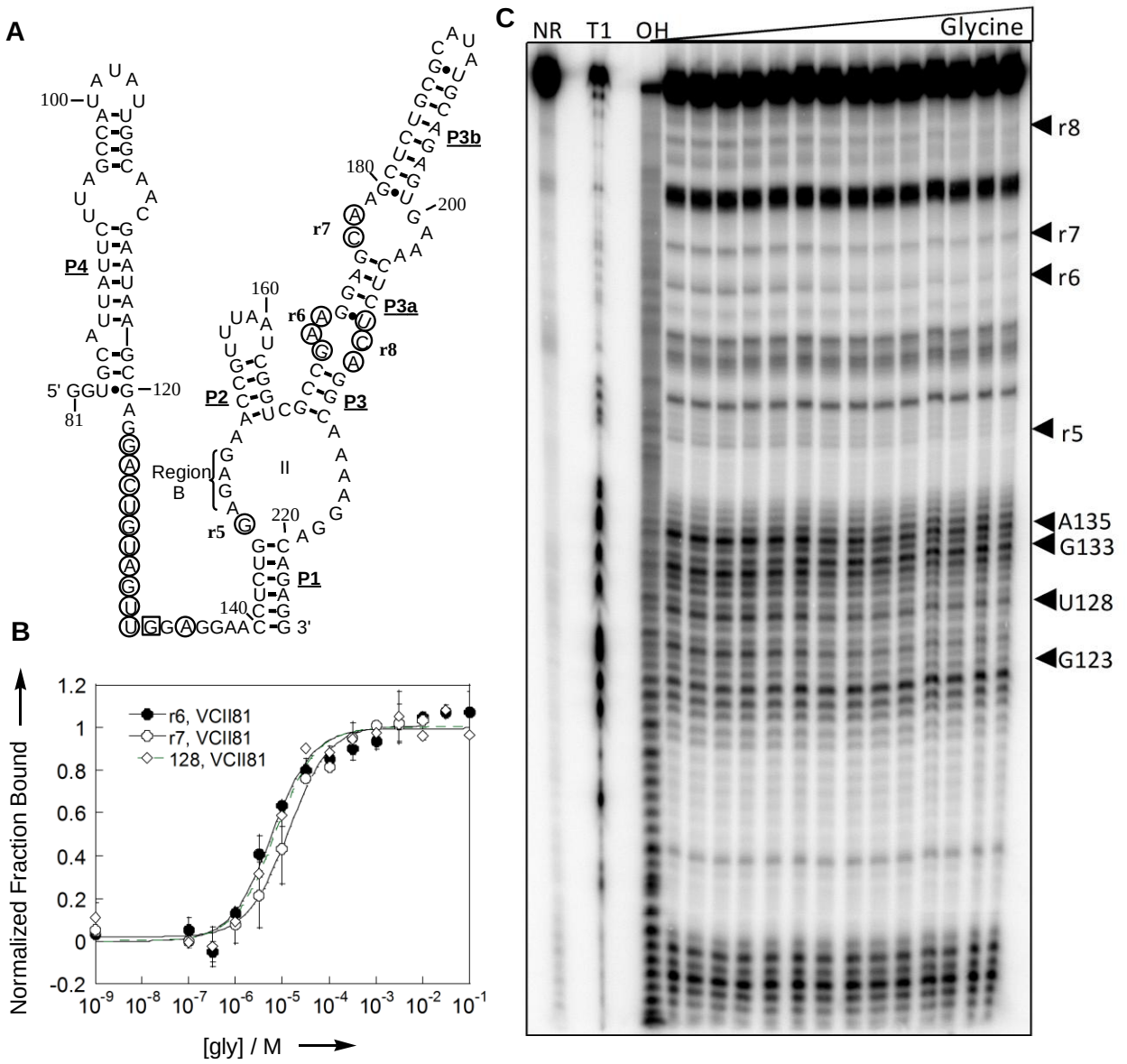


Figure 2.5 VCII81 retains glycine binding affinity of VCII122.

A) Secondary structure of VCII81. r5–r8 and the nucleotides circled or boxed with a square are the same as defined in Fig. 2.1. Additional glycine induced perturbation regions nts 123–133 and nt 135 are also denoted. B) Representative semi-log plots of the normalized fraction of RNA bound versus glycine concentration for r7, r6, and U128 of VCII81. C) In-line probing gel image of VCII81. NR, T1, and OH represent no reaction, partial digestion with RNase T1, and partial digestion with alkali, respectively.

The difference in the glycine binding affinities between VCII66 and VCII81 suggests that the structural elements modulating aptamer II function locate between nt 66 and nt 80, which contains the predicted Region A (nts 75–78, the 3' half of P3a), L3a, and the 5' half of P3a. To test whether L3a and the 5' half of P3a are responsible for modulating aptamer II function, we have prepared a construct VCII73 (nts 73–225), removing the 5' half of P3a and L3a. In-line probing experiments showed that VCII73 binds glycine much more weakly than VCII122 or VCII81 with a K_d value around $(1.5 \pm 0.2) \times 10^3$ μM (Table 2.1, Fig. 2.6), even weaker than VCII66. Similar to VCII66, native gel electrophoresis showed that VCII73 folds into a single species (Fig. 2.4 Lane 3), ruling out of the possibility of a masking effect of a misfolded species. The weaker affinity of VCII73 compared to VCII66 may be due to the fact that the 5' portion of the P3a stem is no longer present in VCII73 (as it is in VCII66) to competitively duplex with Region A, allowing Region A to inhibit aptamer II function even further through better interaction with Region B.

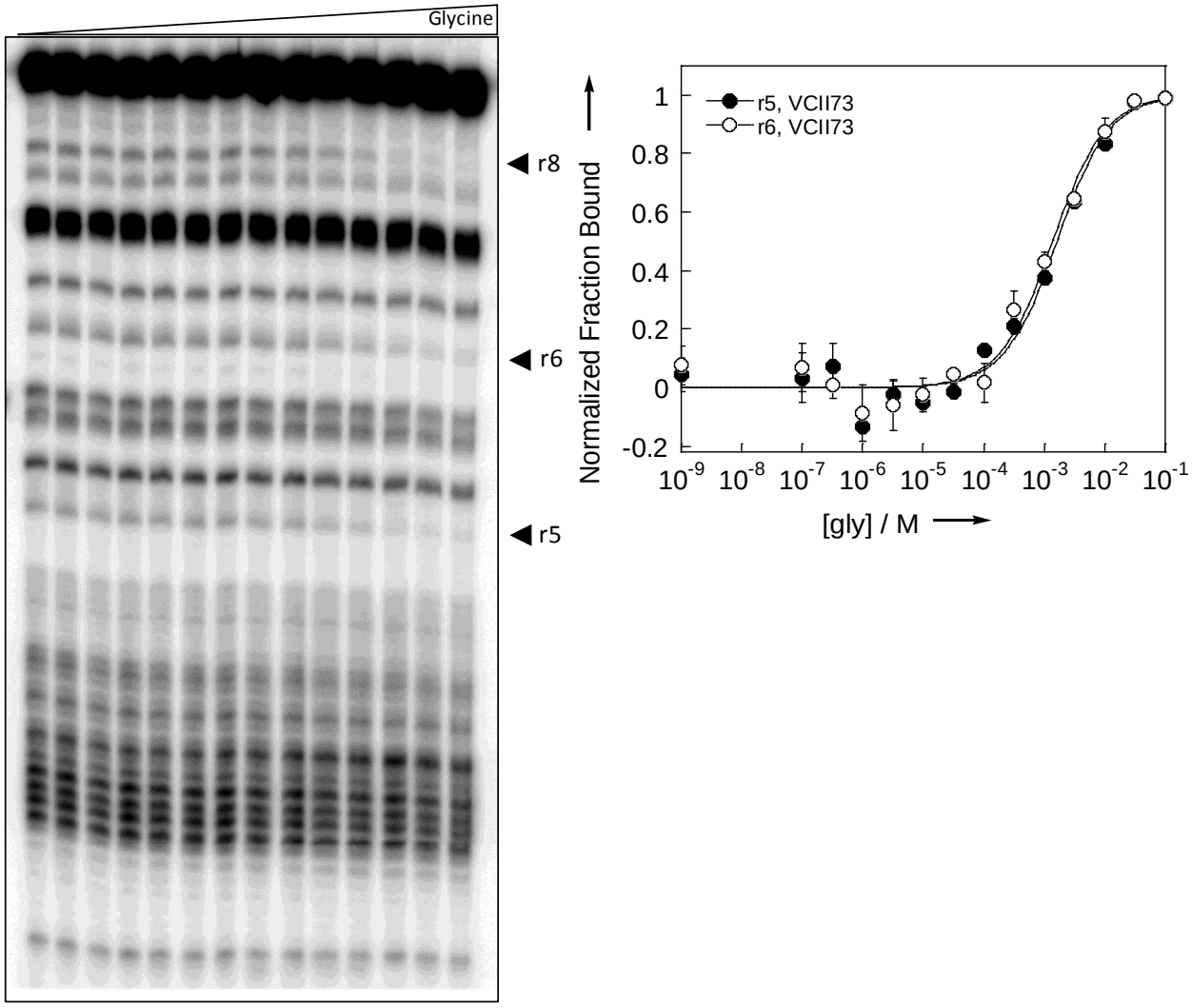


Figure 2.6 Representative in-line probing gel image and semi-log plots of the normalized fraction of RNA bound versus glycine concentration of VCII73.

r5-r8 indicate regions defined in Fig. 2.1.

Another way to verify the inhibitory effect of Region A on aptamer II glycine binding is to directly mutate Region A in VCII66 or VCII73 to prevent the interaction between Region A and Region B and analyze whether there is any rescue in aptamer II glycine affinity. We therefore prepared mutants VCII73M1 (nts 73–225, CUUU75-78AAAA) and VCII66P3a (nts 66 to 225, GAAG66-69UUUU, CUUU75-78AAAA). Here VCII66P3a contains extra mutations GAAG66-69UUUU to maintain possible P3a duplex formation. We then analyzed the glycine binding

properties of these constructs with in-line probing assays. Unfortunately both constructs exhibit two glycine induced transitions, making it difficult to evaluate the glycine affinity of these constructs. Native gel electrophoresis showed that in addition to the band at the expected location, a new slower migrating band was observed for each of these two constructs (Fig. 2.4 Lane 2 and Lane 4). We attribute this second species to either a simply misfolded or oligomeric form that may exhibit glycine induced transition at a higher glycine concentration. In-line probing data fit quite well with the equation describing two independent glycine-folded species (Methods section). The correctly folded species showed a glycine affinity of 15.5 ± 4.2 μM for VCII66P3a (Fig. 2.7) and 34 ± 32 μM for VCII73M1 (Fig. 2.8), both exhibiting a significant rescue from their parent constructs (Table 2.1).

Collectively, these results confirmed the predicted Region A (nts 75–78) to be the structural element responsible for inhibiting aptamer II glycine affinity. In addition, Region A is solely responsible for the inhibition without the need for loop L3a or P3a duplex formation.

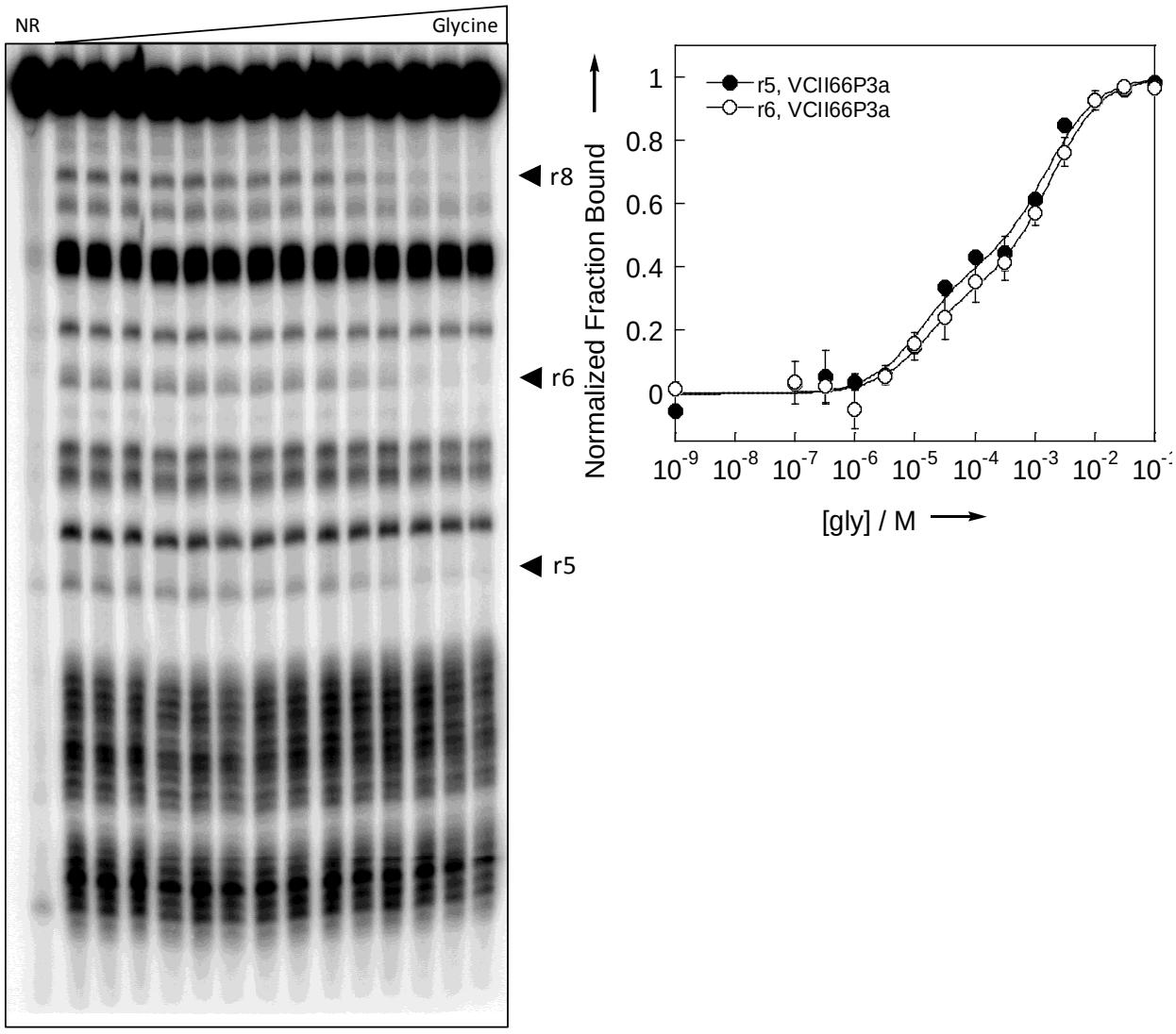


Figure 2.7 Representative in-line probing gel image and semi-log plots of the normalized fraction of RNA bound versus glycine concentration of VCII66P3a.

NR represents no reaction. r5-r8 indicate regions defined in Fig. 2.1.

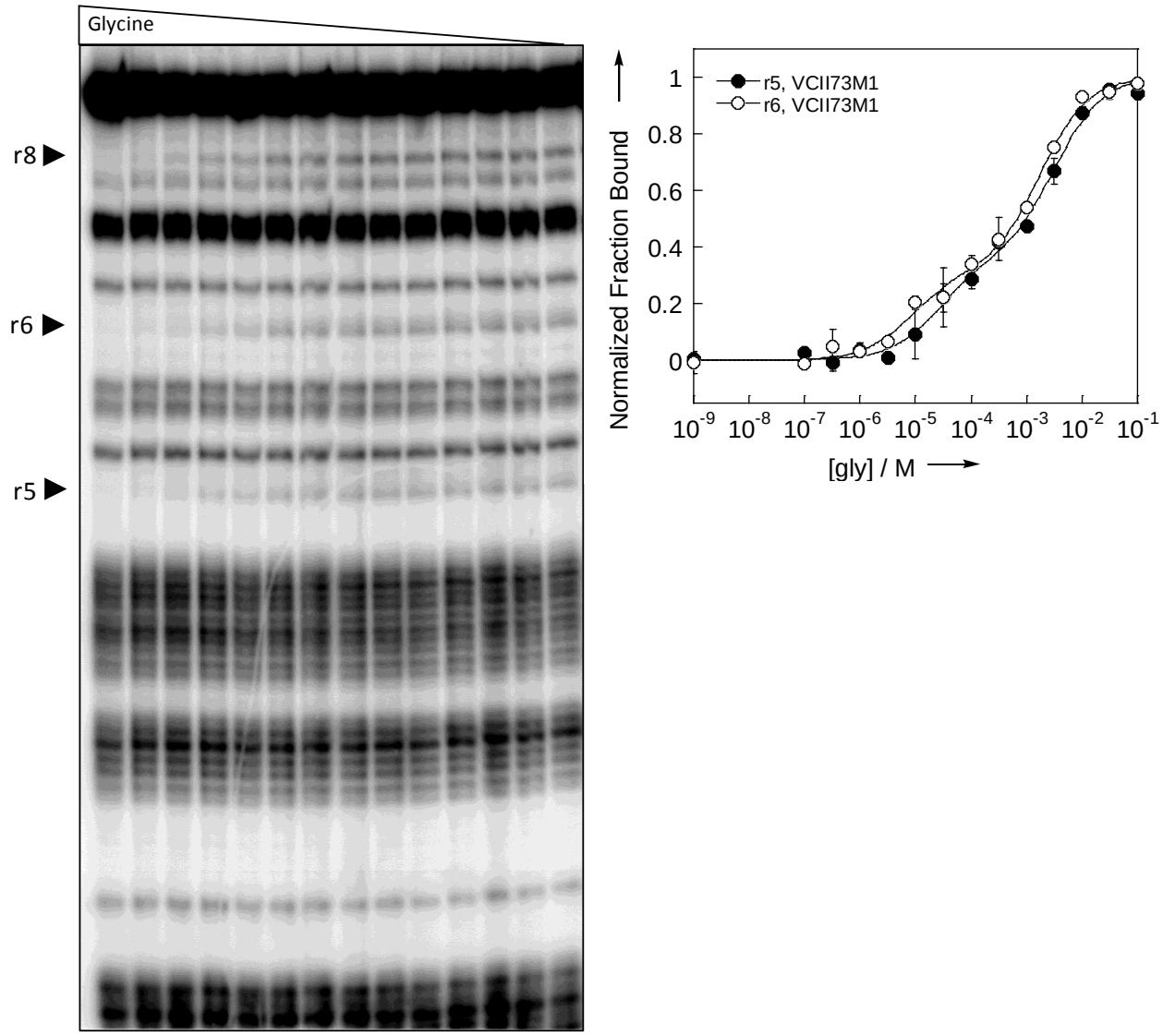


Figure 2.8 Representative in-line probing gel image and semi-log plots of the normalized fraction of RNA bound versus glycine concentration of VCII73M1.

r5-r8 indicate regions defined in Fig. 2.1.

2.2.4 Characterization of the interaptamer effect of Region A in double aptamer glycine riboswitches

Having established the modulating role of Region A on the glycine binding affinity of aptamer II in the single aptamer constructs, next we investigate its role on aptamer II in the

double aptamer glycine riboswitches. According to the crystal structure of the glycine riboswitches (18, 19), the corresponding Region A directly participates in the glycine binding pocket of aptamer I. This means direct mutation of Region A on the wild-type double aptamer VCIII glycine construct will likely diminish or disable aptamer I glycine binding, making it difficult to precisely dissect the effect of Region A on aptamer II glycine affinity as deleterious mutation in aptamer I was demonstrated to reduce the glycine affinity in aptamer II previously (6, 13, 51). To circumvent this complication, we have examined the effect of Region A on the double-aptamer VC glycine riboswitch (VCIII) in the presence of G17C mutation, which rendered aptamer I incapable of binding glycine regardless of the presence or absence of Region A. As reported previously (6, 13, 51), mutation in VCIIIG17C disables aptamer I function and greatly reduced the glycine affinity of aptamer II. If Region A remains functional and modulates the glycine binding activity of aptamer II in VCIIIG17C, one would expect that a mutation in Region A would remove this modulation and rescue the glycine affinity in aptamer II. We therefore prepared VCIIIG17CM1 (nts 1-225, G17C, CUUU75-78AAAA) and analyzed its glycine binding properties. Native gel showed that VCIIIG17CM1 folded into a single species similar to VCIIIG17C (Fig. 2.4 Lane 7 and Lane 8) and in-line probing experiments showed that mutation in Region A had a modest rescuing effect (~2 fold) and increased the glycine affinity of aptamer II in VCIIIG17CM1 to $(1.3 \pm 0.5) \times 10^3$ μM (Fig. 2.9A, Fig. 2.10) from $(2.3 \pm 0.5) \times 10^3$ μM in VCIIIG17C. In addition, we prepared VCIIIG17CP3a (nts 1-225, G17C, GAAG66-69UUUU, CUUU75-78AAAA) which contains additional mutations to maintain a possible P3a duplex. Native gel showed that VCIIIG17CP3a folded into a single species (Fig. 2.4 Lane 6) and in-line probing experiments showed that VCIIIG17CP3a bound glycine with a modestly rescued affinity of $(1.2 \pm 0.2) \times 10^3$ μM (Fig. 2.11), similar to VCIIIG17CM1.

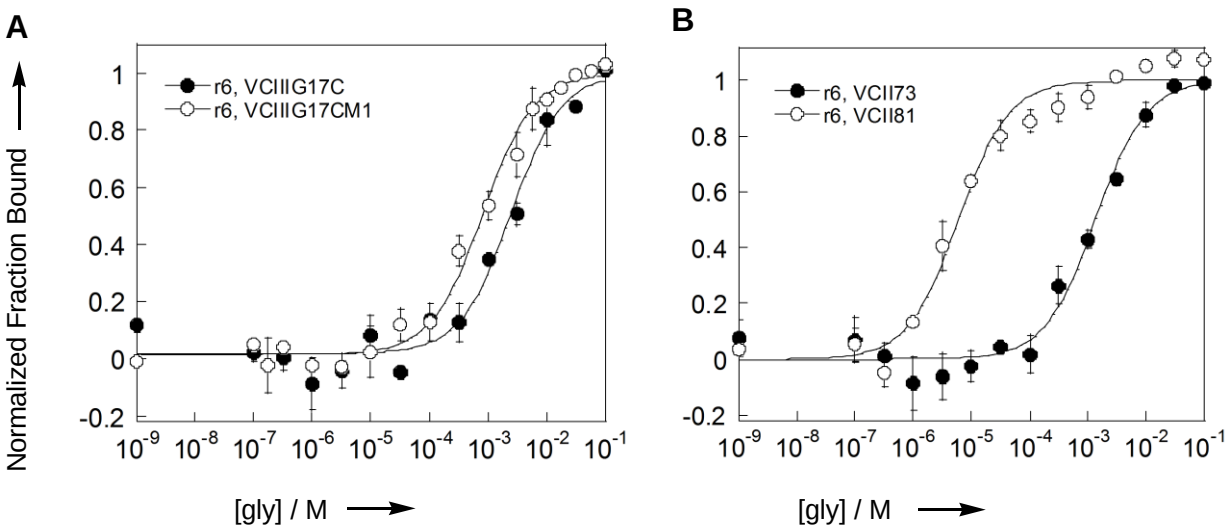


Figure 2.9 Effect of Region A on the glycine binding affinity of aptamer II.

Representative semi-log plots of the normalized fraction of RNA bound versus glycine concentration for r6 of VCIIIG17CM1 compared to that of VCIIIG17C (A) and VCII73 compared to VCII81 (B).

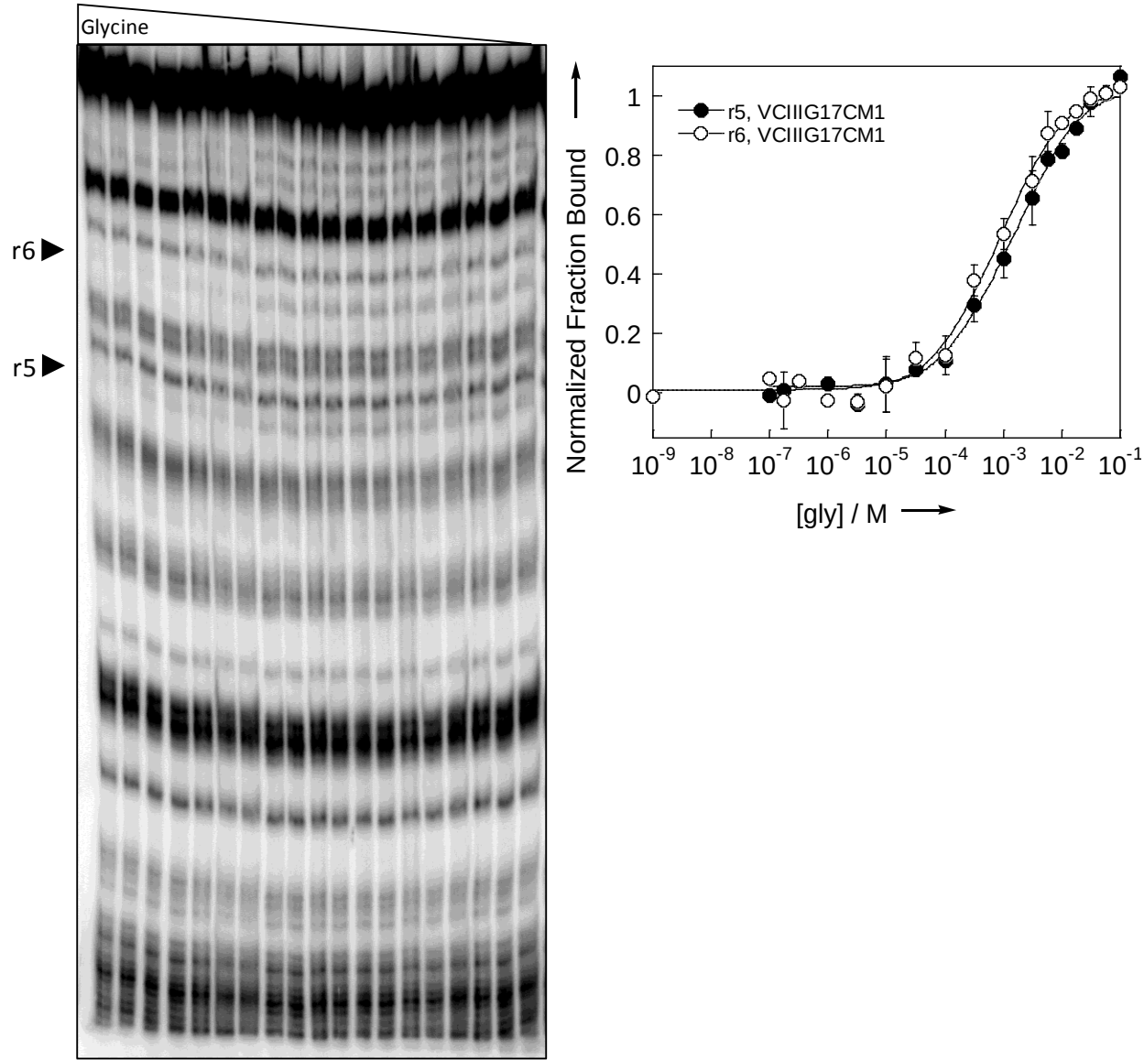


Figure 2.10 Representative in-line probing gel image and semi-log plots of the normalized fraction of RNA bound versus glycine concentration of VCIIIG17CM1.

r5-r6 indicate regions defined in Fig. 2.1.

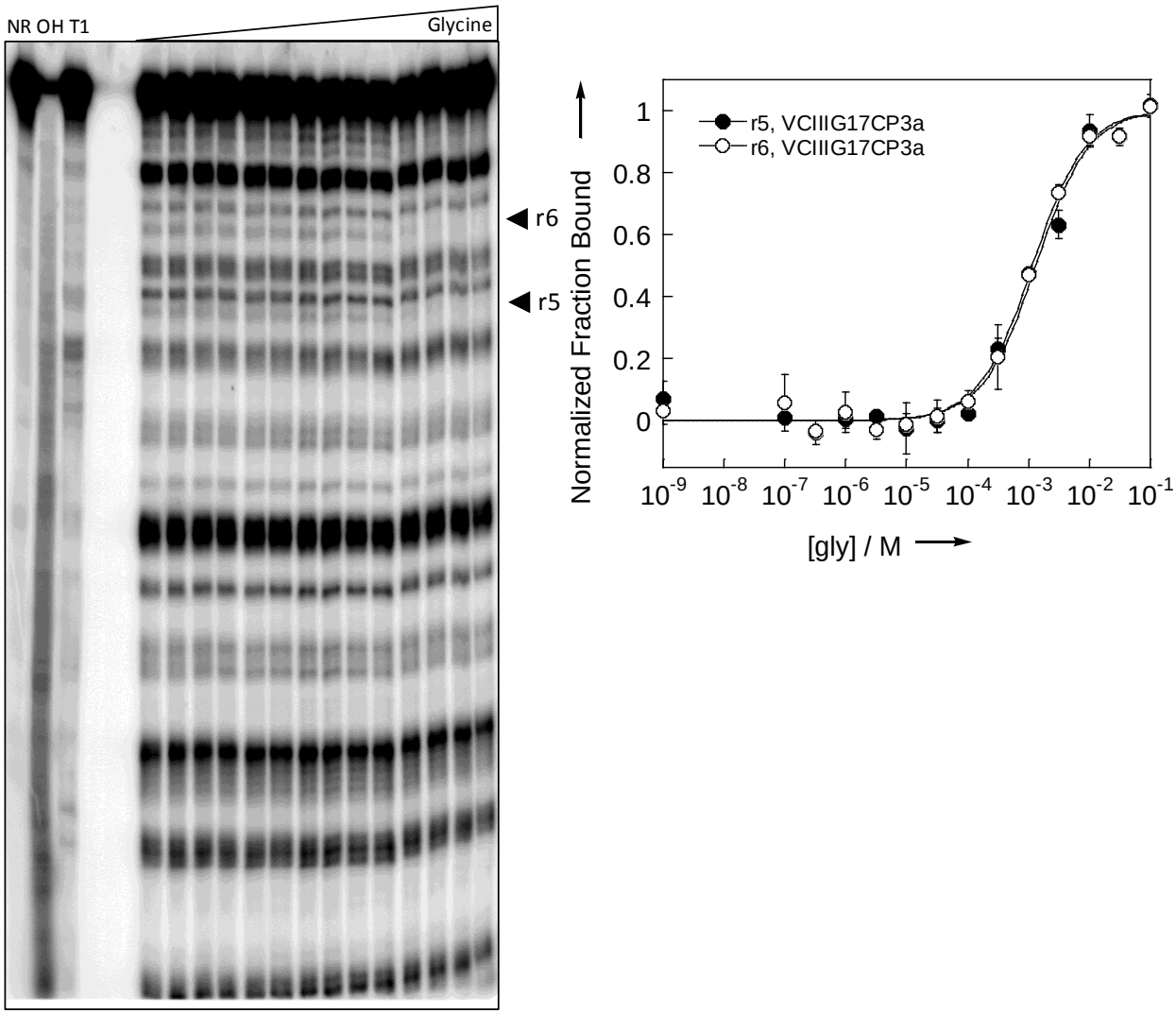


Figure 2.11 Representative in-line probing gel image and semi-log plots of the normalized fraction of RNA bound versus glycine concentration of VCIIIG17CP3a.

NR, T1, and OH represent no reaction, partial digestion with RNase T1, and partial digestion with alkali, respectively. r5-r6 indicate regions defined in Fig. 2.1.

The drastically different modulating effect of Region A in the two versions of constructs, VCIIIG17C vs. VCIIIG17CM1/VCIIIG17CP3a (~2 fold modulation, Fig. 2.9A, Table 2.1) and VCII73 versus VCII81 (~200 fold modulation, Fig. 2.9B, Table 2.1), shows that Region A modulates aptamer II activity in a context dependent manner and truncation has a large effect on the interaptamer interaction and glycine binding. In the context of the double-aptamer glycine

riboswitches, Region A plays a partial but not a sole role in the interaptamer inhibition of aptamer II. What else can cause the interaptamer inhibition of aptamer II in the double-aptamer glycine riboswitches? It is possible that when the aptamer I glycine binding site is not occupied, interaptamer interaction is not formed properly for optimum glycine binding in the aptamer II site. This is not completely surprising as we have demonstrated that glycine binding acts synergistically with interaptamer interaction in the glycine riboswitch (51). The smaller rescuing effect of Region A mutation in the double-aptamer glycine riboswitch may be more biological relevant to the naturally occurring glycine riboswitch, where the full-length P0-containing VC glycine riboswitch does not exhibit glycine binding cooperativity (51, 53). However, this result also shows that the shortened VC glycine riboswitch VCII73 is more suitable for future engineering into a novel genetic circuit.

2.2.5 Allosteric inhibition of aptamer II function by Region A may be rescued by a DNA oligonucleotide

Our finding of the inhibitory role of Region A on aptamer II glycine affinity in the shortened VC glycine riboswitch is intriguing, for the ~200 fold effect provides a good window for potential gene expression control. However, in order to harness this effect, we need to be able to reverse this inhibition by an exogenous effector, a signal that can be sensed by an *in vitro* or *in vivo* system based upon this in the future. Therefore, we investigated whether the modulation effect of Region A on aptamer II can be rescued by such a simple input signal. Our previous analysis indicates that Region A may regulate aptamer II function by duplexing with sequences on aptamer II; unsurprisingly, this suggests that a short nucleic acid strand complementary to Region A may competitively remove the inhibitory effect and rescue the glycine binding affinity of aptamer II. To test this, the glycine binding affinity of VCII73 in the presence of complementary oligodeoxynucleotide (ACCTGAAAGATCAGAT, Fig. 2.12A) was determined with in-line probing. Results showed that in the presence of 12 μM complementary oligodeoxynucleotide, VCII73 exhibited a similar glycine perturbation pattern as VCII81 and

bound glycine with an affinity of 12.5 ± 5.8 µM (Figures 2.12B and 2.12C). This corresponds to a 120-fold increase of glycine affinity from that of VCII73 and an almost complete restoration of the glycine affinity of VCII81.

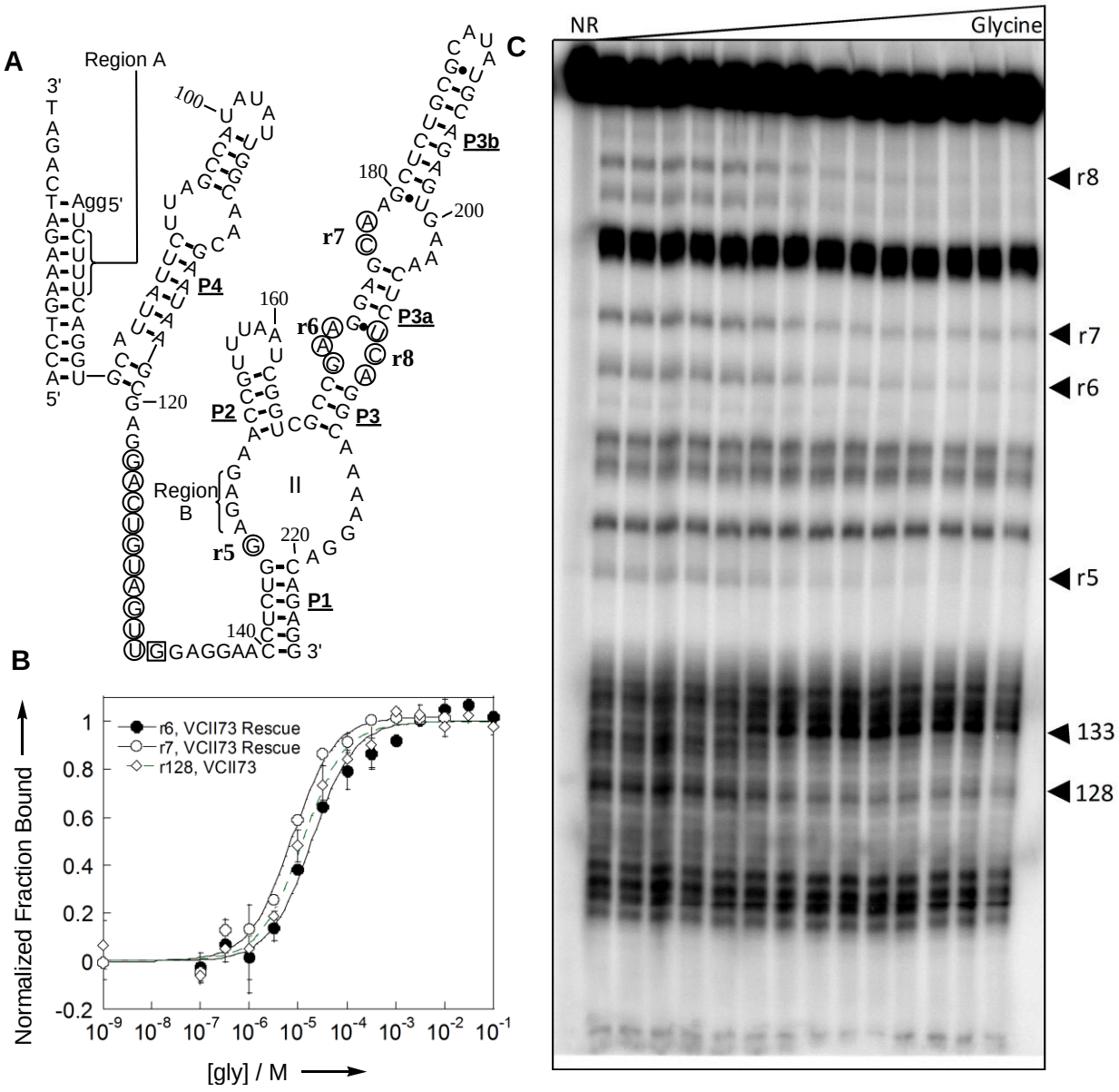


Figure 2.12 The glycine binding affinity of VCII73 can be rescued by the addition of a complementary oligodeoxynucleotide.

A) Secondary structure of VCII73 when annealed to complementary oligodeoxynucleotide ACCTGAAAGATCAGAT. r5–r8 and the nucleotides circled or boxed with a square are the same as defined in Fig. 2.1. Additional glycine induced perturbation region nts 123–133 are also

denoted as defined in Fig. 2.1. B) Representative semi-log plots of the normalized fraction of RNA bound versus glycine concentration for r7, r6, and U128 of VCII73 rescued by complementary oligodeoxynucleotide ACCTGAAAGATCAGAT. C) In-line probing gel image of VCII73 rescued by complementary oligodeoxynucleotide ACCTGAAAGATCAGAT. NR represents no reaction.

2.3 Discussion

Glycine riboswitches are most often located upstream of the *gcvTHP* operon controlling the expression of the glycine cleavage system (6, 7), which breaks down glycine and produces carbon dioxide, ammonia, and 5-10-methylenetetrahydrofolate in the presence of Tetrahydrofolate (8). 5-10-methylenetetrahydrofolate can also serve as a source of methyl groups for methionine, serine, and 2'-deoxythymidine (9). Glycine riboswitches have also been identified upstream of the *glcB* gene in the most abundant marine bacterium *Candidatus* Pelagibacter ubique (12), controlling the expression of malate synthetase which makes malate from exogenous glycine derivatives. These show that naturally occurring glycine riboswitches are important regulators of carbon metabolism in bacteria. Besides the natural gene control functions, glycine riboswitches have been used for controllable overproduction of recombinant proteins using the inexpensive glycine inducer (54). This has demonstrated application of the glycine riboswitch as a general genetic control element. However, the ligand binding cooperativity in the P0-lacking glycine riboswitch suggests that this system may be able to offer more than just a basic metabolite-sensing genetic control gate to synthetic biology.

In this work we have identified a structural element (Region A) located in aptamer I that regulates the ligand binding affinity of aptamer II in the shortened glycine riboswitches, VCII66 and VCII73. Removal of Region A from VCII73 resulted in a construct VCII81 which showed fully restored (~200 fold rescue) glycine binding affinity. The interaction partner of Region A in aptamer II is proposed as Region B (nts 147-150) for this aptamer II modulation. However, this is not experimentally tested because the proposed Region B is highly conserved, and the

mutations in Region B are likely to diminish the glycine affinity of aptamer II, making any rescue of aptamer II's glycine affinity due to the disruption of Region A and B interaction unobservable. Interestingly, a short oligodeoxynucleotide complementary to Region A and neighboring nucleotides can rescue the glycine binding activity of VCII73 by ~120 fold. Given the demonstrated importance of the gene control functions of natural and synthetic glycine riboswitches, our findings could pave the path for developing a new genetic circuit sensing two mechanistically distinct signals, a short nucleic acid strand and small molecule metabolite (glycine).

2.4 Materials and Methods

2.4.1 DNA constructs of the RNAs

The VC glycine riboswitch genes were generated by recursive PCR (55) from DNA oligonucleotides (Integrated DNA Technologies) with EcoRI restriction site and T7 promoter sequence engineered at the 5' terminus, and EarI and HindIII sites engineered at the 3' terminus similar to that described previously (56). Double digestion of the riboswitch DNA genes with *Eco*RI and *Hin*dIII (New England Biolabs) was then performed in the supplied NEB buffer 2 at 37°C for 1 hour. Double digested riboswitch DNA genes and pUC19 vector were ligated together using T4 DNA ligase (New England Biolabs) in the supplied T4 DNA ligation buffer at 16°C with overnight incubation to generate glycine riboswitch plasmids as reported previously (42). The glycine riboswitch mutants were constructed either by a similar method or by QuikChange site-directed mutagenesis (Stratagene) from the corresponding wild type plasmids and proper DNA oligonucleotide primers. These plasmids were transformed into JM109 cells, amplified, purified with Qiagen plasmid preparation kits, and sequenced by Eton Bioscience Inc.

In vitro transcription for the preparation of RNAs

Riboswitch plasmids were linearized by incubation with EarI (New England Biolabs) at 0.2 unit per µg of plasmid in the supplied NEB buffer 4 at 37°C overnight. RNAs were

transcribed in 1 ml reaction mixture containing 40 mM Tris-HCl (pH 7.9), 2 mM spermidine-3HCl, 10 mM DTT, 25 mM $MgCl_2$, 5 mM each of the nucleotide triphosphates, 50 µg/ml linearized plasmid, 0.01 unit/µl 5 PRIME^{TM} stop RNase inhibitor RX (5 PRIME life science), 8 miliunit/µl thermostable inorganic pyrophosphatase (New England Biolabs), and 45 µg/ml T7 RNA polymerase (expressed and purified in-house) and incubated overnight at 37°C. The transcription mixture was incubated with 1 mM $CaCl_2$ and RNase-free DNase I (Promega) at a concentration of 0.5 unit per µg of DNA plasmid at 37°C for 30 minutes. The RNA transcripts were purified by 6-8% denaturing polyacrylamide gel electrophoresis, isolated by UV shadowing and excised, electroeluted into TE buffer (10 mM Tris-HCl, pH 8, 1 mM EDTA), PCA extracted, ethanol precipitated, and resuspended in TE buffer. RNA concentrations were determined by UV absorbance at 260 nm wavelength with extinction coefficients calculated by OligoCalc (57).

5' end ^{32}P labeling of RNAs

In a typical procedure, 40 pmol RNA transcript was dephosphorylated by incubation with 1 unit of shrimp alkaline phosphatase (Fermentas Life Sciences) in the supplied phosphatase buffer in a total volume of 10 µl at 37 ^{o}C for 30 min. After incubation at 65 ^{o}C for 15 min to inactivate the phosphatase, 1 µl sterile H_2O, 1.5 µl 10x T4 polynucleotide kinase buffer (New England Biolabs), 2 µl γ-^{32}P ATP (Perkin Elmer), and 0.5 µl 10 unit/µl T4 polynucleotide kinase (New England Biolabs) were added and the reaction mixture was incubated at 37 ^{o}C for 30 min for RNA phosphorylation. For RNA transcripts with the 5' end buried by or too close to a duplex region, to improve the labeling yield, the transcripts were first annealed to a DNA oligonucleotide to free the 5' end before dephosphorylation and phosphorylation reactions. For example, to radiolabel the 5' end of VCII81 (nts 81-225), 40 pmol VCII81 was first incubated with 1.05 eq of an annealing oligonucleotide (AACTACAGTCCTCGCTTATTC) at 95 ^{o}C for 1 min, then at room temperature for 10 min before dephosphorylation and phosphorylation reactions. After phosphorylation, ^{32}P-labeled RNAs were purified by 6-8% denaturing polyacrylamide gel electrophoresis, visualized on an x-ray film, excised, and passively eluted in

300 µl TE buffer on a rotator at 4 ^{o}C overnight. After centrifugation, the radiolabeled RNAs were PCA extracted twice, precipitated with ethanol, and redissolved in TE buffer.

2.4.2 In-line probing assays and quantification

In-line probing assays were carried out similar to the reported procedure (28). Briefly, ~50 kcpm ^{32}P labeled RNA transcript was incubated at 25 ^{o}C for ~48 h in 50 mM Tris-HCl (pH 8.3), 20 mM $MgCl_2$, 100 mM KCl under various glycine concentrations. For the in-line probing assays of VCII73 in the presence of complementary oligodeoxynucleotide (ACCTGAAAGATCAGAT), 12 µM complementary oligodeoxynucleotide was first annealed with 50 kcpm 5' ^{32}P radiolabeled VCII73 by incubation at 95 ^{o}C for 1 min and then room temperature 5 min before addition of $MgCl_2$ and KCl. After 48 h incubation at 25 ^{o}C, spontaneously cleaved RNA fragments were resolved by 8% denaturing PAGE, dried, exposed to PhosphorImager screens, scanned by PhosphorImager SI (Molecular Dynamics), and quantified by ImageQuant v5.2 (Molecular Dynamics). For VC glycine riboswitches and mutants, the following regions are quantified if contained on the constructs and well-resolved: U74 (r2), A121-G123 (r3), nts 124-133 (for VCII81, and VCII73 when rescued by a complementary oligodeoxynucleotide), G136-G137 (r4), G146 (r5), G170-A172 (r6), C177-A178 (r7), and U207-C208. To control for loading differences, the band intensities for tight binding constructs (K_d < 100 µM, including VCII66P3a and VCII73M1) were normalized to reference bands (U156-U161). The reference bands do occasionally show small perturbation which has little effect on the glycine affinity of the tight binding constructs but can notably affect weak binders (K_d > 500 µM); to avoid this, weak binders were not normalized. K_d values of the individual glycine-perturbed regions were determined by non-linear regression fitting of each plot using the following equation in KaleidaGraph software v3.09 (Synergy Software): fraction bound = $m_1 \times [gly]/(K_d + [gly]) + m_2$. The reported binding affinities were calculated as average K_d values obtained from the individual glycine-perturbed regions with curve fit coefficients better than 0.97 in at least two independent trials. For constructs that exhibit two glycine induced transition due to two

independent folding species, the following equation was used: fraction bound = $m_1 \times [gly]/(K_{d1} + [gly]) + m_2 \times [gly]/(K_{d2} + [gly]) + m_3$.

2.4.3 Native gel electrophoresis

5' radiolabeled RNA was folded by the following procedure: 1 µl 20 µM RNA was added to 7 µl 10 mM Tris-HCl (pH 7.5) and incubated at room temperature for 2 min. The VCII73 rescue DNA oligo, where used, was also added to the mixture and incubated at either room temperature or 95 ^{o}C for 2 min. 1 µl 100mM $MgCl_2$ and 1 µl 0.1 M glycine were added to the mixture, followed by incubation at room temperature for 20 min. This mixture was then combined with 2 µl native gel loading dye (50%v/v glycerol, 0.1%w/v xylene cyanol) and loaded onto a 6% native polyacrylamide (29:1 acrylamide:bisacrylamide) gel in TB buffer (89 mM Tris base and boric acid) containing 5 mM $MgCl_2$ and 5 mM glycine. The electrophoresis setup was prerun with TB buffer containing 5 mM $MgCl_2$ for two hours at constant 25 W and conducted for additional 3.5 hours after sample loading at 4^{o}C with constant 250 V. The gel was dried, exposed to PhosphorImager screens, scanned with PhosphorImager SI (Molecular Dynamics), and analyzed by ImageQuant v5.2 (Molecular Dynamics).

CHAPTER 3 SPECIFIC RNA-BINDING ANTIBODIES WITH A FOUR AMINO ACID CODE

3.1 Introduction

Recent advances in genomic and next-generation sequencing have greatly accelerated the discovery of non-protein coding RNAs (ncRNAs) (64, 65). Revealed by high-throughput methods, the observation that 90% of the human genome is transcribed while only 2% is protein-encoding suggests that the human transcriptome likely provides a rich spectrum of non-coding sequences to expand the RNA function repertoire (66). For example, small ncRNAs such as microRNAs have greatly shaped our view of posttranscriptional gene regulation (67).

The structurally more complex ncRNAs are RNAs with larger sizes ranging from about one hundred bases to several kilobases, including long non-coding RNAs (lncRNAs) from the eukaryotic genomes (66, 68), as well as riboswitches (1) and the exceptional structured ncRNAs (69) from bacteria. Aided by computational methods (66) and experimental approaches such as cDNA cloning and gene tiling arrays (70), thousands of large ncRNAs are reported each year (68). They play key roles in a wide range of functions, including transcriptional regulation, translational control, epigenetics, stem cell biology, cancer, signaling, and brain function (1, 68, 71, 72).

To carry out their biological functions, the large ncRNAs typically adopt extensive secondary and tertiary structures and often interact with a partner, such as a protein, RNA, ribonucleoprotein complex, metal ion, or metabolite (48, 68, 73). The structured nature of most large ncRNAs renders them suitable targets for specific recognition by well-characterized proteins, such as antibody fragments (42, 74). We are particularly interested in using specific antibody recognition to assist the structural and functional studies of ncRNAs for the following reasons.

First, antigen binding fragments (Fabs) when complexed with RNAs may help elucidate RNA structures (74, 75). Elucidating the structures of ncRNAs with unknown functions may

provide important clues towards revealing their functions and the molecular mechanisms behind their functions. However, detailed tertiary structure of large ncRNAs often requires high-resolution crystal structures that are generally difficult to obtain (42, 76, 77). Chaperone-assisted RNA crystallography (CARC) represents one promising method which employs an antibody fragment to bind RNAs, aiding in crystallization through extensive crystal contact participation as well as facilitating the structure solving process by providing superior initial phasing information (74, 75).

Secondly, targeting structured ncRNAs with antibodies (or antibody fragments) may provide potential therapeutics for cancer or other human diseases involving RNA causative agents. Numerous lncRNAs have been shown to lead to tumorigenesis. The lncRNA HOX transcript antisense RNA (HOTAIR), for instance, targets polycomb repressive complex 2 (PRC2) genome-wide to alter H3K27 methylation and gene expression patterns, resulting in increased cancer invasiveness and metastasis in breast carcinomas (78). A new class of ncRNAs, the transcribed ultraconserved regions (T-UCRS), has been shown to be abnormally expressed in a number of human cancers such as leukemia, colorectal carcinoma and hepatocellular carcinoma (79-82). Besides carcenogenicity, large ncRNAs are also implicated in non-neoplastic diseases, such as Alzheimer's disease (83), neurodeveloplemtal disorders (84), and imprinting disorders (85). Targeting these disease-associated RNAs with antibody fragments may provide diagnostic or therapeutic tools in the future.

In addition, RNA-binding antibodies may be used to help discover interaction partners of ncRNAs by immunoprecipitating the RNAs directly or to determine cellular distribution of ncRNAs through imaging fluorescently labeled RNA-binding antibodies (74, 86).

Recent advances in antibody phage display have led to facile production of antibodies binding to proteins in vitro without the expensive and laborious hybridoma approach (87-89). Using a phage platform that displays Fab libraries with naïve synthetic repertoires, we have recently established a general approach to obtain Fabs binding to structured RNAs (42, 74, 75).

Using the YSGX and YSGRX libraries, we have obtained Fabs that specifically recognize and co-crystallize the P4-P6 domain of the self-splicing Group I intron from *Tetrahymena* (74) and the synthetic class I ligase (75).

However, neither the YSGX nor YSGRX library was designed specifically for RNA recognition: their success in RNA antibody production has been limited with much potential for improvement. Although successful with the P4-P6 domain, the YSGX library failed to produce RNA-binding Fabs for a variety of RNA targets (75). The YSGRX library does produce RNA-binding Fabs for many RNA targets, however, often their affinity and specificity are relatively low, requiring affinity-maturation before the Fabs can be used effectively in downstream biochemical experiments (75). For instance, the best Fab selected directly from the YSGRX library binds the class I ligase with an affinity of 338 nM. It was affinity-matured to an affinity of 35 nM through error-prone PCR Fab library production followed by additional rounds of selection before it was used to co-crystallize the class I ligase (75).

Here we describe the design, construction, selection, and characterization of a new Fab library designed specifically for RNA recognition. We aimed to create an antibody library with a large randomized surface area and a chemical diversity tailored for RNA-binding. Because libraries with strictly limited chemical diversity are functional (87, 90-92) and they allow the sampling of a larger sequence space than libraries more chemically diverse, we produced an antibody library, YSGR Min library, with strictly limited chemical diversity to Tyr, Ser, Gly, and Arg amino acids spread out on all six CDR loops. Our results showed that although the amino acid composition is extremely limited, the YSGR Min library produced Fabs successfully for all three RNA targets used in selection, and their affinities and specificities are comparable to or slightly better than those from the YSGX and YSGRX library which employs a much richer amino acid composition. This library will serve as a basis for supplementing and testing the effect of additional amino acids for future library improvement.

3.2 Results

3.2.1 Design of the RNA-targeting Fab library with a minimal amino acid composition

Our immediate goal is to search for and test the Fab variable sites that are suitable for RNA recognition. As the YSGX and YSGRX libraries, combined together, have reasonable successes in RNA recognition (74, 75), they naturally become a good starting point for the new library. In both YSGX and YSGRX libraries, only solvent exposed residues in four out of six complementarity determining regions (CDRs) were randomized, including CDR-L3, CDR-H1, CDR-H2, and CDR-H3, while CDR-L1 and CDR-L2 were held constant (74, 75, 87). Similar to that in the YSGX and YSGRX libraries, a total of 15 solvent exposed sites were chosen as variable sites in CDR-L3, CDR-H1, and CDR-H2 for our minimal library. Solvent exposed positions in CDR-H3 were replaced by random-sequence loops of three different lengths. In addition, six structurally important residues (framework residues) in these CDRs were given a small subset of variability as designed in both YSGX and YSGRX libraries (74, 75, 87). Two framework residues, heavy chain Ile34 (located in CDR-H1) and Gly55 (CDR-H2), were completely fixed in the YSGR Min library as they are almost fully conserved in P4-P6 and ligase I binding Fabs (74, 75) and other RNA-binding Fabs (unpublished results).

In the Fab-RNA binding interface, we found CDR-L1 and CDR-L2 play an important role in RNA recognition (74), suggesting that including CDR-L1 and CDR-L2 in the randomization may improve RNA binding efficiency in the library. We examined all the residues in CDR-L1 and L2 defined by the Kabat definition (93) and their flanking regions in the crystal structure of Fab2-P4-P6 with PyMOL software (The PyMOL Molecular Graphics System, Version 0.99rc6, Schrödinger, LLC). Solvent exposed and potential contact residues were identified as positions 28-32 in CDR-L1 and positions 49-50, 52-53, 55-56, and 60-61 in CDR-L2.

Adding all the variable sites from six CDRs gives a total of 37 randomized locations in the library design, which correspond to a theoretical library size of 4.9×10^{55} if we use NNK (N =

A, C, G, or T; K = G or T) to encode for all 20 amino acids types at each variable site. This is drastically larger than the practical size (~ 10^{10-11}) of a typical phage displayed library achievable in the laboratory and the large gap reduces the effectiveness of the constructed library (92). To decrease the size gap and increase the potency of the library, we will adopt the reduced chemical diversity design in which variable sites are randomized with only a small fraction of the natural spectrum of the amino acids. This reduced chemical diversity design has been successfully employed in the Fab libraries for protein targets using either highly frequent amino acids from the Kabat database (88, 94) or in two extreme cases, a tetranomial code (Tyr, Ala, Asp, Ser) (90) and a binomial code (Tyr and Serine) (91).

Inspired by the success of the reduced chemical diversity and minimal diversity in the Fab libraries for protein targets, we have decided to use an extremely simplified amino acid composition for our library. This library design tests the minimal requirements of Fab-RNA binding interfaces. One advantage of this extremely simplified library is that if it is reasonably effective, it can serve as a basis to introduce and test more amino acid composition in future library designs. We chose a tetranomial code (Tyr, Ser, Gly, Arg) in the design of our library, named YSGR Min. Tyr and Ser were chosen as they are the most abundant amino acids in the selected RNA-binding Fabs binding to the P4-P6 domain (74), the class I ligase (75), and other RNA targets (J. Y., unpublished results). Gly was chosen because it is known to be flexible and advantageous in the CDR loops to achieve many efficacious conformations for antigen binding (92); indeed, it is relatively abundant in the Fab-RNA interfaces (74, 75). Both Arg and Lys were abundant in the Fab-RNA interfaces as well as other protein-RNA interfaces (95, 96), however, only Arg was chosen because the longer and more flexible Lys may be less beneficial (bulky) to the antigen binding interfaces.

3.2.2 Construction of the YSGR Min library

The YSGR Min library was constructed from a single Fab framework (P4-P6 binding Fab2) using Kunkel mutagenesis (97). As Kunkel mutagenesis generally retains a significant

amount of the unmutated template gene, to prevent the overpopulation of the parent Fab in the library, we first generated a DNA clone with two TAA stop codons added in all six CDR regions from P4-P6 binding Fab2 using Kunkel mutagenesis. After verifying stop codon insertion in all CDRs via sequencing, we made the single-stranded DNA (ssDNA) of this stop template clone and used it as the template for library construction (Figure 3.1).

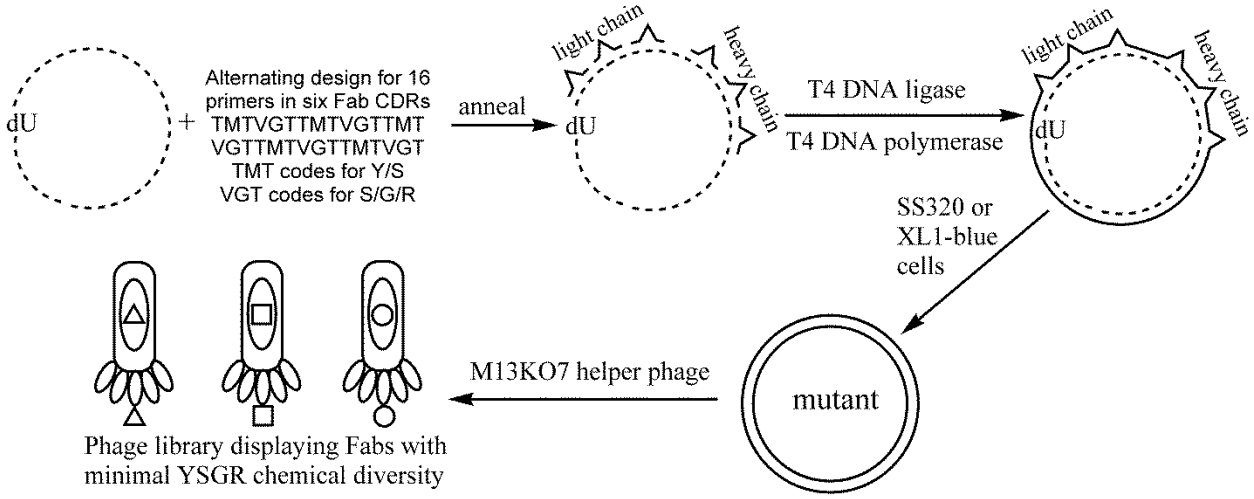


Figure 3.1 Construction of the YSGR Min phage displayed antibody library through oligonucleotide-directed mutagenesis.

Sixteen synthetic oligonucleotide primers (two for each of CDR-L1, L2, L3, H1, H2, and six for CDR-H3) are annealed to the dU-ssDNA template (dashed circle). These oligonucleotides contain mutations that encode the library diversities which are flanked by perfectly complementary sequences. Chemical diversities (Y/S/G/R) are encoded by two alternating degenerate codons (TMT for Y/S and VGT for S/G/R) in pairs of Kunkel primers. Heteroduplex dsDNA is enzymatically synthesized by T7 DNA polymerase and T4 DNA ligase, then transformed into E. coli SS320 or XL1-blue cell hosts where the dU-ssDNA is converted into T-containing DNA that is fully complementary to the mutant sequence (circle with solid line).

The designed diversity (YSGR) was incorporated into the Fab genes through oligonucleotide primers similarly via Kunkel mutagenesis (98). For synthetic ease and

economical reasons, we employed the degenerate codon method to establish sequence diversity in the primer design. However, there is no single degenerate codon encoding all four YSGR residues without introducing detrimental extra codons such as cysteines and stop codons. To exclude these from our design, we utilized two separate degenerate codons, TMT (M = A or C) to encode equal portions of Tyr and Ser, and VGT (V = A, C, or G) to encode equal portions of Ser, Gly, and Arg. These two codons were applied in an alternating fashion in the randomized CDR sequence, and furthermore, oligonucleotide primer pairs were used to apply TMT and VGT codons to both odd and even numbered positions. For example, for positions 28-32 of CDR-L1, primers containing either TMTVGTTMTVGTTMT or VGTTMTVGTTMTVGT were used in the library construction. This alternating approach has a serendipitous advantage in that it eliminates consecutive arginines. This is advantageous because Fabs produced by the YSGRX library frequently contain consecutive Arg in CDR-H3, which often increase non-specific binding to RNAs (75).

For the most flexible and important CDR-H3, we have designed three different lengths for the peptide chain sequence to increase the effectiveness of the library. This increased the total number of primers to an unprecedented 16 in a one-pot Kunkel reaction. Our initial attempt in the Kunkel reaction resulted in low yield in the desired product. Reaction optimizations for annealing procedure, primer-to-template ratio, enzyme concentration, and incubation temperature and duration were carried out to maximize the reaction yield. In the optimized annealing procedure, we mixed all sixteen phosphorylated primers with single-stranded DNA template (total primer concentration at a given CDR location/template concentration = 3) and incubated the mixture at 95 ^{o}C for 2 min, 50 ^{o}C for 3 min and then room temperature for 5 min. After annealing, ATP, dNTP, DTT, T4 DNA ligase, and T7 DNA polymerase were added sequentially and the mixture was incubated at room temperature overnight. After the polymerization and ligation step, the reaction mixture was purified and electroporated into

electrocompetent SS320 cells to obtain the YSGR Min library (99) with an apparent diversity of 3.2 x 10^9. This library was then used in the following selection and characterization (Figure 3.2).

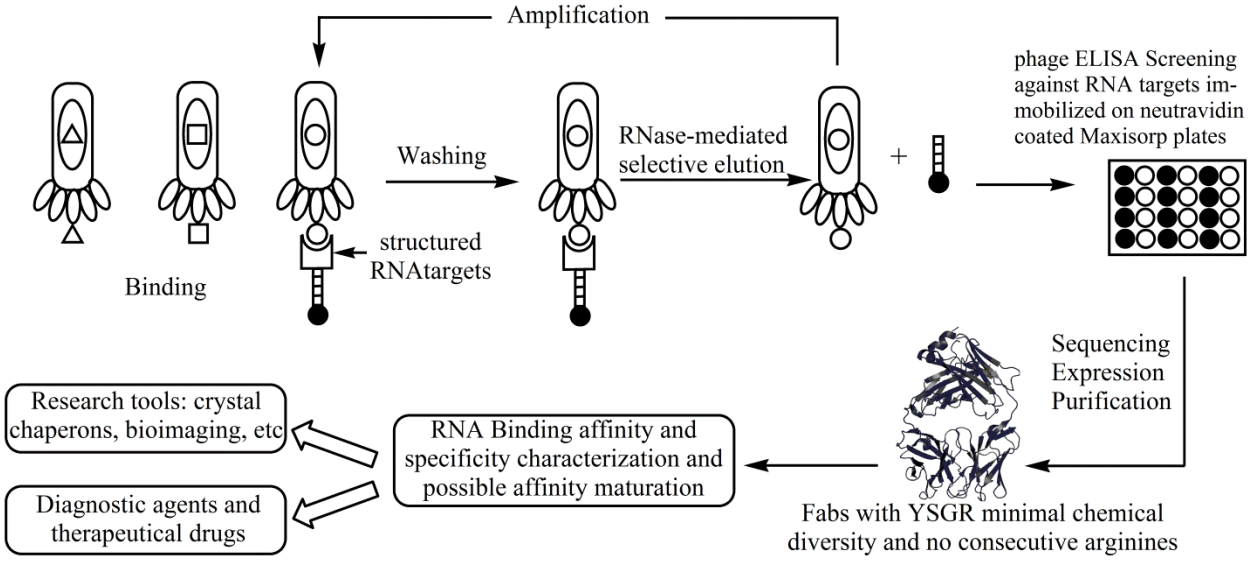


Figure 3.2 RNA-binding Fab selection, expression, and application.

A library of bacteriophage displaying Fabs were allowed to bind biotinylated RNA target complex immobilized on the streptavidin-coded magnetic beads. After washing, specific binders were selectively eluted by degradation of the RNA target with RNase A. Individual clones can be screened by phage Enzyme-Linked Immunosorbent Assay (ELISA) with biotinylated RNA target complex immobilized on neutravidin-coated Maxisorp plates. Positive clones can be sequenced, expressed, and purified to generate soluble Fabs. These specific Fabs are characterized to verify desired binding affinity and specificity and used in basic biochemical research or therapeutic drug development targeting structured RNAs.

3.2.3 Selection of the YSGR Min library against three RNA targets

After the YSGR Min Fab library was constructed, it was screened against three RNA targets: the P4-P6 domain (P4-P6)(74, 100), the *Fusobacterium nucleatum* glycine riboswitch (FNI-II)(13, 51), and the *Vibrio cholerae* glycine riboswitch (VCI-II)(6). The P4-P6 domain is an individual structural domain of the self-splicing *Tetrahymena* Group I intron and folds into a well-

defined three-dimensional structure. FNI-II is a *cis*-acting non-coding RNA that switches its structure upon binding glycine molecules and controls downstream gene expression in *Fusobacterium nucleatum*. FNI-II folds into a compact three-dimensional structure and its crystal structure was determined by the Strobel lab (19). VCI-II is similar to FNI-II in its function; however, it is significantly larger with several extra domains and does not yet have a crystal structure. For P4-P6 and VCI-II, we have selected Fabs from the YSGX or YSGRX libraries (42, 74), which will allow us to compare the affinities and specificities of these Fabs to those obtained from the YSGR Min library.

In order to immobilize the RNA targets for selection (Figure 3.2), we extended the 3' end of all three RNA targets with a 29 nucleotide tag sequence to prepare the corresponding RNA targets for selection: P4-P6Tag, FNI-IITag, and VCI-IITag. These RNA targets were then annealed with 5' biotinylated DNA, allowing immobilization onto streptavidin coated magnetic beads for selection (74). In the first round, the biotinylated DNA annealed RNA targets were immobilized on the streptavidin coated magnetic beads before binding to the YSGR Min library in the presence of RNase inhibitor. After brief washing steps, the phage bead complex was directly used to infect XL1-blue cells for reamplification for future rounds of selection. In the second round, purified phage pools were first incubated with streptavidin beads to remove the streptavidin binders and the supernatant was then used to bind biotinylated DNA annealed RNA targets (50 nM) in the presence of 50 µg/ml *E. coli* tRNA mixture and RNase inhibitor before being mixed with and captured on fresh aliquot of streptavidin beads. After extensive washing steps, biotinylated RNase A was used to selectively elute the RNA binding phage, which was re-amplified and purified for future rounds. In the third round and later rounds, selection was carried out similar to the second round except that RNA target concentration was at 5 nM and the *E. coli* tRNA competitor concentration was 500 µg/ml, which was used as a stringent condition to increase the affinity and specificity of the selected Fabs. In the second and later rounds, selection was carried out in parallel with a control experiment in which no RNA target

was present. The number of phage in the input, output, and control experiments were titered on the LB ampicillin plates. The enrichment number was calculated as output phage numbers in the target selection divided by that in the non-target control experiment, and the reduction ratio was calculated as the ratio of input phage number divided by the output phage number in the control experiment. These two selection parameters were used to monitor the selection process and summarized in Table 3.1. As the selection proceeds, the enrichment gradually increases, reaching about two orders of magnitude when the selection was stopped. The background reduction ratio remained relatively constant, indicating credible enrichment numbers. No more than five rounds were needed for each RNA target, showing the effectiveness of the YSGR Min library.

Table 3.1 Phage display selection results of YSGR Min library against RNA targets.

RNA targets[a]	Round #	Enrichment	Reduction Rate
P4-P6	2	11.87	7.01×10^5
	3	269.67	4.66×10^6
FNI-II	2	.77	2.6×10^7
	3	8.12	3.54×10^7
	4	34.42	9.06×10^4
	5	75.6	2.62×10^4
VCI-II	2	1.61	8.89×10^5
	3	6.84	3.11×10^7
	4	176.97	9.27×10^4

[a] P4-P6: ΔC209 P4-P6 domain from *Tetrahymena* Group I intron. FNI-II: the glycine riboswitch from *Fusobacterium nucleatum*. VCI-II: the glycine riboswitch from *Vibrio cholerae*.

*3.2.4 Screening individual phage clones by Enzyme-Linked Immunosorbent Assay (*ELISA*)*

After successful phage selection, individual phage clones were screened against RNA targets by Enzyme-Linked Immunosorbent Assay *(*ELISA, Figure 3.2). We adapted the phage ELISA procedure published by Sidhu *et al.* to be suitable for RNA binding phage displayed Fab screening (98). The phage ELISA screening was performed in 96 well Maxisorp immunoplates. To ensure the RNA targets are intact and correctly folded, instead of directly immobilizing the targets on the plate as for protein targets, we immobilized the neutravidin first on the Maxisorp

plate and then bound the biotinylated DNA annealed RNA targets to the neutravidin on the plate. Phage solutions grown from individual clones were then added to the RNA bound wells as well as neutravidin coated control wells in the presence of vanadyl ribonucleotide complex and sheared salmon DNA. Here vanadyl ribonucleotide complex is used as an inexpensive substitute for RNase inhibitor and the sheared salmon DNA is used as a nucleic acid competitor. After washing steps, the plate bound phage were then recognized by anti-M13 antibody/horseradish peroxidase conjugate, detected by 3,3',5,5'-tetramethylbenzidine (TMB) solution, quenched by phosphoric acid to give yellow color, and quantified spectrophotometrically by a microtiter plate reader. A high ratio between the target well and the control well indicates positive, specific binding of the RNA targets. Ten positive clones for each of the three targets were sent for sequencing. The sequences of the unique clones are shown in Table 3.2. For target P4-P6, besides obtaining a sequence as designed in the library (Fab8), we also obtained the parent Fab2 and a partially mutated Fab11. We attribute this to a trace amount of contamination of Fab2 in the stop template clone which was not detected by sequencing. For FNI-II and VCI-II targets, we obtained multiple unique sequences which were all as designed in the library.

Table 3.2 Sequences and affinities of the selected Fabs[a].

RNA Targets	Fab Clones[b]	CDR-L1	CDR-L2	CDR-L3	CDR-H1	CDR-H2	CDR-H3	K_d (nM)
P4-P6	2 (6)[c]	SVSSA	YSASSLYSGVPSR	SYSSPI	NLYSSSI	YISSSYGYTY	RAAGMSTY-GF	66±13
	8 (1)	YGYYRS	YSASGLYRGVPSR	GYRSPV	NLGSGYI	SYRPSSGSTR	SYSSRYSY-AM	54±2
	11 (3)	SVSSA	YSASSLYSGVPSR	SYSSPI	NLYSSSI	SRSPRSGGTS	RAAGMSTY-GF	22±3
FNI-II	10 (2)	GYSYS	YGAYRLSRGVPYG	YSSGPV	NISYGSI	SSYSRYGRTY	SRSRSSY--AM	≥1.5μM
	12 (1)	SRYRY	RSASSLSSGVPSY	YSYRLI	NIGSRSI	SGSSRYGGTY	SGSGSSSRYGM	nd[d]
	14 (1)	YGSGY	SGASGLSGGVPSG	SYRYPF	NVSGSRI	SGSSGYGGTS	SRSRYSYRYGL	≥1.5μM
	15 (2)	GYSYS	YGAYRLYGGVPSR	SSRSLV	NISSYSI	SRSPRSGRTS	YRYRSGS--GL	600±180
	16 (2)	YSSRY	YGAYGLYRGVPYR	SYSYPI	NLGYRYI	SGSSGSGSTY	GSSYSYRY-AF	160±12
	18 (1)	GYGSR	YSASRLSRGVPSS	YSYSLL	NFSSGYI	YGSPSSGRTS	SSYRSRYG-AM	≥2.0μM
VCI-II	24 (1)	YRYRS	SGAYGLYRGVPSR	YRSGLV	NVSSSSI	GSRSSSGSTG	SSYGSRSS-AM	≥1.5μM
	25 (2)	YRYRS	SGAYSLYRGVPYR	YSSGLI	NFSSGSI	GSGPSRGSTS	SRSRSRSR-AM	333±110
	27 (2)	SGSRS	YGASSLYGGVPSR	YSYGLL	NFRSRYI	GSGPSRGYTG	GSRYRYGSRGM	559±208
	29 (2)	YSYRS	YRASRLYGGVPSR	RSSYPV	NFSGSSI	YGSPGYGRTS	SSYGSRSGYAM	39±2
	30 (1)	SGSSY	YRASSLSSGVPSR	YSSRLL	NIYRYGI	YGSPGSGRTY	RYRYRYRS-GL	22±1
	31 (1)	RYRSG	SGAYGLYSGVPSR	YGSGLI	NFRYRSI	SRSSRSGRTS	SRSGYSSG-AL	408±139

[a] Tyr, Ser, Gly, and Arg residues are shaded in yellow, red, green, and pink, respectively. Unmutated residues are shaded in grey. Framework residues with limited randomization are unshaded. Dashes indicate gaps in the alignment.
[b] Numbers in parentheses indicate the frequency of the Fab sequences from the 10 sequenced clones for each target.
[c] This Fab is identical to the parent Fab.
[d] Binding is not detected.

3.2.5 Expression and characterization of the Fabs

Phage ELISA identifies positive Fab clones binding to stationary RNA targets, which may be complicated by Fab-phage fusion or the expression level of Fabs, making it difficult to obtain accurate measurements of the binding affinities and specificities. To overcome these difficulties, unique clones identified by phage ELISA and sequencing were expressed as soluble Fabs for further biochemical characterization (42). Stop codon (TAG) was introduced to the phagemid through Kunkel mutagenesis to break the fusion between Fab gene and protein III, a coat protein of the M13 phage. The clones carrying the TAG mutation were identified by colony PCR. These resulting expression clones were transformed into *E. coli* 34B8 strain for expression followed by purification on protein A affinity column with published protocols (42).

We first used the Electrophoretic Mobility Shift Assay (EMSA) on native gel to analyze the binding between P4-P6 and its selected Fabs (Figure 3.3A). Results showed that both Fab8

and Fab11 form a discrete and slower-migrating complex with P4-P6 similar to the parent Fab2, suggesting that they may bind P4-P6 with high affinities and specificities. As glycine is needed to keep VCI-II and FNI-II in their native folding states, to analyze the binding between the glycine riboswitch binding Fabs, we used an analogous modified native gel assay, termed Polyacrylamide Coelectrophoresis (PACE) in which a saturating amount of glycine was added in the polyacrylamide gel to maintain glycine riboswitch structure. Results showed that Fab16 forms a discrete complex with FNI-II glycine riboswitch (Figure 3.3B) and Fab29 and Fab30 form a discrete complex with VCI-II glycine riboswitch (Figure 3.3C), indicating these Fabs may bind their targets with high affinities and specificities. However, multiple Fabs selected against VCI-II and FNI-II glycine riboswitches do not form a discrete complex under gel electrophoresis, indicating they are either weak binders or non-specific binders.

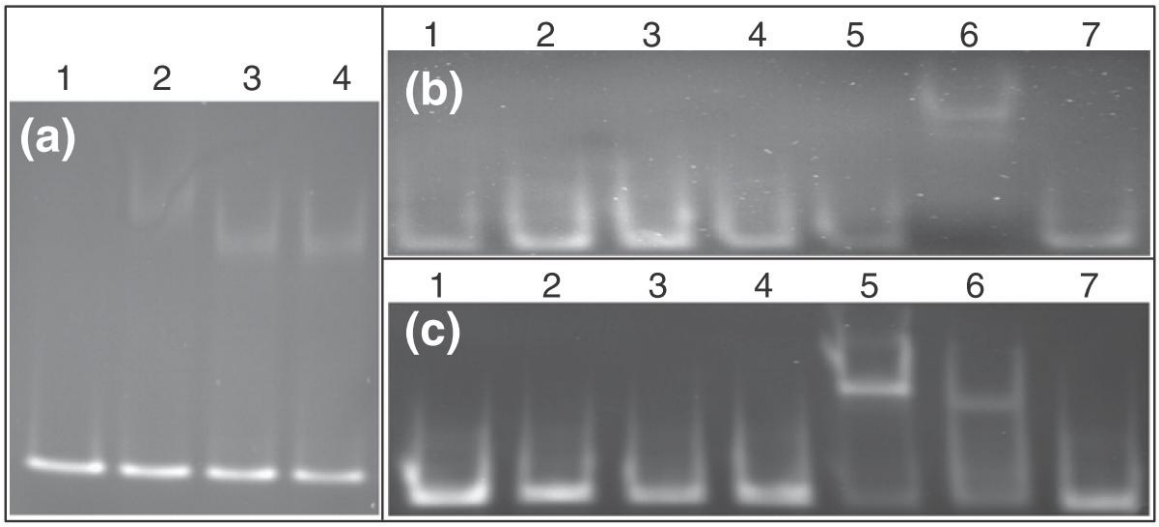


Figure 3.3 Gel images of native electrophoretic mobility shift assay (EMSA) and polyacrylamide coelectrophoresis.

A) Fabs binding to P4-P6. Lane 1, P4-P6 only; Lanes 2-4, 2 µM of P4-P6 with 1.6 µM of Fab8, Fab2, and Fab11, respectively. B) Fabs binding to FNI-II. Lane 1, FNI-II only; Lanes 2-7, 2 µM of FNI-II with 2 µM of Fab10, Fab12, Fab14, Fab15, Fab16, and Fab18, respectively. C) Fabs binding to VCI-II. Lane 1, VCI-II only; Lanes 2-7, 2 µM of VCI-II with 2 µM of Fab24, Fab25, Fab27, Fab29, Fab30, and Fab31, respectively.

We next employed filter binding assays to determine dissociation constants (K_ds) for binding between Fabs and their ^{32}P-labelled RNA targets. Results showed that both Fab8 and Fab11 bind P4-P6 with high affinities, 54 nM and 20 nM (Figure 3.4A, Table 3.2), respectively, comparable to or slightly better than the parent Fab2 with an affinity of 66 nM. Out of six Fabs selected for FNI-II glycine riboswitch, one Fab (Fab12) did not show detectable binding; three Fabs (Fab10, Fab14 and Fab18) showed binding affinities larger than 1 µM; one Fab (Fab15) showed binding affinity in the high nanomolar range. These results are consistent with the native gel results in which they do not form a discrete and slower-migrating complex with FNI-II. In contrast, Fab16 binds FNI-II with midnanomolar affinity (160 nM, Figure 3.4B, Table 3.2), consistent with complex formation in the native gel assay. For VCI-II glycine riboswitch target, one Fab (Fab24) showed a binding affinity lower than 1 µM and three Fabs (Fab25, Fab27 and Fab31) showed binding affinities in the high nanomolar range, consistent with the lack of complex formation in the native gel assay. However, as the native gel results predicted, Fab29 and Fab30 bind VCI-II glycine riboswitch with high affinities, 39 nM and 22 nM, respectively (Figure 3.4C, Table 3.2).

Combining the results of the native gel shift assay and the filter binding assay, the YSGR Min library has successfully produced at least one Fab for each of the three RNA targets that shows tight binding affinity and forms a specific complex with its target.

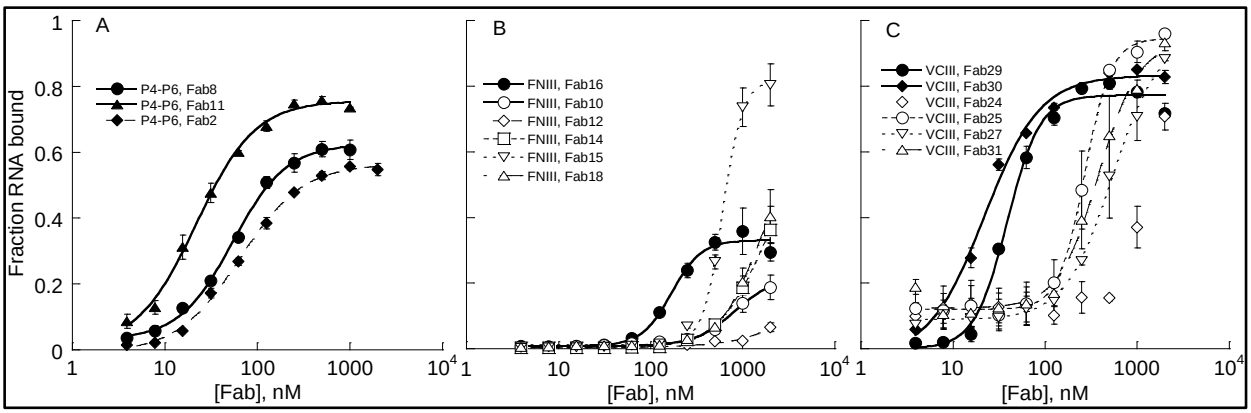


Figure 3.4 Binding curves for the selected Fabs with their cognate RNA targets, P4-P6 (A), FNI-II (B), and VCI-II (C).

Fabs that form discrete bands with their cognate RNAs on native gel are represented by solid symbols; others use open symbols. These fraction binding data are not normalized for direct comparison to the raw binding data in the specific binding assay reported in Table 3.3.

3.2.6 Specificity Assay

An antibody library designed for RNA targets could be potentially very useful for RNA crystallization, oncogenic RNA recognition and neutralization, and other biochemical applications. However, the nature of the molecular recognition requires the Fab obtained from the library to be not only a tight binding one, but also highly specific as well. Our native gel shift assay and filter binding assay showed that we have obtained tight-binding Fabs for all three RNA targets from the YSGR Min library. However, the specificity of these Fabs is not rigorously assessed. Currently there is no quantitative way for measuring the specificity of an RNA-binding Fab. We have now designed a competitive filter binding assay for this purpose.

In the competitive filter binding assay, a reference point is first determined which measures the percent binding of the ^{32}P labeled cognate RNA to Fab at Fab concentration of $2K_d$. Here Fab concentration of $2K_d$ is chosen instead of K_d to ensure a detectable percent binding value as Fabs with low specificity or low affinity often lead to unexpectedly low binding values possibly due to low accuracy in the determination of the binding affinity. By definition, this

reference point should be around 67% of the maximum fraction of binding (unnormalized), empirically high enough to give a consistent signal. Next, similar percent binding data was obtained with the competition of each of the seven non-radiolabeled non-cognate RNAs present at a concentration equal to $2K_d$. The percent binding values obtained with seven competitors were then averaged, divided by the reference point, and averaged over two independent experiments to give Fab Specificity Index (FSI) for a given Fab target pair (Table 3.3). This FSI value ranges from 0 to 1, with 0 indicating completely non-specific and 1 indicating extremely specific within the choice of competitive RNAs.

Table 3.3 Specificity binding raw data and Fab specificity index values.

Targets[a]	Fab Clones	Competitive RNAs[a]								FSI[b]
		P4-P6	FNI-II	VCI-II	ExD1	AMTHF	DASAH	tRNA	CBATHF	
P4-P6	2	0.386[c]	0.363	0.375	0.310	0.341	0.320	0.325	0.321	0.888
		0.364	0.333	0.323	0.328	0.343	0.335	0.315	0.324	
	8	0.479	0.466	0.474	0.441	0.452	0.366	0.388	0.403	0.926
		0.449	0.453	0.463	0.412	0.442	0.424	0.447	0.382	
	11	0.501	0.396	0.456	0.407	0.420	0.399	0.369	0.387	0.821
		0.476	0.407	0.409	0.430	0.434	0.411	0.378	0.315	
FNI-II	15	0.352	0.570	0.249	0.0966	0.223	0.208	0.189	0.336	0.371
		0.397	0.704	0.260	0.0873	0.183	0.163	0.179	0.344	
	16	0.671	0.539	0.780	0.514	0.563	0.513	0.484	0.560	1.045
		0.654	0.549	0.679	0.529	0.605	0.505	0.434	0.462	
VCI-II	25	0.123	0.0586	0.355	0.0169	0.0620	0.0563	0.0937	0.0844	0.255
		0.203	0.134	0.423	0.0916	0.154	0.0977	0.155	0.0861	
	27	0.168	0.0922	0.210	0.0419	0.0936	0.0899	0.119	0.139	0.424
		0.163	0.101	0.312	0.0572	0.113	0.0789	0.114	0.118	
	29	0.348	0.305	0.338	0.343	0.336	0.297	0.282	0.266	0.972
		0.402	0.409	0.362	0.323	0.405	0.327	0.354	0.374	
	30	0.297	0.232	0.221	0.231	0.0860	0.229	0.216	0.212	0.933
		0.234	0.265	0.248	0.208	0.192	0.212	0.232	0.209	
	31	0.259	0.192	0.290	0.135	0.187	0.159	0.199	0.254	0.683
		0.376	0.277	0.396	0.214	0.269	0.248	0.274	0.245	
	18	0.418	0.293	0.442	0.228	0.261	0.253	0.249	0.265	0.683
		0.469	0.315	0.453	0.289	0.416	0.257	0.294	0.278	
	20	0.0583	0.0987	0.138	0.0375	0.101	0.0718	0.154	0.0733	0.622
		0.0541	0.112	0.132	0.0407	0.0807	0.0823	0.143	0.0674	

[a] P4-P6, FNI-II, and VCI-II are the same as described in Table 3.1. ExD1: exD1 domain from ai5γ Group II intron. AMTHF: *Alkaliphilus metalliredigens* tetrahydrofolate riboswitch. DASAH: *Dechloromonas aromatica* SAH riboswitch. tRNA: *E coli* tRNA mixture. CBATHF: the tetrahydrofolate riboswitch from *Clostridium botulinum A*.

[b] FSI, Fab Specificity Index. For a given pair of Fab and RNA target, FSI value = (sum of unshaded raw data/7)/(shaded reference data) for each replicate. The two replicates are then averaged to produce the given value.

[c] Shaded data: reference point, no competitive RNA was added.

Table 3.3 shows the results of the competitive filter binding assay of the selected Fabs. The reference point of the Fabs for P4-P6 is very close to 67% of the maximum fraction of binding as expected (Table 3.3, Figure 3.4A). For instance, the maximum fraction of Fab8 binding to P4-P6 is around 0.61 (Figure 3.4A), the reference point of ~0.45 corresponds to 74% of the maximum fraction binding, close to 67% expected. With the addition of competitive RNAs, the percent binding values were only reduced marginally for all seven competitor RNAs. The FSI values of both Fab8 and Fab11 are close to one, comparable to the value of Fab2, a Fab that was co-crystallized with P4-P6 previously, confirming their high binding specificities.

Two FNI-II binding Fabs (Fab15 and Fab16) have affinities better than 1 μM and their binding specificity values were determined. With the addition of competitive RNAs, the changes in the percent binding values are very different for these two Fabs (Figure 3.4B). For Fab16, the percent binding values barely changed compared to its reference point while the values for Fab15 were reduced significantly. The FSI values for Fab16 and Fab15 are 1.05, and 0.37, respectively, indicating that Fab16 is highly specific while Fab15 is not, consistent with our native gel shift assays.

Five of the VCI-II glycine riboswitch binding Fabs with affinities better than 1 μM were analyzed by the competitive filter binding assays. Results showed that Fab25 and Fab27 have low specificities, with FSI values of 0.26 and 0.42, respectively. Fab31 has a medium FSI value of 0.68, indicating Fab31 as a reasonably specific Fab. Perhaps the lack of a specific complex formation of Fab31 with VCI-II in the native gel experiment is due to its low binding affinity. On the other hand, both Fab29 and Fab30 are highly specific with FSI values close to one. We have also measured the FSI values of Fab18 and Fab20, which were selected against VCI-II from the YSGRX library (42). Results showed that Fab18 and Fab20 have medium specificities with FSI values of 0.68 and 0.62, respectively. These values are significantly lower than that of Fab29 and Fab30, indicating that the YSGR Min library can produce exceptionally specific RNA-binding Fabs, even better than the YSGRX library. As both Fab18 and Fab20 have consecutive

arginines in CDR-H3 while Fab29 and Fab30 do not (42), these higher specificity values of Fab29 and Fab30 may indicate the success of the alternating codon design in avoiding consecutive arginine stretches in the Fab CDRs of the YSGR Min library.

3.3 Discussion

In this work, we have constructed the YSGR Min library with minimal chemical diversity in an antibody Fab framework. Compared to previous Fab libraries (YSGX and YSGRX libraries) that have been tentatively used for RNA targets, the YSGR Min library is uniquely different in the following aspects: 1) the YSGR Min library expands the variable sites to include all six CDRs versus only four CDRs in the other two libraries; 2) the chemical diversity of the YSGR Min library is restricted to include only Tyr, Ser, Gly and Arg; 3) the minimal YSGR chemical diversity is applied to all six CDRs while the YSGX and YSGRX libraries use binary code (Tyr and Ser only) for CDR-L3, H1 and H2, and 19 amino acid types (no Cys) for CDR-H3; 4) the alternating codon design eliminates the consecutive arginines in the CDRs while the YSGRX library frequently produced stretches of consecutive arginines in CDR-H3. The YSGR Min library was selected against three RNA targets and has yielded high-affinity and high-specificity Fabs for each of them. This suggests that the minimal chemical diversity is sufficient to make tight and specific contact with RNA targets, different from a much broader amino acid spectrum used in common RNA binding protein interfaces (95, 96, 101, 102). These findings may provide a unique approach to designing future RNA-binding proteins. Using a competitive filter binding assay, we have defined the specificity of the high-affinity RNA binding Fabs and found that these Fabs have higher specificity compared with the Fabs selected from the YSGRX library. This correlates well with the lack of consecutive arginine stretches in the CDR and suggests the superiority of the alternating codon design over the uniform codon design. The detrimental effect of a large amount of arginines at the Fab-antigen interface is also consistent with the findings of Birtalan et al. (103), in which the authors reported that arginine content clearly correlates with the increased non-specific binding between synthetic Fabs and protein

antigens. High specificity is very important to achieve high homogeneity in complex formation with the high sample concentrations used in crystal screening. This may partially explain the difficulty in obtaining crystal formation using Fab18 and Fab20 in co-crystallizing VC glycine riboswitch (J. Y., unpublished results). High specificity is also very important for future diagnostic tool or drug design targeting oncogenic RNAs, where off-target effects generally are detrimental. We have noticed that besides the high-affinity and high-specificity Fabs obtained for each target, we have also obtained weak binding and low specificity Fabs for two of the targets. We suspect that this is due to oversimplified codon design. With the current YSGR minimal codon design as a working basis, additional amino acids can be supplemented and tested for their effect on RNA binding affinity and specificity.

3.4 Materials and Methods

3.4.1 DNA constructs of the RNAs

The *Fusobacterium nucleatum* glycine riboswitch (FNI-II), *Vibrio cholerae* glycine riboswitch (VCI-II), *Alkaliphilus metalliredigens* tetrahydrofolate riboswitch (AMTHF), *Clostridium botulinum A* tetrahydrofolate riboswitch (CBATHF) and *Dechloromonas aromatica* SAH riboswitch (DASAH) genes (104) were generated by recursive PCR (55, 105) from DNA oligonucleotides (Integrated DNA Technologies) with EcoRI restriction site engineered at the 5' and EarI and HindIII sites at the 3' (6, 13). For RNAs used as Fab selection targets, an extra 29 nucleotide tag (GCTTTAGGTCGACTCTAGAGGATCCCCGG) was added to the 3' of the gene during recursive PCR. RNA genes and pUC19 vector were double digested with both EcoRI and HindIII restriction enzymes and ligated together with T4 DNA ligase to generate plasmids pVCI-II and pFNI-II according to standard protocols (42). These plasmids and plasmids containing *Tetrahymena* ΔC209 P4-P6 group I intron domain(100) and ai5γ exD1 group II intron domain (106) were transformed in JM109 cells, amplified, and linearized by EarI restriction enzyme to generate transcription templates.

3.4.2 In vitro transcription for the preparation of RNAs

RNAs were transcribed in 40 mM Tris-HCl (pH 7.9), 2 mM spermidine-3HCl, 10 mM DTT, 25 mM $MgCl_2$, and 5 mM each of the nucleotide triphosphates for 4 h at 37°C. All the RNAs were purified by 6% polyacrylamide gel electrophoresis, isolated with UV shadowing and excised, electroeluted into TE buffer (10 mM Tris-HCl, pH 8, 1 mM EDTA), PCA extracted, ethanol precipitated, and resuspended in TE buffer. RNA concentrations were determined by UV absorbance at 260 nm wavelength. E coli tRNA mixture purchased from Sigma-Aldrich was used directly as a competitor in the competitive filter binding assay without purification.

3.4.3 Preparation of the stop template for library construction

The phagemid of P4-P6 Fab2, designed to display bivalent, humanized Fab on the surface of M13 bacteriophage has been described previously (42). Electroporation of this phagemid into CJ236 cells (uracil deglycosidase deficient) generated uracil-containing ssDNA by following the procedure as described (42). The following Kunkel primers were designed to insert two stop codons (underlined) into each of the six CDRs through Kunkel mutagenesis: (97)

ACCTGCCGTGCCAGTCAGTAAGGCCAAGACGGCCTATAAGTAGCCTGGTATCAACAG

(CDR-L1),

GCTCCGAAGCTTCTGATTTAAGGCCAAGACGGCCTATAATTCTCTGGTAGCCGTTCC

(CDR-L2),

ACTTATTACTGTCAGCAATAAGGCCAAGACGGCCTATAAACGTTCGGACAGGGTACC

(CDR-L3),

GCAGCTTCTGGCTTCAACTAAGGCCAAGACGGCCTATAACACTGGGTGCGTCAGGCC

(CDR-H1),

GGCCTGGAATGGGTTGCATAAGGCCAAGACGGCCTATAATATGCCGATAGCGTCAAG

(CDR-H2),

GTCTATTATTGTGCTCGCTAAGGCCAAGACGGCCTATAAGACTACTGGGGTCAAGGA

(CDR-H3). After phosphorylation using T4 Polynucleotide kinase, these primers were annealed with uracil-containing ssDNA of P4-P6 Fab2, extended with T7 DNA polymerase, ligated with T4 DNA ligase, and transformed into XL1-blue cells (uracil glycosidase containing strain) to generate a phagemid encoding Fab stop template (pFyes1) following the procedure as described (42). Small scale DNA was purified using Qiagen's miniprep kit and sent for sequencing to confirm the sequence (Eton Bioscience Inc.).

3.4.4 Construction of the YSGR Min library

Single-stranded DNA encoding Fab stop template (pFyes1) was prepared similar to the procedure described above. The following 16 Kunkel primers were designed to incorporate designed diversity (underlined, where M = A or C; V = A, C, or G) into all six CDRs through large scale Kunkel mutagenesis: (97, 98)

ACCTGCCGTGCCAGTCAGTMTVGTTMTVGTTMTGTAGCCTGGTATCAACAG (CDR-L1),

ACCTGCCGTGCCAGTCAGVGTTMTVGTTMTVGTGTAGCCTGGTATCAACAG (CDR-L1),

GCTCCGAAGCTTCTGATTVGTTMTGCAVGTTMTCTCVGTTMTGGAGTCCCTVGTTMTTTCT CTGGTAGCCGTTCC (CDR-L2),

GCTCCGAAGCTTCTGATTTMTVGTGCATMTVGTCTCTMTVGTGGAGTCCCTTMTVGTTTCT CTGGTAGCCGTTCC (CDR-L2),

ACTTATTACTGTCAGCAATMTVGTTMTVGT*CYGNTT*ACGTTCGGACAGGGTACC (CDR-L3),

ACTTATTACTGTCAGCAAVGTTMTVGTTMT*CYGNTT*ACGTTCGGACAGGGTACC (CDR-L3),

GCAGCTTCTGGCTTCAAC*NTT*TMTVGTTMTVGTATTCACTGGGTGCGTCAGGCC (CDR-H1),

GCAGCTTCTGGCTTCAAC*NTT*VGTTMTVGTTMTATTCACTGGGTGCGTCAGGCC (CDR-H1),

GGCCTGGAATGGGTTGCAVGTTMTVGT*YCT*TMTVGTGGCTMTACTVGTTATGCCGATAGC GTCAAG (CDR-H2),

GGCCTGGAATGGGTTGCATMTVGTTMT*YCT*VGTTMTGGCVGTACTTMTTATGCCGATAGC GTCAAG (CDR-H2),

GTCTATTATTGTGCTCGCVGTTMTVGTTMTVGTTMTVGT*GSTWTK*GACTACTGGGGTCAAG

GA (CDR-H3),

GTCTATTATTGTGCTCGCTMTVGTTMTVGTTMTVGTTMT*GSTWTK*GACTACTGGGGTCAAG

GA (CDR-H3),

GTCTATTATTGTGCTCGCVGTTMTVGTTMTVGTTMTVGTTMT*GSTWTK*GACTACTGGGGTC

AAGGA (CDR-H3),

GTCTATTATTGTGCTCGCTMTVGTTMTVGTTMTVGTTMTVGT*GSTWTK*GACTACTGGGGTC

AAGGA (CDR-H3),

GTCTATTATTGTGCTCGCVGTTMTVGTTMTVGTTMTVGTTMTVGT*GSTWTK*GACTACTGGG

GTCAAGGA (CDR-H3),

GTCTATTATTGTGCTCGCTMTVGTTMTVGTTMTVGTTMTVGTTMT*GSTWTK*GACTACTGGG

GTCAAGGA (CDR-H3). In these sequences, the designed codons for variable sites are underlined (M = A or C; V = A, C, or G). Codons of the structural residues important for CDR chain conformation integrity, shown in italics, are randomized to encode the naturally abundant amino acid types (Y = C or T; N = A, T, C, or G; S = C or G; W = A or T; K = G or T), similar to that described by Fellouse et al.(87). These primers were phosphorylated individually using T4 polynucleotide kinase (NEB) as described (42). In the optimized Kunkel mutagenesis, the reaction mixture contains 0.33 µg of each of the phosphorylated Kunkel primers for the light chain CDRs and CDR-H1 and CDR-H2, 0.11 µg of each of the six Kunkel primers for CDR-H3, 20 µg of the ssDNA of the Fab stop template (pFyes1), 50 mM Tris-HCl (pH 7.5), and 10 mM $MgCl_2$ in a total 250 µl volume. This mixture was incubated at 95 ^{o}C for 2 min, 50 ^{o}C for 3 min and then room temperature for 5 min. After annealing, 1 µl of 100 mM ATP, 10 µl of 25 mM dNTPs, 15 µl 100 mM DTT, 2400 units of T4 DNA ligase, and 30 units of T7 DNA polymerase were added sequentially and the mixture was incubated at room temperature overnight. After the polymerization and ligation step, the reaction mixture were purified on a QIAquick column (Qiagen), eluted with 35 µl ultrapure H_2O, electroporated into electrocompetent SS320 cells,

and co-transfected with M13KO7 helper phage to obtain the YSGR Min library according to the published procedure (98).

3.4.5 Phage display selection for RNAs

For selection, the 3' tagged RNAs were first incubated with 1.1 eq of 5'-biotinylated DNA oligonucleotide, 5'-ACCGGGGATCCTCTAGAGTC-3', in the presence of 10 mM Tris-HCl (pH 7.5) and 200 mM NaCl at 90 ^{o}C for 1 min and room temperature for 10 min to form the annealed RNA target complex. In the first round, 0.5 nmol of the RNA target complex was immobilized on streptavidin-coated magnetic beads (Promega), blocked with biotin, and incubated with ~$10^{12\text{-}13}$ cfu (colony forming unit) of phage for 15 min in 1 ml Binding Buffer (PBS [8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.2, 137 mM NaCl, 3 mM KCl], 0.05% Tween 20, 2.5 mM EDTA, 12.5 mM $MgCl_2$; for glycine riboswitches, 1 mM glycine was added to Binding Buffer) supplemented with 0.1 mg/ml BSA, 0.1 mg/ml streptavidin, and 1 unit/µl RNase inhibitor (Promega). The beads were then separated from the solution on a magnet, washed twice with Binding Buffer, and directly used to transduce the XL1-blue cells to amplify the phage (74). In the subsequent rounds, purified phage were first incubated with streptavidin beads in Binding Buffer for 30 min to remove the bead binders and the supernatant was used in the subsequent selection on a KingFisher magnetic particle processor (Thermo Electron Corporation). Phage (~$10^{10\text{-}11}$ cfu) were incubated for 15 min with 50 nM of the RNA target complex in 100 µl Binding Buffer, supplemented with 0.1 mg/ml BSA, 1 unit/µl RNase inhibitor, and 50 µg/ml *E. coli* tRNA mixture. Streptavidin magnetic beads were then added to the solution for 15 min to allow the capture of the RNA target complex together with the bound phage. The beads were then blocked with 50 µM biotin, washed five times with Binding Buffer, and eluted in 50 µl elution buffer (PBS, 5% glycerol, 1 µg/ml biotinylated RNase A). The biotinylated RNase A was removed from the eluted phage by incubation with streptavidin beads. After each round of selection, recovered phage were amplified as described previously (74, 98).

3.4.6 Phage Enzyme-Linked Immunosorbent Assay (ELISA) for RNAs

Phage ELISA for RNA-binding Fabs was adapted from the procedure described by Sidhu et al.(98). Forty-eight individual colonies were picked from a fresh LB/Amp plate (LB agar and 100 µg/ml ampicillin), inoculated in 400 µl of 2YT medium containing 100 µg/ml ampicillin and 10^{10} pfu/ml M13KO7 helper phage in a 96-well deep well plate, and grown at 37 ^{o}C overnight with shaking at 300 rpm. The deep well plate was then centrifuged for 15 min at 3500 rpm to pellet the cells. The supernatant was diluted three fold to prepare a phage solution in Binding Buffer supplemented with 0.5% (w/v) BSA, 100 µg/ml sheared salmon sperm DNA, 0.67 mg/ml heparin, and 1 mM vanadyl ribonucleoside complexes. A 96-well Maxisorp plate was coated with 100 µl of 2 µg/ml neutravidin in 100 mM sodium bicarbonate coating buffer (pH 9.6) overnight at 4 ^{o}C. The coating solution was removed and the Maxisorp plate was blocked for 1 h with 200 µl/well of 1% (w/v) BSA in PBS. After the blocking solution was removed, the Maxisorp plate was washed with PBS with 0.05% (v/v) Tween 20, then incubated with 100 µl/well 25 nM RNA target complex in Binding Buffer for 30 min at room temperature. For each well containing RNA target complex, a control well was prepared by incubation in Binding Buffer. The Maxisorp plate was then washed with Binding Buffer and incubated with 100 µl/well phage solution at room temperature for 30 min. After washing with Binding Buffer, the Maxisorp plate was incubated with 100 µl/well anti-M13/HRP conjugate (diluted 5000x in Binding Buffer supplemented with 0.5% (w/v) BSA and 1 mM vanadyl ribonucleoside complexes) at room temperature for 30 min. After another washing step with Binding Buffer, the Maxisorp plate was incubated with 100 µl/well Ultra TMB-ELISA Substrates (Pierce) for 5-10 min, quenched with 100 µl/well 1 M phosphoric acid, and read spectrophotometrically at 450 nm in a microplate reader.

3.4.7 Fab expression and purification

A starter culture was inoculated with Fab plasmid containing *E. coli* 34B8 strain cells and was grown at 30 ^{o}C and 300 rpm overnight in 25 ml 2YT with 100 µg/ml ampicillin. 5ml of the

overnight starter culture was used to subculture 500 ml CRAP media (27 mM $(NH_4)_2SO_4$, 2.4 mM sodium citrate-$2H_2O$, 14 mM KCl, 5.4 g/liter yeast extract, 5.4 g/liter HyCase SF Casein, 0.11 M MOPS–NaOH buffer [pH 7.3], 0.55% [w/v] glucose and 7 mM $MgSO_4$) supplemented with 100 µg/ml of ampicillin and grown at 30 ^oC and 300 rpm for 24 h. Cells were spun down at 6000~8000 rpm for 10 min at 4 ^oC and the pellets were frozen at -20 ^oC overnight.

The frozen cell pellets were thawed and lysed with lysozyme (0.3 mg/ml) and Dnase I (2.5mg/30ml) in 30 ml lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 0.5 mM PMSF) with vigorous stirring for 1 h at room temperature. Cell debris was centrifuged down and the supernatant was loaded onto a protein A column. The column was washed with 50 column volumes (CV) of 10 mM Tris pH 7.5, 500 mM NaCl and eluted with 0.1 M citric acid, pH 3. The eluate was collected in 3 ml fractions in tubes prefilled with 1.5 ml 1 M Tris pH 8.0 mixed immediately after elution to neutralize the solution. The eluate was then analyzed on a 10% SDS-PAGE gel, and the fractions of interest were pooled together.

The pooled fractions were buffer exchanged into Low Salt NaOAc Buffer (10 mM NaOAc, pH 5.0, 50 mM NaCl) and loaded on a High S resin column packed in house. The column was washed and eluted with Low Salt NaOAc Buffer, and salt concentration was increased stepwise (from 50 mM to 500 mM NaCl). The eluate was analyzed on a 10% SDS-PAGE gel and the fractions of interest were pooled together. The pooled Fab solution was buffer exchanged into Fab Storage Buffer (10 mM Tris pH 7.5, 50 mM NaCl), aliquoted and stored at -20 ^oC.

3.4.8 Native Electrophoretic Mobility Shift Assay (EMSA) and Polyacrylamide Coelectrophoresis (PACE)

Native gel electrophoresis and polyacrylamide coelectrophoresis were carried out using a procedure adapted from that described previously (51). 20 pmol of RNA in 6.5 µl of 10 mM Tris-HCl (pH 7.5) was incubated at 95 ^oC for 2 min and room temperature for 2 min before the addition of 0.1 µmol $MgCl_2$ and 1.5 µmol NaCl (for glycine riboswitches, 0.1 µmol glycine was

also added at this stage). After further incubation for 20 min, 10 µg heparin and 20 pmol of corresponding Fab was added to prepare a 10 µl mixture, which was incubated at room temperature for an additional 30 min. After homogenizing with 2 µl of native loading dye (50% glycerol, 0.1% xylene cyanol), the mixture was resolved via a 6% native PAGE in TB buffer containing 10 mM $MgCl_2$ (for glycine riboswitches, PACE solution also contains 5 mM glycine). Electrophoresis was carried out with TB buffer containing 10 mM $MgCl_2$ at room temperature for 1 h under approximately 12.5 V per cm of gel height.

3.4.9 Filter binding assay

The binding constants of Fab with ΔC209 P4-P6 and other RNAs were determined by filter binding assay. A constant concentration (~0.3 nM) of the 5'-^{32}P-labeled RNA was incubated at 50 ^{o}C for 10 min in PEM buffer (PBS, 0.1 mM EDTA, 10 mM $MgCl_2$; for glycine riboswitches, 10 mM glycine was added to the PEM buffer) and cooled to room temperature for 10 min. The RNA was incubated for 30 min with Fabs ranging from 4 to 2000 nM in 40 µl PEM buffer supplemented with 1 mg/ml heparin. A BA85 nitrocellulose (S&S) and Hybond N+ filter (Amersham Pharmacia) were placed in the Dot-Blot apparatus (Bio-Rad) and the wells were pre-equilibrated with 100 µl PEM buffer. The Fab-RNA mixture was applied to the Dot-Blot apparatus and rinsed with 100 µl PEM buffer five wells at a time. Both filters were air dried and exposed to PhosphorImager screens, and the amount of RNA bound to each of the filters was quantified using ImageQuant (Molecular Dynamics). Binding constants were calculated by fitting the data to the following equation: fraction bound = $M \times [Fab]^n/(K_d^n + [Fab]^n)$, where K_d is the binding constant; M is the maximum fraction of RNA bound at the highest Fab concentration; and n is the Hill coefficient.

3.4.10 Competitive filter binding assay

To determine the FSI value of each Fab, a filter binding assay was first carried out to determine the binding affinity (K_d) of the Fab binding to its cognate RNA. In the competitive filter

binding assay, a constant concentration (~0.3 nM) of the 5'-^{32}P-labeled cognate RNA was incubated at 50 ^{o}C for 10 min in PEM buffer (PBS, 0.1 mM EDTA, 10 mM $MgCl_2$; for glycine riboswitches, 10 mM glycine was added to the PEM buffer) and cooled to room temperature for 10 min. The RNA was incubated for 30 min with its selected Fab at a concentration equal to $2K_d$ in 40 µl PEM buffer. In seven other reactions, the mixture was incubated with $2K_d$ concentration of one of the seven non-radiolabeled non-cognate RNAs from the following list: P4-P6, FNI-II, VCI-II, ExD1, AMTHF, DASAH, tRNA, CBATHF. A BA85 nitrocellulose (S&S) and Hybond N+ filter (Amersham Pharmacia) were placed in the Dot-Blot apparatus (Bio-Rad) and the wells were pre-equilibrated with 100 µl PEM buffer. The Fab-RNA mixture was applied to the Dot-Blot apparatus and rinsed with 100 µl PEM buffer five wells at a time. Both filters were air dried and exposed to PhosphorImager screens, and the amount of RNA bound to each of the filters was quantified using ImageQuant (Molecular Dynamics). The percent binding of the reaction without non-cognate RNA competitor was used as the reference point. The percent binding values obtained with seven non-cognate RNA competitors were then averaged, divided by the reference point, and averaged over two independent experiments to give FSI value for a given Fab RNA target pair, where 0 is completely nonspecific and 1 is completely specific.

CHAPTER 4 TOWARDS A HIGH RESOLUTION CRYSTAL STRUCTURE OF THE GLYCINE RIBOSWITCH

4.1 Introduction

Our understanding of RNA's role in life processes has expanded greatly since the initial discovery of structured cleaving ribozymes over three decades ago. Besides canonical coding mRNA, non-coding RNAs exist with many manifestations mediating genetic expression and regulation such as histone modification, DNA methylation, metabolite sensing, post transcriptional gene silencing, intron splicing, trans RNA processing, and ribozymes involved in rolling circle replication (47, 107, 108). Non-coding RNAs have also been implicated in disease processes such as cancer, autism, obesity, Alzheimer's, and atherosclerosis (109-119).

RNA often assumes complex tertiary structures to execute its functions by interacting with proteins, other RNA, DNA, or small molecules; tRNA is a prime example of an RNA that does not encode information but rather folds into a complex tertiary structure to execute its function. Therefore, determining high resolution RNA structures will blaze the trail towards further understanding RNA mediated genetic regulation and potential diagnostic and clinical applications. For example, such structures allow understanding of RNA catalysis, methods of specific and high affinity metabolite binding, and interaction with other molecules in life processes, i.e. translation assemblies. Far more protein structures, however, have been determined than RNA; of the 81,103 x-ray diffraction structures currently in the PDB, less than 2.5% contain RNA. This reflects the inherent difficulty of RNA crystallography and current lack of a systematic method to accomplish RNA crystallization.

Crystal packing of RNA is difficult because of the large amount of anionic surface area; the repetitive anionic phosphate backbone can create imperfectly replicable packing, slipping slightly out of register. RNA's limited chemical diversity relative to proteins (4 bases with relatively similar structure vs. 20 amino acids with a larger variety of structures, more potential

for interactions) is also seen as an impediment to packing (77, 120). Both chemicals and RNases degrade RNA, making the precursor crystal solution more heterogeneous, impeding high order ergo quality crystal formation. RNA structure is also more flexible than a protein of similar mass, perhaps again due to anionic repulsion, which also is counterproductive to a regular crystalline lattice. In addition to crystal formation difficulties, phasing is also problematic. Unlike selenomethionine mutant facilitated phasing in protein crystallography(120), RNA crystallography often relies on the large atoms from heavy metal 'soak and pray' methods to obtain phase information by methods such as MAD/SAD (multi-/single wavelength anomalous dispersion). Other options for phasing include large bromine in the form of brominated/iodated uridine in sequences (34) and a three base pair heavy metal binding motif engineered to facilitate heavy metal ion binding (121). Obtaining crystals can be challenging enough, but these phasing approaches often require making multiple batches of crystals to obtain complete diffraction data/phase information which is another issue if the crystals are not isomorphic/metal inclusion is sporadic.

Fab chaperone assisted RNA crystallography (CARC) is an RNA crystallization method which uses Fab (antigen binding antibody fragments) as chaperones. Fab CARC provides surface protein chemical diversity for crystal contact formation and decreases surface anionic content to promote packing while also addressing phasing issues, as its presence will provide a large structure for molecular replacement phasing. A library of Fabs fused to a phage coat protein is used in phage display selection against RNA to isolate Fabs binding to a particular RNA target, an in vitro technique required since an RNA antigen cannot be injected into a host for hybridoma development of antibodies due to degradation. Fab CARC successfully crystallized *Tetrahymena* Group I intron P4-P6 domain (PBD: 2R8S)(74) and Class I Ligase (3IVK)(75) structures. Fab mediated crystal contacts provided 61% and 59% of the total solvent accessible surface area buried in crystal contacts for P4-P6 and the Ligase, respectively. Moreover, crystal structures of the Class I Ligase chaperoned with Fab protein proved much

less arduous to phase than the U1A version (75). Fabs are also beneficial as they can be selected against an unengineered structure using the phage display technique, requiring little previous knowledge of its functional residues; such knowledge is required if mutational approaches to crystallization/phasing are employed. Raising Fab against novel RNA in cells could allow specific RNA immunostaining and identification of disease markers, to name a few other possibilities.

Well structured RNAs for which crystal structures can be obtained (as opposed to unstructured RNAs which may still have a biological function) are comprised of more than just the previously mentioned tRNA and Group I/Group II introns; another class of structured RNA is the riboswitch. Riboswitches are structured motifs usually upstream of start codons that bind a small molecule (often metabolite) and undergo a conformational change which, most frequently, inhibits or activates translation or transcription, occasionally with protein assistance (47). In addition to the basic research goal of elucidating the exact atomic mechanism of gene regulation, a riboswitch structure is useful for applied research, allowing structure based design of new antibiotics. The basis for functioning of the antimicrobial compound pyrithiamine has been shown to be the TPP riboswitch, lysine analogues target the lysine riboswitch, and roseoflavin targets riboflavin riboswitches; riboswitch mutants that no longer effectively regulate translation show resistance (122). Since initial realizations of the mechanism of these previous antibiotics, analogues for c-di-GMP and guanine have been designed for their riboswitches (123-126). A trial testing *Staphylococcus aureus* bovine intramammary infection with the guanine analogue 2,5,6-triaminopyrimidine-4-one showed some positive results (127). The glycine riboswitch is an especially novel gene regulator with its dual aptamer configuration, the purpose of which is still uncertain (6, 51). It is upstream of the gcvT operon, involved in glycine cleavage; glycine is an important part of bacterial metabolism, making it a reasonable target for antibiotics. Accordingly, the last portion of this research attempts to crystallize the glycine riboswitch.

The crystal structure of the dual aptamer truncated FNI-II riboswitch (19) and of an individual aptamer from species *Vibrio cholerae* (VC) (18), were obtained by other laboratories. The FNI-II glycine riboswitch (truncated glycine riboswitches) behaves much differently than the FNLD (contains leader sequence, full length riboswitches) riboswitch, however. Studied more in depth in the VC version of the truncated riboswitch (VCI-II) (51), a mutation in one aptamer that prevented it from binding glycine also weakened the other aptamer's ability to bind glycine, leading to the idea of glycine binding cooperativity. However, such is not the case with VCLD; even with one aptamer's mutation-induced lack of glycine binding, the other was still able to bind glycine. Furthermore, the full length glycine riboswitches (VC, FN, and *Bacillus subtilis*) bind glycine with a lower K_d of about 5 µM (now similar for these species, which was not the case for truncated versions); the kink-turn may further preorganize the glycine binding sites. Because of this new data, a new structure of the glycine riboswitch is obviously needed. FNLD is chosen due to smaller size since smaller targets crystallize more readily.

The approach that yielded the FNI-II structure involved mutating the P2 loop of aptamer I, which influenced the following work. The Class I Ligase structure's Fab BL3-6 bound to a portable RNA loop which was mutated into the same location in FNLD as was done for FNI-II. Several Fab and RNA combinations based on this, along with hundreds of crystallization conditions, have been attempted. Preliminary crystals have been diffracted but to insufficient resolution. Several further approaches are also posited to attempt to achieve a sufficiently high enough resolution crystal.

4.2 Results

The FNI-II structure was obtained using a smaller chaperone protein, the U1A protein. It should be noted that it is only about 11 kiloDaltons (kD), while a Fab is 48 kD, much larger, therefore facilitating phasing by molecular replacement of its large known structure. The U1A protein has a small sequence of RNA that is its cognate ligand. In the FNI-II structure (19), aptamer I's P2 loop was replaced with this based upon previous information that mutating this

location should be innocuous, and this construct is what allowed FNI-II cocrystallization with U1A. Based upon the success of this approach, a similar approach was taken in the hopes of expediting crystallization. The aforementioned Class I Ligase crystal's cognate Fab BL3-6 actually bound to a loop of the ligase as well, creating a portable Fab/ligand nucleotide sequence pair. This GAAACAC loop sequence was mutated into the same location in aptamer I P2, and the RNA construct FNLDP2 was made in crystallization quantity, followed by large scale expression of BL3-6. After expression, BL3-6 was purified through protein A affinity, High S cation exchange gravity flow column chromatography, and finally, Superdex 200 gel filtration with an AKTAprime Plus. All of these steps purify the Fab, as having a pure sample is very important for packing into a regular crystalline lattice. This purification is also important to remove RNases that are present from the *E. coli* 34B8 expression system from which Fab is obtained by cell lysis. Fab is tested for nuclease activity by incubation with non cognate ^{32}P radiolabeled RNA, fragments of which are then separated by denaturing polyacrylamide gel electrophoresis (dPAGE). This is visualized by autoradiography and subsequently quantified.

Once Fab and RNA were purified, the pair was checked for complexation with a polyacrylamide native gel and the 176 nM K_d was determined by filter binding (dot blot) assay. The native complexation gel was done with stoichiometric excesses of Fab (1, 1.1, 1.2, 1.3, 1.4 equivalents; concentrations of Fab and RNA are determined with computational constants and may be slightly incorrect) from which the lowest amount that gave as complete a shift as a possible was chosen as the crystallization stoichiometry (Fig 4.1). With this stoichiometry a Fab/RNA complex solution was formed, and screening trays were set up. Screening comprised of 96 well sitting drop plates of Natrix 1&2, Crystal Screen 1&2, and Index 1&2 crystal solution kits from Hampton Research; duplicates were made for 15 and 4 °C for a total of nearly 600 conditions.

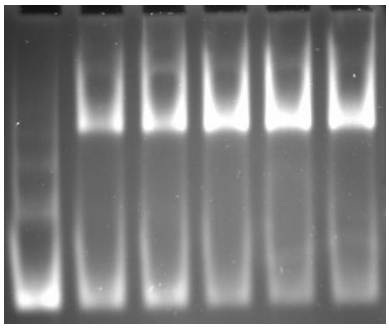


Figure 4.1 Native complexation gel of FNLDP2 and BL3-6.

The first lane is folded RNA, followed by 1, 1.1, 1.2, 1.3, and 1.4 equivalents of BL3-6. The lanes containing Fab show a higher band, the Fab RNA complex.

From these, two initial conditions were obtained, one at each temperature. After about one month, Index condition D3 at 4 °C contained a crystal shower, resembling rectangles (Fig. 4.2 left). At 15°C, Index F9 showed needle clusters after about two and a half weeks (Fig. 4.2 right). These two conditions were used as a starting point for variations in the percent of their respective precipitants as well as variations upon their pHs in 24 well hanging drop optimization trays. Once a course optimization was done (ex. pH increments of 0.5 and ± several percent precipitant on either side of the initial condition) and a condition was chosen from this, a fine optimization with smaller increments was done to further refine conditions. During the course of optimizations, a new batch of BL3-6 needed to be made, which resulted in some repetition of effort. Crystal reproducibility was sporadic, however (perhaps in part due to the new batch of Fab), with different formations of needle clusters for F9 or a much longer incubation time needed for the D3 condition. Consternatingly, also the F9 needles often had loosely aggregated quasi-crystal ‘blobs’/spherulites in addition to or in place of the needles clusters themselves, a situation which is not noted as helpful for crystallization by crystallographers. Some optimization trays were done altering the ratio of well solution to complex solution; typically, they are 1:1, but 2 µl complex:1 µl well solution and 0.5 µl complex:1 µl well solution were also used. Half of an optimization tray was made for the D3 condition to contain complex with 10mg/ml, slightly lower

than the usual 12mg/ml, and the other half had a slightly stoichiometrically altered ratio of Fab to RNA in the complex solution. A few conditions related to F9 showed similar crystal formation at a later date, and an optimization tray was made exploring this. Unfortunately, none of these and other variations attempted lead to any definite improvement in reproducibility or morphology (single crystals are usually best for diffraction, but the needle clusters are far from this ideal).

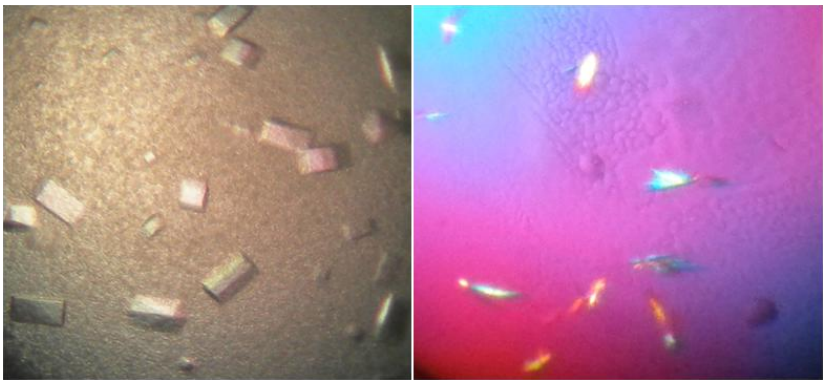


Figure 4.2 Initial crystal hits.

Left, initial rectangular Index D3 crystals surrounded by precipitate. Right, initial Index F9 needle crystals with phase separation. This picture does not show many quasi-crystal 'blobs' or spherulites also present in the F9 condition, though one is obvious in the lower right corner.

From these roughly two dozen optimization trays along the ideas listed above, there was a D3 tray from which large rectangular crystals were harvested and sent for initial diffraction testing to Brookhaven National Laboratory courtesy of Annie Heroux. They contained a large quantity of water (~70%) and diffracted only weakly, roughly an 8 Å resolution (about 4-5 Å too high for current standards). Another tray was set up with similar conditions after this result, and the relative dimensions of the rectangular crystals which grew were altered. These were harvested under two conditions, one with normal cryoprotectant solution, the second under conditions that attempted to dehydrate the crystals to remove excess water and improve diffraction. Post crystallization dehydration treatment has been noted to improve the diffraction

of some crystals (128, 129). Showing how difficult it is to judge a crystal by its appearance, despite the different relative dimensions, this second batch of crystals diffracted equal to or just slightly better than (6-7Å) the previously diffracted batch, including the dehydrated crystals, for which no observable change was noted during dehydration (i.e. cracking) as has been for some crystals that underwent successful, resolution improving dehydration.

The addition of additives is known to assist crystal formation. As such, additive screens for both of these conditions were performed using Hampton Research's additive screen kit. For the F9 condition, the additive screen did indeed result in more frequent crystal formation. Unfortunately, however, most conditions still contained quasi crystals, and the overall morphology was similar to that seen previously, various types of needle clusters ill suited for diffraction. The additive screen for the D3 condition showed better results. Many different wells contained crystals, from rectangular showers of a multitude of smaller crystals, wells with a far smaller number but larger size of crystals, intermediates to these, and even some wells where the morphology changed to rod clusters. This additive screen was also encouraging since the time for initial D3 crystal formation was on the order of weeks instead of months. These underwent diffraction testing, at best resulting in ~5.3 Å, and some of them are pictured in Fig. 4.3. As was the case previously, here none of the dehydration conditions (one with EtOH, the other with increased precipitant) produced any discernible resolution improvement relative to those simply harvested in cryoprotectant.

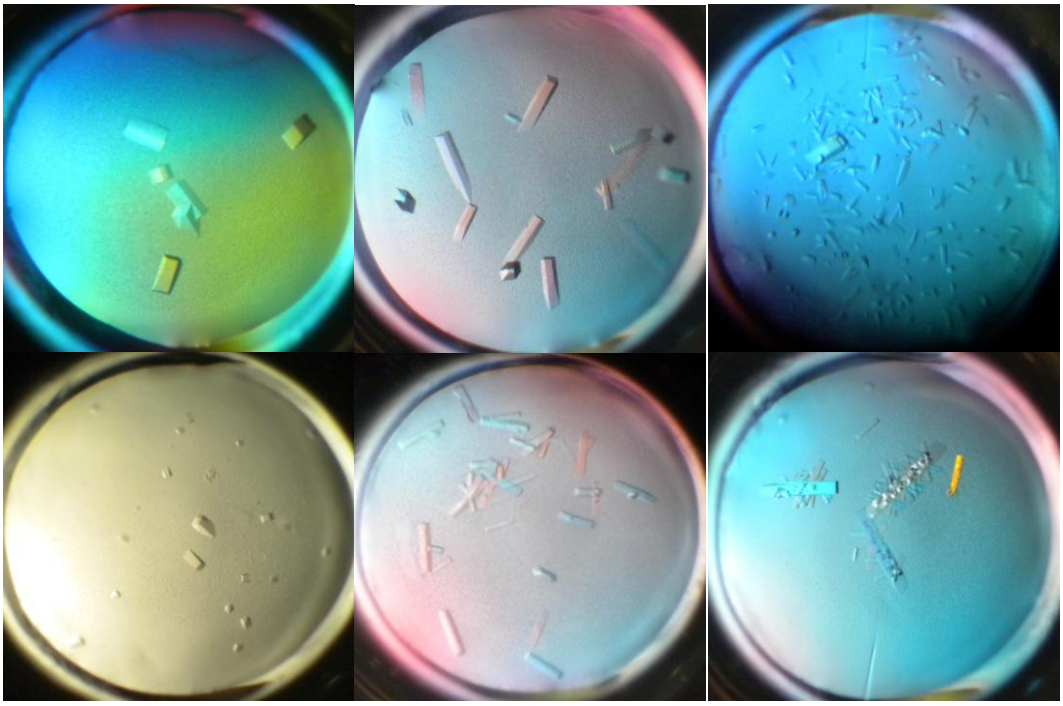


Figure 4.3 Several different D3 additive crystal drops.

Concurrently with the additive tray approach, four mutants of the P2 base paired stem below the loop in the FNLD RNA were made, adding or removing one or two base pairs to create FNLDP2b, FNLDP2c, FNLDP2d, and FNLDP2e RNAs. Altering stem length changes the orientation of Fab relative to the RNA target due to helical rotation, altering three dimensional structures, and has proven helpful in crystallization (130) (also suggested by Barbara Golden in personal communication). Though the lengthened mutants did not prove very fruitful, the shortened ones did show formation of a handful of crystals, mostly from FNLDP2d, though one interesting morphology was obtained through FNLDP2e. In addition, based upon work by others in the laboratory, a Fab mutant was made called BL3-6 BS1 for use in crystallization. BS1 stands for beta sheet 1 mutant, mutations created to strengthen a beta sheet contact present in the P4-P6 Fab 2 crystal structure. BL3-6 BS1 combined with FNLDP2 created many small showers; these have yet to be explored extensively in additional trays.

4.3 Discussion and Future Directions

Though currently no useful diffraction data has been obtained, especially given the increase in crystal formation due to the additive screen, there are still many efforts that can be attempted, some of which have been mentioned in context in the previous section. One approach would be to slightly elevate the temperatures of the additive screen tray conditions which showed showers of a multitude of crystals, as increasing the temperature may slow the crystal formation to allow a smaller number but larger size of crystal to be formed, though the more common remedy to this is seeding, mentioned below.

The F9 condition with additive screen well F10 seemed to reduce needle clusters to individual needles, possibly useful for seeding. Seeding is the process of inserting a small crystal (mactroseeding) or small crystal pieces (microseeding, streak seeding) into a crystal drop in an effort to grow one or a few large crystals for diffraction, as crystals need to be large enough (perhaps about 100 microns) to allow collection of enough data. This approach might also be taken for the D3 wells as well, especially with some of the new rod clusters, a different

morphology which grew somewhat smaller crystals in a shower. This approach may further be advisable due to the large D3 unit cell (~120x120x915), as one potential remedy to a larger unit cell crystal with weak diffraction is to make a larger crystal containing more unit cells in hopes of increasing diffraction (increasingly larger crystals could be seeded to further increase their size). Lastly, there were multiple small showers for especially the FNLDP2d stem mutants, which may be useful in crystal seeding.

As previously mentioned, dehydration or other post crystallization treatments may also allow a crystal that does not quite diffract to high enough resolution to increase a few Angstroms to sufficient resolution. Mutations have been done to the Group I intron to improve its crystallizability as well(128, 130-132); a similar approach might be taken with this riboswitch.

Several different Fabs were also raised against FNI-II. The YSGR Minimal library produced FNI-II Fab 16, and FNI-II Fabs 4, 8, and 9 were selected from the YSGRX library. These could all be checked for complexation with FNLD constructs (16 and 4 are first choice due to lower K_d values), then used in crystal trays if the Fabs bind. Another Fab approach is surface entropy reduction (SER) of Fabs, or mutating flexible, protruding surface residues into smaller, less obtruding residues. This approach was done for the Fab 2 P4-P6 crystal by another individual in this laboratory, showing encouraging results in formation of crystals in additional conditions (additional 'hits') (42). SER super mutant alanine or serine (SMA/SMS) versions of any of the above five Fabs (4, 16, 8, 9, and BL3-6) may prove helpful in crystallization. SMA/SMS Fab expression plasmids of BL3-6, Fab 4, and Fab 16 have already been created. Along similar lines, the previously mentioned beta sheet 1 (BS1) mutation designed to improve the P4-P6/Fab 2 crystal's beta sheet interaction has already shown some promise in screening trays on the Fab BL3-6 (namely, BL3-6 BS1/FNLDP2). The Fab 2 BS1/P4-P6 complex actually produced solvable diffraction data which is why the BS1 mutations were chosen over BS2, though a BL3-6 BS2 Fab expression plasmid has also already been created. If additional optimization, additive, and seeding approaches do not provide well resolved

diffracting crystals with any of the five RNA stem construct variations in combination with Fab BL3-6 BS1 (or any of its other mutants), BS1 and BS2 mutations could also be done to any of the other Fabs (4,8,9,16).

4.4 Materials and Methods

4.4.1 RNA Preparation

The previously constructed glycine riboswitch FNI-II was generated by annealing DNA oligonucleotides (Integrated DNA Technologies) and amplified by recursive PCR (55) with an EcoRI restriction site and T7 promoter sequence engineered at the 5' terminus, and EarI and HindIII sites engineered at the 3' terminus. Double digested (EcoRI/HindIII, New England Biolabs protocol, 37°C for 1 hour, NEB buffer 2) riboswitch DNA inserts and pUC19 vector were ligated together (T4 DNA ligase, New England Biolabs protocol, 16°C overnight) to generate the plasmid. FNLD was previously constructed from FNI-II (51) by Quikchange and used as a template for FNLDP2, FNLDP2b, FNLDP2c, FNLDP2d, FNLDP2e which were constructed by QuikChange site-directed mutagenesis (Stratagene) from the FNLD plasmid and proper DNA oligonucleotide primers (see primer table below). The plasmids were transformed into JM109 cells, amplified and purified with Qiagen's miniprep plasmid preparation kit, and their sequences were confirmed by sequencing (Eton BioScience Inc.). More of the correct resulting plasmids were produced with Qiagen's maxiprep kit and were then linearized with EarI (New England Biolabs, 0.2 unit per µg of plasmid, NEB buffer 4, 37°C overnight) for transcription by T7 RNA polymerase.

Construct	Primer	Sequence (added Fab binding loop in bold, inserted base pairs red)
FNLDP2	FNLDP2Fwd	CGGATGAAGATATGAGGAGAGATTTCAT**GAAACAC**ATGAAACACCGAAGAAGTAAATC
	FNLDP2Rvs	GATTTACTTCTTCGGTGTTTCAT**GTGTTTC**ATGAAATCTCTCCTCATATCTTCATCCG
FNLDP2b	FNLDP2bFwd	CGGATGAAGATATGAGGAGAGATTTCATC**GAAACAC**GATGAAACACCGAAGAAGTAAATC
	FNLDP2bRvs	GATTTACTTCTTCGGTGTTTCATC**GTGTTTC**GATGAAATCTCTCCTCATATCTTCATCCG
FNLDP2c	FNLDP2cFwd	GATGAAGATATGAGGAGAGATTTCATAC**GAAACAC**GTATGAAACACCGAAGAAGTAAATC
	FNLDP2cRvs	GATTTACTTCTTCGGTGTTTCATAC**GTGTTTC**GTATGAAATCTCTCCTCATATCTTCATC
FNLDP2d	FNLDP2dFwd	CGGATGAAGATATGAGGAGAGATTTCA**GAAACAC**TGAAACACCGAAGAAGTAAATC
	FNLDP2dRvs	GATTTACTTCTTCGGTGTTTCA**GTGTTTC**TGAAATCTCTCCTCATATCTTCATCCG
FNLDP2e	FNLDP2eFwd	CGGATGAAGATATGAGGAGAGATTTC**GAAACAC**GAAACACCGAAGAAGTAAATC
	FNLDP2eRvs	GATTTACTTCTTCGGTGTTTC**GTGTTTC**GAAATCTCTCCTCATATCTTCATCCG

RNAs were transcribed in 1 ml reaction mixture containing 40 mM Tris-HCl (pH 7.9), 2 mM spermidine-3HCl, 10 mM DTT, 25 mM $MgCl_2$, 5 mM each of the nucleotide triphosphates, 50 µg/ml linearized plasmid, 0.01 unit/µl 5 PRIME^{TM} stop RNase inhibitor RX (5 PRIME life science), 8 mili unit/µl thermostable inorganic pyrophosphatase (New England Biolabs), and 45 µg/ml T7 RNA polymerase (expressed and purified in-house) for 4 h to overnight at 37°C. The transcription mixture was incubated with 1 mM $CaCl_2$ and RNase-free DNase I (Promega) at a concentration of 0.5 unit per µg of DNA plasmid at 37°C for 30 minutes. The RNA transcripts were purified by 6% to 8% denaturing PAGE, excised under UV TLC plate shadowing, electroeluted into ~15 ml TE buffer (10 mM Tris-HCl, pH 8, 1 mM EDTA) in dialysis tubing, Amicon concentrated, PCA then CA extracted, ethanol precipitated, and redissolved in 1M pH 6.5 Na cacodylate buffer to ~14-24mg/ml concentration (needed for later dilution in crystallization). RNA concentrations were determined by UV absorbance at 260 nm wavelength with extinction coefficients calculated by OligoCalc (57). After solvation, crystallization RNAs

were benchtop centrifuged and filtered through a Millipore 0.22 µm Ultrafree microcentrifuge tube device to remove any precipitate/particulate and flash frozen with liquid nitrogen for storage at -80 °C in small aliquots.

4.4.2 Fab Expression and Purification

Fab BL3-6 plasmid was provided by Dr. Piccirilli, University of Chicago, and electroporated into 34B8 cells. A starter culture was inoculated and grown at 30 ^{o}C and 300 rpm overnight in 25 ml 2YT with 100 µg/ml ampicillin. 5ml of the overnight starter culture was used per 500 ml subculture of CRAP media (27 mM $(NH_4)_2SO_4$, 2.4 mM sodium citrate-$2H_2O$, 14 mM KCl, 5.4 g/liter yeast extract, 5.4 g/liter HyCase SF Casein, 0.11 M Mops– NaOH buffer [pH 7.3], 0.55% [w/v] glucose and 7 mM $MgSO_4$) supplemented with 100 µg/ml of ampicillin and grown at 30 ^{o}C and 300 rpm for 24 h. Cells were spun down at 6000~8000 rpm for 10 min at 4 ^{o}C and the pellets were frozen at -20 ^{o}C overnight. Over the course of all of the crystal trays, several dozen liters were grown in a total of three batches.

The frozen cell pellets were thawed and lysed with lysozyme (9mg/30ml), Dnase I (2.5mg/30ml) in 30 ml lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 0.5 mM PMSF) per 500ml pellet which was completely resuspended before vigorous stirring for 1 h at room temperature. Cell debris was centrifuged down and the supernatant was loaded onto a protein A column at 4 ^{o}C which was previously equilibrated with 20CV of TBS (10 mM Tris pH 7.5, 150 mM NaCl). The column was washed with 50 column volumes (CV) of 10 mM Tris pH 7.5, 500 mM NaCl and eluted with 0.1 M citric acid, pH 3. The eluate was collected in 3 ml fractions in tubes prefilled with 1.5 ml 1 M Tris pH 8.0, mixed immediately after elution to neutralize the solution, and analyzed on a 10% SDS-PAGE gel. The fractions of interest were pooled together.

The pooled fractions were dialyzed into Low Salt NaOAc Buffer (10 mM NaOAc, pH 5.0, 50 mM NaCl, 4L per eluate from two 5mL protein A columns divided into three aliquots) and loaded at 4 ^{o}C on a High S resin gravity flow econo column packed in house previously washed with 20CV of Low Salt NaOAc Buffer. The column was washed and eluted with Low Salt NaOAc

Buffer and stepwisely increased salt concentration (90% low salt, 10% high salt 500mM NaCl buffer, 80/20, etc., from 50 mM to 500 mM NaCl). Fractions were checked for protein at 280nm, (extinction coefficients obtained from the expasy.org ProtParam tool) the eluate was analyzed on a 10% SDS-PAGE gel, and the fractions of interest were pooled together in an Amicon so that no more than 5mg was in 250µl for quantitative recovery from a gel filtration column.

Before loading on a gel filtration column, a portion of the Fab was diluted to about 0.4mg/ml (approximate High S fraction concentration) in one tube and 6mg/ml (approximate crystallization drop concentration) in another tube (final incubated concentrations) in an assay with 0.5µl ~10kcpm radiolabeled (see procedure below) non cognate RNA (0.5 hours, 37 °C). Final volume of 9.5 µl is accomplished with Fab storage buffer (10 mM Tris pH 7.5, 50 mM NaCl). TE buffer and Fab storage buffer with RNA only are each done in an additional tube for comparison, RNase contamination control. Fragments are electrophoretically size separated by 8% polyacrylamide gel, the gel is dried, exposed to PhosphorImager cassettes overnight, and quantified with ImageQuant software (Molecular Dynamics). The rule of thumb is if the Fab sample has less than ~10% additional degradation relative to the lanes that contain buffer only, it is then purified by gel filtration. Otherwise, another round of High S is performed.

Fab gel filtration is performed with a Superdex 200 10x300mm column connected to an AKTAprime plus at 4 ^{o}C for final polishing, removing proteases/RNases and increasing Fab homogeneity. Based upon manufacturer's guidelines for maximum recovery, we load ~250 µl and ~5 mg into a 0.5 ml loop that has been first benchtop centrifuged and transferred into a new microcentrifuge tube to remove precipitate. The sample is monitored at 254 nm. At a flow rate of 0.4 ml/min., we collect ~0.5 ml fractions that begin to elute ~45 min. Useful fractions are concentrated and buffer exchanged into Fab storage buffer with an Amicon centrifugal concentrating device (~14-24 mg/ml), aliquoted in small fractions, and immediately frozen in $N_{2(l)}$ for storage at -80 ^{o}C.

4.4.3 5' end ^{32}P labeling of RNAs

In a typical procedure, 40 pmol RNA transcript was dephosphorylated by incubation with 1 unit of shrimp alkaline phosphatase (Fermentas Life Sciences) in the supplied phosphatase buffer in a total volume of 10 µl at 37 ^{o}C for 30 min. After incubation at 65 ^{o}C for 15 min to inactivate the phosphatase, 1 µl sterile H_2O, 1.5 µl 10× T4 polynucleotide kinase buffer (New England Biolabs), 2 µl γ-^{32}P ATP (Perkin Elmer), and 0.5 µl 10 unit/µl T4 polynucleotide kinase (New England Biolabs) were added and the reaction mixture was incubated at 37 ^{o}C for 30 min for RNA phosphorylation. For RNA transcripts with the 5' end buried by or too close to a duplex region, to improve the labeling yield, the transcripts were first annealed to a DNA oligonucleotide to free the 5' end before dephosphorylation and phosphorylation reactions. After phosphorylation, ^{32}P-labeled RNAs were purified by 6-8% denaturing polyacrylamide gel electrophoresis, visualized on an x-ray film, excised, and passively eluted in 300 µl TE buffer on a rotator at 4 ^{o}C overnight. After centrifugation, the radiolabeled RNAs were PCA extracted twice, EtOH precipitated, and redissolved in TE buffer.

4.4.4 Native Electrophoretic Mobility Shift Assay (EMSA) and Polyacrylamide Coelectrophoresis (PACE)

Before setting up each new crystal tray with a new Fab/RNA pair, a native gel was done to confirm their complexation and determine complex (Fab:RNA) stoichiometry. Native gel electrophoresis/polyacrylamide coelectrophoresis was carried out using a procedure adapted from that described previously.(51) 20 pmol of RNA in 6.5 µl of 10 mM Tris-HCl (pH 7.5) was incubated at 95 ^{o}C for 2 min and room temperature for 2 min before the addition of 0.1 µmol $MgCl_2$ and 1.5 µmol NaCl (for glycine riboswitches, 0.1 µmol glycine was also added at this stage). After further incubation for 20 min, 10 µg heparin and 20 pmol of corresponding Fab was added to prepare a 10 µl mixture, which was incubated at room temperature for an additional 30 min. After homogenizing with 2 µl of native loading dye (50% glycerol, 0.1% xylene cyanol), the mixture was resolved via a 6% native PAGE in TB buffer containing 10 mM $MgCl_2$ (for glycine

riboswitches, PACE solution also contains 5 mM glycine). Electrophoresis was carried out with TB buffer containing 10 mM $MgCl_2$ at room temperature for 1 h under approximately 12.5 V per cm of gel height.

4.4.5 Filter binding/dot blot assay

The binding constant of Fab BL3-6 with FNLDP2 was determined by filter binding assay. A constant concentration (~0.3 nM) of the 5'-^{32}P-labeled RNA was incubated at 50 ^{o}C for 10 min in PEM buffer (PBS, 0.1 mM EDTA, 10 mM $MgCl_2$; for glycine riboswitches, 10 mM glycine was added to the PEM buffer) and cooled to room temperature for 10 min. The RNA was incubated for 30 min with Fabs ranging from 4 to 2000 nM in 40 µl PEM buffer supplemented with 1 mg/ml heparin. A BA85 nitrocellulose (S&S) and Hybond N+ filter (Amersham Pharmacia) were placed in the Dot-Blot apparatus (Bio-Rad) and the wells were pre-equilibrated with 100 µl PEM buffer. The Fab-RNA mixture was applied to the Dot-Blot apparatus and rinsed with 100 µl PEM buffer five wells at a time. Both filters were air dried and exposed to PhosphorImager screens, and the amount of RNA bound to each of the filters was quantified by using ImageQuant (Molecular Dynamics). Binding constants were calculated by fitting the data to the following equation: fraction bound = $M \times [Fab]^n/(K_d^n + [Fab]^n)$, where K_d is the binding constant; M is the maximum fraction of RNA bound at the highest Fab concentration; and n is the Hill coefficient.

4.4.6 Fab surface protein residue mutations (BS1/BS2/SMA/SMS)

The parent phagemid of BL3-6, designed to display bivalent, humanized Fab on the surface of M13 bacteriophage, has been described previously (42). Electroporation of this phagemid into CJ236 cells (uracil deglycosidase/dUPTase deficient) generated uracil-containing ssDNA by following the procedure as described (42). Various Kunkel primers were designed to insert the SMA/SMS/BS1/BS2 (see primer table below, LC light chain, HC heavy chain) mutations through Kunkel mutagenesis (97). After phosphorylation using T4 polynucleotide kinase, these primers were annealed with uracil-containing ssDNA of BL3-6 (or other parent

Fab on which these mutations were desired), extended with T7 DNA polymerase, ligated with T4 DNA ligase, cleaned with PCA extraction/EtOH precipitation, and transformed into XL1-blue cells (uracil glycosidase/dUTPase containing strain) to generate a phagemid following the procedure as described (42). Small scale DNA was purified using Qiagen's miniprep kit and sent for sequence confirmation (Eton Bioscience Inc.).

Fab	Primer	Sequence
BL3-6 BS1	LCE123A	CTTCATCTTCCCGCCATCTGATGCCCAGTTGAAATCT GGAACTGC
	Bsheet1 K214A, D216T, K217A, K218T, E220A, stop codon	CACAAGCCCAGCAACACCGCCGTCACCGCCACCGTT GCCCCCAAATCTTGTGACAAAACTCACACATAGGGC CGGCCCTCTGGTTCC
BL3-6 BS2	LCE123A	CTTCATCTTCCCGCCATCTGATGCCCAGTTGAAATCT GGAACTGC
	Bsheet2 K214A, D216A, K217A, K218A, E220A, stop codon	CACAAGCCCAGCAACACCGCCGTCGCCGCCGCCGT TGCCCCCAAATCTTGTGACAAAACTCACACATAGGG CCGGCCCTCTGGTTCC
BL3-6 SMA	LCE123A	CTTCATCTTCCCGCCATCTGATGCCCAGTTGAAATCT GGAACTGC
	LCK190A	AGCAGACTACGAGAAACACGCCGTCTACGCCTGCGA AGTCA
	HCSM1stop K217A, K218A, E220A, stop codon	AGCAACACCAAGGTCGACGCCGCCGTTGCCCCCAAA TCTTGTGACAAAACTCACACATAGGGCCGGCCCTCT GGTTCC
BL3-6 SMS	LCE123S	CTTCATCTTCCCGCCATCTGATAGCCAGTTGAAATCT GGAACTGC
	LCK190S	AGCAGACTACGAGAAACACAGCGTCTACGCCTGCGA AGTCA
	HCSM2stop K217S, K218S, E220S, stop codon	AGCAACACCAAGGTCGACAGCAGCGTTAGCCCCAAA TCTTGTGACAAAACTCACACATAGGGCCGGCCCTCT GGTTCC

4.4.7 Crystallization complexation

Before addition to crystal trays with crystallization solutions, Fab and RNA samples undergo a folding and complexation procedure. First, a 6-8% native polyacrylamide gel (EMSA described previously) is done with slightly differing RNA:Fab ratios (1:1, 1:1.1, 1:1.2, 1:1.3, 1:1.4) in individual lanes to determine the best

stoichiometry for each batch/new combination of Fab and RNA. Once this is determined, a large batch of complex with the empirically best ratio can be made for crystal trays. Buffers, microcentrifuge tubes, and pipette tips must be as RNase free as possible so that there is as little RNase present as possible in the drops during the potentially long incubation times. A 12mg/ml Fab/RNA complex is used (made in ~15% volume excess for pipette error), which is usually combined with one equivalent of well solution, though instances of 1:2 and 2:1 sample:well solution ratio have been attempted on occasion but did not produce useful crystals. This 12mg/ml complex is converted to moles for a 1:1 ratio, then whatever slight stoichiometric excess of Fab has been determined by native gel is added in addition to this value. $MgCl_2$ has a final concentration in the complex solution of 20mM for RNA structure, 10mM pH 7.5 Tris HCl is used as a buffer and adjustable volume (as Fab and RNA concentrations vary from batch to batch), 50mM NaCl for Fab structure (Fab solvated in Fab storage buffer which contains 50mM NaCl already, so this must be subtracted from the final complex volume), 0.5U/µl final concentration of Superase RNase inhibitor (20U/µl tube, purchased through Invitrogen AM2696, made by Ambion). For the glycine riboswitch 5mM glycine final concentration is also used. The final volume required depends upon the number of trays being set up; for example, for one 96 well sitting drop screening tray, 55.2 µl would be required (48 µl*1.15 excess pipette error volume), as 0.5 µl well solution is combined with 0.5 µl complex in each well. The complexing and folding procedure for a crystallization batch of RNA consists of adding calculated required volumes of RNA and Tris into a microcentrifuge tube; vortexing and centrifuging down; heating to 95 °C for 2 min.; cooling at room temperature (RT) 2 min.; adding Mg, NaCl,

and glycine; vortexing and centrifuging down; RT 20 min.; adding Superase and Fab; vortexing and centrifuging down; 30 min. RT; centrifuge in a benchtop centrifuge for 5 min. at max speed to pellet any particulates; and lastly, careful pipetting to remove complex to a new tube without disturbing any pellet, ready for addition to a tray. Hanging drop trays are usually 1 µl complex: 1 µl well solution on the cover slip, double volume from sitting drop.

4.4.8 Crystal screening conditions

Screening (FNLDP2, FNLDP2b, FNLDP2c, FNLDP2d, FNLDP2e, each with BL3-6, and FNLDP2 with BL3-6 BS1) comprised of 96 well sitting drop plates of Natrix 1&2, Crystal Screen 1&2, and Index 1&2 crystal solution kits from Hampton Research; duplicates were made for 15 and 4 °C for a total of nearly 600 conditions. Initially 100 µl, then 80 µl well solution in later trays from a kit was placed in the main well, and up to the entire three sitting drops in one well were used, 0.5 µl complex solution combined with the same volume of well solution. Plates were sealed with Crystal Clear tape and left to incubate.

From the BL3-6 trays, two initial conditions were obtained, one at each temperature. After about one month, Index condition D3 at 4 °C (0.1 M HEPES pH 7, 30% v/v Jeffamine ED-2001 pH 7; no crystals at 15°) contained a crystal shower, resembling rectangles. At 15°C, Index F9 (0.2 M ammonium sulphate, 0.1 M Tris pH 8.5, 25% w/v polyethylene glycol 3350; 4°, very small rosettes/spherulites) showed needle clusters after about two and a half weeks. These two conditions were used as a starting point for variations in the percent of their respective precipitants as well as variations upon their pHs in 24 well hanging drop optimization trays with 500 µl well

solution. Components for these solutions were either ordered from Hampton Research (usually precipitants for which solvation is difficult) or made in the laboratory and sterile filtered with a luer syringe 0.22 micron filter. Once a course optimization was done (ex. pH increment 0.5 and ± several percent precipitant on either side of the initial condition) from which a condition was chosen, a fine optimization with smaller increments (pH increment 0.1 and smaller variations in precipitant) was done to further refine conditions. Hanging drops were 1µl complex: 1µl well solution usually, though 2:1 and 1:2 ratios were tried but did not produce better results. Due to sporadic reproducibility and long incubation times for crystals, Hampton Research's additive screen kit was used with one specific optimization condition for both F9 and D3, added according to manufacturer's instructions (10% of well solution) and set up in 96 well trays to incubate. This produced marked increases in speed (the D3 condition went from months to weeks) and reproducibility of crystal formation, and as previously noted, some of the D3 rectangular rods did change to rod clusters, though for the most part, the basic crystal morphology did not change with additives, and F9 conditions still contained some spherulites.

Initial (8-9Å) D3 optimized crystals were harvested, cryoprotected, and flash frozen in liquid nitrogen. Upon discovery of a large unit cell (~70% water), the subsequent batch (6-7Å at best, others same resolution as previous) was harvested, and some were treated similarly while others were incubated in a solution containing EtOH for 0.5-1 hours before flash freezing in an attempt to decrease water content, the efficacy of which is uncertain from diffraction data. The most recent, proliferous batches of crystals from the additive screens were harvested in three conditions: glycerol

cryoprotectant (0.075 M HEPES pH 7, 28.5% Jeffamine ED-2001, 19% glycerol, 50 mM $MgCl_2$, 15 mM glycine, 10% v/v additive), solution with EtOH (0.075 M HEPES pH 7, 29% Jeffamine ED-2001, 3% glycerol, 50 mM $MgCl_2$, 15 mM glycine, 10% v/v additive, 15% v/v EtOH) (dehydration inspired by the Group I ribozyme)(128), and solution with increased precipitant (0.075 M HEPES pH 7, 38% Jeffamine ED-2001, 50 mM $MgCl_2$, 15 mM glycine, 10% v/v additive) (dehydration inspired by the *glms* ribozyme)(129).

REFERENCES

1. Breaker RR (2011) Prospects for riboswitch discovery and analysis. *Mol Cell* 43(6):867-879.
2. Roth A & Breaker RR (2009) The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem* 78:305-334.
3. Cheah MT, Wachter A, Sudarsan N, & Breaker RR (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* 447(7143):497-500.
4. Wachter A, *et al*. (2007) Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell* 19(11):3437-3450.
5. Breaker RR (2010) Riboswitches and the RNA World. *Cold Spring Harb Perspect Biol*.
6. Mandal M, *et al.* (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* 306(5694):275-279.
7. Barrick JE, *et al.* (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci U S A* 101(17):6421-6426.
8. Fujiwara K, Okamura-Ikeda K, & Motokawa Y (1984) Mechanism of the glycine cleavage reaction. Further characterization of the intermediate attached to H-protein and of the reaction catalyzed by T-protein. *J Biol Chem* 259(17):10664-10668.
9. Birdsall DL, Finer-Moore J, & Stroud RM (1996) Entropy in bi-substrate enzymes: proposed role of an alternate site in chaperoning substrate into, and products out of, thymidylate synthase. *J Mol Biol* 255(3):522-535.
10. Pizer LI & Potochny ML (1964) Nutritional and Regulatory Aspects of Serine Metabolism in Escherichia Coli. *J Bacteriol* 88:611-619.

11. Ravnikar PD & Somerville RL (1987) Genetic characterization of a highly efficient alternate pathway of serine biosynthesis in Escherichia coli. *J Bacteriol* 169(6):2611-2617.

12. Tripp HJ, *et al.* (2009) Unique glycine-activated riboswitch linked to glycine-serine auxotrophy in SAR11. *Environ Microbiol* 11(1):230-238.

13. Kwon M & Strobel SA (2008) Chemical basis of glycine riboswitch cooperativity. *Rna* 14(1):25-34.

14. Erion TV & Strobel SA (2011) Identification of a tertiary interaction important for cooperative ligand binding by the glycine riboswitch. *Rna* 17(1):74-84.

15. Trausch JJ, Ceres P, Reyes FE, & Batey RT (2011) The structure of a tetrahydrofolate-sensing riboswitch reveals two ligand binding sites in a single aptamer. *Structure* 19(10):1413-1423.

16. Lipfert J, *et al.* (2007) Structural transitions and thermodynamics of a glycine-dependent riboswitch from Vibrio cholerae. *J Mol Biol* 365(5):1393-1406.

17. Lipfert J, Sim AY, Herschlag D, & Doniach S (2010) Dissecting electrostatic screening, specific ion binding, and ligand binding in an energetic model for glycine riboswitch folding. *Rna* 16(4):708-719.

18. Huang L, Serganov A, & Patel DJ (2010) Structural insights into ligand recognition by a sensing domain of the cooperative glycine riboswitch. *Mol Cell* 40(5):774-786.

19. Butler EB, Xiong Y, Wang J, & Strobel SA (2011) Structural basis of cooperative ligand binding by the glycine riboswitch. *Chem Biol* 18(3):293-298.

20. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13):3406-3415.

21. Gardner PP, *et al.* (2011) Rfam: Wikipedia, clans and the "decimal" release. *Nucleic Acids Res* 39(Database issue):D141-145.

22. Klein DJ, Schmeing TM, Moore PB, & Steitz TA (2001) The kink-turn: a new RNA secondary structure motif. *Embo J* 20(15):4214-4221.

23. Cruz JA & Westhof E (2011) Sequence-based identification of 3D structural modules in RNA with RMDetect. *Nat Methods* 8(6):513-521.

24. Goody TA, Melcher SE, Norman DG, & Lilley DM (2004) The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *Rna* 10(2):254-264.

25. Liu J & Lilley DM (2007) The role of specific 2'-hydroxyl groups in the stabilization of the folded conformation of kink-turn RNA. *Rna* 13(2):200-210.

26. Gordon L, Chervonenkis AY, Gammerman AJ, Shahmuradov IA, & Solovyev VV (2003) Sequence alignment kernel for recognition of promoter regions. *Bioinformatics* 19(15):1964-1971.

27. Regulski EE & Breaker RR (2008) In-line probing analysis of riboswitches. *Methods Mol Biol* 419:53-67.

28. Soukup GA & Breaker RR (1999) Relationship between internucleotide linkage geometry and the stability of RNA. *Rna* 5(10):1308-1325.

29. Sclavi B, Sullivan M, Chance MR, Brenowitz M, & Woodson SA (1998) RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* 279(5358):1940-1943.

30. Latham JA & Cech TR (1989) Defining the inside and outside of a catalytic RNA molecule. *Science* 245(4915):276-282.

31. Brenowitz M, Chance MR, Dhavan G, & Takamoto K (2002) Probing the structural dynamics of nucleic acids by quantitative time-resolved and equilibrium hydroxyl radical "footprinting". *Curr Opin Struct Biol* 12(5):648-653.

32. Wakeman CA & Winkler WC (2009) Structural probing techniques on natural aptamers. *Methods Mol Biol* 535:115-133.

33. Khvorova A, Lescoute A, Westhof E, & Jayasena SD (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat Struct Biol* 10(9):708-712.

34. Martick M & Scott WG (2006) Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* 126(2):309-320.

35. Cilley CD & Williamson JR (1999) PACE analysis of RNA-peptide interactions. *Methods Mol Biol* 118:129-141.

36. Montange RK & Batey RT (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* 441(7097):1172-1175.

37. Lee ER, Baker JL, Weinberg Z, Sudarsan N, & Breaker RR (2010) An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science* 329(5993):845-848.

38. Blouin S & Lafontaine DA (2007) A loop loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *Rna* 13(8):1256-1267.

39. Schroeder KT, Daldrop P, & Lilley DM (2011) RNA tertiary interactions in a riboswitch stabilize the structure of a kink turn. *Structure* 19(9):1233-1240.

40. Turner B, Melcher SE, Wilson TJ, Norman DG, & Lilley DM (2005) Induced fit of RNA on binding the L7Ae protein to the kink-turn motif. *Rna* 11(8):1192-1200.

41. Turner B & Lilley DM (2008) The importance of G.A hydrogen bonding in the metal ion- and protein-induced folding of a kink turn RNA. *J Mol Biol* 381(2):431-442.

42. Ravindran PP, Heroux A, & Ye JD (2011) Improvement of the crystallizability and expression of an RNA crystallization chaperone. *J Biochem* 150(5):535-543.

43. Wickiser JK, Winkler WC, Breaker RR, & Crothers DM (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell* 18(1):49-60.

44. Smith KD & Strobel SA (2011) Interactions of the c-di-GMP riboswitch with its second messenger ligand. *Biochem Soc Trans* 39(2):647-651.

45. Wang B, Kitney RI, Joly N, & Buck M (2011) Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nat Commun* 2:508.

46. Khalil AS & Collins JJ (2010) Synthetic biology: applications come of age. *Nat Rev Genet* 11(5):367-379.

47. Serganov A & Nudler E (2013) A decade of riboswitches. *Cell* 152(1-2):17-24.

48. Serganov A & Patel DJ (2012) Metabolite recognition principles and molecular mechanisms underlying riboswitch function. *Annu Rev Biophys* 41:343-370.

49. Bauer G & Suess B (2006) Engineered riboswitches as novel tools in molecular biology. *Journal of Biotechnology* 124(1):4-11.

50. Isaacs FJ, Dwyer DJ, & Collins JJ (2006) RNA synthetic biology. *Nat Biotechnol* 24(5):545-554.

51. Sherman EM, Esquiaqui J, Elsayed G, & Ye JD (2012) An energetically beneficial leader-linker interaction abolishes ligand-binding cooperativity in glycine riboswitches. *Rna* 18(3):496-507.

52. Kladwang W, Chou FC, & Das R (2012) Automated RNA structure prediction uncovers a kink-turn linker in double glycine riboswitches. *J Am Chem Soc* 134(3):1404-1407.

53. Baird NJ & Ferre-D'Amare AR (2013) Modulation of quaternary structure and enhancement of ligand binding by the K-turn of tandem glycine riboswitches. *Rna* 19(2):167-176.

54. Phan TT & Schumann W (2007) Development of a glycine-inducible expression system for Bacillus subtilis. *J Biotechnol* 128(3):486-499.

55. Prodromou C & Pearl LH (1992) Recursive PCR: a novel technique for total gene synthesis. *Protein Eng* 5(8):827-829.

56. Sherman EM, Holmes S, & Ye JD (2014) Specific RNA-Binding Antibodies with a Four-Amino-Acid Code. *J Mol Biol* 426(10):2145-2157.

57. Kibbe WA (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* 35(Web Server issue):W43-46.

58. Perutz MF (1989) Mechanisms of cooperativity and allosteric regulation in proteins. *Q Rev Biophys* 22(2):139-237.

59. Brunel C, Marquet R, Romby P, & Ehresmann C (2002) RNA loop-loop interactions as dynamic functional motifs. *Biochimie* 84(9):925-944.

60. Hermann T & Patel DJ (1999) Stitching together RNA tertiary architectures. *J Mol Biol* 294(4):829-849.

61. Pleij CW, Rietveld K, & Bosch L (1985) A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res* 13(5):1717-1731.

62. Puglisi JD, Wyatt JR, & Tinoco I, Jr. (1988) A pseudoknotted RNA oligonucleotide. *Nature* 331(6153):283-286.

63. Puglisi JD, Chen L, Blanchard S, & Frankel AD (1995) Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science* 270(5239):1200-1203.

64. Taft RJ, Pheasant M, & Mattick JS (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 29(3):288-299.

65. Eddy SR (2001) Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* 2(12):919-929.

66. Washietl S, *et al*. (2012) Computational analysis of noncoding RNAs. *Wiley Interdiscip Rev RNA* 3(6):759-778.

67. Chen K & Rajewsky N (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* 8(2):93-103.

68. Novikova IV, Hennelly SP, & Sanbonmatsu KY (2012) Sizing up long non-coding RNAs: Do lncRNAs have secondary and tertiary structure? *Bioarchitecture* 2(6).

69. Weinberg Z, Perreault J, Meyer MM, & Breaker RR (2009) Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature* 462(7273):656-659.

70. Mattick JS (2005) The functional genomics of noncoding RNA. *Science* 309(5740):1527-1528.

71. Qureshi IA, Mattick JS, & Mehler MF (2010) Long non-coding RNAs in nervous system function and disease. *Brain Res* 1338:20-35.

72. Ponting CP, Oliver PL, & Reik W (2009) Evolution and functions of long noncoding RNAs. *Cell* 136(4):629-641.

73. Wan Y, Kertesz M, Spitale RC, Segal E, & Chang HY (2011) Understanding the transcriptome through RNA structure. *Nat Rev Genet* 12(9):641-655.

74. Ye JD, *et al.* (2008) Synthetic antibodies for specific recognition and crystallization of structured RNA. *Proc Natl Acad Sci U S A* 105(1):82-87.

75. Koldobskaya Y, *et al.* (2011) A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination. *Nat Struct Mol Biol* 18(1):100-106.

76. Ke A & Doudna JA (2004) Crystallization of RNA and RNA-protein complexes. *Methods* 34(3):408-414.

77. Golden BL & Kundrot CE (2003) RNA crystallization. *J Struct Biol* 142(1):98-107.

78. Tsai MC, Spitale RC, & Chang HY (2011) Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res* 71(1):3-7.

79. Esteller M (2011) Non-coding RNAs in human disease. *Nat Rev Genet* 12(12):861-874.

80. Braconi C, *et al.* (2011) Expression and functional role of a transcribed noncoding RNA with an ultraconserved element in hepatocellular carcinoma. *Proc Natl Acad Sci U S A* 108(2):786-791.

81. Calin GA, *et al.* (2007) Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* 12(3):215-229.

82. Lujambio A, *et al.* (2010) CpG island hypermethylation-associated silencing of non-coding RNAs transcribed from ultraconserved regions in human cancer. *Oncogene* 29(48):6390-6401.

83. Faghihi MA, *et al.* (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med* 14(7):723-730.

84. Martinez F, *et al.* (2010) Enrichment of ultraconserved elements among genomic imbalances causing mental delay and congenital anomalies. *Bmc Med Genomics* 3.

85. Horsthemke B & Wagstaff J (2008) Mechanisms of imprinting of the Prader-Willi/Angelman region. *Am J Med Genet A* 146A(16):2041-2052.

86. Giepmans BN, Adams SR, Ellisman MH, & Tsien RY (2006) The fluorescent toolbox for assessing protein location and function. *Science* 312(5771):217-224.

87. Fellouse FA, *et al.* (2007) High-Throughput Generation of Synthetic Antibodies from Highly Functional Minimalist Phage-Displayed Libraries. *J. Mol. Biol.* 373(4):924-940.

88. Miller KR, *et al.* (2012) T cell receptor-like recognition of tumor in vivo by synthetic antibody fragment. *PLoS One* 7(8):e43746.

89. Miersch S & Sidhu SS (2012) Synthetic antibodies: concepts, potential and practical considerations. *Methods* 57(4):486-498.

90. Fellouse FA, Wiesmann C, & Sidhu SS (2004) Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc Natl Acad Sci U S A* 101(34):12467-12472.

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

92. Sidhu SS & Kossiakoff AA (2007) Exploring and designing protein function with restricted diversity. *Curr Opin Chem Biol* 11(3):347-354.

93. Kabat EA & Wu TT (1971) Attempts to locate complementarity-determining residues in the variable positions of light and heavy chains. *Ann N Y Acad Sci* 190:382-393.

94. Sidhu SS, *et al.* (2004) Phage-displayed antibody libraries of synthetic heavy chain complementarity determining regions. *J Mol Biol* 338(2):299-310.

95. Agris PF, *et al.* (1999) 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(4):425-435.

96. Draper DE (1999) Themes in RNA-protein recognition. *J Mol Biol* 293(2):255-270.

97. Kunkel TA (1985) Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection. *P Natl Acad Sci USA* 82(2):488-492.

98. Sidhu SS, Lowman HB, Cunningham BC, & Wells JA (2000) Phage display for selection of novel binding peptides. *Methods Enzymol* 328:333-363.

99. Ye J-D, Archer J, & Sherman EM (2013) 61/783,906.

100. Juneau K, Podell E, Harrington DJ, & Cech TR (2001) Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of RNA--solvent interactions. *Structure* 9(3):221-231.

101. Treger M & Westhof E (2001) Statistical analysis of atomic contacts at RNA-protein interfaces. *J Mol Recognit* 14(4):199-214.

102. Jeong E, Kim H, Lee SW, & Han K (2003) Discovering the interaction propensities of amino acids and nucleotides from protein-RNA complexes. *Mol Cells* 16(2):161-167.

103. Birtalan S, *et al.* (2008) The intrinsic contributions of tyrosine, serine, glycine and arginine to the affinity and specificity of antibodies. *J Mol Biol* 377(5):1518-1528.

104. Wang JX, Lee ER, Morales DR, Lim J, & Breaker RR (2008) Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. *Mol Cell* 29(6):691-702.

105. Lemay JF, *et al.* (2011) Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS Genet* 7(1):e1001278.

106. Fedorova O & Pyle AM (2008) A conserved element that stabilizes the group II intron active site. *Rna* 14(6):1048-1056.

107. Sabin LR, Delas MJ, & Hannon GJ (2013) Dogma derailed: the many influences of RNA on the genome. *Mol Cell* 49(5):783-794.

108. Serganov A & Patel DJ (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet* 8(10):776-790.

109. Pavon-Eternod M, *et al.* (2009) tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Res* 37(21):7268-7280.

110. Gutschner T & Diederichs S (2012) The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol* 9(6):703-719.

111. Nakatani J, *et al.* (2009) Abnormal behavior in a chromosome-engineered mouse model for human 15q11-13 duplication seen in autism. *Cell* 137(7):1235-1246.

112. Ziats MN & Rennert OM (2013) Aberrant expression of long noncoding RNAs in autistic brain. *Journal of molecular neuroscience : MN* 49(3):589-593.

113. Powell WT, *et al.* (2013) A Prader-Willi locus lncRNA cloud modulates diurnal genes and energy expenditure. *Hum Mol Genet*.

114. Sibbritt T, Patel HR, & Preiss T (2013) Mapping and significance of the mRNA methylome. *Wiley Interdiscip Rev RNA* 4(4):397-422.

115. Hulsmans M & Holvoet P (2013) MicroRNAs as Early Biomarkers in Obesity and Related Metabolic and Cardiovascular Diseases. *Current pharmaceutical design*.

116. Tan L, Yu JT, Hu N, & Tan L (2013) Non-coding RNAs in Alzheimer's disease. *Molecular neurobiology* 47(1):382-393.

117. Schonrock N & Gotz J (2012) Decoding the non-coding RNAs in Alzheimer's disease. *Cellular and molecular life sciences : CMLS* 69(21):3543-3559.

118. Wei Y, Nazari-Jahantigh M, Neth P, Weber C, & Schober A (2013) MicroRNA-126, -145, and -155: a therapeutic triad in atherosclerosis? *Arteriosclerosis, thrombosis, and vascular biology* 33(3):449-454.

119. Flowers E, Froelicher ES, & Aouizerat BE (2013) MicroRNA regulation of lipid metabolism. *Metabolism: clinical and experimental* 62(1):12-20.

120. Ferre-D'Amare AR & Doudna JA (2000) Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J Mol Biol* 295(3):541-556.

121. Keel AY, Rambo RP, Batey RT, & Kieft JS (2007) A general strategy to solve the phase problem in RNA crystallography. *Structure* 15(7):761-772.

122. Blount KF & Breaker RR (2006) Riboswitches as antibacterial drug targets. *Nat Biotechnol* 24(12):1558-1564.

123. Furukawa K, *et al.* (2012) Identification of ligand analogues that control c-di-GMP riboswitches. *ACS Chem Biol* 7(8):1436-1443.

124. Shanahan CA & Strobel SA (2012) The bacterial second messenger c-di-GMP: probing interactions with protein and RNA binding partners using cyclic dinucleotide analogs. *Organic & biomolecular chemistry* 10(46):9113-9129.

125. Mulhbacher J, *et al.* (2010) Novel riboswitch ligand analogs as selective inhibitors of guanine-related metabolic pathways. *PLoS pathogens* 6(4):e1000865.

126. Kim JN, *et al.* (2009) Design and antimicrobial action of purine analogues that bind Guanine riboswitches. *ACS Chem Biol* 4(11):915-927.

127. Ster C, *et al.* (2013) Experimental treatment of Staphylococcus aureus bovine intramammary infection using a guanine riboswitch ligand analog. *Journal of dairy science* 96(2):1000-1008.

128. Guo F, Gooding AR, & Cech TR (2004) Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell* 16(3):351-362.

129. Klein DJ & Ferre-D'Amare AR (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* 313(5794):1752-1756.

130. Adams PL, *et al.* (2004) Crystal structure of a group I intron splicing intermediate. *RNA* 10(12):1867-1887.

131. Golden BL, Kim H, & Chase E (2005) Crystal structure of a phage Twort group I ribozyme-product complex. *Nat Struct Mol Biol* 12(1):82-89.

132. Adams PL, Stahley MR, Kosek AB, Wang J, & Strobel SA (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature* 430(6995):45-50.