AMERICAN UNIVERSITY OF BEIRUT

LOCAL EVOLUTIONARY LANDSCAPES OF λ N-BOXB AND HK022 NUN-BOXB INTERACTIONS

by
NINA EMAD KHALDIEH

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Science
to the Department of Biology
of the Faculty of Arts and Sciences
at the American University of Beirut

Beirut, Lebanon
January 2019

AMERICAN UNIVERSITY OF BEIRUT

LOCAL EVOLUTIONARY LANDSCAPES OF λ N-BOXB AND HK022 NUN-BOXB INTERACTIONS

by

NINA EMAD KHALDIEH

Approved by:

Colin Smith

Dr. Colin Smith, Professor — Advisor
Biology

Dr. Heinrich Dohna, Assistant Professor — Member of Committee
Biology

Dr. Zakaria Kambris, Associate Professor — Member of Committee
Biology

Date of thesis defense: January 25, 2019

AMERICAN UNIVERSITY OF BEIRUT

THESIS RELEASE FORM

Student Name:

Khaldieh Nina Emad

(✓) Master's Thesis ( ) Master's Project ( ) Doctoral Dissertation

[ ] I authorize the American University of Beirut to: (a) reproduce hard or electronic copies of my thesis, dissertation, or project; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

[X] I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of it; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes
after:

One ---- year from the date of submission of my thesis, dissertation, or project.

Two ---- years from the date of submission of my thesis, dissertation, or project.

Three -X- years from the date of submission of my thesis, dissertation, or project.

Nina Khaldieh 5 Feb, 2019

Signature Date

ACKNOWLEDGMENTS

I would like to express my deep gratitude to my advisor Colin Smith for all his guidance, advice, and useful critiques. I am also very fortunate to have met and worked with Ingrid Ghattas. I am truly grateful for all her encouragement and emotional support. I would like to thank my thesis committee members: Zakaria Kambris and Heinrich Dohna for their insightful questions and comments.

AN ABSTRACT OF THE THESIS OF

Nina Emad Khaldieh for Master of Science
Major: Biology

Title: Local Evolutionary Landscapes of λ N-boxB and HK022 Nun-boxB Interactions

Expression of λ delayed-early genes requires overcoming transcription terminators through an antitermination process initiated by an arginine-rich motif (ARM) binding a regulatory RNA in a λ N ARM-boxB interaction. HK022 Nun protein competes with N to bind boxB and induces premature λ transcription termination. Although structural data support very similar N ARM-boxB and Nun ARM-boxB interactions, recently published mutagenesis data suggests Nun-NusG contacts through non-conserved but required Asp26, Trp33, and Arg36 residues embedded in a functionally important exposed hydrophobic ridge. This prompted a further examination of the interactions to elucidate their differences. In this study, mutational analyses and specificity assessments of boxBs with expanded and unexpanded loops shows Nun recruits boxB variants that N cannot, affirming that their binding modes are distinct. Scanning mutagenesis of N ARM revealed N does not have a restriction on Thr5, Ala12, and Gln15 at positions equivalent to the restriction on Asp26, Trp33, and Arg36 of Nun. Complex ARM hybrids having a specific combination of N and Nun residues characterized Nun's recognition strategy as complex, involving a combination of multiple requirements, and likely not modular. In contrast, N's recognition strategy is simple, more robust (resistant to changes in its amino acid sequence), and with no equivalent dependence on NusG-CTD contacts. The results elucidate the difference between N and Nun recognition strategies, allow speculations on NusG localization within the termination and the antitermination complexes, and address the possible genetic drift of N and Nun binding modes, as well as boxB sequence, along neutral networks.

CONTENTS

ACKNOWLEDGMENTS..... v

ABSTRACT..... vi

LIST OF ILLUSTRATIONS..... x

LIST OF TABLES..... xi

LIST OF ABBREVIATIONS..... xii

Chapter

I. INTRODUCTION 1

A. Overview..... 1

B. RNA-Protein Interactions 1

1. Arginine-Rich Motifs..... 3
2. RNA Structure 6
3. GNRA-Tetraloop Motif 7
4. RNA-Peptide Adaptive Recognition 9

C. Neutral Evolution 10

D. Bacteriophage Transcription Antitermination 12

1. λN ARM-boxB Interaction 16

E. Bacteriophage Nun Termination 18

F. The Experimental Approach and Aims of This Study 20

II. MATERIALS AND METHODS..... 24

A. Constructing λ boxB expanded loop plasmid library (LL4NL)..... 24

B. Chemically Competent E Coli Cells 25

1. Preparation 25

2. Transformation …… 26

C. Screening for N-Antitermination with X-Gal Plate Assays …… 26

D. Screening for Nun-Termination with XP Plate Assay…… 28

E. DNA Extraction …… 29

F. DNA Column Purification and Sequencing …… 30

G. β-Galactosidase N-Antitermination Solution Assays …… 30

H. Alkaline Phosphatase Solution Assay for Nun-Termination …… 32

I. N-deficient λ complementation assay …… 32

J. Wild type λ exclusion assay …… 33

H. Screening Libraries of N 2-18 ARM …… 34

III. RESULTS…… 35

A. Nun ARM, unlike N ARM, recognizes non-GNRA-like pentaloops … 36

B. Nun's relaxed requirements for boxB stem sequence contribute to boxB Nun-specificity …… 40

C. BoxB loop expansion is a mechanism to evolve specificity to Nun but not to N …… 44

D. Comparing the mutability of N and Nun ARMs …… 47

E.λ N recognition strategy cannot be recapitulated in the termination context …… 50

F.Gradual mutation of λ N ARM towards Nun ARM …… 54

IV. DISCUSSION …… 61

A. The Difference in N and Nun contacts with NusG influence the ARM-boxB interaction …… 61

B. The difference in tolerance to boxB-loop expansion reflects differences in spatial requirements …… 62

C. Neutral Evolution of boxB specificity …… 64

D. N ARM sequence is very robust and doesn't have an equivalent to Nun's DWR requirements …… 65

E. The neutrality of Trp18 and Tyr39 interchangeability depends on the sequence context …… 66

F. Nun ARM function is not modular …… 67

G. Speculations on NusG-CTD within the Termination and the Antitermination Complexes …… 68

H. The neutral networks of N and Nun recognition strategies are not proximate in sequence space in the termination context …… 69

I. Further experiments and unresolved questions …… 70

Appendix

I. CONSTRUCT SEQUENCES…… 72

II. PLATE ASSAY RESULTS…… 74

III. SOLUTION ASSAY RESULTS…… 76

REFERENCES

ILLUSTRATIONS

Figure		Page
1.	The diversity of recognition strategies of ARM-RNA interactions	5
2.	Diagram of common RNA secondary structural elements	6
3.	GNRA tetraloop fold and GNRA-like pentaloop fold	8
4.	Diagram of phenotypic neutral networks in genotypic sequence space	12
5.	Genetic map of phage lambda and transcription Nun-Termination/N-Antitermination Complex	14
6.	NMR structural models of lambdoid bacteriophage N ARMs binding the major grooves of their cognate boxB hairpins	15
7.	Comparison of NMR structural models and sequences of λN ARM-boxB and Nun ARM-boxB	17
8.	Rigid-body docking of NusG-CTD and Nun ARM-boxB NMR structural models	20
9.	Two-plasmid reporter system	23
10.	Secondary structures and activities of loop-mutant boxB sequences	39
11.	Secondary structures and activities of stem-mutant boxB sequences	41
12.	Secondary structures and activities of boxB pentaloops fused to P22 boxB right stem	43
13.	Secondary structures and activities of expanded-loop boxB sequences	46
14.	Scanning mutagenesis profiles of N and Nun ARMs	48
15.	λ N and HK022 Nun alignment and activity	54
16.	Distribution of boxB specificities	62

TABLES

Table		Page
1.	λ N complementation and wt λ exclusion functions of Nun-N hybrids	52
2.	Activity of LN ARM gradually mutating towards Nun ARM	56
3.	Alanine mutagenesis of MDRG sequence	58
4.	Activity of cNNun ARM incrementally mutating towards Nun ARM	60

ABBREVIATIONS

%	Percent
µl	Microliter
A	Adenine or Alanine
BIV	Bovine immunodeficiency virus
C	Cytosine or Cysteine
D	Aspartate
DNA	Deoxyribonucleic acid
E	Glutamate
Et al.	Et allii (and others)
F	Phenylalanine
G	Guanine, Guanosine, or Glycine
GNRA	G is Guanine; N is any base (Guanine, Uracil, Adenine, or Cytosine); R is purine (Guanine or Adenine); A is Adenine
HIV	Human immunodeficiency virus
I	Isoleucine
IPTG	Isopropyl β-D-thiogalactoside
K	Lysine
L	Leucine
LacZ	β-galactosidase gene
mg	Milligram
min	Minutes
ml	Milliliter
mM	Millimolar
N	Asparagine or any base
NMR	Nuclear magnetic resonance
nut	N-utilization
OD	Optical Density
ONPG	Ortho-Nitrophenyl Galactoside
PDB	Protein database
PNPP	Para-nitrophenyl phosphate
Q	Glutamine
R	Arginine or purine
RNA	Ribonucleic acid
RNAP	RNA polymerase
rpm	Revolution per minute
RRE	Rev-response element
S	Serine
SDS	Sodium dodecyl sulfate
T	Thymine or Threonine
W	Tryptophan
λ	Bacteriophage lambda
ϕ21	Bacteriophage phi21

CHAPTER I

INTRODUCTION

A. Overview

The general aim of this thesis is to explore the evolution of ARM-RNA recognition in the context of λ N-boxB and HK022 Nun-boxB interactions. Can different recognition strategies and specificities evolve by genetic drift? λ N and HK022 Nun proteins interact with RNAP and an identical set of host-encoded factors, yet they mediate contrary outcomes by nucleating the assembly of transcription antitermination and termination complexes respectively. The assembly of both complexes is initiated by the recognition of boxB RNA sequence by N and Nun ARMs. The separate NMR structural models of λ N ARM- and HK022 Nun ARM-boxB complexes suggested N ARM and Nun ARM recognition strategies are indistinguishable. Unanticipated by published structural and mutagenesis data, Tawk et al. (2015) observed immutability of non-conserved residues in Nun ARM (D26, W33, and R36) that are part of its exposed hydrophobic ridge and had been proposed to contact NusG-CTD (Faber et al., 2001; Tawk et al., 2015). Puzzlingly, this holds true in the termination and the antitermination contexts, which suggests that NusG-CTD binding is required not just for Nun termination, but for Nun ARM interaction with boxB. Thus, it remains unclear to what extent Nun and N recognition strategies are alike or distinct.

B. RNA-Protein Interactions

1. Arginine-Rich Motifs

RNA-protein interactions underlie many essential cellular and viral structural and regulatory processes. Various classes of conserved RNA-binding motifs have been identified (Harrison, 1991). One widely found class is the arginine-rich motif (ARM). ARMs are found mediating essential molecular processes in transcription, translation, and RNA processing (Weiss and Narayana, 1998). Such processes are highly specific and are governed by a strict spatial-temporal regulation. Although considered conserved motifs, these short sequences (10-20 residues long) are classed by the only one feature: a preponderance of arginine residues. The long and flexible arginine side chains contribute to ARMs' conformational plasticity, their guanidinium moeities can engage in as many as five hydrogen bonds each, and their positive charges neutralize backbone phosphate anions.

ARMs are an attractive model to study peptide-RNA recognition. They bind small RNA structures with high affinity and specificity, and are frequently found to behave as independent functional units within the larger protein they are embedded in (Smith et al., 2000b). Besides the advantage of their small size which allows a detailed molecular dissection of their interaction, they have been found to bind their RNA targets with a wide diversity of conformations (Figure 1). Bacteriophage λN and HIV-1 Rev peptides bind as bent and continuous α-helices respectively (Tan et al. 1993; Battiste et al. 1996; Legault et al. 1998), BIV Tat peptide binds as a β- hairpin (Chen and Frankel 1995; Puglisi et al. 1995), and HIV Tat peptide binds as an extended conformation (Calnan et al. 1991; Aboul-ela et al. 1995).

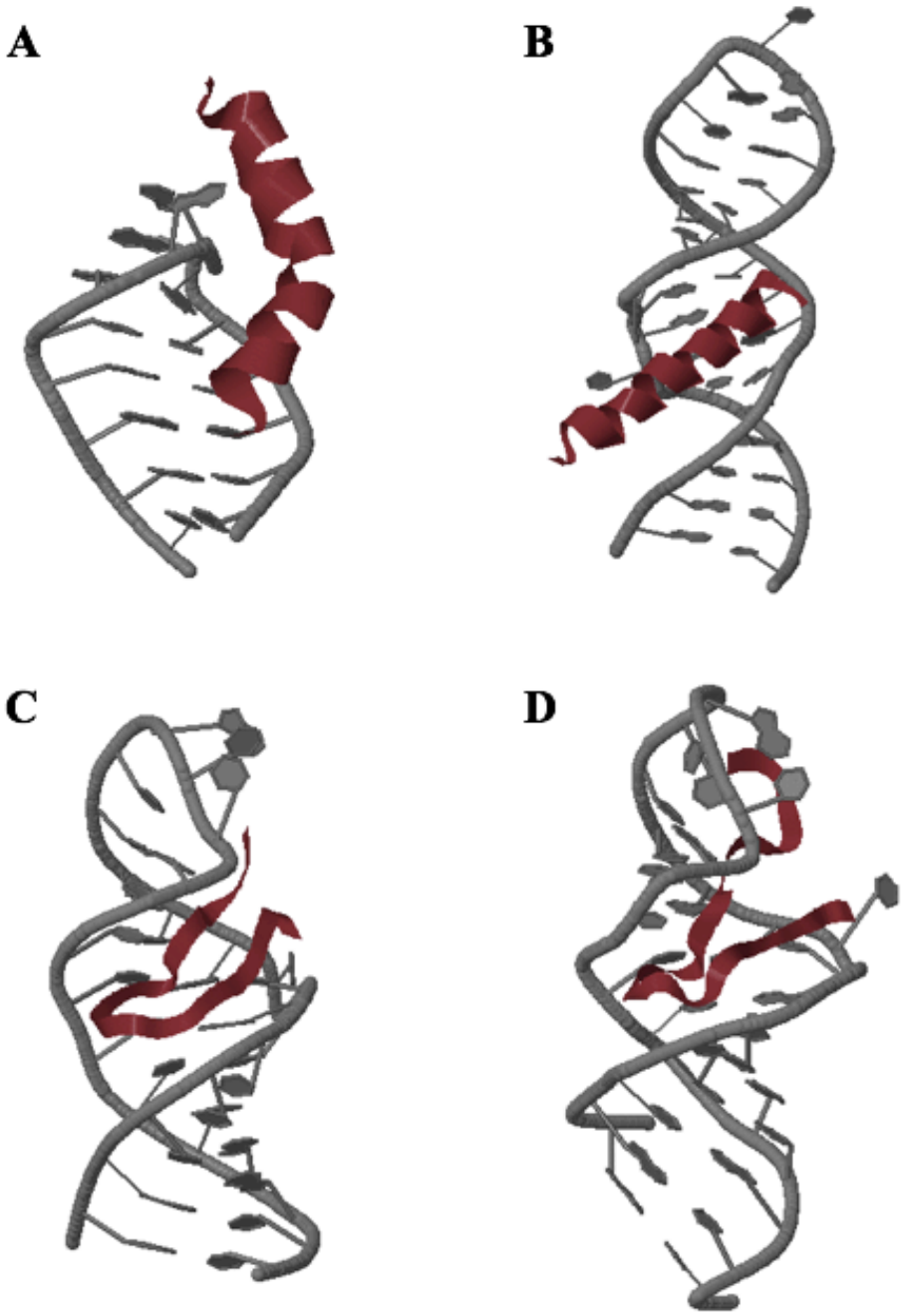


Figure 1. The diversity of recognition strategies of ARM-RNA interactions. (A) λN ARM-boxB (Scharpf et al., 2000) (PDB 1QFQ). (B) HIV-1 Rev-RRE IIB (Battiste et al., 1996) (PDB 1ETF). (C) BIV Tat-TAR (Puglisi et al., 1995) (PDB 1MNB). (D) JDV Tat-BIV TAR (Calabro et al., 2005) (PDB 1ZBN).

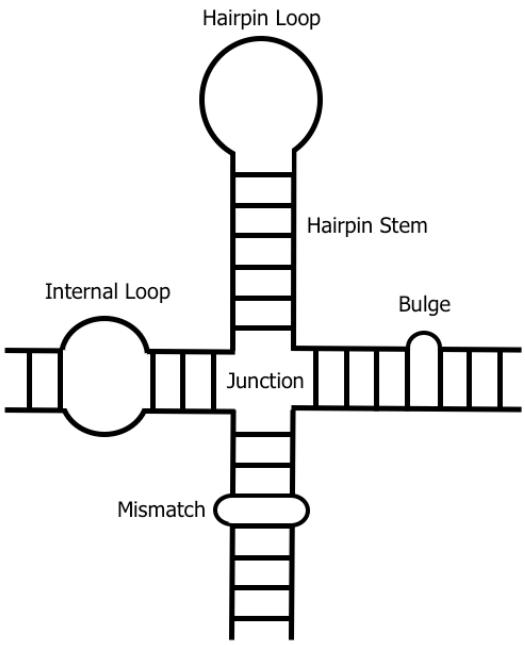


Figure 2. Diagram of Common RNA Secondary Structural Elements.

2. RNA Structure

RNA molecules reveal a rich repertoire of conserved secondary structural elements and motifs that matches their functional diversity. Canonical RNA base pairs commonly stack into regular A-form double helices (Draper, 1999; Hermann *et. al.*, 1999; Strobel *et. al.*, 2007; Weiss *et. al.*, 1998). The surface of the shallow minor groove is wide and easily accessible for peptide secondary structures, but the deep major groove is narrow, which limits its potential as a peptide-binding site. This limitation is overcome by non-canonical base pairs: the RNA helical flexibility accommodates mismatches that widen the groove enough for peptides to penetrate (Draper, 1999). Base triplets, hairpins, internal loops, bulges (multi-base and single-base), junctions (three-way and four-way) are other structural elements (Figure 2) that

interrupt the A-form geometry of RNA helices allowing specific recognition by peptides (Wyatt *et. al.*, 1989; Varani, 1997; Leulliot *et al.,* 2001; Hermann *et al.,*1999).

3. GNRA-tetraloop motif

Among known RNA structural motifs, the GNRA-tetraloop motif is of particular interest to this thesis (Figure 3A). GNRA is the consensus sequence of tetraloop-motifs very commonly found capping RNA W-C helices and less frequently next to non-canonical motifs. By inverting the direction of the phosphodiester backbone, these motifs hairpin structures. One prominent feature is their ability to engage with peptide and RNA tertiary contacts to mediate the assembly of ribonucleoprotein complexes (Correll and Swinger, 2003). The thermodynamic stability of a GNRA tetraloop is imparted by intramolecular interactions: π-π stacking of the aromatic rings of N, R, and A, hydrogen bonding between G and R, non-canonical base pairing of G and A (Heus & Pardi,1991), and hydrogen bonding of the amino proton of the former with the phosphate oxygen of the latter (Jucker et al., 1995). The GNRA motif has a great potential for specific recognition by proteins and RNAs (Wool et al., 1992; Doherty et al., 2001; Nissen et al., 2001). The different functional groups of N, R, and A which are accessible in the minor groove, and the great variation in their orientation among GNRA sequence variants, present opportunities for specific recognition. The sheared geometry of the GA base pair presenting intermolecular hydrogen bonding surfaces allows specific tertiary contacts as well (Hermann *et. al.*, 1999). This common structural motif is not limited to tetraloops only. GNRA-like folds were found embedded in larger loops (Huang et al., 2005). The NMR structures of N-boxB interactions of λ and P22 reveal pentaloops adopting the GNRA fold by extruding

the fourth base of λ (Figure 3B) and the third base of P22 (Legault et al., 1998; Cai et al., 1998).

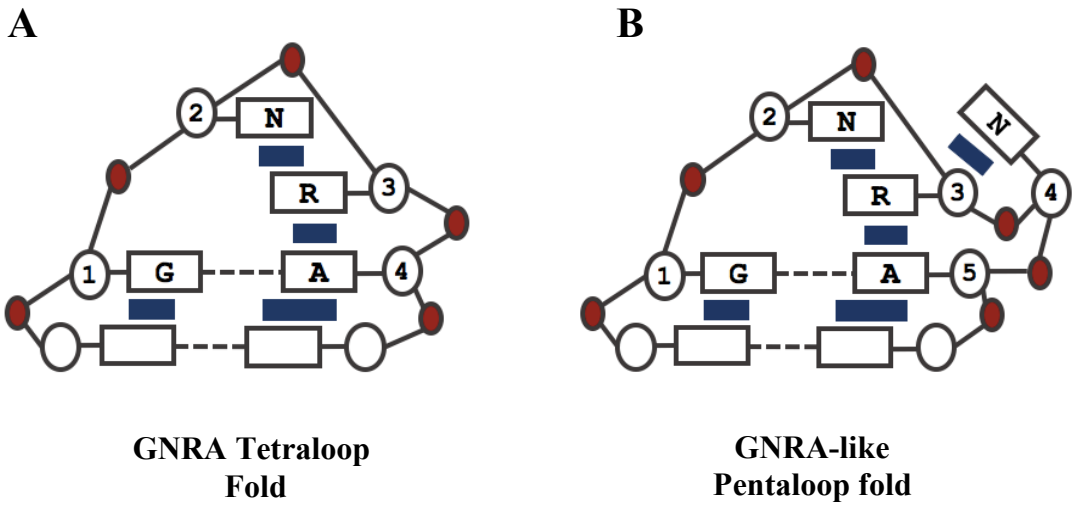


Figure 3. (A) Diagram showing GNRA tetraloop motif (Heus and Pardi, 1991). (B) Diagram showing λ boxB pentaloop motif (Legault et al., 1998). Important structural features are shown, and the base requirements of the folds are summarized. In (A) and (B) the white boxes, open circles, and red circles represent bases, ribose groups, phosphate moieties, respectively. Stacking and base-pairing interactions are represented by dark boxes and dotted lines respectively. The figure has been redrawn from Legault et al., 1998.

4. RNA-peptide Adaptive Recognition

ARM-RNA recognition is mediated by an extensive set of interactions: salt bridges, aromatic π-π stacking, hydrogen bonding, Van der Waal interactions. Studies of RNA-peptide recognition strategies were based on model ARM-small RNA interactions derived from larger viral regulatory systems. Well established models include bacteriophage N ARM interaction with hairpin boxB structure in regulating transcription of delayed early viral genes, HIV Rev ARM interaction with RRE for regulating nuclear export of viral transcripts, and the BIV Tat-TAR interaction. Structural insights derived from the study of these models are presumably generalizable to other peptide-RNA interactions across many biological contexts.

An important theme among studied ARM-RNA interactions is induced fit binding. Conformational changes, adjustments to more thermodynamically favorable conformations, take place in one or both binding partners. Such adaptive binding interactions range from few minor changes in bond angles to large-scale changes involving most bonds. This adaptive remodeling of one or both binding partners could be important for creating recognition surfaces for further binding of components within multimeric complexes. For instance, the free λ N ARM is totally disordered in solution (Mogridge et al. 1998; Van Gilst et al., 1997). Folding into a stable bent alpha helical structure occurs only upon binding to the RNA boxB hairpin. The latter also becomes more structured upon binding and undergoes further stabilization of its sheared GA base pair (Su et al., 1997b; Mogridge et al., 1998). This ARM-induced folding of boxB creates an important recognition site for further host factor binding. The RNA-binding domain of HIV Rev peptide that binds its cognate RNA target in an α-helical conformation but binds a different RNA aptamer (selected in vitro) in an extended

conformation (Ye et al., 1999). This emphasizes the role of RNAs as structural scaffolds that dictate peptide folding and the high structural flexibility and adaptability of arginine-rich peptides to their RNA contexts. A very different example of adaptive binding was observed with the bacteriophage MS2 coat protein-RNA interaction. In this case, the RNA component undergoes remodeling while the protein retains its conformation as revealed in the crystal structures of three different complexes between the protein and three RNA aptamers (Valegard et al., 1997; Convery et al., 1998; Rowsell and Stockley, 1998)

E. Neutral Evolution

Refinements of Kimura's neutral theory of molecular evolution (Kimura, 1968,1991) such as the near-neutral theory (Ohta, 2002) and neutral networks theories (Huynen, 1996; Fay, 2002) assert that there are more genotypes than phenotypes by many orders of magnitude. As a consequence, genotypes mapping to the phenotype form neutral networks that permeate large regions of sequence space. These networks are so extensive that it is possible to transit from one phenotype to another through single-step mutations without loss-of-fitness intermediates (Figure 4). Computational studies testing neutral theory predictions have employed RNA molecules due to their simple genotype-phenotype relationships and the computationally predictable secondary structures. The revealed neutral networks and robustness of RNA secondary structures elucidates how continuous gradual genotypic changes can underlie discontinuous changes in phenotype (Fontana et al., 1992; Grüner et al., 1996; Schuster et al., 1994). Neutral networks have also been discovered for protein structures as well (Babajide et al., 1997). Conservation of RNA and protein structure as sequences change supports

evolution along neutral networks (Gutell et al., 1993; Huynen et al., 1998). One particularly interesting study (Schultes and Bartel., 2000) examined the proximity of ribozyme neutral networks. Two evolutionary unrelated ribozymes, each catalyzing a different reaction, were found to be connected by a continuous path of neutral point mutations without loss of function. At the junction of the two neutral networks, the intersection RNA sequence is capable to adopt both ribozyme structures and was bifunctional.

The remarkable diversity of recognition strategies of RNA-ARM interactions raises the question of what potential mechanisms underlie the evolution of these distinct binding modes. Retroviral and lambdoid phage ARM-boxB interactions are convenient models to study the evolution of binding specificities and recognition strategies of RNA-peptide interactions. These small interactions are amenable to detailed investigation of their fitness landscape topologies within meaningful biological contexts. The evolution of their remarkable diversity has been successfully addressed within the framework of neutral network theories (Smith et al. 1998; Smith et al. 2000, Iwazaki et al. 2005; Sugaya et al. 2008; Possik et al. 2013; Abdallah & Smith, 2015). The arginine-rich RNA-binding domains of Jembrana disease virus JDV Tat protein recognizes two different viral RNA sites by employing different sets of amino acids and adopts very different conformations upon binding in each RNA context (Smith et al., 2000a). When JDV Tat ARM binds HIV's TAR sequence, it folds into a beta-hairpin, whereas when it binds to BIV's TAR, it adopts an extended conformation. Most remarkably, the JDV Tat requires contacts with cyclin T1 protein to bind with high affinity to on RNA site but not to the other. This is an example of a "chameleon" protein sequence that exists at the intersection of the neutral networks of two very

distinct recognition strategies and can thus engage in both.

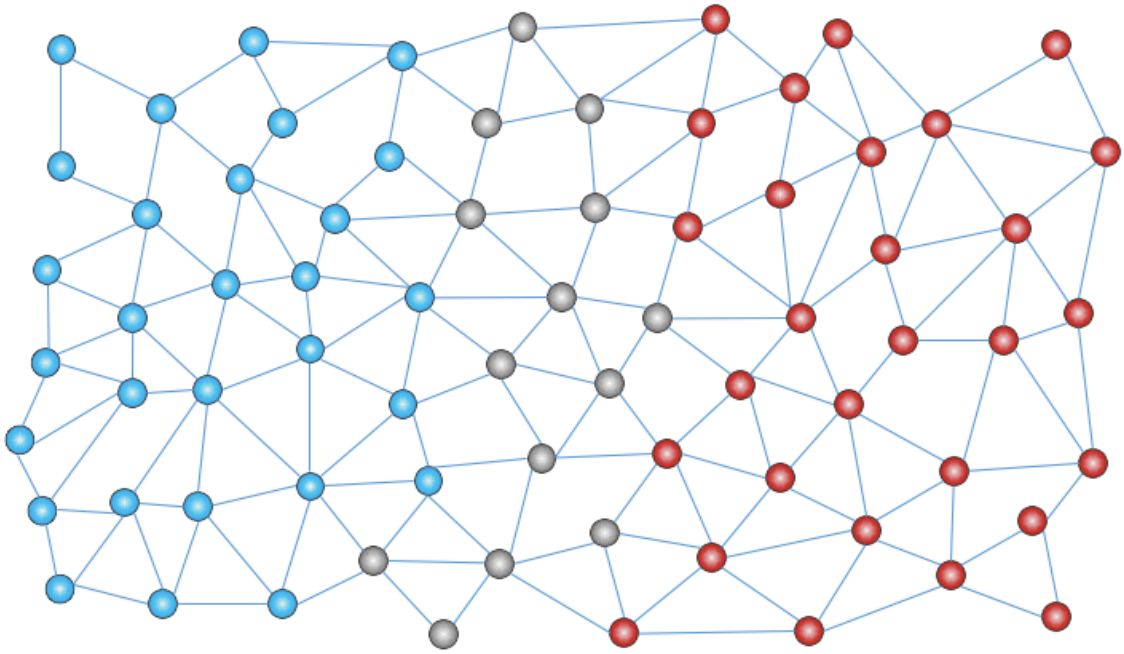


Figure 4. Diagram of Phenotypic Neutral Networks in Genotypic Sequence Space. Each circle is a genotype. Genotypes separated by one mutational step are connected by lines. Molecular phenotypes (structure, function, recognition strategies) are represented by the different colors. Red and blue inter-connected circles represent the neutral networks of two different molecular phenotypes. At the intersection of two neutral networks are genotypes that manifest both phenotypes (grey circles).

C. Bacteriophage Transcription Antitermination.

After bacteriophage λ infects its host, *Escherichia coli*, the bacterial RNA polymerase (RNAP) initiates RNA synthesis at the two λ early promotors pL and pR of the left and right early operons producing the immediate-early transcripts (Figure 5A). Subsequent expression of delayed early genes requires overcoming intrinsic transcription termination signals through a transcription antitermination process that allows RNAP to read through a series of Rho-dependent and Rho-independent

terminators (Gottesman et al., 1980; Weisberg and Gottesman, 1999). The antitermination process is initiated by binding of N protein, an immediate early gene product, to boxB RNA sequence in the nascent transcript of the *nut* site (N-utilization site). In addition to boxB, the *nut* site includes boxA sequence separated by a linker (Olson et al., 1984; Salstrom and Szybalski, 1978). The N-boxB complex nucleates the assembly of host factors NusA, NusG, NusE, NusB and RNAP into a termination-resistant transcription complex (Figure 5B).

This processive antitermination, which is necessary for λ lytic development (Friedman, 1966; Roberts, 1966), is also shared by the lambdoid phages P22 and φ21 where the arginine-rich N protein of each phage binds with high affinity and specificity its cognate boxB sequence (Figure 6) (Austin *et. al.*, 2003). In these phages, N antitermination is type specific: N protein of one phage does not complement an N-deficient strain of another phage (Dambly & Couturier, 1971; Hilliker & Botstein, 1976), and noncognate N ARM-boxB interactions are of low affinity *in vitro* (Cilley & Williamson, 1997; Austin et al., 2003). Interestingly, this high specificity in recognition can become relaxed with only few single mutations in boxB or ARM sequences (Cocozaki et al., 2008a, 2008b).

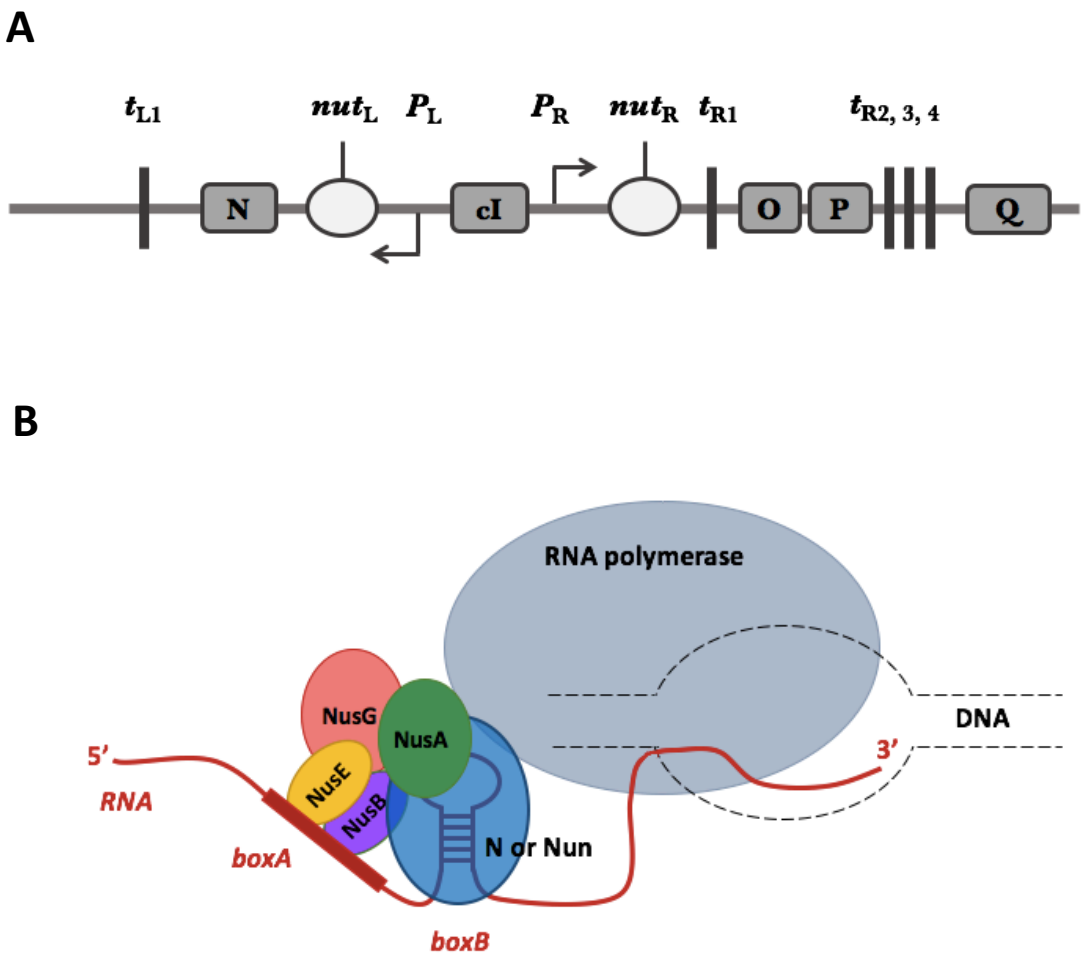


Figure 5. (A) Schematic diagram of the genetic map of phage lambda. (B) Transcription Nun-Termination/N-Antitermination Complex. Binding of the N/Nun ARM to the hairpin structure in the nascent immediate-early transcripts of the left and right operons regulates the switch to delayed-early gene expression.

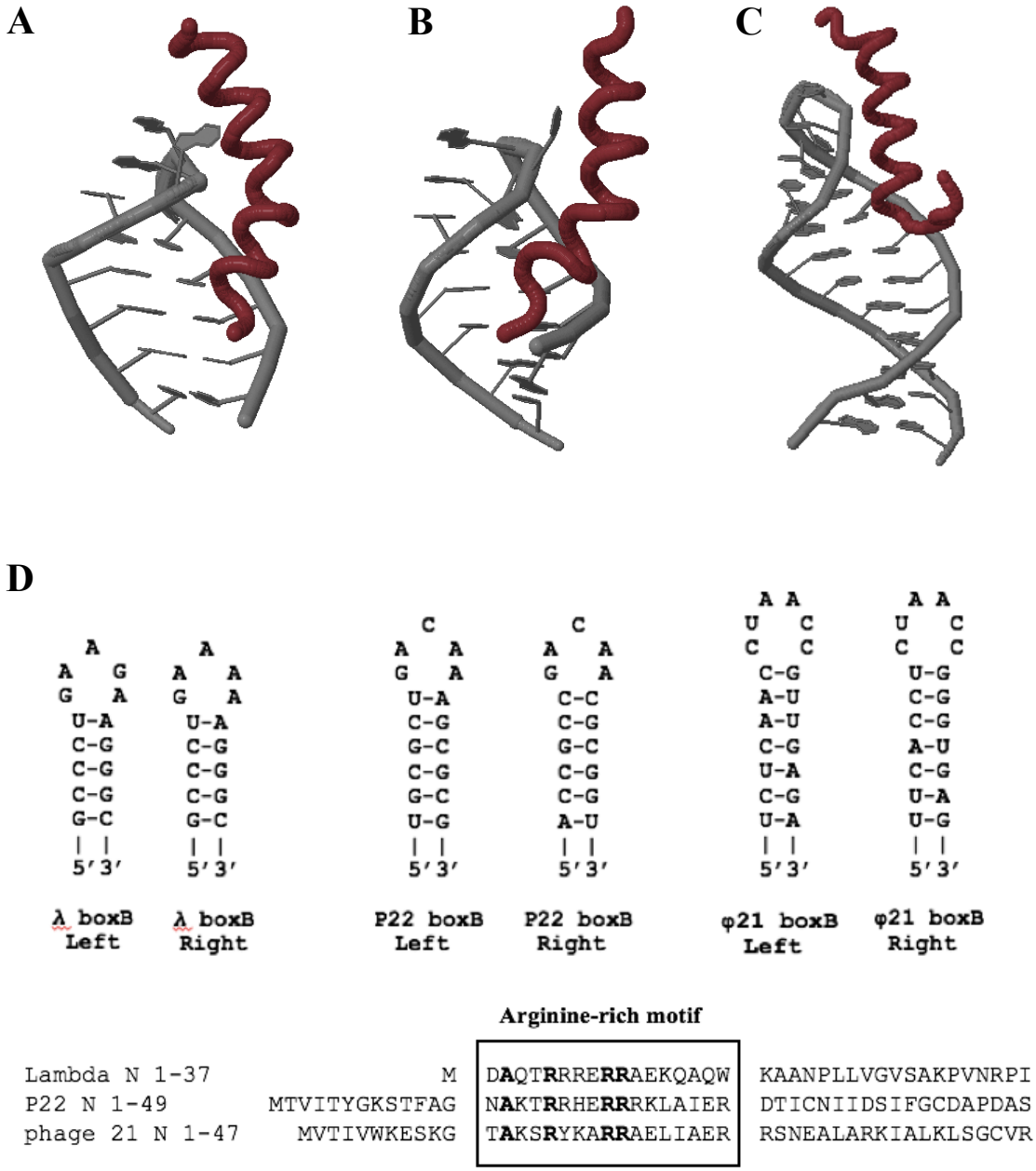
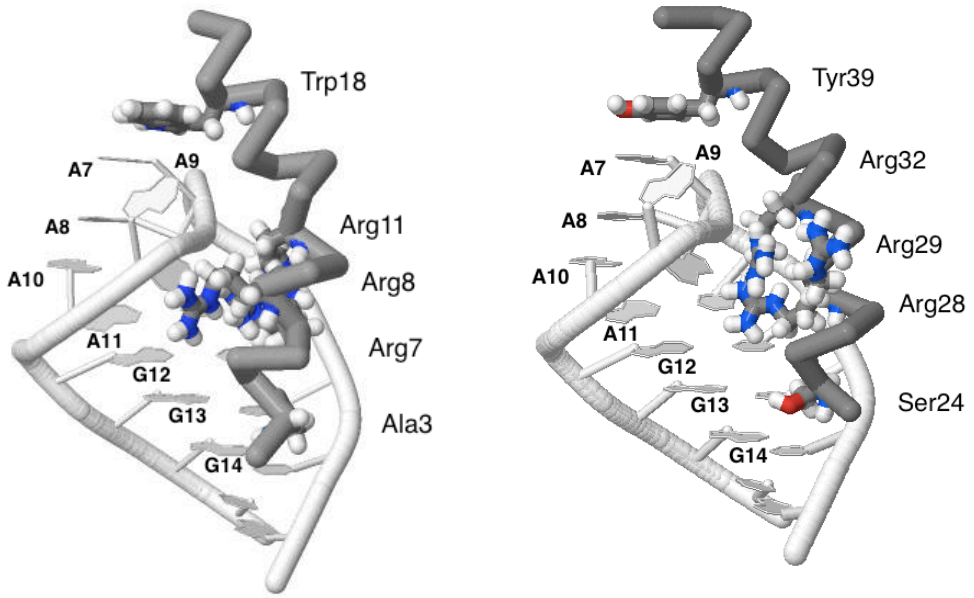


Figure 6. NMR structural models of lambdoid bacteriophage N ARMs binding the major grooves of their cognate boxB hairpins. (A) λN-boxB (PDB 1QFQ). (B) P22 N-boxB (PDB 1A4T). (C) φ21 N-boxB (PDB 1NYB). (D) Folded secondary structures of λ, P22, and φ21 left and right boxB sequences (upper panel). Aligned N ARM sequences with conserved residues shown in bold (lower panel).

1. λ N ARM-boxB interaction

The disordered N protein (107 amino acids) binds through its amino terminal ARM (the first 22 residues) to the boxB hairpin of the nut site. The NMR structural model of λ N ARM-boxB complex (Figure 7A, left) reveals a bent α-helix with a turn capping its amino terminus. Remarkably, boxB pentaloop folds into the common GNRA tetraloop motif (Figure 3B) (Heus and Pardi, 1991) by extruding the fourth base. BoxB stem is half a turn of a regular A-form helix. N ARM helix does not penetrate the major groove, it binds only the first three loop nucleotides and the 5' strand of the stem (Legault et al., 1998; Scharpf et al., 2000). The binding interactions is mainly a recognition of the negatively charged topology of the boxB hairpin through hydrophobic and electrostatic interactions, with only few base-specific contacts. Most prominently, Ala3 and Trp18 are involved in well defined interactions that anchor the helix to boxB at the amino and carboxy termini respectively. Ala-3 methyl group is buried in the hydrophobic pocket formed by the bases and riboses of C4 and C5. Trp18 indole ring is involved in a stacking interaction with the apical loop base A7 and contributes to specific recognition of λ boxB since an equivalent interaction is not present in P22 and φ21 homologous N ARMs (Lazinski et al., 1989). The positively charged side chains of Arg and Lys residues are engaged in electrostatic interactions with the negatively charged phosphate backbone of boxB 5' strand. Unlike the predominance of electrostatic and hydrophobic interactions, few intermolecular hydrogen bonds are proposed: hydrogen bonding of Gln4 and Arg7 with U5 and G6. This emphasizes the importance of the sheared G:A and the "loop closing" U-A base pairs for specific recognition of λ boxB structure.

A



λN ARM-boxB **Nun ARM-boxB**

B

```
HK022 Nun 19-41 DRG LTSRDRRIARWEKRIAY AL-
Lambda N 1-20       MDAQTRRRERRAEKQAQW KA-
```

Figure 7. (A) Comparison of NMR structural models of λN ARM-boxB (PDB 1QFQ; Scharpf et al., 2000) and Nun ARM-boxB (PDB 1HJI, Faber et al., 2001) interactions. ARMs are shown as grey backbones and boxBs as labelled white cartoons. N and Nun important and equivalent residues are shown as sticks and labelled. (B) Nun ARM and λN ARM sequences aligned. Conserved residues are underlines, residues involved in similar important interactions are shown in bold.

D. Bacteriophage Nun Termination

Bacteriophage HK022 Nun protein has no function essential for the phage lytic or lysogenic development. Its only known function is to block λ phage superinfection by terminating transcription of λ left and right operons (Robert et al.1987; Robledo et al. 1990; Sloan and Weisberg 1993). Similar to λ N, Nun specifically binds boxB stem-

loop structure in the nascent early λ left and right transcripts through its ARM (20-44) at its amino terminus (Chattopadhyay et al. 1995). Nun competes with λ N for their common binding site, and engages with RNAP and the same set of host factor (NusG, NusA, NusE, NusB) to actively mediates premature transcription termination (Figure 5B) (Robert et al., 1987; Robledo et al., 1990, Watnick and Gottesman, 1998). A recently published structural paper (Kang et al., 2017) elucidates the mechanism by which Nun interacts with RNAP to crosslink it to the nucleic acid scaffold and consequently block forward and reverse translocation, but the precise localization of host factors within the termination complex remains unclear. Similarly in N and Nun, the RNA-binding function is attributed to the arginine-rich amino terminal domain, while the antitermination/termination function is attributed to the carboxy terminal activation domain (Weisberg et al., 1999).

The NMR structural model of Nun ARM-boxB interaction reveals conformations and set of interactions very similar to those of λN ARM-boxB model (Figure 7A, right) (Faber et al., 2001; Legault et al., 1998). Nun ARM adopts a similar α-helix bent at Ala31 and Arg32 residues. BoxB shows conformations in its stem and loop regions identical to its structure when bound to N. Nun Ser24 and Tyr39 interactions are equivalent to the interactions of Ala3 and Trp18 in λN. R28, R29, and R32 of Nun ARM are the equivalents of R7, R8, R11 of λN ARM (Faber et al., 2001; Legault et al., 1998). Henthorn and Friedman (1996) examined different hybrid combinations of Nun and N domains and reported a complete interchangeability, and thus functional modularity, of their RNA-binding domains. Mutagenesis studies report similar effects of boxB mutations on N and Nun activities (Chattopadhyay, 1995b). This suggested N ARM and Nun ARM recognition strategies are indistinguishable.

Unanticipated by published structural and mutagenesis data, Tawk et al. (2015) observed unexpected immutability of non-conserved residues in Nun ARM (D26, W33, and R36) that are part of its hydrophobic ridge and had been proposed to contact NusG-CTD. Attempted rigid-body docking identified a plausible interaction between Nun ARM hydrophobic ridge and NusG-carboxy terminal domain (NusG-CTD) (Tawk et al., 2015). The suggested interaction nestles Trp33 between NusG Phe144 and Phe165. This matches Trp33 immutability in Nun and the isolation of NusG mutations at Phe144 and Phe165 positions that abolish Nun's, but not N's, function (Burova et al., 1999; Mooney et al., 2009). Puzzlingly, Nun preserves its D26 and W33 requirements when Nun ARM is fused to N activation domain and examined for N-antitermination function which suggests the reliance of Nun ARM on contacts with NusG for a productive interaction with boxB (Tawk et al., 2015).

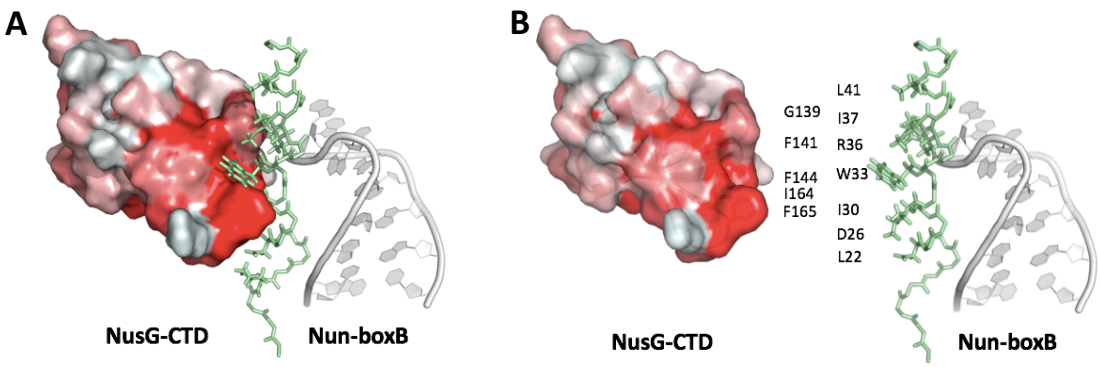


Figure8. Rigid-body docking of nusG-CTD and Nun ARM-boxB NMR structural models (Tawk et al., 2015). (A) NusG-CTD hydrophobicity is shown in red. Nun ARM (green) and boxB (white) are represented as cartoon structures. (B) NusG-CTD is shown as a gold cartoon. Nun ARM-boxB are shown as in panel (A). NusG and Nun residues at the suggested interface are shown as sticks with labels.

E. The experimental approach and aims of this study

The general aim of this thesis is to explore the evolution of ARM-RNA interactions. Can different recognition strategies and specificities evolve by genetic drift along neutral networks, or are other routes necessary, such as loss-of-fitness intermediates, recombination, or simultaneous mutations? The interactions of λ N-ARM and HK022 Nun-ARM with the hairpin boxB RNA structure were employed to examine in detail potential evolutionary routes. Despite the similarity of the NMR structural models of N ARM-boxB and Nun ARM-boxB interactions, Tawk et al., (2015) revealed distinct requirements in Nun ARM residues that are suggested to contact NusG-CTD in the termination and antitermination contexts. Nothing supports analogous contacts of NusG-CTD with N ARM. This prompted a further examination of the complexes to elucidate their differences, examine whether the suggested difference affects the mutability of boxB sequence, and test the possibility of transition from one recognition strategy to another by incremental mutational steps.

- Specific Aim 1: Explore the potential of boxB loop expansion as a mechanism to evolve specificity to N or Nun. A boxB library with a hexaloop having the four apical nucleotides randomized (LL4NL) will be screened with N and Nun using the double-plasmid reporter system (Figure 9). BoxB clones with interesting activities will be sequenced, ordered as oligonucleotides, and examined again to confirm activity. N-antitermination activity is screened by beta-galactosidase expression. Nun-termination activity is screened by alkaline phosphatase expression.

- Specific Aim 2: Characterize a panel of pentaloop-boxB mutants for specificity to N and Nun. An existing panel of loop and stem mutants will be screened for N-antitermination and Nun-termination activities with the double plasmid reporter system (Figure 9). Differences in tolerated mutations in boxB sequence will reflect differences in the recognition strategies of N and Nun.
- Specific Aim 3: Determine the mutational profile of N ARM. Residue-by residue mutagenesis will be done for N ARM residues 2-18. For every codon position, a separate pBRN plasmid library was made. Each plasmid library will be transformed into *E. coli* and 50 clones will be assessed for ability to complement the growth of an N-deficient λ strain. The proportion of functional clones will reflect the importance of the examined position. Importantly, this will allow a direct comparison of requirements between N and Nun ARMs and aid the design of Specific Aim 4.
- Specific Aim 4: Design, construct, and screen hybrid ARMs fused to N or Nun. To characterize differences between N and Nun recognition strategies, complex ARM mutants having a specific combination of N and Nun residues will be designed. Determining what changes in N ARM makes it functional in Nun, and what changes in Nun ARM keeps it functional in N will reveal to what extent N and Nun interactions with boxB are distinct or similar. Designed hybrid ARMs fused to N and Nun activation domains will be screened with *in vivo* plaque assays. N-antitermination function will be examined with N-complementation assays, and Nun-termination function will be examined with lambda infection exclusion assays.

The results should reveal the the local evolutionary landscape of N-boxB and

Nun-boxB interactions, elucidate the difference between N and Nun recognition strategies, and examine how λ can escape HK022 through the mutational potential of boxB sequence. The findings will test the validity of addressing the evolution of recognition strategies within the neutral theory framework.

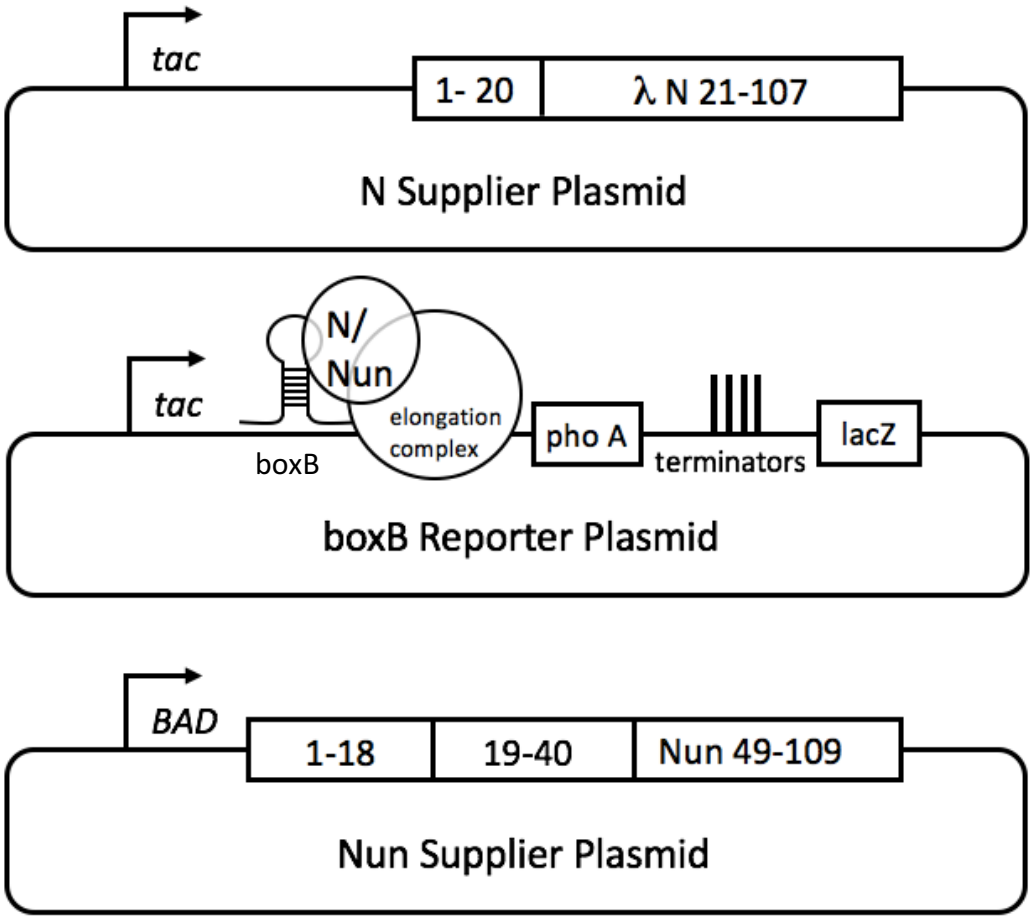


Figure 9. Diagram redrawn from (Tawk et al., 2015) showing the two-plasmid reporter system developed by N. Franklin (1993). To screen for N-antitermination the upper and middle plasmids are used. N protein is provided by the N supplier plasmid (pBRN). The N gene has NcoI-BsmI sites that flank the RNA binding domain to allow cloning. BoxB reporter plasmid (pACnut) expresses λ left *nut* site, phoA reporter gene, transcription terminators, and lacZ reporter gene. PstI and BamHI sites (not indicated) that flank the nut site are used for cloning nut sites with boxB mutations. When N binds to boxB hairpin of the nascent nut transcript, it mediates the assembly of host factors on RNAP which reads through terminator sequences and expresses lacZ reporter gene. To screen for Nun-termination, boxB reporter plasmid is used with Nun supplier plasmid (lower) (pBADcasNun) which has NcoI-BsmI sites for cloning. Nun binds to boxB and prematurely terminates transcription downstream of phoA gene leading to reduced phoA expression.

CHAPTER II

MATERIALS AND METHODS

A. Constructing λ BoxB Expanded Loop Plasmid Library (LL4NL)

Single stranded synthetic oligonucleotides of the following sequence were ordered from TIB-MOLBIOL (Germany): 5'CTGCAGGTCGACGCTCTTAAAAA TTAAGCCCTGNNNNAAGGGCAGCATTCAAAGCAGGGATCC3' for the LL4NL library. The sequence is randomized at 4 apical nucleotide positions of boxB hexaloop such that A, T, G, or C nucleotides were inserted with equal probabilities. The complementary sequence was synthesized with a primer extension reaction using the following boxA primer sequence: 5'-GTCGACGCTCTTAAAAATTAA-3'. The standard 500 μl primer extension reaction was used. A mixture of the template, primer, buffer, $MgCl_2$, and water in a tube was incubated for 5 min at 95°C. The tube was then placed on ice to add dNTP, DTT, and Taq polymerase. The following thermocycler program was used to carry out the reaction: 5 min at each temperature starting with 40°C with incremental increase of 5°C to reach 70°C, followed by 20 min at 72 °C.

The double-stranded oligonucleotides were purified by adding an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and transferring the supernatant after centrifugation. The supernatant was mixed with an equal volume of chloroform and spun down. The upper layer containing the DNA was retained and precipitated with a mixture of NaOAc and 100% ethanol. The obtained DNA pellet was resuspended in pure sterile water and stored at -20°C. After digesting the DNA library with PstI and BamHI restriction enzymes, the inserts were purified with phenol:chloroform extraction

again and cloned into pACnut backbone which was cut with the same restriction enzymes.

B. Chemically Competent *E. coli* Cells

1. Preparation

Chemically competent E coli N567 cells carrying pBRNKH, pBRNRev, pBADwtNun, pBADcasRevNun, or no plasmid were prepared similarly. Cells to be made competent were available as glycerol stocks. LB agar plates with the appropriate antibiotic were streaked and incubated overnight at 34°C. Isolated colonies were picked and inoculated into 3 LB medium (with or without antibiotic). The culture tubes were incubated at 37°C with 200 rpm shaking for 16 hours. The following day, 1 ml of the grown bacterial culture was added to 100 ml of LB medium (with or without antibiotic). The new culture was incubated at 37°C with 200 rpm shaking for a limited time until an OD_{600} of 0.38-0.40 was reached. The culture was shaken for few minutes in an ice water bath to ensure rapid cooling and block further growth. The 100ml culture was divided into two 50 ml samples in pre-chilled conical tubes which were centrifuged at 4500 rpm for 12 min at 2-4°C. The supernatant LB was discarded and the bacterial pellets were resuspended in cold solution of $CaCl_2$ (60 mM $CaCl_2$, 15% glycerol, 10 mM PIPES (piperazine-N, N'-bis-2-ethanesulfonic acid) pH 7). Resuspensions were centrifuged for 8-10 min at 2-4°C and 2500 rpm. All the following steps were done under strict cold conditions to ensure higher cell competency. The obtained pellets were resuspended again in cold 10 ml of $CaCl_2$ and kept on ice for 30 min. After a last round of centrifugation at 2-4°C and 2500 rpm for 5 min, each pellet was resuspended in 2 ml of cold $CaCl_2$. The final resuspension was divided into 250-300 μl aliquots in prechilled

1.5 ml Eppendorf tubes and stored at -70°C. The competency of the cells was assessed by transforming with DNA plasmids of known concentration and calculating number of colonies obtained per μg DNA.

2. Transformation

Chemically competent cells were thawed on ice. Dilution with cold $CaCl_2$ was done according to desired colony yield. A volume of 50-100 μl cells was added to a 1.5 ml tube containing 1-10 μl DNA. All tubes were kept on ice. The cells and DNA were mixed and incubated on ice for 20-30 min. This was followed by heat shock at 37°C for 3 min and immediate 7 min incubation on ice. A volume of 950 μl LB was added and mixed and then incubated at 37°C for 1-2 hours according to the kind of plasmid used for transformation. Using sterile glass beads, 50-100 μl of cells were spread on agar plates (with appropriate antibiotic) and incubated at 34°C for overnight growth.

C. Screening for N-Antitermination with X-Gal Plate Assays

To screen for boxB mutants active with N, Franklin's (1993) double-plasmid reported system was used (Figure 9). One plasmid supplies λ N protein (N sequence variants and Nun-N fusions) and a second plasmid expresses λ *nut* site, phoA reporter gene, transcription terminators, and lacZ reporter gene (Figure 9). Cloning boxB LL4NL library into pAC reporter plasmid was done with PstI and BamHI sites. Cloning N ARMs into pBRN supplier plasmid was done with NcoI and BsmI sites. Wild-type N gene has an NcoI site; K19N had been substituted in N to create a BsmI site. Upon transcription of the *nut* site, N binds to boxB hairpin and nucleates the assembly of a complete antitermination complex that reads through the downstream terminators and

expresses beta-galactosidase. X-Gal (5-bromo-4-chloro-3-indolyl –D-galactoside) is a chromogenic substrate of beta-galactosidase that turns into an insoluble indigo pigment upon hydrolysis. This allows screening for the level of beta-galactosidase activity, and thus N-boxB interaction affinity, through observing the intensity of blue color of colonies grown on XGal-containing agar plates.

In this study, LL4NL boxB library was cloned into pAC reporter plasmid using PstI and BamHI restriction sites. HIV Rev ARM-RRE heterologous interaction was used as a control. LL4NL was screened by transforming competent cells carrying supplier plasmids (pBRNKH or pBRNRev) with pACnut LL4NL library, pACnutLL(λ wildtype left boxB), pACnutLR (λ wildtype right boxB), or pACflRRE (HIV-1 RRE) and plating on tryptone agar medium (5 g/l NaCl, 10 g/l tryptone, 15 g/l Agar) containing 100 μg/ml ampicillin, 11.3 μg/ml chloramphenicol, 80 μg/ml X-Gal, and 0.1 mM IPTG (isopropyl-D-thiogalactoside). IPTG was used to induce P_{Tac} promotors of supplier and reporter plasmids. After an overnight incubation of transformation plates at 34°C, the plates were kept at room temperature and colony color was examined after 24 hours. Colonies with intense blue color were picked and used to grow cell cultures to extract pAC plasmids carrying the active boxB sequence variants.

D. Screening for Nun-Termination with XP Plate Assay

To screen for boxB mutants active with Nun, the same double-plasmid reporter system was used. One plasmid supplies wild type Nun protein, and a second plasmid expresses λ *nut* site, phoA reporter gene, transcription terminators, and lacZ reporter gene (Figure 9). Upon transcription of *nut* site, Nun binds to boxB hairpin and nucleates the assembly of a complete termination complex that prematurely terminates

transcription between the nut site and phoA reporter gene. As a result, neither phoA nor lacZ is expressed. Active Nun-boxB interactions can be monitored by the reduction in alkaline phosphatase activity. The levels of alkaline phosphatase can be monitored colorimetrically by supplying the solid medium with the chromogenic substrate X-phosphate (XP) which produces an insoluble indigo pigment upon hydrolysis. HIV Rev ARM-RRE heterologous interaction was used as a control. LL4NL boxB library was screened by transforming competent cells carrying supplier plasmids (pBADwtNun or pBADcasRevNun) with pACnut LL4NL library, pACnutLL (λ $boxB_{left}$), pACnutLR (λ $boxB_{right}$), or pACflRRE (HIV-1 RRE) and plating on tryptone agar medium (5 g/l NaCl, 10 g/l tryptone, 15 g/l Agar) containing 100 μg/ml ampicillin, 11.3 μg/ml chloramphenicol, 50 μg/ml XP, and 0.5 mM IPTG (isopropyl-D-thiogalactoside). IPTG was used to induce P_{Tac} promotors of the reporter plasmid. After an overnight incubation of transformation plates at 34°C, the plates were kept at room temperature and colony color was examined in a 18-20 hours window. Colonies with pale blue color were picked and used to grow cell cultures to extract pAC plasmids carrying the active boxB sequence variants. The paler the blue color, the more active the Nun-boxB interaction.

E. DNA Extraction

An isolated colony carrying the desired plasmid is inoculated into 5 ml LB liquid media containing the appropriate antibiotic and incubated for 16 hours at 37°C with 220 rpm shaking. The next day, cells are spun down at 4°C, 4500 rpm, for 10-12 min. The pellet is resuspended in 150 μl cold SET solution (20% sucrose, 50 mM Tris pH 8.0, 50 mM EDTA). The resuspension was transferred to a 1.5 ml tube placed on

ice. A volume of 350 µl of freshly made lysis buffer (0.2 M NaOH and 1% SDS) was added and the contents were mixed gently by inverting the tube multiple times. The tube was kept on ice for precisely 5 min followed by addition of 250 µl of cold 3 M KOAc. The contents were mixed gently and the tube was kept 10 min on ice. The tube was centrifuged for 10-12 min at 14000 rpm and the supernatant was transferred to labelled 1.5 ml tube containing 600 µl isopropanol. The contents were mixed well by vortexing and left at room temperature for more than 15 min. The tube was spun down for 10 min at 14000 rpm and the supernatant was discarded. To ensure the complete removal of alcohol, the tube was spun down again for 30-40 s and the last drops of isopropanol were pipetted out. The DNA pellet was left to dry under the hood for 20-45 min and then resuspended in 40-50 µl of sterile H2O containing 1-2 µl of 20 mg/ml RNase A. After 2-3 µl of the sample was run in a 1% agarose gel by electrophoresis to ensure the presence of DNA, the sample was stored at -20°C.

F. DNA Column Purification and Sequencing

GE Healthcare DNA purification kit (United Kingdom) was used for DNA purification. To 35-40 µl DNA sample, 500 µl capture buffer was added and mixed well by vortexing. The mixture was then transferred to the purification column and centrifuged for 1 min at 10000 g. The eluent was removed. A volume of 750 µl of washing buffer (Tris-HCl pH 8.0, 1 mM EDTA, 80% ethanol) was added to the column and and kept at room temperature for 2 min before centrifuging 1 min at 10000 g. The eluent was discarded and the washing step was repeated. The tube was centrifuged again at 15000 g for 30-40 s to ensure complete removal of washing buffer. The DNA was collected from the column membrane in a labelled tube by adding 40 µl of the

elution buffer followed by centrifugation for 1 min at 5000 g. DNA samples of 20 µl were loaded on a 96-well plate and the proper primer supplied in a separate 500 µl tube. The samples were sent to Macrogen Korea for sequencing.

G. β-Galactosidase N-Antitermination Solution Assays

pACnut plasmids with known boxB sequences (designed boxB sequences, sequences from an existing panel, and boxB sequences from LL4NL library screening) were used to transform competent cells carrying plasmids supplying wild-type N (pBRNKH) or Rev ARM-N fusion (pBRNRev). Average-size blue colonies from XGal plates were inoculated into 3 ml tryptone liquid media containing 100 µg/ml ampicillin, 11.3 µg/ml chloramphenicol, and 0.1 mM IPTG. Culture tubes were incubated for 16 hours at 30°C with 200 rpm shaking. The next day, 250-500 µl of cells (depending on expected level of activity) were added to 2 ml tubes containing 500-750 µl of Z buffer (Miller, 1992) (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 2% SDS, and 3.5 µl/ml Z buffer of β-mercaptoethanol). The cells were lysed by adding 40 µl of chloroform and vortexing for 10 s. The total volume in every tube was 1 ml. The reaction tubes were put for 8-10 min in a water bath set to 28°C. To start the reactions, 200 µl of 4 mg/ml ONPG (ortho-nitro-phenol- galactopyranoside) were added to every tube. ONPG is a chromogenic substrate of β-Galactosidase enzyme that turns yellow after cleavage. Upon faint yellow coloring, the reactions were stopped by adding 200 µl of 1 M Na_2CO_3. An accurate record was kept for reaction start and end time points. After centrifugation of debris, the yellow color intensity was measured at OD_{420} and culture growth was measure at OD_{600}. To account for cellular debris affecting measurement, reactions were measure at OD_{550} to be subtracted from OD_{420}

measurements. Calculation of β-Galactosidase units was done with the following formula (Miller, 1992):

$$100\times\frac{(\mathrm{OD}420-1.75\mathrm{OD}550)}{t\times v\times OD600}$$

where t is reaction time in minutes and v is the used volume of cell culture.

H. Alkaline Phosphatase Solution Assay for Nun-Termination

pACnut plasmids with known boxB sequences (designed boxB sequences, sequences from an existing panel, and boxB sequences from LL4NL library screening) were used to transform chemically competent cells that carry pBADcasNun or pBADcasRevNun plasmids. Transformants were grown overnight at 34°C on tryptone plates containing ampicillin 100 μg/ml, chloramphenicol 11.3 μg/ml, and X-Phospate 50 μg/ml. One day before the experiment, average four isolated colonies were picked from each plate and inoculated into tryptone liquid media (with ampicillin 100 μg/ml, chloramphenicol 11.3 μg/ml, and 100 mM IPTG). Cultures were grown for 12 hours at 34°C with 220 rpm shaking. After 12-hour growth, 0.5 ml of each cell culture was spun down in 1.5 ml tubes and washed with 700 μl resuspension buffer (10mM Tris-HCl pH 8.0, 0.1M NaCl). Tubes were spun down again and the pellets resuspended in 1 ml resuspension buffer. Of the 1 ml resuspension, 900 μl were used to measure OD_{600}, and 10 μl were transferred to 2 ml tubes containing 990 μl Tris Buffer (1M Tris-HCl pH 8.0). To lyse the cells, 60 μl chloroform and 30 μl 0.1% SDS were added. Samples were vortexed for 10 s and then put in a 28°C water bath for 5-7 min. Reactions were started by adding 100 μl PNPP (para-nitro-phenylphosphate 4 mg/ml) which is a chromogenic

substrate of he alkaline phosphatase enzyme. Upon faint yellow coloring of the samples, reactions were stopped by adding 100 µl cold 1 M K_2HPO_4. The start and end time points were accurately recorded. After centrifugation of cell debris, the OD_{420} and OD_{550} of every sample were measured. Alkaline phosphatase units were calculated following this formula (Brickman and Beckwith, 1975):

$$1000\times\frac{OD420-1.75OD550}{time\times OD600}\times dilution\ factor$$

I. N-deficient λ complementation assay:

N567 *E. coli* were transformed with plasmid DNA expressing the N protein variant to be examined. Transformed cells are plated on solid LB medium containing 100 µg/ml ampicillin. Colonies from overnight plates were picked and used to grow separate cultures in LB liquid media (with ampicillin) at 37°C with 200 rpm shaking. After 16 hours, the OD_{600} of the cultures was measured before centrifugation and resuspension of cells in 10 mM $MgSO_4$. The volume of $MgSO_4$ is calculated such that the OD_{600} of the resuspension is approximately 2.0. Cultures were kept at 4°C. To screen a particular N protein for antitermination function, 50 µl of the cell culture of that clone is added to 50 µl of N-deficient λ strain (diluted with SM buffer to produce around 500 pfu). The N gene of this phage has two amber mutations making it nonfunctional. This λ strain is referred to as $\lambda^{Namb7Namb53}$ (Franklin, 2004). The tube was incubated for 20 min at 37°C to allow λ adsorption. After incubation, 1 ml of 48-49°C top agar (tryptone, NaCl, 0.7% agar) was quickly added to the tube and the content was spread by swirling on a small tryptone agar plate with or without 0.3 mM IPTG. The plate was left at room temperature for 10 min for the top agar to dry and then incubated

for around 12 hours at 37°C. Assessment of plaque formation was done using W1485 *E. coli* amber suppressor strain as a control. Plaque size assessment was done relative to N567 cells with pBRNKH.

J. Wild type λ exclusion assay

The same procedure for λ complementation assay was followed but with few variations. Cells transformed with plasmid supplying Nun and N-Nun fusion proteins were infected with wild type λ strain to screen for Nun's ability to exclude λ superinfection. N567 cells and cells with pBADcasNun were used as controls to assess plaque.

H. Screening Libraries of N 2-18 ARM

Singly randomized codon libraries of positions 2-18 of λN ARM were cloned into pBRN plasmids. The library for each codon position was made and screened separately. Chemically competent N567 *E. coli* cells were transformed with the cloned libraries and grown on LB agar plates with ampicillin (100 μg/ml). Random 50 clones were picked and tested for N-deficient λ complementation assay as described earlier.

CHAPTER III

RESULTS

λ N and HK022 Nun proteins interact with RNAP and an identical set of host-encoded factors, yet they mediate contrary outcomes by nucleating the assembly of transcription antitermination and termination complexes, respectively. The assembly of both complexes is initiated by the recognition of boxB RNA sequence by N and Nun ARMs. The separate NMR structural models of λ N ARM- and HK022 Nun ARM-boxB complexes reveal similarly bent alpha-helices binding the major groove of boxB. BoxB adopts an almost identical conformation with both ARMs: a regular A-form stem and a GNRA-like pentaloop with the fourth nucleotide extruded (Legault et al., 1998; Scharpf et al., 2000; Faber et al., 2001). Mutagenesis studies report similar effects of boxB mutations on N and Nun activities (Chattopadhyay, 1995b). This suggested N ARM and Nun ARM recognition strategies are indistinguishable. Unanticipated by published structural and mutagenesis data, Tawk et al. (2015) observed unexpected immutability of non-conserved residues in Nun ARM (D26, W33, and R36) that are part of its hydrophobic ridge and are proposed to contact NusG-CTD. Attempted docking of NusG-CTD to Nun ARM was successful and matched published mutagenesis data of NusG residues with Nun DWR requirements that fall at their interaction interface (Tawk et al., 2015). Interestingly, lambda N ARM has very different residues at the corresponding positions and lacks an equivalent hydrophobic ridge in its α-helical conformation. This suggests NusG-CTD does not bind N ARM. Yet it is puzzling that Nun preserves D26 and W33 requirements to mediate antitermination when fused to the N activation domain. This suggests that, unlike N,

Nun ARM relies on its contacts with NusG-CTD to productively bind boxB which is at the heart of the difference in their recognition strategies. Thus, it remains unclear to what extent these binding modes are alike or distinct. Results from this thesis confirm and characterize the difference between N and Nun binding modes through an analysis of the mutational potential of the two ARMs and their binding target boxB, and test the validity of neutral evolutionary theories when applied to RNA-protein interactions.

Considering that differences in recognition strategies employed by N and Nun within the transcription elongation complex could affect which boxB mutations maintain productive binding, a panel of boxB mutants from previous projects was examined for N and Nun activities. BoxB sequences were ordered as DNA oligonucleotides, cloned into pAC reporter plasmids, and tested by lacZ and phoA solution assays to determine their specificities to N and Nun respectively. Additional boxB mutants were designed guided by published mutagenesis and NMR structural data. Such an *in vivo* comparison of the mutational potential of boxB with Nun and lambda N, which has not been published before, identified differences in boxB sequence requirements. BoxBs that work with N but not with Nun, and vice versa, confirmed that Nun and N employ distinct recognition strategies.

G. **Nun ARM, unlike N ARM, recognizes non-GNRA-like pentaloops:**

According to published mutagenesis data (Cilley & Williamson, 1997; Chattopadhyay, 1995a; Salstrom et al., 1978; Szybalski, 1987; Mogridge et al., 1995; Tan and Frankel, 1995; Doelling and Franklin, 1989), N strictly requires the sheared base pair G:A for productive binding. Any mutation at positions 1 and 5 of the loop abolishes N binding to boxB *in vitro* and eliminates antitermination activity *in vivo*.

Two hydrogen bonds between boxB G1 and N's R7 and Gln4 are suggested by Scharpf et al. (2000). G_1substitutions are associated with reduced peptide induced-fit (Sue et al., 1997). Our results confirmed this restriction by the loss of N antitermination activity upon mutating G_1A_5 to G_1U_5, G_1G_5 , or A_1A_5 (Figure 10; mutants LL10U, LL10G, LRG6A). Although Chattopadhyay et al. (1995a) report wild type levels of *in vitro* binding affinities between N and LL10U and LL10G, these mutants were not active in our *in vivo* double-plasmid reporter system. These discrepancies imply that high affinity interactions *in vitro* do not necessarily correlate with *in vivo* antitermination activity.

Although the sheared G:A base pair of the boxB GNRA tetraloop fold is important for defining the orientation of the loop relative to the stem (Su et al., 1997), Nun allows substitutions of G:A with the wobble pair G-U and the non-canonical base pair G:G (Figure 10; LL10U and LL10G). It would seem that the overall recognition strategy of Nun either limits or tolerates the expected changes in the orientations and stability of those mutant boxB loops. Similarly to N, G_1 to A mutation completely abolished Nun function (LRG6A). This is in accord with G_1 being critical for Nun *in vivo* activity (Robert et al., 1987) despite the absence of contacts between Nun and G1 base (Faber et al., 2001).

Positions 2 and 4 of the loop were reported to favor purine nucleotides; pyrimidines maintain good binding affinity *in vitro* but significantly reduce N antitermination *in vivo* (Chattopadhyay, 1995a, 1995b; Doelling and Franklin, 1989). Data herein (Figure 10) show N tolerates pyrimidines at both position, but C less than U (L7U, L7C, D20, L9U, and L9C). With Nun, all nucleotide substitutions maintained a wild type level of activity confirming *in vitro* data (Chattopadhyay, 1995b).

In vitro binding affinities of N and Nun are greatly reduced with pyrimidines at position 3 of the loop (Cilley and Williamson, 1997; Chattopadhyayet al., 1995b; Tan and Frankel, 1995). *In vivo* studies confirm that for N although there is no data reported for Nun (Chattopadhyayet al., 1995a; Doelling & Franklin, 1989). This purine requirement was tested with the mutants D18 and D22: both have the pyrimidine C in place of A3 in LR and LL respectively. This mutation was tolerated in Nun more than in N. The reason might be the absence of any Nun contacts to the base A8 (Faber et al., 2001), whereas Scharpf et al. (2000) NMR data suggest hydrogen bonding between the guanidine group of N R8 and the base A8.

Some boxB loop variants with multiple mutations were tested (Figure 10). LL7G9C and L7G8G9 mutants were functional with N and Nun as expected since positions 2 and 4 tolerate all nucleotide substitutions. LL6A8U9U mutant did not function with N and Nun. This could be mainly because of the untolerated A_1A_5 mutation as in LRG6A. L8G9GU10U worked with Nun but not with N. This was consistent with Nun's tolerance of G:U wobble pair adjacent to the loop.

Taken together, these results show that Nun's recognition strategy, unlike N's, does not strictly adhere to GNRA-like requirements in the boxB pentaloop.

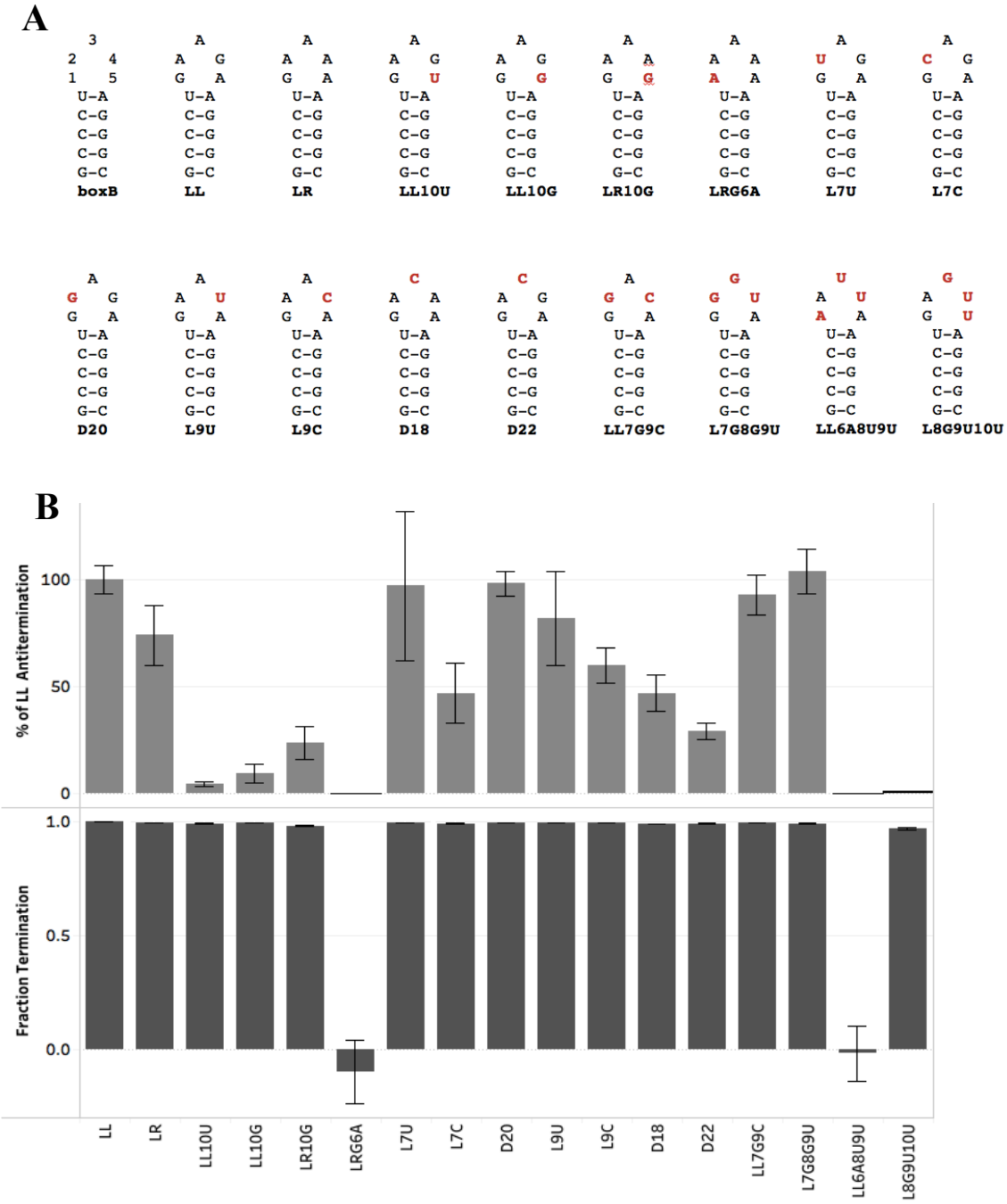


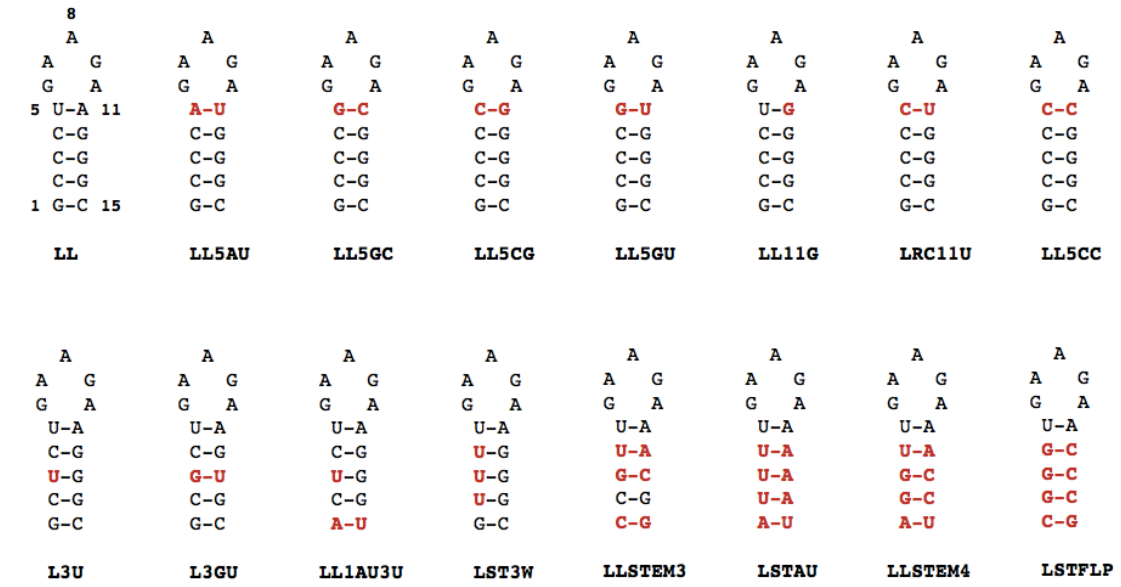
Figure 10. (A) Secondary structures of loop-mutant boxB sequences. The first structure is a reference for loop nucleotide numbering. Sequence deviations from λ $boxB_{left}$ or $boxB_{right}$ are shown in bold red. The reference name of each sequence is shown in bold under the hairpin structure. (B) % of LL Antitermination is: ß-galactosidase activity of the mutant boxB relative to λ $boxB_{left}$ (light gray bars). Fraction termination (dark grey bars) is: phoA units of wtNun-mutant boxB/phoA units of RevN-mutant boxB. Error bars are standard deviations from the mean values of 4 replicates.

B. Nun's relaxed requirements for boxB stem sequence contribute to boxB Nun-specificity

NMR structural models use an identical A-form helical conformation of boxB stem in its complexes with Nun and N ARMs (Legault et al., 1998; Faber et al., 2001). The NMR paper (Legault et al., 1998) support the identity of the C-G base pairs in the boxB stem is not critical for N as long as proper stem structure is maintained by Watson-crick base pairing. However, (Chattopadhyay et al., 1995a) work shows that both sequence and structure are important for N function. BoxB stem mutagenesis with Nun activity had not been examined closely yet. In this *in vivo* reporter system, it seems that for Nun, as long as base pairing is maintained in the stem, termination function is achieved with no regard to sequence (Figure 11; LLSTEM3, LSTAU, LLSTEM4, LSTFLP). Disrupting one base pair in the 5 base pair stem does not abolish Nun termination but disrupting 3 base pairs does (L3U, L3GU, LL1AU3U & LST3W). This implies that the interaction of Nun ARM with the stem relies mainly on recognition of its negatively charged topology.

In contrast, for proper N-mediated antitermination, a recognition of both stem structure and stem sequence matters (Figure 11). Disrupting 3 W-C base pairs completely abolished antitermination (LST3W). The behavior of L3U, L3GU, and LL1AU3U mutants reveals that the pyrimidine identity of C_3 stem nucleotide is of significance to N functionC_3 to U substitution increased N activity above wild type levels, whereas $C_3$to G mutation greatly reduced N antitermination. In LSTAU, four U-A stem base pairs instead of G-C completely abolished antitermination. It's likely that the weaker base pairing affected stem stability. The low activity of LSTAU, LLSTEM4 and LSTFLP demonstrate that stem sequence is important for N function.

A



B

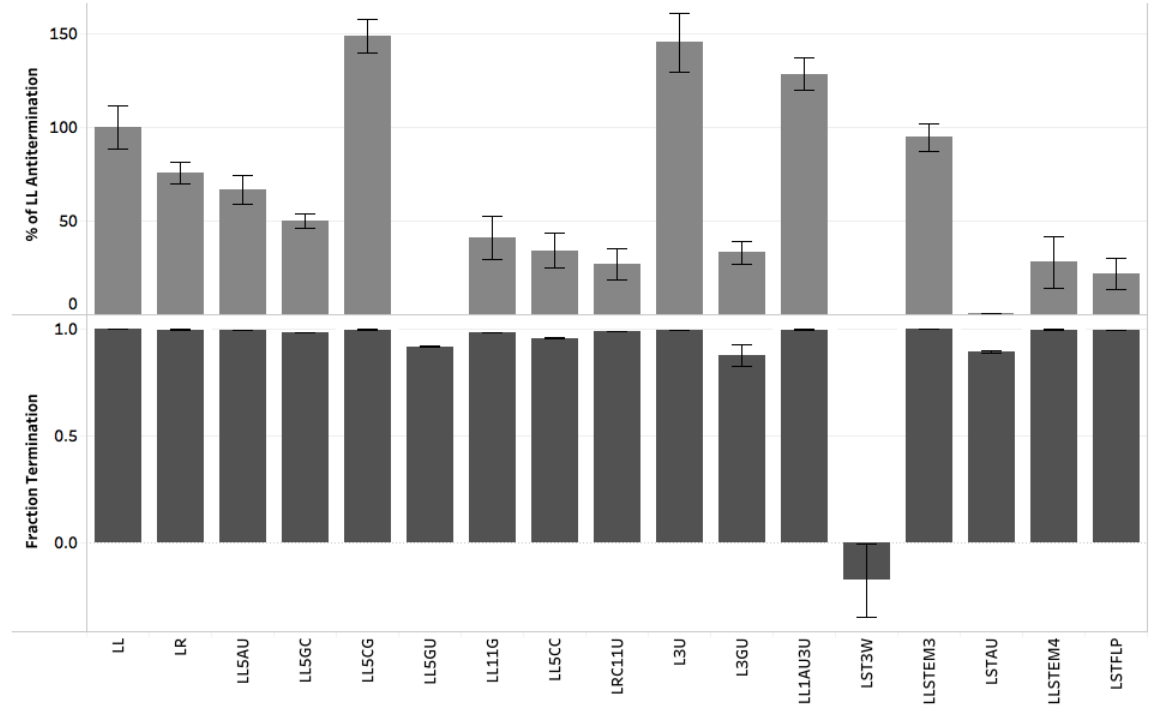


Figure 11. (A) Secondary structures of stem-mutant boxB sequences. The first structure is a reference for nucleotide numbering. Sequence deviations from λ $boxB_{left}$ or $boxB_{right}$ are shown in bold red. The reference name of each sequence is shown in bold under the hairpin structure. (B) % of LL Antitermination (light grey bars) is: ß-galactosidase activity of the mutant boxB relative to λ $boxB_{left}$. Fraction termination (dark grey bars) is: phoA units of wt Nun-mutant boxB/phoA units of RevN-mutant boxB. Error bars are standard deviations from the mean values of 4 replicates.

N Gln-4 and Arg-7 hydrogen bond with the U of the U-A base pair, and the aliphatic portion of Arg-7 is involved in hydrophobic contacts with U (Legault et al.,1998). Changing U-A to A-U is reported by (Tan and Frankel, 1995) to reduced N binding to boxB by more than 20 fold. In our study, flipping this loop-closing base pair U-A did not have a significant effect on N antitermination activity (LL5AU). Mutating U-A to GC (LL5GC) decreased N activity in half, while a C-G (LL5CG) mutation unexpectedly lead to a significant increase in N activity, higher than that of N ARM-LL interaction. This means that the weaker base pairing of U with A adjacent to the loop is not particularly important for loop function. A G:U base pair at the same position (LL5GU) completely abolished N activity whereas a U:G (LL11G) maintained an intermediate level of antitermination (41.5%) likely because wild-type contacts with U were maintained. Substituting U-A with C:C or C-U (LL5CC & LRC11U) caused approximately 3-fold decrease in N activity. This indicates that for N's recognition strategy paired-bases are preferred adjacent to the loop. Non-paired bases reduce N's activity. The Watson-Crick geometry of the loop-closing base pair contributes to stem stability or is specifically critical for loop function. In Nun, Faber et al. (2001) suggested R28 makes contacts with U. The same U-A mutations above were examined with Nun and all had no effect on its termination function, regardless of maintaining Watson-Crick base pairing or not. Thus, Nun's binding mode, unlike N's, has no specific requirements at this base pair position indicating its different contribution to loop and stem configurations.

A set of boxBs from Cocozaki et al. (2008a) work with P22 boxB right (PR) stems fused to wild type and mutant λ boxB pentaloops was screened (Figure 12). PR stem differs from λ stem by a flipped GC at the third base pair position, and an extra

UG base pair at the stem base. N tolerates PR stem well with LL and LR loops (D4 and D2), as well as with a G substitution at position 2 or 3 in the loop (D3 and L2), but not with a pyrimidine substitution at position 3 or 4 (D1 and L1). Thus, the PR stem reduces N's tolerance of pyrimidines in the loop. Although flipping UA loop-closing base pair of the lambda stem (see LL5AU Figure 11) had no significant effect on N activity, the same mutation in PR stem greatly abolished N antitermination (D7). All the mentioned mutants were very active with Nun thus D1, D7, and L1 are Nun-specific boxBs.

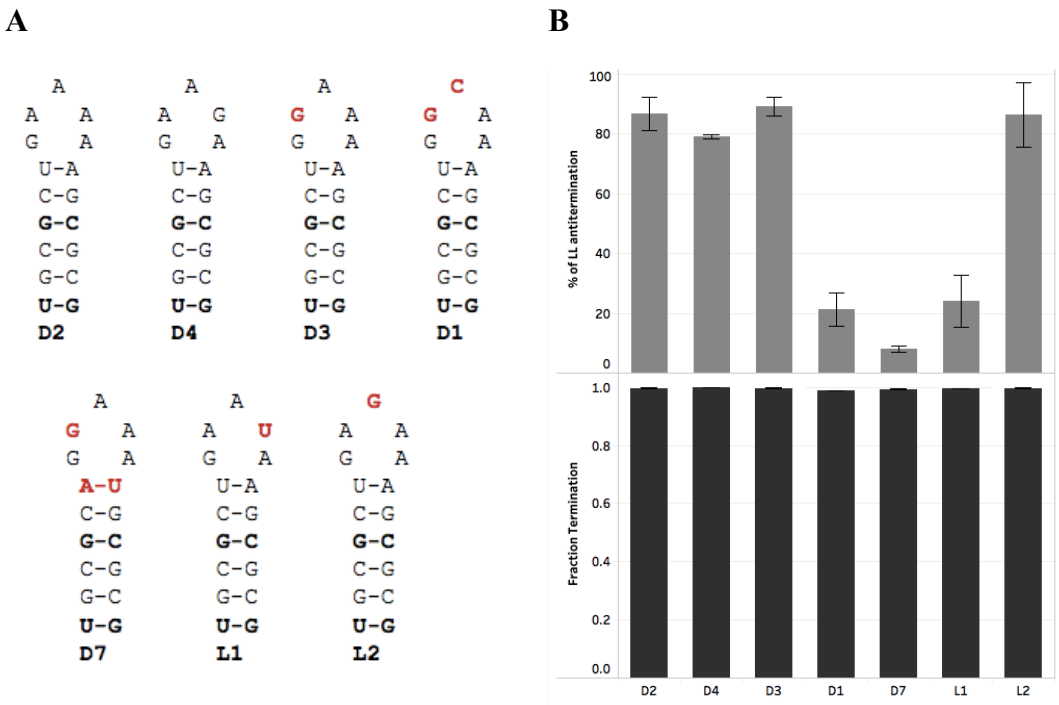


Figure 12. (A) Secondary structures of boxB pentaloops fused to P22 boxB right stem. The first structure is a reference for nucleotide numbering. Sequence deviations from λ $boxB_{left}$ or $boxB_{right}$ are shown in bold red. The reference name of each sequence is shown in bold under the hairpin structure. (B) Percentage of LL antitermination (light grey bars) is: ß-galactosidase activity of the mutant boxB relative to λ $boxB_{left}$. Fraction termination (dark grey bars) is: phoA units of wtNun-mutant boxB/phoA units of RevN-mutant boxB. Error bars are standard deviations from the mean values of 4 replicates.

C. BoxB loop expansion is a mechanism to evolve specificity to Nun but not to N

GNRA tetraloops are highly stable RNA motifs that interact with other RNAs and proteins. This explains their frequent occurrence in RNA structures (Antao *et. al.*, 1992; Hermann *et. al.*, 1999). When the lambda N ARM-boxB NMR structure was determined, it revealed the first structural example of a pentaloop folding into a GNRA tetraloop motif. This was achieved by extruding the fourth nucleotide (Legault et al., 1998). GNRA motifs have been observed in bigger loops (Huang et al., 2005; Klein et al., 2001) which suggested that any loop sequence following the consensus sequence GNR(N)xA, where x={1,2,3,..?}, is capable of folding into the GNRA motif. One aim of this study is to examine the mutational potential of boxB with respect to evolving specificity to N or Nun by the expansion of the hairpin loop.

Since the GNRA motif can be found embedded in larger loops, a hexaloop-boxB plasmid library was constructed in which the 4 apical loop nucleotides were randomized. The sheared G_1:A_6 base pair was not randomized because it is critical for N binding. This library should contain 256 unique sequences. Colonies were screened in *E. coli* N567 lambda N on Xgal and in *E. coli* N567 Nun on XP plates. Blue colonies on Xgal plates and grey-pale blue colonies on XP plates were picked and used to isolate the reporter plasmids. *E. coli* RevN cells were used as a negative control. After further separate confirmation of activity, tested clones were sequenced. Twelve unique sequences were ordered as DNA oligonucleotides to be used for antitermination and termination sollution assays.

Few hexaloop-boxB sequences were active with N (Figure 13). A larger panel of boxBs was active with Nun. At loop positions 2, 3, and 4 were found only purines. All four nucleotides were found at position 5. Thus, all the examined functional

hexaloops conform to a $G_1R_2R_3R_4N_5A_6$ consensus sequence. A designed hexaloop with C_3 was functional with Nun, but not with N: this supports a possible $G_1R_2N_3R_4N_5A_6$ consensus sequence for Nun. The LL4NL library screening was not exhaustive, so whether pyrimidines work at the second and fourth loop positions remains unresolved. This is consistent with the previously mentioned pentaloop mutants D1and D22 that have a similar cytosine substitution. This corroborates Nun's higher tolerance for pyrimidines at the third nucleotide position of the boxB loop. LLL2 is a designed heptaloop-boxB with two adenine insertions. This mutant was active with Nun. It is possible that the folded heptaloop is extruding two nucleotides which is accommodated in the termination complex more than the antitermination complex.

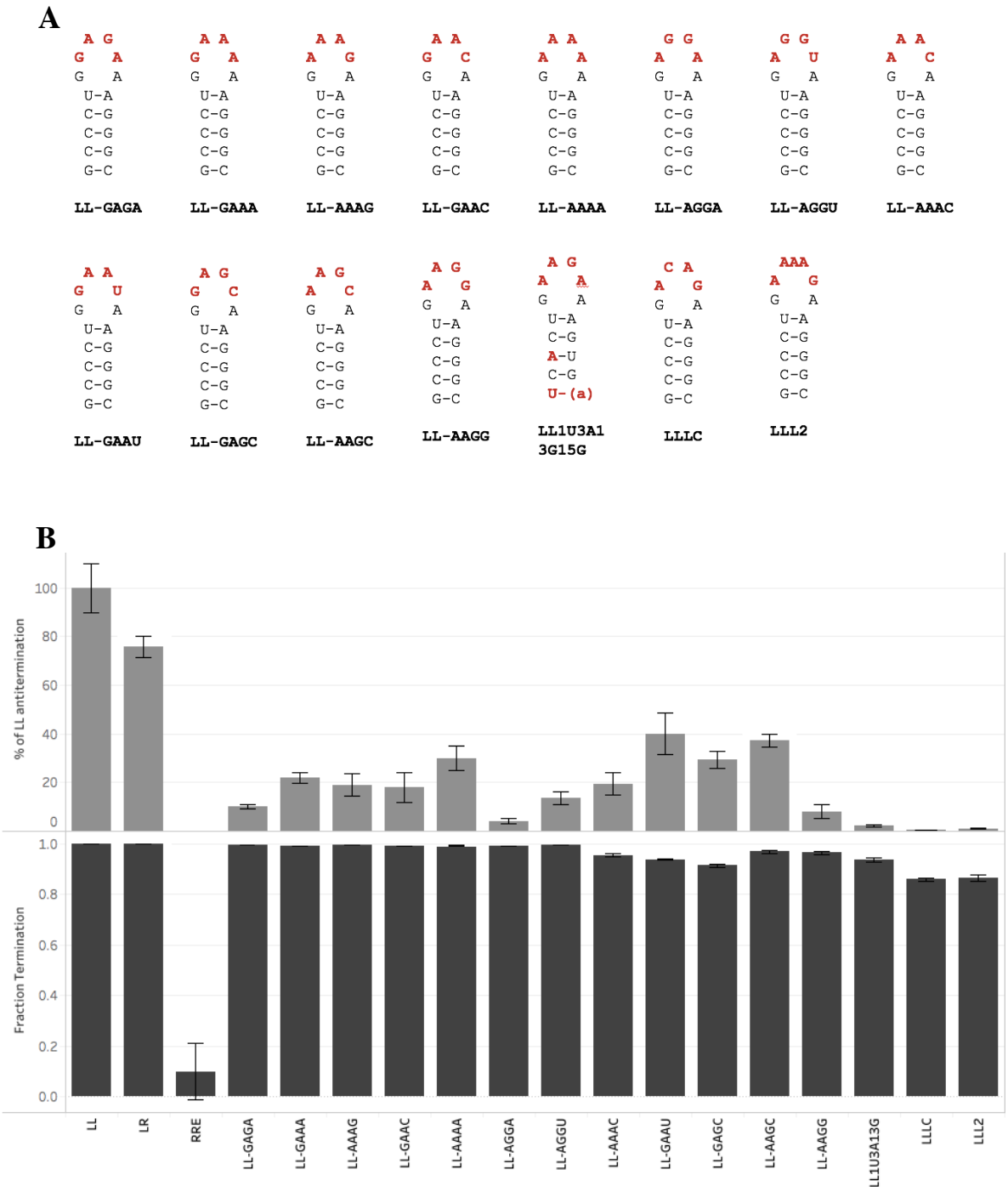


Figure 13. (A) Secondary structures of expanded-loop boxB sequences. The first structure is a reference for nucleotide numbering. Sequence deviations from λ $boxB_{left}$ or $boxB_{right}$ are shown in bold red. The reference name of each sequence is shown in bold under the hairpin structure. (B) % of LL Antitermination (light grey bars) is: ß-galactosidase activity of the mutant boxB relative to λ $boxB_{left}$. Fraction termination (dark grey bars) is: phoA units of wtNun-mutant boxB/phoA units of RevN-mutant boxB. Error bars are standard deviations from the mean values of 4 replicates.

D. Comparing the mutability of N and Nun ARMs

Although N ARM has been closely examined with mutagenesis and biophysical studies, there are still gaps in its mutagenesis data and discrepancies with the two published NMR N ARM-boxB structural models (Legault et al., 1998, Scharpf et al., 2000, Franklin, 1993). Published mutational analysis reports that lambda N residues A3, R7, R8, R11, and W18 are not mutable (Franklin, 1993), but NMR data reveal boxB bases making specific contacts to D2, T5, R6, R7, E9, E13, Q17, and W18 (Legault et al.,1998; Schärpf et al., 2000). To resolve these discrepancies and to directly compare N ARM requirements to those of Nun ARM, the mutational profile of N ARM was determined (Figure 14). Seventeen N ARM libraries, each with a singly randomized codon position, allowed complete mutagenesis one residue at a time. The libraries were cloned into pBRN plasmids. Fifty clones were randomly selected from each library and screened for antitermination activity by *in vivo* N-complementation plaque assays. The percentage of active clones at each position revealed which residues are important for N ARM function. Differences in the pattern of critical residues between N and Nun ARM reflected the differences in their binding mode to their common target boxB.

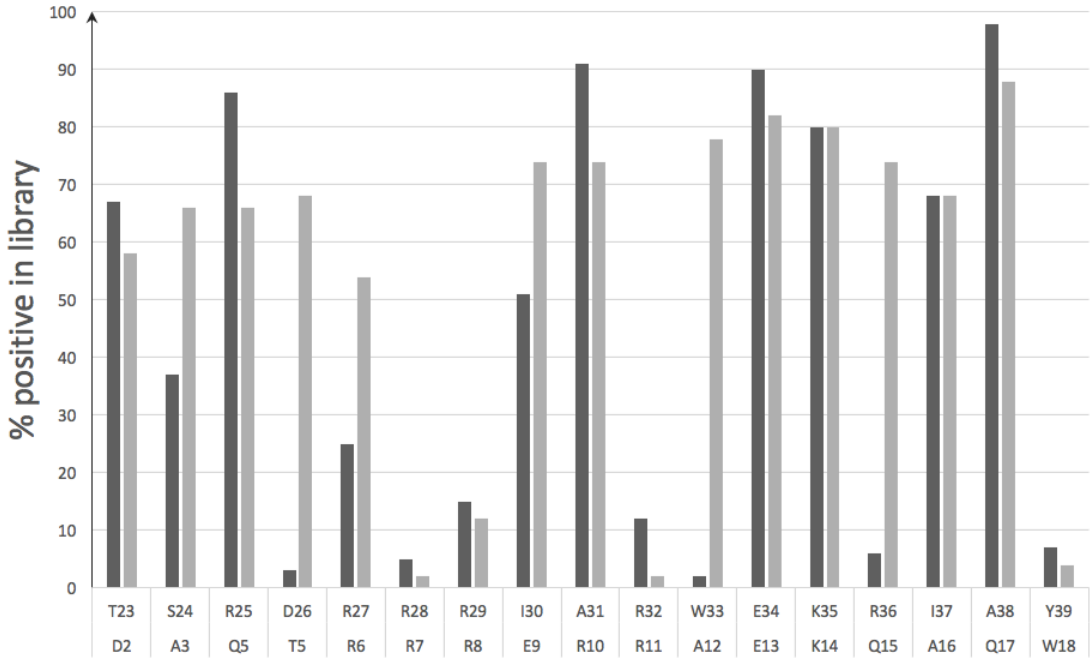


Figure 14. Scanning mutagenesis profiles of N and Nun ARMs spanning λ ARM positions D2 to W18 and the equivalent positions in Nun ARM T23 to Y39. The x-axis represents the percentage of active ARM single mutants at each position: light grey bars for λ and dark grey bars for Nun. The y-axis represents the aligned Nun (upper) and λ (lower) ARM residues. Nun ARM mutagenesis data is from (Tawk et al., 2015).

According to NMR data (Legault et al., 1998), Ala-3 is involved in a sequence non-specific hydrophobic interaction at the base of the boxB stem. Ala-3 methyl group is buried in a hydrophobic pocket formed by the riboses and bases of C4 and C5 boxB stem nucleotides. According to the mutational profile determined in this study, Ala-3 position is highly mutable. This contradicts the reported significant contribution of Ala-3 to the N ARM peptide-boxB complex stability. It has been shown that Ala-3 can only be substituted by a serine without disrupting boxB binding and even an Ala-3 to Ser substitution showed a 20% decrease in binding affinity in gel mobility shift assays (Su et al., 1997a; Franklin, 1993). Interestingly, alanine is a conserved residue in the homologous N ARMs of the closely related bacteriophages λ, P22 and φ21, and the λ N

ARM Ala-3 equivalent in Nun ARM is the Ser-24 residue. The analogous Nun ARM mutational profile, determined in the Nun termination context by *in vivo* λ exclusion plaque assay (Tawk et al., 2015), shows a higher restriction at position Ser-24 yet Trp, Tyr, Cys, and Ala substitutions were all functional. Although active clones were not sequenced at this position in N ARM, given a high percentage (66%) of active clones, it is plausible to assume that a wide variety of amino acid substitutions are expected to work.

R7, R8, and R11 are immutable positions in N ARM, whereas R6 and R10 are relatively more tolerant of mutation. These results match very well with NMR structural data (Legualt et al., 1998) and published mutational analysis of these positions (Franklin,1993). According to the NMR structural model, R6 and R10 interact with the boxB stem through electrostatic interactions only, and thus can be replaced by lysines. R7, R8, and R11 interact with the boxB loop and the U-A loop-closing base pair. Lysine substitutions at those three arginine positions were not functional, indicating that their role goes beyond their positive charge contribution. (Franklin, 1993; Su et al., 1997a). The equivalents of R7, R8, and R11 residues in Nun are also arginines (R28, R29, and R32 respectively): they interact similarly with boxB and are comparably immutable (Faber et al., 2001; Tawk et al., 2015).

In the Trp18X library, only 4% of the tested clones were active indicating a very restricted position. Trp-18 is involved in an important hydrophobic interaction, stacking its indole ring upon the base of the second nucleotide (A7) of the boxB loop. All possible 19 mutations at this positions were examined before by Sue et. al., and all reduced binding affinity. Of these substitutions, Tyr and Phe were the most neutral due to their aromatic side chains (Su et al., 1997a). The high restriction at this position is

shared by the Nun ARM equivalent Y39 yet Y39 can be substituted by Lys, Phe, and His (Tawk et al., 2015).

Comparison of the mutational profiles of Nun and N ARMs brings notice to the greater variability in the degree of restriction of Nun ARM residues ranging from strictly immutable to highly tolerant of substitution. In N ARM, most positons are highly robust to amino acid changes. Most importantly, the equivalent positions of Nun Asp-26, Trp-33, and Arg-36 residues are Thr-5, Ala-12, Gln-15: although critical to Nun function, the three positions are highly mutable in N. Distinct patterns of requirements are clearly detected between N and Nun ARMs. This supports well the suggested difference in the way N and Nun ARM engage with boxB for productive binding. In general, the data agrees with published mutagenesis data except for Ala-3. Although Gln-4, Lys-14, and Gln-15 are reported by NMR structural data to be involved in specific interactions to boxB, this specificity is not reflected as restricted tolerance for mutations in our profile. These N ARM scanning mutagenesis data confirmed Franklin's results (1993) within our system and allowed a direct comparison of N and Nun requirements. Determining N ARM mutational profile was valuable for the design of the series of incremental hybrids.

E. λ N recognition strategy cannot be recapitulated in the termination context

In related bacteriophage N proteins, antitermination function has been attributed to the activation domains at their carboxy termini, which are interchangeable, whereas specificity of interaction with *nut* sites has been attributed to the arginine-rich regions at their amino termini (Lazinski et al., 1989; Franklin, 1993). N and Nun amino acid sequences were inspected by Henthorn and Friedman (1996). By examining different hybrid combinations of N and Nun similar regions, they assigned Nun-

termination and N-antitermination functions to the carboxy domains (activation domains) and revealed the interchangeability, and thus modularity, of Nun and N RNA-binding domains at the amino termini. The work of a previous lab member (Tawk et al., 2015) showed that Nun ARM fused to N activation domain mediates antitermination. However, replacing Nun ARM with N ARM residues failed to support termination (casNun and LN in Table 1).

To address this discrepancy, it was hypothesized that this is a consequence of the cloning strategy which did not take differences in ARM length into account. casNun ARM is longer than wtNun ARM by one residue at its carboxy terminus, and by two residues at its amino terminus; these changes were necessary to create restriction sites for cloning. Nun ARM (casNun) is 4 residues longer than N ARM (LN) at its carboxy terminus, and 2 residues longer at its amino terminus. Such a difference might be affecting the positioning of the ARM residues at their interface with boxB residues. To test this, we designed alignN and cNNun ARMs. The former is totally aligned with wtNun ARM, the latter with casNun. However, when cloned into pBADcasNun both ARMs failed to exclude lambda infection in an *in vivo* plaque assay, indicating their inability to mediate Nun termination (Table 1)

Notably, Nun has a stretch of 21 residues at the amino terminus of its ARM. Whereas in N, the first residue of the ARM is also the first residue of the peptide (Figure 15). Henthorn and Friedman (1996) reported fusing N ARM to Nun activation domain, without any Nun residues amino to N ARM, yielded a functional N-Nun

Table 1. λ N^- complementation and wtλ exclusion functions of Nun-N hybrids

ARM[a]		N-Complementation	Lambda Exclusion
		N^- λ plaques	wtλ plaques
wtN	-MDAQTRRRERRAEKQAQWK-		
LN	-MDAQTRRRERRAEKQAQWN-	Y	Y
wtNun	-DRGLTSRDRRRIARWEKRIAYA-	Y	N
casNun	-MDRGLTSRDRRRIARWEKRIAYANA-	Y	N
alignN	-MRGMDAQTRRRERRAEKQAQWA-	N	Y
cNNun	-MDRGMDAQTRRRERRAEKQAQWANA-	N	Y

[a] The designed ARMs were fused to the N activation domain and examined for supporting growth of N-deficient λ. The same ARMs were fused to the Nun activation domain and examined for the ability to exclude wt λ lytic growth.
[b] N567 cells carrying the indicated hybrid ARM-N fusions were infected with a λ strain deficient in N without IPTG induction. N, N^- λ plaques were not observed; Y, N^- λ plaques were observed.
[c] N567 cells carrying the indicated hybrid ARM-Nun fusions were infected with wild type λ. Functional Nun fusions prevent the formation of λ plaques. N, wtλ plaques were not observed; Y, wt λ plaques were observed.

hybrid protein that supported Nun termination. The cloning strategy herein retains the extra stretch of residues at the amino terminus of Nun when fusing ARMs to Nun activation domains. It is possible these residues are either interfering with proper N ARM orientation with its boxB interface or affecting N ARM stability. Thus, the cloning strategy was altered to recapitulate Henthorn and Friedman's construct within our system. However, this N-Nun fusion (Del-18) failed to exclude lambda infection (Figure 7B). These results suggest that Nun's RNA-binding domain is not as modular as N's. This can be a consequence of differences in constraints and requirements imposed by the ribonucleoprotein complexes rather than a consequence of the cloning strategy. The ARM-boxB interaction is in the context of numerous interpeptide and host factor contacts which differ between the termination and antitermination complexes.

Surprisingly, the Nun-aligned N ARMs, when cloned into pBRN plasmids, failed to complement N-deficient lambda strains (Table 1) even after inducing the promoter with IPTG to increase the level of the expressed protein (supplementary material). It is intriguing how casNun worked when cloned into pBRN, whereas cNNun which is totally aligned with casNun, didn't work in the same antitermination context. Horiya et al., (2009) suggest the distance between N ARM and the remainder of the protein is very strict. Insertion of residues in-between reduced antitermination activity, presumably due to strict spatial requirements involving the formation of N ARM-boxB-NusA ternary complex (Horiya et al., 2009). This spatial constraint is perhaps relaxed when N ARM is replaced with Nun ARM, possibly due to difference in NusA cooperative interaction with Nun ARM-boxB complex. Thus results here confirm the sensitivity of N-binding mode to ARM length variations in the antitermination context.

A

```
                                1
N    ........................MDAQTRRRERRAEKQAQWKAANPLLV
Nun  MLMVKKTIYVNPDSGQNRKVSDRGLTSRDRRRIARWEKRIAYALKNGVTP
     1

N    GVSAKPVNRPILSLNRKPKSRVESALN...PIDLTVLAE..YHKQIESNL
Nun  GFNAID.DGPEYKINEDPMDKVDKALATPFPRDVEKIEDEKYEDVMHRVV

                                        107
N    QRIERKNQRTWYSKPGERGITCSGRQKIKGKSIPLI
Nun  NHAHQRNPNKKWS.......................
                 112
```

B

Protein			λ plaques in exclusion assay
Nun	Nun BD	Nun AD	N
λ N	λ N BD	λ N AD	Y
LN-Nun	λ N BD	Nun AD	Y
Del-18	λ N BD	Nun AD	Y

Figure 15. (A) λN and HK022 Nun amino acid sequences aligned (Henthorn and Friedman, 1996). Identical amino acids are in shown in bold. Each sequence is numbered separately. (B) Lambda infection exclusion mediated by differently cloned ARMs. N, wt λ plaques were not observed; Y, wt λ plaques were observed.

F- Gradual mutation of λN ARM towards Nun ARM

A series of Nun-N hybrid ARMs bearing incremental changes that gradually mutate the N ARM sequence to Nun ARM sequence was designed to examine the possibility of a smooth transition from an N-binding mode to a Nun-binding mode. This

was done through gradually substituting LN ARM residues with DWR residues that are essential to Nun function and creating a Nun-equivalent hydrophobic ridge to enable putative contacts with NusG-CTD which characterize the Nun-binding mode (Table 2). Single substitutions of Thr-5, Ala-12, and Gln-15 to D, W, and R respectively did not affect antitermination function. Normal size lambda plaques were observed in the N-complementation assay. This was expected from the mutational profile: these positions in N, unlike in Nun, are highly mutable (Figure 6). A triple mutation of T512AQ15 to DWR also did not disrupt N function. However, none of these mutant versions of N ARM worked when fused to Nun activation domain. To increase the hydrophobicity at the helical face that bears DWR, an additional double mutation of E9 and A16 to isoleucines was done. The cumulative effect of the five Nun residues did not affect N antitermination function, further confirming the robustness of N ARM sequence. However, the hybrid ARM still did not work fused to Nun, even with an additional mutation of W18 to Y. This last W18Y substitution significantly decreased the size of lambda plaques in N-complementation assay indicating a decrease in antitermination activity. This effect of tryptophan to tyrosine substitution in LNdiwri was not expected considering the functional LN18Y mutant ARM and the proposed similar stacking interaction of their aromatic rings with A7 of boxB loop (Faber et al., 2001; Legault et al., 1998). Nun2338 mutant is an ARM in which the lambda N residues 2 to 17 (DAQTRRRERRAEKQAQ) are replaced by Nun residues 23 to 38 (TSRDRRRIARW EKRIA). This hybrid ARM complemented an N-deficient lambda strain. However, it did not exclude lambda infection even with a subsequent W18Y mutation (Nun2339). Interestingly, upon this further W18Y mutation the plaque size decreased

Table 2. Activity of LN ARM gradually mutating towards Nun ARM.

Hybrid ARM[a]		λ N⁻ plaques[b] no IPTG	λ N⁻ plaques[b] IPTG	wtλ exclusion[c]
LN	MDAQTRRRERRAEKQAQWN A	+++	+++	+++
LNT5D	MDAQDRRRERRAEKQAQWN A	+++	+++	+++
LNA12W	MDAQTRRRERRWEKQAQWN A	+++	+++	+++
LNQ15R	MDAQTRRRERRAEKRAQWN A	+++	+++	+++
LN18Y	MDAQTRRRERRAEKQAQYN A	+++	+++	+++
LN5D12W15R	MDAQDRRRERRWEKRAQWN A	+++	+++	+++
LNdiwri	MDAQDRRRIRRWEKRIQWN A	+++	+++	+++
LNdiwriy	MDAQDRRRIRRWEKRIQYN A	+	++	+++
Nun2338	MTSRDRRRIARWEKRIAWN A	+++	+++	+++
Nun2339	MTSRDRRRIARWEKRIAYN A	+	++	+++
Nun2340	MTSRDRRRIARWEKRIAYANA	-	+	++
casNun	MDRGLTSRDRRRIARWEKRIAYANA	+	++	-
cNun39W[d]	MDRGLTSRDRRRIARWEKRIAWANA	na	na	-

[a] The designed ARMs were fused to the N activation domain and examined for supporting growth of N-deficient λ. The same ARMs were fused to the Nun activation domain and examined for the ability to exclude wt λ lytic growth.
[b] N567 cells carrying the indicated hybrid ARM-N fusions were infected with a λ strain deficient in N without or with 300 M IPTG induction. +++, wild type-size plaques with LN being the reference; ++, small plaques; + tiny plaques; - no plaques observed.
[c] N567 cells carrying the indicated hybrid ARM-N fusions were infected with wild type λ. Functional Nun fusions prevent the formation of λ plaques. Plaque size is reported as explained in footnote *b*.
[d] cNun39W was not cloned into the N activation domain. na stands for not available.

significantly in the N-complementation assay mirroring LNdiwri to LNdiwriy W18Y substitution effect. Inserting an alanine residue after Y18 in Nun2339, aligns the carboxy terminal of the ARM with casNun. The obtained hybrid ARM Nun 2340 fails to support antitermination (fused to N) but shows an increase in termination activity (fused to Nun) which manifests as a decrease in the plaque size in a lambda exclusion assay. All (Table 2) results showed that gradually mutating N ARM toward Nun ARM did not affect N's antitermination function but did not enhance N ARM ability to mediate Nun's termination function.

That no member of this series of incremental transitions (Table 2), especially mutants Nun2338 and Nun2339, had Nun termination activity prompted the examination of the MDRG sequence which is at the amino terminus of Nun ARM, for any hidden requirements (Table 3). These residues were not included in the Nun ARM mutational profile (Figure 6). Scanning alanine mutagenesis was used to test the effect of every residue separately. None of the single alanine mutants affected termination activity. However, when all four residues were mutated to alanine, the construct failed to exclude lambda infection. It is very likely that this stretch of amino acids is involved in interpeptide stabilizing contacts which explains why deleting it is compromising to Nun ARM-boxB mediated termination. This might also explain why the set of hybrid ARMs in Table 2, which all lacked the MDRG sequence, did not work in the Nun-termination context no matter how close the ARM sequence was to Nun ARM sequence.

Table 3. Alanine mutagenesis of MDRG sequence.

ARM[a]		wt λ exclusion[b]
casNun	M**DRGLTSRD**RRR**IA**R**W**EK**RIAYA**NA	-
NunD19A	M**A**RGLTSRDRRR**IA**R**W**EK**RIAYA**NA	-
NunR20A	M**DA**GLTSRDRRR**IA**R**W**EK**RIAYA**NA	-
NunG21A	M**DRA**LTSRDRRR**IA**R**W**EK**RIAYA**NA	-
NunL22A	M**DRGA**TSRDRRR**IA**R**W**EK**RIAYA**NA	-
Nun19A22	M**AAAA**TSRDRRR**IA**R**W**EK**RIAYA**NA	++

[a] casNun ARMs with alanine substitutions (indicated in bold black) were fused to Nun activation domain and tested for in vivo exclusion of wild type lambda infection.
[b] N567 cells carrying the indicated hybrid ARM-Nun fusions were infected with wild type λ. Functional Nun fusions prevent the formation of λ plaques. – no plaques observed; ++ small plaques observed.

To avoid the compromising effect of MDRG deletion, the same incremental changes of (Table 2) were done starting with cNNun sequence (aligned to casNun). The set of cNNundwr to cNNundiwriy ARMs (Table 4) represent the same mutations as in LNdwr to LNdiwriy (Table 2). A triple mutation to substitute TAQ with the Nun-required DWR residues was not enough to make the hybrid ARM mediate lambda lytic growth exclusion (cNNundwr). Interestingly, upon the further increase of hydrophobicity through one leucine and two isoleucine substitutions, lambda plaques decreased considerably in size, indicating an increase in the termination activity of cNNundiwri, but apparently not above a certain threshold to yield a full exclusion of lambda infection. This confirms the importance of the hydrophobic ridge for a productive Nun ARM-boxB interaction. Notably, a single-residue scanning mutagenesis

would naturally miss the importance of the combined effect of multiple residues as with the combined effect of hydrophobic side chains at one face of the Nun ARM helix. A further substitution of Trp to Tyr (cNNunldiwiriy) contributed to a complete absence of plaques in the lambda exclusion assay indicating a fully functional hybrid ARM-boxB interaction. Again, the observed effect of Trp to Tyr substitution was not expected with reference to NMR structural models of λN ARM-boxB and Nun ARM-boxB interactions. The examined hybrid ARM in (Table 4) revealed that incremental addition of Nun-required residues gradually increased the hybrid ARM ability to exclude lambda infection in the presence of the MDRG required sequence.

The same set of hybrid ARMs (Table 4) was tested for ability to complement N-deficient lambda strain growth, by fusion to N activation domain. As mentioned before, N ARM aligned to casNun ARM failed to complement N-deficient λ growth. That did not change with the subsequent mutations from cNNundwr to cNNundirwiy. Although this further confirms that N-binding mode does not tolerate variations in ARM length, it is quite puzzling how casNun works when fused to N while the very closely resembling sequence cNNundiwriy does not.

Table 4. Activity of cNNun ARM gradually mutating towards Nun ARM.

ARM[a]		N-deficient λ[b]		wtλ
		no IPTG	IPTG	exclusion[c]
cNNun	MDRGMDAQTRRRERRAEKQAQWANA	-	-	+++
cNNundwr	MDRGMDAQDRRRERRWEKRAQWANA	-	-	+++
cNunldiwri	MDRGLDAQDRRRIRRWEKRIQWANA	-	-	+
cNNunldiwriy	MDRGLDAQDRRRIRRWEKRIQYANA	-	-	-
casNun	MDRGLTSRDRRRIARWEKRIAYANA	+	+	-
cNun39W	MDRGLTSRDRRRIARWEKRIAWANA	na	na	-

[a] The designed ARMs were fused to the N activation domain and examined for supporting growth of N-deficient λ. The same ARMs were fused to the Nun activation domain and examined for the ability to exclude wt λ lytic growth.
[b] N567 cells carrying the indicated hybrid ARM-N fusions were infected with a λ strain deficient in N without or with 300 M IPTG induction. +++, wild type-size plaques with LN being the reference; ++, small plaques; + tiny plaques; - no plaques observed.
[c] N567 cells carrying the indicated hybrid ARM-N fusions were infected with wild type λ. Functional Nun fusions prevent the formation of λ plaques. Plaque size is reported as explained in footnote *b*.
[d] cNun39W was not cloned into the N activation domain. na stands for not available.

CHAPTER IV

DISCUSSION

A. The Difference in N and Nun contacts with NusG influence the ARM-boxB interaction

As visualized in (Figure 16), Nun and N bind distinct but overlapping panels of boxB mutants. Selecting for active boxBs N or Nun found many active with both, some active with Nun only, and none active with N only. Most tested boxB variants are very active with Nun with a fraction termination values above 0.9. Of the examined panel, only few boxB mutants were not functional in the termination context. Distinctively, boxB mutants screened for N antitermination revealed a distribution over a wider range of activity. Identifying Nun-specific boxBs supports a model in which the differences in N and Nun contacts with NusG influence the ARM-boxB interaction. Considering the mutual structural rearrangements that occur upon binding of Nun or N ARM to boxB (Faber et al., 2001; Legault et al., 1998), and the energetic consequences of such adjustments, it is conceivable that further conformational changes occur upon the putative binding of NusG-CTD to the Nun-ARM alpha helix, affecting to some extent the ARM-boxB interface. In the absence of an equivalent NusG binding to N ARM at the same site, the panel of boxBs capable of recruiting lambda N and HK022 Nun would be different.

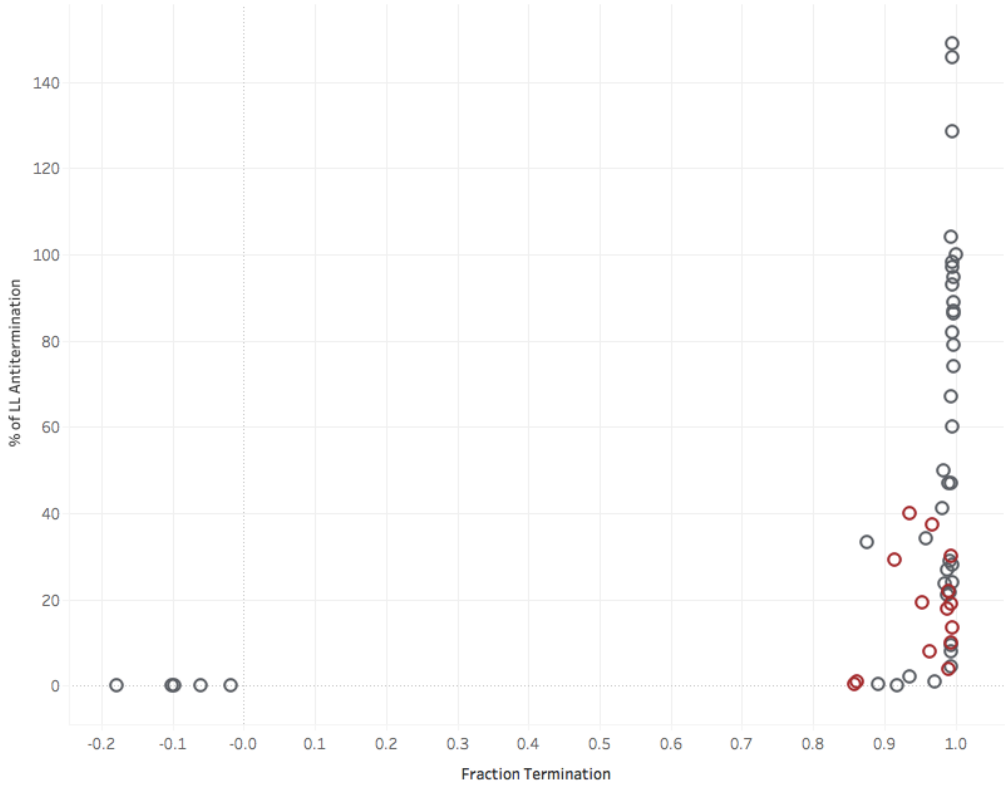


Figure 16. Distribution of boxB specificities. Every circle is a boxB mutant. Red circles are expanded-loop boxBs. Grey circles are boxBs with pentaloops.

B. The difference in tolerance to boxB-loop expansion reflects differences in spatial requirements

A hexaloop boxB library with the four apical nucleotides randomized was screened for N and Nun activities. Many Nun-specific expanded loops were discovered, and all functional hexaloops are related to relaxed specificity boxBs by single insertions. Only few hexaloop mutants were moderately to weakly functional with N, and the N-functional hexaloops did not achieve high levels of antitermination activity, the highest is 40% of the antitermination on LL. The only designed heptaloop was completely nonfunctional with N. These results show that boxB Pentaloop expansion does not favor N-mediated antitermination. This is in agreement with Cocozaki et al. (2008a) report of loss of λ and P22 N antitermination activity with hexa- and heptaloop

P22 boxBs.

The GNRA-tetraloop fold is crucial for productive N ARM-boxB interaction. BoxB pentaloop adopts this structural motif by extruding the fourth adenosine base (Legault et al., 1998; Scharpf et al., 2000). Additionally, NusA interaction with the extruded base is critical for N antitermination (Watnick and Gottesman, 1999; Mogridge et al., 1995). It is plausible to speculate that hexa- and heptaloops maintain the essential the GNRA-tetraloop motif by protruding the extra bases. The occurrence of GNRA-like tetraloop folds within hexa-and heptaloops by base extrusion is well supported in literature (Huang et al., 2005; Klein et al., 2001). The loss of antitermination activity upon loop expansion can be explained by the extruded nucleotides interference through steric hindrance with proper NusA recognition of the boxB-N ARM complex. Published data support strict spatial requirements for the formation of a stable ternary complex between N ARM-boxB and NusA (Horiya et al., 2009).

Although boxB loop adopts a similar fold in its complex with Nun ARM (Faber et al., 2001), it is not clear whether the extruded base is as critical an anchorage site for NusA in Nun termination as in N antitermination. Nun tolerance to loop expansion, and presumably base extrusion, might be attributed to differences in NusA interaction with ARM-boxB between the termination and antitermination complexes or perhaps to more relaxed spatial requirements and constrains in the vicinity of the extruded loop bases.

C. Neutral Evolution of boxB specificity

The neutral evolution of RNA secondary structures has been extensively

studied computationally but with very limited experimental examination within biologically meaningful contexts. For small RNA hairpin sequences that interact with ARMs, switching of specificity by few single step mutations has been experimentally demonstrated (Iwazaki et al., 2005; Smith et al., 1998; Cocozaki et al., 2008a). Data herein demonstrated that single-nucleotide mutations, substitutions or insertions, are capable of changing a boxB sequence with relaxed specificity to a highly specific variant. This validates RNA molecular speciation through genetic drift along neutral networks.

In the biological context, boxB cannot evolve specificity to Nun. BoxB is a genetic element that belongs to λ left and right operons. A productive N ARM-boxB interaction is indispensable for lambda lytic growth and replication on which the propagation of boxB sequence depends on. A λ strain with Nun-specific boxB is necessarily defective in replicating itself. This renders the evolution towards Nun-specificity non-selected.

The absence of N-specific boxBs reveals that the ability of bacteriophage λ to escape HK022 Nun is constrained by the topology of the local fitness landscape of the N ARM-boxB interaction set by the wild type N ARM sequence. For λ phage to escape Nun termination, change must occur at the level of the N sequence which might consequently change this disadvantageous fitness topology. Notably, Franklin's genetic analysis (1993) reported the absence of N ARM mutations capable of compensating the compromising effect of boxB loop mutations. Coevolution of N protein and boxB sequence would extend the sampling of sequence space beyond this experimentally examined local evolutionary landscape of λ N ARM-boxB interaction.

D. N ARM sequence is very robust and doesn't have an equivalent to Nun's DWR requirements

This study corroborates lambda N ARM is highly mutable. In general, the results are consistent with the importance of Arg-7, Arg-8, Arg-11, and Trp-18 suggested by *in vitro* and *in vivo* analysis (Legault et al., 1998, Scharpf et al., 2000, Franklin, 1993; Su et al., 1997a). Yet, the high mutability of Ala-3 is in conflict with the reported essential hydrophobic interaction that anchors the ARM at the base of boxB stem. As observed in the scanning mutagenesis profiles, N ARM is more robust to mutations compared to Nun ARM. In addition to common expected requirements, Nun ARM has restrictions that are not present in N ARM. Most importantly, lambda ARM scanning mutagenesis confirmed the high mutability of Thr-5, Ala-12, and Gln-15 residues, the equivalent positions of the immutable Asp-26, Trp-33, and Arg-36 Nun residues. As first noted in Faber et al. (2001), the most distinctive feature of Nun ARM is the amphipathic nature of its alpha helix, where the most hydrophobic residues are clustered on the side facing away from boxB. Nun ARM D26W33R36 requirements intriguingly align with the hydrophobic ridge. The residues contributing to the hydrophobicity are mainly Leu-41, Ile-37, Trp-33, Ile-30, Leu-22. Although Leu-41 was not included in the Nun ARM scanning mutagenesis (Tawk et al., 2015), Ile-37, Ile-30, Leu-22, excluding the critical Trp-33, are individually mutable. The examined set of N-Nun hybrids ARMs confirmed Nun's D26W33R36 requirements. The importance of the Nun's hydrophobic ridge is also established by showing that adding the critical DWR residues to N ARM is not enough to elicit the exclusion of lambda lytic growth when fused to Nun. Data herein shows that a considerable decrease in lambda plaque size, in an *in vivo* lambda exclusion assay, is observed only after three amino acid substitutions in cNNunDWR introduce the hydrophobic Leu22, Ile30, and

Ile37 residues. Differentially, N ARM helix is hydrophilic with no equivalent requirements on the same face of the helix.

E. The neutrality of Trp18 and Tyr39 interchangeability depends on the sequence context

N Trp-18 and Nun Tyr-39 appear to participate in the same essential stacking interaction with the apical pentaloop base of boxB (Legault et al., 1998; Scharpf et al., 2000; Faber et al., 2001). Both positions are highly restrictive in their respective ARMs (Figure 14). A single Trp to Tyr substitution in λN ARM has no observable effect on the plaque size in N-complementation assay (LN18Y in Table 2). Likewise, the substitution of Tyr-39 with Trp in casNun is neutral in the lambda exclusion assay (cNun39W in Table 2). However, data herein showed that the interchangeability of N Trp-18 and Nun Tyr-39 is not independent of sequence changes at other positions in the ARM. A Trp to Tyr substitution in the context of LNdiwri and LNun2339 hybrid ARM sequences was compromising to antitermination activity (Table 2). Similarly, a Trp to Tyr substitution in cNNunldiwiriy rendered the hybrid N-Nun ARM incapable of fully mediating lambda growth exclusion (Table 4). Xia et al., (2003) observed that the fully stacked conformation of λN Trp-18 is essential for N antitermination and is dependent on the sequence context wherein changes at positions 14 and 15 can cause partial or complete unstacking of the Trp aromatic side chain. The examined specific hybrid ARMs either suggest subtle differences in the structural roles of Trp-18 in N and Tyr-39 in Nun, possibly imposed by the difference in termination and antitermination complexes, or that the compromising W-Y mutations is the result of complex effects of multiple mutations at neighboring positions that affect the extent of stacking of their aromatic side chains with boxB loop nucleotides.

F. Nun ARM function is not modular

Data herein support that Nun ARM function in the Nun-termination complex is likely not modular. λ N ARM did not work in Nun despite all considerations for phasing differences between N and Nun ARMs (Table 1 and Figure15B). Loss of function was also observed with Nun fusions to different binding domains: φ21 N ARM, P22 N ARM, and P22 ARM mutant with relaxed specificity (Cocozaki et al., 2008b; Tawk, 2011). It seems that heterologous ARMs are not tolerated in Nun. This is in marked contrast to the modularity of λN-boxB antitermination-mediating interaction which can be replaced with a diversity of heterologous arginine-rich peptide-RNA interactions: U1A protein interaction with U1 hairpin II, HIV Rev ARM interaction with RRE, BIV Tat ARM interaction with TAR. Indeed, λ antitermination system has been extensively used to examine binding affinities between many heterologous arginine-rich peptides and their RNA targets (Harada et al., 1996; Harada and Frankel, 1999; Wilhelm and Vale, 1996; McColl et al., 1999; Peled-Zehavi et al., 2003). This stark difference in modularity might be due to requirements imposed by the Nun-termination complex that all the examined heterologous ARMs do not satisfy, presumably the ability to bind NusG.

G. Speculations on NusG-CTD within the Termination and the Antitermination Complexes

Nun ARM relies on NusG-contacting residues D26W33R36 embedded in a functionally important hydrophobic ridge. This holds true in the termination and the antitermination contexts since the immutability of D26 and W33 is preserved when Nun

ARM is fused to N (Tawk et al., 2015). This suggests that NusG-CTD binding is required not just for Nun termination, but for Nun ARM interaction with boxB. Nun ARM employs the same binding strategy that is NusG-dependent for a productive binding interaction with boxB. It is possible that the contacts with NusG are needed to stabilize the Nun ARM-boxB interaction by covering the solvent-exposed hydrophobic site, or that NusG binding elicits subtle, but essential, conformational rearrangements in Nun ARM, boxB, or both, to successfully mediate subsequent assembly of host factors within the termination or antitermination complexes. Interestingly, lambda N ARM alpha helix is hydrophilic and has very different and highly mutable residues at the positions corresponding to DWR. It's worth noting that Nun ARM-NusG-CTD docking is abolished upon substituting Nun W33 with Ala which is the N lambda residue at the equivalent ARM position (Tawk et al., 2015). This suggests that NusG-CTD does not bind N ARM. Thus, N's recognition strategy, unlike Nun's, is NusG-independent.

In the light of all the mentioned above, it is tempting to speculate about the localization and role of NusG within the termination and antitermination complexes. It could be that NusG-CTD is available for binding in both N-antitermination and Nun-termination complexes, and that Nun ARM binding to, and thus sequestering, NusG-CTD is necessary for termination but not important for antitermination. This would rationalize why Nun ARM-NusG binding does not disrupt the antitermination process, and how all tested N ARMs fail to support termination (most likely due to inability to sequester NusG-CTD). This agrees well with the isolation of many NusG mutants that abolish Nun termination (Burova et al., 1999; Mooney et al., 2009) and the inability to find any NusG point mutants that abolish N antitermination.

Progress has been made in the elucidation of the structural basis of Nun-

termination and N-antitermination (Said et al., 2017; Kang et al., 2017), yet NusG interaction with N and Nun are still not resolved. NusG consists of two structurally independent domains, NTD and CTD, separated by a flexible linker (Mooney et al., 2009). The most recent crystal structure of the N antitermination complex does not have NusG (Said et al., 2017). The complete electron cryo-microscopic structure of the antitermination complex (Said et al., 2017) localizes NusG-NTD, but not NusG-CTD, in agreement with the published NusG-RNAP crystal structure (Liu and Steitz, 2017). Importantly, NusG-CTD is revealed bound to NusE (Said et al., 2017). Still, it is conceivable that the NusG's flexible linker would allow NusG-CTD to reach N ARM-boxB at a certain stage of the catalytic cycle of the transcription elongation complex (TEC). Structural models of the TEC are static snapshots of one state, other states presumable exist.

H. The neutral networks of N and Nun recognition strategies are not proximate in sequence space in the termination context

The examined set of hybrid ARMs (Table 2) revealed the possibility of a smooth incremental transition of wild type N ARM sequence from N residues to Nun residues in the context of N antitermination as long as wild type N ARM length is maintained. However, it is unclear whether there was a switch from N-binding mode (with hydrophilic helix) to Nun-binding mode (hydrophobic helix with NusG-contacting residues), or that was merely the very high robustness of N-binding mode to mutations. casNun is the only ARM sequence that was found functional in both contexts. Addressing this within the framework of neutral networks theories, casNun is at the intersection of the neutral networks of N and Nun recognition strategies. Such smooth transition was not possible within the termination context without complex

mutational leaps. This is not surprising: Nun's recognition strategy is complex and involves a combination of multiple requirements while N's recognition strategy, in comparison, is simpler and more robust. Within the termination context, the neutral networks of N and Nun recognition strategies are found very distant in sequence space. Although the sample of hybrid ARMs was very limited, it seems unlikely that neutral networks of Nun and N recognition strategies intersect in their local evolutional fitness landscape set by the wild type lambda boxB LL and LR sequences. The genetic recombination that takes place among the large populations of lambdoid phages (Campbell, 1994; Casjens, 2005) might allow a more extensive and diverse sampling of sequence space compared to mutations through continuous small incremental steps. The complexity of the difference between the ribnucleoprotein termination and antitermination machinery renders the neutral networks approach very complicated. It is also plausible that the evolvability of RNA-ARM recognition strategies differs when the sequences evolve within the same biochemical context (lambdoid phage antitermination) versus across different contexts (Nun termination and N antitermination). The evolution of ARM-RNA interaction specificities faces constraints when the interaction switches contexts as is the case with N and Nun ARM-boxB interactions.

I. Unresolved questions and planned experiments

This limited exploration of the local evolutionary landscapes of Nun ARM- and N ARM-boxB interactions can aid the design of experiments for further examination of the complexes. For instance, to confirm that NusG binding to Nun ARM has conformational implications that encompass the ARM-boxB interaction, the identified set of Nun-specific boxBs can be screened with casNun fused to N. If Nun

ARM indeed employs the same NusG-dependent recognition strategy in the termination and antitermination contexts, then the Nun-specific boxBs should be active with this casNun-N fusion when screened for antitermination activity. In the series of hybrid ARMs LN to Nun2340 that progress gradually from N-like ARM sequence to Nun-like ARM sequence without loss of antitermination function, it remains unclear whether there is a switch from an N-binding mode to a Nun-binding mode. This can also be resolved by screening for switch in boxB specificity using the Nun-specific boxB panel.

APPENDICES

I. Construct Sequences

Table 5. Sequences of the synthetic oligonucleotides

Protein	NcoI-BsmI casette
λ N (LN)	5'-C ATG GAT GCA CAA ACA CGC CGC CGC GAA CGT CGC GCA GAG AAA CAG GCT CAA TGG AAT GCA-3'
HK022 Nun	5'-C ATG GAT AGA GGT CTT ACA TCT CGA GAC AGG AGG AGA ATA GCG AGA TGG GAA AAA AGG ATA GCA TAT GCG AAT GCA-3'
LNT5D	5'-C ATG GAT GCA CAA GAC CGC CGC CGC GAA CGT CGC GCA GAG AAA CAG GCT CAA TGG AAT GCA-3'
LNA12W	5'-C ATG GAT GCA CAA ACA CGC CGC CGC GAA CGT CGC TGG GAG AAA CAG GCT CAA TGG AAT GCA-3'
LNQ15R	5'-C ATG GAT GCA CAA ACA CGC CGC CGC GAA CGT CGC GCA GAG AAA AGG GCT CAA TGG AAT GCA-3'
LN18Y	5'-C ATG GAT GCA CAA ACA CGC CGC CGC GAA CGT CGC GCA GAG AAA CAG GCT CAA TAT AAT GCA-3'
LN5D12W15R	5'-C ATG GAT GCA CAA GAC CGC CGC CGC GAA CGT CGC TGG GAG AAA AGG GCT CAA TGG AAT GCA-3'
LNdiwri	5'- C ATG GAT GCA CAA GAC CGC CGC CGC ATA CGT CGC TGG GAG AAA AGG ATA CAA TGG AAT GCA-3'
LNdiwriy	5'-C ATG GAT GCA CAA GAC CGC CGC CGC ATA CGT CGC TGG GAG AAA AGG ATA CAA TAT AAT GCA-3'
Nun2338	5'-C ATG ACA TCT CGA GAT CGC CGC CGC ATT GCG CGC TGG GAA AAA CGC ATT GCA TGG AAT GCA-3'
Nun2339	5'-C ATG ACA TCT CGA GAT CGC CGC CGC ATT GCG CGC TGG GAA AAA CGC ATT GCA TAT AAT GCA-3'
Nun2340	5'-C ATG ACA TCT CGA GAT CGC CGC CGC ATT GCG CGC TGG GAA AAA CGC ATT GCA TAT GCG AAT GCA-3'

cNun39W	5'-C ATG GAT AGA GGT CTT ACA TCT CGA GAC AGG AGG AGA ATA GCG AGA TGG GAA AAA AGG ATA GCA TGG GCG AAT GCA-3'
cNNun	5'-C ATG GAT AGA GGT ATG GAT GCA CAA ACA AGG AGG AGA GAA CGT AGA GCA GAA AAA CAG GCT CAA TGG GCG AAT GCA-3'
cNNundwr	5-C ATG GAT AGA GGT ATG GAT GCA CAA GAC AGG AGG AGA GAA CGT AGA TGG GAA AAA AGG GCT CAA TGG GCG AAT GCA-3'
cNunldiwri	5-C ATG GAT AGA GGT CTT GAT GCA CAA GAC AGG AGG AGA ATA CGT AGA TGG GAA AAA AGG ATA CAA TGG GCG AAT GCA
cNunldiwriy	5-C ATG GAT AGA GGT CTT GAT GCA CAA GAC AGG AGG AGA ATA CGT AGA TGG GAA AAA AGG ATA CAA TAT GCG AAT GCA

II- Plate Assay Results

Table 6. Antitermination and termination activities of mutant boxBs

	XGal		XP	
boxB	N	RevN	Nun	RevNun
LL	5	0	1	5
LR	5	0	1	5
RRE	0	3	5	5
LRG6A	0	0	4-5	5
L7U	4	0	2	5
L7C	4	0	2	5
D20	5	0	1.5	5
D22	4	0	3.5	5
L9U	4	0	1-2	5
L9C	4	0	1	5
LR10G	3	0	3.5-4	5
LL10G	4	0	3	5
LL10U	4	0	3.5-4	5
LL7G9C	4	0	1-2	5
L7G8G9U	4.5	0	2-3	5
LL6A8U9U	0	0	5	5
L8G9U10U	2	0	4	5
LL11G	5	0	4	5
LL5AU	4.5	0	3.5-4	5
LL5GC	5	0	4	5
LL5GU	0	0	5	5
LL5CC	5	0	4-5	5
LRC11U	4	0	3.5-4	5
L3U	4	0	3	5
L3GU	5	0	4.5	5
LST3W	0	0	na	5
LSTFLP	3.5	0	2-3	5
LSTAU	2	0	5	5
LLSTEM4	3.5	0	1-2	5
LL5CG	4.5	0	2	5
LLSTEM3	4	0	1-2	5
LL1AU3U	4	0	2	5
LR5CU6C	0	0	4-5	5
LR5AG6	0	0	4-5	5
LL1U3A13G	2	0	2.5	5
LLLC	0	0	5	5

LLL2	1	0	4	5
D1	3-3.5	0	3	4
D2	na	na	1	4
D3	4	0	1	5
D4	4	0	1-1.5	4
D7	3	0	1.5	5

III- Solution Assay Results

Table 7. Antitermination and termination activities of mutant boxBs

boxB	% of LL Antitermination	Fraction Termination
LL	10 ± 6.529	1 ± 0.001
LR	74 ± 14.225	0.997 ± 0.001
LRG6A	0.2 ± 0.016	-0.098 ± 0.139
L7U	97.2 ± 34.818	0.995 ± 0.001
L7C	47 ± 13.820	0.994 ± 0.001
D20	98.323 ± 5.704	0.996 ± 0.000
D18	47 ± 8.570	0.99 ± 0.001
D22	29.083 ± 3.840	0.992 ± 0.001
L9U	82 ± 22.207	0.996 ± 0.001
L9C	60 ± 8.479	0.996 ± 0.001
LR10G	23.6 ± 7.847	0.984 ± 0.003
LL10G	9.3 ± 4.412	0.994 ± 0.001
LL10U	4.6 ± 1.067	0.993 ± 0.001
LL7G9C	93 ± 9.400	0.995 ± 0.001
L7G8G9U	104 ± 10.507	0.993 ± 0.001
LL6A8U9U	0 ± 0.006	-0.0175 ± 0.121
L8G9U10U	0.9535 ± 0.260	0.97 ± 0.003
LL11G	41.1 ± 11.500	0.982 ± 0.002
LL5AU	67 ± 7.700	0.993 ± 0.000
LL5GC	50 ± 3.900	0.983 ± 0.001
LL5GU	0 ± 0.000	0.917 ± 0.004
LL5CC	34.3 ± 9.400	0.958 ± 0.003
LRC11U	27 ± 8.600	0.988 ± 0.001
L3U	145.7 ± 15.800	0.995 ± 0.000
L3GU	33.2 ± 6.200	0.875 ± 0.051
LST3W	0.2 ± 0.000	-0.178 ± 0.173
LSTFLP	21.8 ± 8.300	0.992 ± 0.002
LSTAU	0.5 ± 0.000	0.892 ± 0.006
LLSTEM4	28.1 ± 13.700	0.995 ± 0.001
LL5CG	149 ± 9.000	0.996 ± 0.001
LLSTEM3	94.84 ± 7.600	0.997 ± 0.000
LL1AU3U	128.5 ± 8.700	0.995 ± 0.001
LR5CU6C	0.22 ± 0.000	-0.102 ± 0.128
LR5AG6	0 ± 0.000	-0.0612 ± 0.102
LL-GAGA	10.1 ± 0.977	0.994 ± 0.001

LL-GAAA	22 ± 2.185	0.99 ± 0.001
LL-AAAG	19 ± 4.541	0.994 ± 0.001
LL-GAAC	18 ± 6.307	0.989 ± 0.000
LL-AAAA	30 ± 4.940	0.993 ± 0.001
LL-AGGA	4 ± 1.129	0.991 ± 0.001
LL-AGGU	13.6 ± 2.650	0.995 ± 0.000
LL-AAAC	19.4 ± 4.504	0.954 ± 0.005
LL-GAAU	40 ± 8.616	0.936 ± 0.002
LL-GAGC	29.3 ± 3.553	0.914 ± 0.006
LL-AAGC	37.3 ± 2.733	0.967 ± 0.007
LL-AAGG	8 ± 2.925	0.963 ± 0.007
LL1U3A13G	2.175 ± 0.428	0.936 ± 0.008
LLLC	0.516 ± 0.041	0.858 ± 0.006
LLL2	1 ± 0.151	0.862 ± 0.013
D1	21.156 ± 5.500	0.988 ± 0.001
D2	86.765 ± 5.600	0.997 ± 0.000
D3	89.031 ± 3.200	0.997 ± 0.001
D4	79 ± 0.800	0.998 ± 0.000
D7	8 ± 1.000	0.993 ± 0.001
L1	23.913 ± 8.700	0.995 ± 0.001
L2	86.344 ± 10.800	0.997 ± 0.002

REFERENCES

Abdallah E. Y., Smith C. A. (2015). Diverse Mutants of HIV RRE IIB Recognize Wild-Type Rev ARM or Rev ARM R35G-N40V. *Journal of Molecular Recognition* 28, 710–721.

Aboul-ela, F., Karn, J., and Varani, G. (1995). The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* 253: 313– 332.

Antao, V. P., and Tinoco, I. Jr. (1992). Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res. 20,* 819-824.

Austin, R. J., Xia, T., Ren, J., Takahashi, T. T., & Roberts, R. W. (2003). Differential modes of recognition in N peptide-boxB complexes. *Biochemistry, 42*(50), 14957- 14967.

Babajide, A., Hofacker, I. L., Sippl, M. J. & Stadler, P. F. (1997) *Fold. Des.* 2, 261–269.

Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., . . . Williamson, J. R. (1996). Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science (New York, N.Y.), 273*(5281), 1547-1551.

Burova E, Hung SC, Chen J, Court DL, Zhou JG, Mogilnitskiy G, Gottesman ME. (1999). Escherichia coli nusG mutations that block tran- scription termination by coliphage HK022 Nun protein. *Mol Microbiol* 31:1783–1793.

Calabro, V., Daugherty, M. D., & Frankel, A. D. (2005). A single intermolecular contact mediates intramolecular stabilization of both RNA and protein. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(19), 6849-54.

Calnan, B.J., Biancalana, S., Hudson, D., and Frankel, A.D. (1991). Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. *Genes & Dev.* 5: 201–210.

Campbell, A. 1994. Comparative molecular biology of lambdoid phages. Annu. Rev. Microbiol. 48:193–222.

Casjens, S. R. 2005. Comparative genomics and evolution of the tailed-bacteriophages. Curr. Opin. Microbiol. 8:451–458.

Chattopadhyay, S., Garcia-Mena, J., DeVito, J., Wolska, K., & Das, A. (1995). Bipartite function of a small RNA hairpin in transcription antitermination in bacteriophage lambda. *Proceedings of the National Academy of Sciences of the United States of America, 92*(9), 4061-4065.

Chattopadhyay, S., Hung, S. C., Stuart, A. C., Palmer, A. G.,3rd, Garcia-Mena, J., Das, A., & Gottesman, M. E. (1995). Interaction between the phage HK022 nun protein and the nut RNA of phage lambda. *Proceedings of the National Academy of Sciences of the United States of America, 92*(26), 12131-12135.

Chen, L. and Frankel, A.D. (1995). A peptide interaction in the major groove of RNA resembles protein interactions in the minor groove of DNA. *Proc. Natl. Acad. Sci.* 92: 5077–5081.

Cilley, C.D., and Williamson, J.R. (1997). Analysis of bacteriophage N protein and peptide binding to boxB RNA using polyacrylamide gel coelectrophoresis (PACE). *RNA 3,* 57-67.

Cocozaki, A. I., Ghattas, I. R., & Smith, C. A. (2008). Bacteriophage P22 antitermination boxB sequence requirements are complex and overlap with those of lambda. *Journal of Bacteriology, 190*(12), 4263-4271. doi:10.1128/JB.00059-08

Cocozaki, A. I., Ghattas, I. R., & Smith, C. A. (2008). The RNA-binding domain of bacteriophage P22 N protein is highly mutable, and a single mutation relaxes specificity toward lambda. *Journal of Bacteriology, 190*(23), 7699-7708. doi:10.1128/JB.00997-08

Convery, M.A., et al., & Stockley, P.G. (1998). Crystal structure of an RNA aptamer-protein complex at 2.8 A resolution. *Nat. Struct. Biol.* 5, 133:139.

Correll, C. C., & Swinger, K. (2003). Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 A resolution. *RNA (New York, N.Y.), 9*(3), 355-363.

Doelling, J.H., and Franklin, N.C. (1989). Effects of all single base substitutions in the loop of boxB on antitermination of transcription by bacteriophage lambda's N protein. *Nucleic Acids Res. 17,* 5565-5577.

Doherty, E.A., Batey, R.T., Masquida, B., and Doudna, J.A. 2001. A universal mode of helix packing in RNA. *Nat. Struct. Biol.* 8: 339– 343.

Draper, D.E. (1999). Themes in RNA-protein recognition. *J. Mol. Biol. 293,* 255-270.

Faber, C., Scharpf, M., Becker, T., Sticht, H., & Rosch, P. (2001). The structure of the coliphage HK022 nun protein-lambda-phage boxB RNA complex. implications for the mechanism of transcription termination. *The Journal of Biological Chemistry, 276*(34), 32064-32070. doi:10.1074/jbc.M102975200

Fay, J.C., Wyckoff, G.J., Wu, C.I., (2002) Testing the neutral theory of molecular evolution with genomic data from Drosophila. *Nature*. 28;415(6875):1024-6.

Fontana, W., Stadler, P. F., Bornberg-Bauer, E. G., Griesmacher, T., Hofacker, I. L., Tacker, M., Tarazona, P., Weinberger, E. D. & Schuster, P. (1992) *Phys. Rev. E* 47, 2083–2099.

Franklin, N. C. (1993). Clustered arginine residues of bacteriophage lambda N protein are essential to antitermination of transcription, but their locale cannot compensate for boxB loop defects. *Journal of Molecular Biology, 231*(2), 343-360. doi:10.1006/jmbi.1993.1287

Gottesman, M. E., Adhya, S., & Das, A. (1980). Transcription antitermination by bacteriophage lambda N gene product. *Journal of Molecular Biology, 140*(1), 57- 75.

Grüner, W., Giegerich, R., Strothmann, D., Reidys, C. M., Weber, J., Hofacker, I. L., Stadler, P. F. & Schuster, P. (1996) *Monatsh. Chem.* 127, 355–389.

Gutell, R., Power, A., Hertz, G. Z., Putz, E. J. & Stormo, G. D. (1993) *Nucleic Acids Res.* 20, 5785–5795.

Harada, K., and Frankel, A.D. (1999) Screening RNA-binding libraries using a bacterial transcription antitermination assay. *Methods Mol Biol* 118: 177–187.

Harada, K., Martin, S.S., and Frankel, A.D. (1996) Selection of RNA-binding peptides *in vivo*. *Nature* 380: 175–179.

Henthorn, K. S., & Friedman, D. I. (1996). Identification of functional regions of the nun transcription termination protein of phage HK022 and the N antitermination protein of phage lambda using hybrid nun-N genes. *Journal of Molecular Biology, 257*(1), 9-20. doi:10.1006/jmbi.1996.0142

Hermann, T., and Westhof, E. (1999). Non-Watson-Crick base pairs in RNA-protein recognition. *Chem. Biol. 6,* R335-43.

Heus, H.A., and Pardi, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. Science *253*, 191–194.

Horiya, S., Inaba, M., Koh, C., Uehara, H., Masui, N.,...Kazuo Harada, K. (2009). Replacement of the λboxB RNA–N peptide with heterologous RNA–peptide interactions relaxes the strict spatial requirements for the formation of a transcription

Huang, H.C., Nagaswamy, U., and Fox, G.E. (2005). The application of cluster analysis in the intercomparison of loop structures in RNA. *RNA 11,* 412-423.

Huynen MA. Exploring phenotype space through neutral evolution. J Mol Evol. 1996 Sep;43(3):165-9.

Huynen, M. A., Doerks, T., Eisenhaber, F., Orengo, C., Sunyaev, S., Yuan, Y. & Bork, P. (1998) *J. Mol. Biol.* 280, 323–326.

Iwazaki, T., Li, X., and Harada, K. (2005). Evolvability of the mode of peptide binding by an RNA. *RNA 11,* 1364-1373.

Jucker, F.M., and Pardi, A. (1995). GNRA tetraloops make a U-turn. RNA *1*, 219–222.

Kang, J.Y., Olinares B. P. D., Gottesman, M., Darst, S.A. (2017) Structural basis of transcription arrest by coliphage HK022 Nun in an Escherichia coli RNA polymerase elongation complex. *Elife* ;6:e25478.

Kimura M. (1968) Related Evolutionary rate at the molecular level. *Nature*. 17;217(129):624-6.

Kimura, M. (1991). Recent development of the neutral theory viewed from the

Wrightian tradition of theoretical population genetics. *Proc. Natl. Acad. Sci. U. S. A. 88,* 5969- 5973.

Klein, D. J., T. M. Schmeing, P. B. Moore, and T. A. Steitz. (2001). The kink-turn: a new RNA secondary structure motif. *EMBO J*. 20:4214– 4221.

Lazinski, D., Grzadzielska, E., & Das, A. (1989). Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. *Cell, 59*(1), 207-218.

Legault, P., J. Li, J. Mogridge, L. E. Kay, and J. Greenblatt. (1998). NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: rec- ognition of a GNRA fold by an arginine-rich motif. *Cell* 93:289–299.

Leulliot, N., & Varani, G. (2001). Current topics in RNA-protein recognition: Control of specificity and biological function through induced fit and conformational capture. *Biochemistry, 40*(27), 7947-7956.

Liu, B., Steitz, A.T. (2017) Structural insights into NusG regulating transcription elongation. *Nucleic Acids Research, Vol. 45, No. 2 968–974*

McColl, D.J., Honchell, C.D., and Frankel, A.D. (1999) Structure-based design of an RNA-binding zinc finger. *Proc Natl Acad Sci USA* 96: 9521–9526.

Mogridge, J., Legault, P., Li, J., Van Oene, M. D., Kay, L. E., & Greenblatt, J. (1998). Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. *Molecular Cell, 1*(2), 265-275.

Mogridge, J., Mah, T.F. & Greenblatt, J. (1995). A protein-RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase of a stable antitermination complex containing the lambda N protein. *Genes Dev*. 9, 2831–2845.

Mooney RA, Schweimer K, Rosch P, Gottesman M, Landick R. 2009. Two structurally independent domains of *E. coli* NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. J Mol Biol 391:341–358.

Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B., and Steitz, T.A. 2001. RNA tertiary interactions in the large ribosomal subunit: The A- minor motif. *Proc. Natl. Acad. Sci.* 98: 4899–4903.

Ohta T. Near-neutrality in evolution of genes and gene regulation. Proc Natl Acad Sci U S A. 2002 Dec 10;99(25):16134-7.

Olson, E. R., Flamm, E. L., & Friedman, D. I. (1982). Analysis of nutR: A region of phage lambda required for antitermination of transcription. *Cell, 31*(1), 61-70.

Peled-Zehavi, H., Horiya, S., Das, C., Harada, K., and Frankel, A.D. (2003) Selection of RRE RNA binding pep- tides using a kanamycin antitermination assay. *RNA* 9: 252–261.

Possik EJ, Bou Sleiman MS, Ghattas IR, Smith CA. 2013. Randomized Codon Mutagenesis Reveals that the HIV Rev Arginine-Rich Motif Is Robust to Substitutions and that Double Substitution of Two Critical Residues Alters Specificity. *Journal of Molecular Recognition* 26, 286-96.

Puglisi, J.D., Chen, L., Blanchard, S., and Frankel, A.D. 1995. Solution structure of a bovine immunodeficiency virus Tat-TAR peptide- RNA complex. *Science* 270: 1200–1203.

Robert, J., Sloan, S. B., Weisberg, R. A., Gottesman, M. E., Robledo, R., & Harbrecht, D. (1987). The remarkable specificity of a new transcription termination factor suggests that the mechanisms of termination and antitermination are similar. *Cell, 51*(3), 483-492.

Robledo, R., Gottesman, M. E., & Weisberg, R. A. (1990). Lambda nutR mutations convert HK022 nun protein from a transcription termination factor to a suppressor of termination. *Journal of Molecular Biology, 212*(4), 635-643.

Rowsell, S. & Stockley, P.G. (1998). Crystal structures of a series of RNA aptamers complexed to the same protein target. Nat Sfruct. Biol. 5, 970-975.

Said, N., Krupp, F., Anedchenko, E., Santos, K. F., Dybkov, O., Huang, Y.H., Wahl, M.C. Structural basis for λN-dependent processive transcription antitermination. *Nature Microbiology* 28;2:17062.

Salstrom, J. S., and W. Szybalski. (1978). Coliphage λ nutL$^-$: a unique class of mutants defective in the site of gene N product utilization for antitermination of leftward transcription. *J. Mol. Biol*. 124:195-221.

Scharpf, M., H. Sticht, K. Schweimer, M. Boehm, S. Hoffmann, and P. Rosch. (2000). Antitermination in bacteriophage lambda. The structure of the

N36 peptide-boxB RNA complex. *Eur. J. Biochem*. 267:2397–2408.

Schultes, E.A., and Bartel, D.P. (2000). One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science 289,* 448-452.

Schuster, P., Fontana, W., Stadler, P.F., and Hofacker, I.L. (1994). From sequences to shapes and back: a case study in RNA secondary structures. *Proc. Biol. Sci. 255,* 279-284.

Smith, C.A., Calabro, V., and Frankel, A.D. (2000a). An RNA-binding chameleon. *Mol. Cell 6,* 1067-1076.

Smith, C.A., Chen, L., and Frankel, A.D. (2000b). Using peptides as models of RNA- protein interactions. *Methods Enzymol. 318,* 423-438.

Smith, C.A., Crotty, S., Harada, Y., and Frankel, A.D. (1998). Altering the context of an RNA bulge switches the binding specificities of two viral Tat proteins. *Biochemistry 37,* 10808-10814.

Strobel, S.A., and Cochrane, J.C. (2007). RNA catalysis: ribozymes, ribosomes, and riboswitches. *Curr. Opin. Chem. Biol. 11,* 636-643.

Su, L., Radek, J. T., Labeots, L. A., Hallenga, K., Hermanto, P., Chen, H., . . . Weiss, M. A. (1997). An RNA enhancer in a phage transcriptional antitermination complex functions as a structural switch. *Genes & Development, 11*(17), 2214- 2226.

Su, L., Radek, J.T., Hallenga, K., Hermanto, P., Chan, G., Labeots, L.A., and Weiss, M.A. (1997). RNA recognition by a bent alpha-helix regulates transcriptional antitermination in phage lambda. *Biochemistry 36,* 12722-12732.

Sugaya M, Nishimura F, Katoh A, Harada K. Tailoring the peptide-binding specificity of an RNA by combinations of specificity-altering mutations. Nucleosides Nucleotides Nucleic Acids. 2008 May;27(5):534-45.

Tan, R., and Frankel, A.D. (1995). Structural variety of arginine-rich RNA-binding peptides. *Proc. Natl. Acad. Sci. U. S. A. 92,* 5282-5286.

Tan, R., Chen, L., Buettner, J. A., Hudson, D., & Frankel, A. D. (1993). RNA recognition by an isolated alpha helix. *Cell, 73*(5), 1031-1040.

Tawk, C.S., Ghattas I.R., Smith C.A. (2015). HK022 Nun Requires Arginine-Rich Motif Residues Distinct from λ N. *Journal of Bacteriology* 197, 3573–3582.

Valegard, K., et al., & Liljas, L. (1997). The three dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. J. Mol. Viol. 270. 724-738.

Van Gilst, M., W.A. Rees, A. Das, and P.H. von Hippel. 1997. Com plexes of N Antiterm ination Protein of phage l with specific and nonspecific RNA target sites on the nascent transcript. *Biochem istry* 36:1514–1524.

Varani, G. (1997). RNA-protein intermolecular recognition. *Accounts of Chemical Research, 30*(5), 189.

Watnick, R. S., & Gottesman, M. E. (1999). Binding of transcription termination protein nun to nascent RNA and template DNA. *Science (New York, N.Y.),286*(5448), 2337-2339.

Weisberg, R. A., & Gottesman, M. E. (1999). Processive antitermination. *Journal of Bacteriology, 181*(2), 359-367.

Weiss, M. A., & Narayana, N. (1998). RNA recognition by arginine-rich peptide motifs. *Biopolymers, 48*(2-3), 167-180. doi:2-8

Wilhelm, J.E., and Vale, R.D. (1996) A one-hybrid system for detecting RNA–protein interactions. *Genes Cells* 1: 317– 323.

Wool, I.G., Gluck, A., and Endo, Y. 1992. Ribotoxin recognition of ribosomal RNA and a proposal for the mechanism of transloca- tion. *Trends. Biochem. Sci.* 17: 266–269.

Wyatt, J.R., Puglisi, J.D., and Tinoco, I.,Jr. (1989). RNA folding: pseudoknots, loops and bulges. *Bioessays 11,* 100-106.

Xia, T., A. Frankel, T. T. Takahashi, J. Ren, and R. W. Roberts. 2003. Context and conformation dictate function of a transcription antitermination switch. Nat. Struct. Biol. 10:812–819.

Ye, X., Gorin, A., Frederick, R., Hu, W., Majumdar, A., Xu, W., . . . Patel, D.

J. (1999). RNA architecture dictates the conformations of a bound peptide. *Chemistry & Biology, 6*(9), 657-669.