AMERICAN UNIVERSITY OF BEIRUT

THREE SUBSTITUTIONS ARE SUFFICIENT TO ALTER THE SPECIFICITY OF RRE IIB TOWARD WILD-TYPE REV OR REV-R35G:N40V

by
EMANE YOUSEF ABDALLAH

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
May 2014

AMERICAN UNIVERSITY OF BEIRUT

THREE SUBSTITUTIONS ARE SUFFICIENT TO ALTER THE SPECIFICITY OF RRE IIB TOWARD WILD-TYPE REV OR REV-R35G:N40V

by

EMANE YOUSEF ABDALLAH

Approved by:

Dr. Colin Smith, Associate Professor
Biology — Advisor

Dr. Elias Baydoun, Professor
Biology — Member of Committee

Dr. Mike Osta, Assistant Professor
Biology — Member of Committee

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

Date of thesis defense: May 5, 2014

AMERICAN UNIVERSITY OF BEIRUT

Student Name:
_____Abdallah_____Emane_____Yousef_____
Last First Middle

[X] 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, **three years after the date of submitting my thesis, dissertation, or project,** 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.

_____

Signature Date

ACKNOWLEDGEMENTS

This work is the result of my persistence and Dr. Colin Smith's guidance. I thank him for his patience. I thank Ingrid Ghattas for her pleasant and valuable discussions. I thank Samer Bu Jawdeh for his support. I also thank Mykal Rizkallah for his stress relief presence.

AN ABSTRACT OF THE THESIS OF

Emane Yousef Abdallah for Master of Science
Major: Biology

Title: Three Substitutions Are Sufficient to Alter the Specificity of RRE IIB toward Wild-Type Rev or Rev-R35G:N40V

The binding of human immune-deficiency virus type 1 (HIV-1) Rev arginine-rich motif (ARM) to stem-loop IIB of Rev-Response Element (RRE) is necessary for viral replication. Despite the essential role of Asn40 in recognition, Rev-R35G:N40V binds RRE IIB well when fused to lambda N protein. Neutral theories of molecular evolution claim that genotypes are sufficiently mutable that new phenotypes occur when genotypes diffusing along neutral mutational paths of one phenotype intersect those of distinct phenotypes. How well neutral theories describe protein-RNA interactions is unclear. The questions raised by Rev-R35G:N40V are: 1) How mutable is the Rev-RRE IIB interaction? and 2) How neutral are evolutionary paths between these distinct specificities? To examine the mutagenic potential of RRE IIB, ten RRE IIB libraries, each completely randomized in a separate over-lapping sub-region, were screened with RNA-binding domain of wild-type Rev and Rev-R35G:N40V using a reporter system based on bacteriophage lambda N-nut antitermination. Flanking regions of RRE IIB are mutable, and the internal loop of RRE IIB allows compensatory mutations that widen the groove. That Rev-R35G:N40V accepts a different panel of RRE variants is consistent with the two ARMs binding RRE IIB using distinct recognition strategies. Many RRE variants obtained showed relaxed specificity with both Rev and Rev-R35G:N40V. From RRE IIB, one substitution was sufficient to confer specificity to wild-type Rev, and three substitutions were sufficient to confer specificity to Rev-R35G:N40V. Retroviruses have high mutation rates and undergo rapid genetic drift; neutral theories are plausible means of generating the diversity of extant arginine-rich motif-RNA recognition strategies.

CONTENTS

ILLUSTRATIONS

TABLES

ABBREVIATIONS

%	Percent
/	Per
µl	Microliter
µM	Micromolar
Å	Angstrom
Arg	Arginine
ARM	Arginine rich motif
Asn	Asparagine
BIV	Bovine immunodeficiency virus
ddH_2O	Double Distilled Water
DNA	Deoxyribonucleic acid
Et al.	Et allii (and others)
E. Coli	*Escherichia Coli*
Fmol	Femtomole
G	Glycine
GNRA	N is adenosine, Cytosine, Guanidine or Uracil
h	hour
HIV	Human Immunodeficiency Virus

IPTG	Isopropyl β-D-Galactoside
JDV	Jembrana Disease Virus
Kb	Kilo base pairs
LacZ	β-galactosidase gene
LANL	Los Alamos National Laboratory
LB	Luria Bertani growth medium
mg	Milligram
min	Minutes
ml	Milliliter
mM	Millimolar
mRNA	Messenger RNA
N	Asparagine
NMR	Nuclear Magnetic Resonance
Nut	N-utilization
OD	Optical density
ONPG	Ortho-nitrophenyl Galactoside
PDB	Protein Data Bank
Pmol	picomole
R	Arginine
RNA	Ribonucleic acid

rpm	Revolution per minute
RRE	Rev-response element
SIV	Simian Immunodeficiency Virus
VV	Visna Virus

CHAPTER 1

INTRODUCTION

A. Overview and Specific Aims

Arginine-rich motifs (ARM) play structural and regulatory roles binding RNA targets in many viral and cellular processes (reviewed in Gesteland et al., 1993; Chen and Varani, 2005). The binding of the ARM of HIV-1 Rev peptide to stem IIB of Rev-Response Element (RRE) is essential for viral replication: it allows the nuclear export of unspliced and singly spliced viral transcripts (reviewed in Fernandes et al., 2012). The Rev ARM binds an internal loop containing non-canonical base pairs in RRE IIB (Bartel et al., 1991), adopts an alpha-helical conformation upon binding, and forms base-specific interactions (Battiste et al., 1996). Despite the essential role of Asn40 in recognition, a double mutant Rev, Rev-R35G:N40V binds RRE IIB well, indicating a distinct binding strategy (Possik et al., 2013). Neutral theories of molecular evolution claim that starting from RNA genotype of one phenotype, sufficient functional single mutants exist that branching paths are created linking distinct phenotypes. A bifunctional intermediate, a sequence that can display two distinct phenotypes, occurs at the intersection of these paths. Speciation occurs upon transit through the intersection sequence from one phenotype to another (reviewed by Kimura, 1991; Fontana and Schuster, 1998a). How Rev-R35G:N40V binds RRE IIB is not known and how well neutral theories describe RNA-protein interactions is unclear. The questions raised by Rev-R35G:N40V are: 1) How mutable is

the Rev-RRE IIB interaction? and 2) How neutral are evolutionary paths between distinct specificities?

The broad aim of this work is to reveal how macromolecular recognition evolves using the HIV-1 Rev ARM-RRE IIB complex as a model system.

Specific Aim 1. Construct RRE IIB plasmid reporter libraries. RRE IIB libraries with randomized regions are constructed and prepared as plasmid reporter system to be used in specific aims 2 and 3. Double stranded DNA is prepared from degenerate, single-stranded oligonucleotides expressing the RRE IIB libraries of interest (figure 1), using primer extension. The double-stranded libraries are ligated into reporter plasmid to replace the λ boxB of the nut site in the RNA reporter plasmid. The use of RRE IIB libraries with completely randomized over-lapping sub-regions reveals the mutability of RRE and allows comparison of the binding strategies of wild-type Rev, Rev17, and Rev double-mutant, Rev-R35G:N40V. Local compensatory mutations explain roles of nearby nucleotides, flanking sequences and covariation due to base stacking.

Specific Aim 2. Isolate RRE IIB active clones that are recognized by wild-type Rev. In order to reveal mutagenic potential of RRE IIB, the RRE IIB libraries are screened with wild-type Rev ARM using a two-plasmid λ N-antitermination system (figure 2).This is a standard approach, successfully implemented in previous projects, where the peptide retains its RNA-binding activity when separated from the rest of the protein. The reporter plasmids with the RRE IIB libraries are transformed into supplier competent N567 *E. coli* cells that contain plasmids with

the wild-type Rev ARM replacing the RNA-binding domain of lambda N protein (Harada and Frankel, 1991). Active clones are screened and isolated by β-galactosidase colony color assay on solid media. The activity of the β-galactosidase gene: LacZ, which is transcribed past terminators upon Rev ARM-RRE IIB binding in the reporter plasmid, reflects the affinity of the interaction.

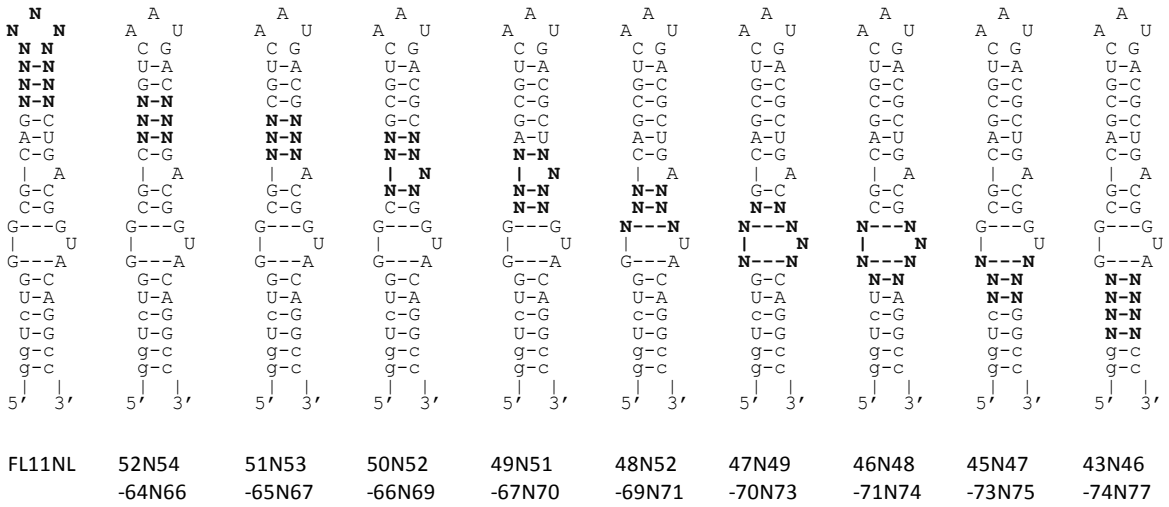


Figure 1. Planned RRE IIB libraries. Randomized bases are shown in bold, with N being any of the four nucleotides.

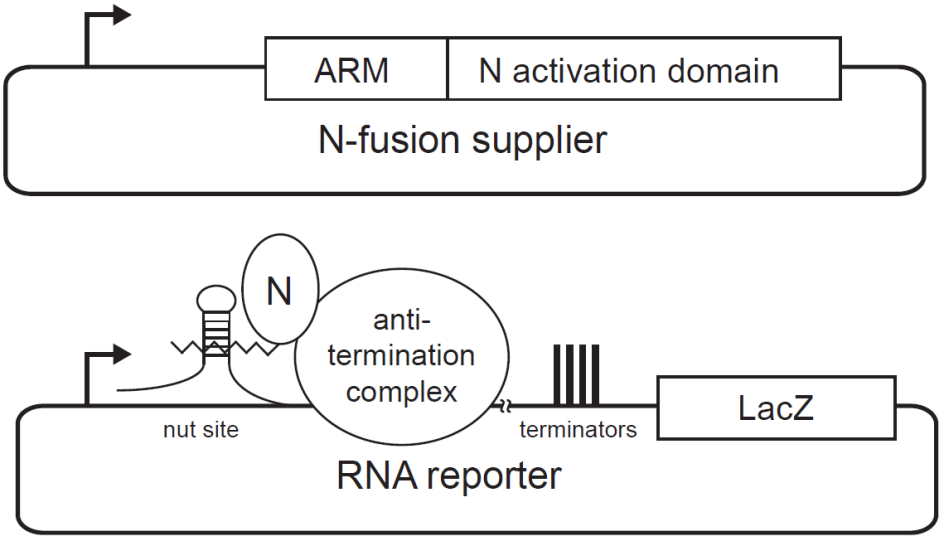


Figure 2. The two-plasmid antitermination system for detecting peptide-binding RRE IIB active clones. In the N-fusion supplier plasmid, arginine-rich motifs replace the RNA-binding domain of lambda N protein which are expressed under the control of the tac promoter. In the RNA reporter plasmid, RRE IIB libraries replace boxB hairpin in the lambda nut left site, followed by transcription terminators and the reporter gene LacZ. The binding of ARM to RRE IIB results in transcription anti-termination and reporter gene expression.

Specific Aim 3. Isolate RRE IIB active clones that are recognized by Rev-R35G:N40V. To isolate functional RRE IIB variants allowed by Rev-R35G:N40V; the RRE IIB libraries are screened with Rev-R35G:N40V using the same two-plasmid λ N-antitermination system of specific aim 2. The reporter plasmids with the RRE IIB libraries were transformed into supplier competent N567 *E. coli* cells that contain plasmids with Rev-R35G:N40V replacing the RNA-binding domain of lambda N protein. Active clones are screened and isolated by β-galactosidase colony color assay. RRE IIB variants recognized by Rev-R35G:N40V may reveal how RRE IIB can mutate and still bind Rev and may shed light on the basis of specificity. RRE IIB active clones of specific aim 2 and 3 reveal the mutagenic potential of Rev-RRE IIB interaction whose details will be examined in specific aim 4.

Specific Aim 4. Characterize RRE IIB variants. The activity and specificity of isolates is confirmed by rescreening with wild-type Rev, Rev-R35G:N40V and bovine immunodeficiency virus (BIV) Tat that binds specifically to BIV TAR RNA. In order to identify the sequences of verified RRE IIB active clones isolated in specific aims 2 and 3, isolates of interest are column purified and sequenced. RRE IIB sequences obtained are folded, examined, and analyzed in comparison to structural models. The activity of the sequenced variants is determined by β-galactosidase solution assay. Important active clones are reconstructed to confirm observations. Functional mutant RRE IIB sequences reveal how different partners restrict RRE IIB mutations and the basis of distinct recognition strategies between

wild-type Rev and Rev-R35G:N40V. Analysis of findings in light of available structural models helps interpreting any conserved residues and compensatory mutations as well as suggesting possible structural explanations. The panel of single mutations assists in tracing an evolutionary path for specific aim 5.

Specific Aim 5. Examine evolutional potential of Rev-RRE interaction. To explore the evolutionary landscape between RRE's distinct specificities and to check their neutrality, paths connecting mutants obtained are traced. Comparing RRE IIB variants to clinical sequences addresses possible restriction to sequence variation in a complete functional system and the selection exerted by the viral context, such as the env open reading frame on natural RRE IIB.

B. RNA-Protein Recognition

1. RNA Structure

RNA-protein recognition is essential for many biological processes including regulation of gene expression (reviewed in Gesteland et al., 1993). RNA-protein recognition involves both nucleobase-specific and nucleobase non-specific interactions facilitated by diverse RNA secondary structures. Complementary Watson-Crick base pairs separated by single-stranded regions including apical and internal loops, bulges, junctions and pseudoknots define the secondary structure of RNA molecules (figure 3A). Canonical A-form double helix structural element of RNA is a poor target for protein recognition (reviewed in Gesteland et al., 1993). The major groove of the A-form helix is narrow and deep making it inaccessible for protein secondary structure and unavailable for base-

specific recognition. The minor groove however is wide and shallow exposing inaccessible canonical base pairs for intermolecular contacts or specific hydrogen-bonding interactions (reviewed in Gesteland et al., 1993). As a consequence, many RNA-protein interactions and other RNA-ligand interactions rely upon special distortions in the A-form geometry fitting proteins and exposing nucleotides for base specific contacts (reviewed in Hermann and Westhof, 1999).

Non-canonical base-pairs are pair wise combinations other than Watson-Crick pairings and are hallmark sites for RNA-protein interactions (reviewed in Hermann and Westhof, 1999) (figure 4). Non-Watson Crick base pairs present hydrogen bond donors and acceptors allowing specific hydrogen-bonding interactions. They also distort regular A-form helix to widen the major groove and to allow binding of proteins (Hermann and Westhof, 1999). Tandem stacks of two non-canonical base pairs; G48•G71 and G47• A73, are found in the Rev peptide-binding site of the HIV-1 RRE IIB (Battiste et al., 1996) that widen the major groove and facilitate the binding of Rev (figure 3B).

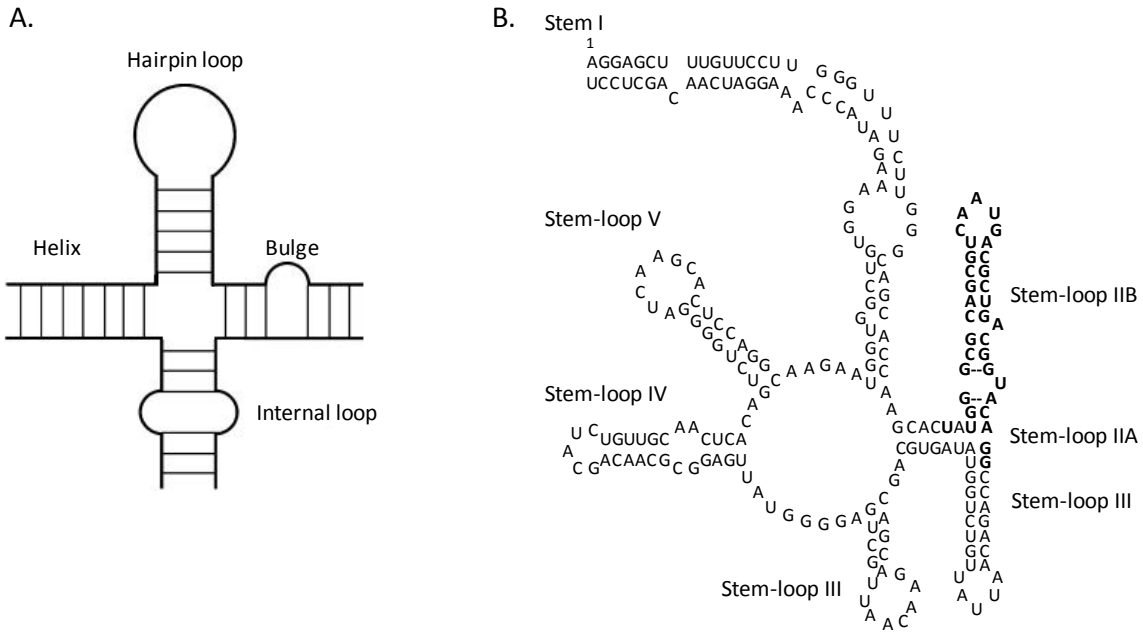


Figure 3. RNA secondary structures. A) Common RNA secondary structures. B) Secondary structure of the HIV-1 Rev response element (RRE). The RRE is redrawn from Malim et al. (1998) as a 234-nucleotide structure. The RRE stem-loop IIB is shown in bold. The two non-canonical purine-purine base pairs (G47•A73 and G48•G71) are drawn with dashes between the bases; these dashes also reflect the widening of the major groove of the helix in this region.

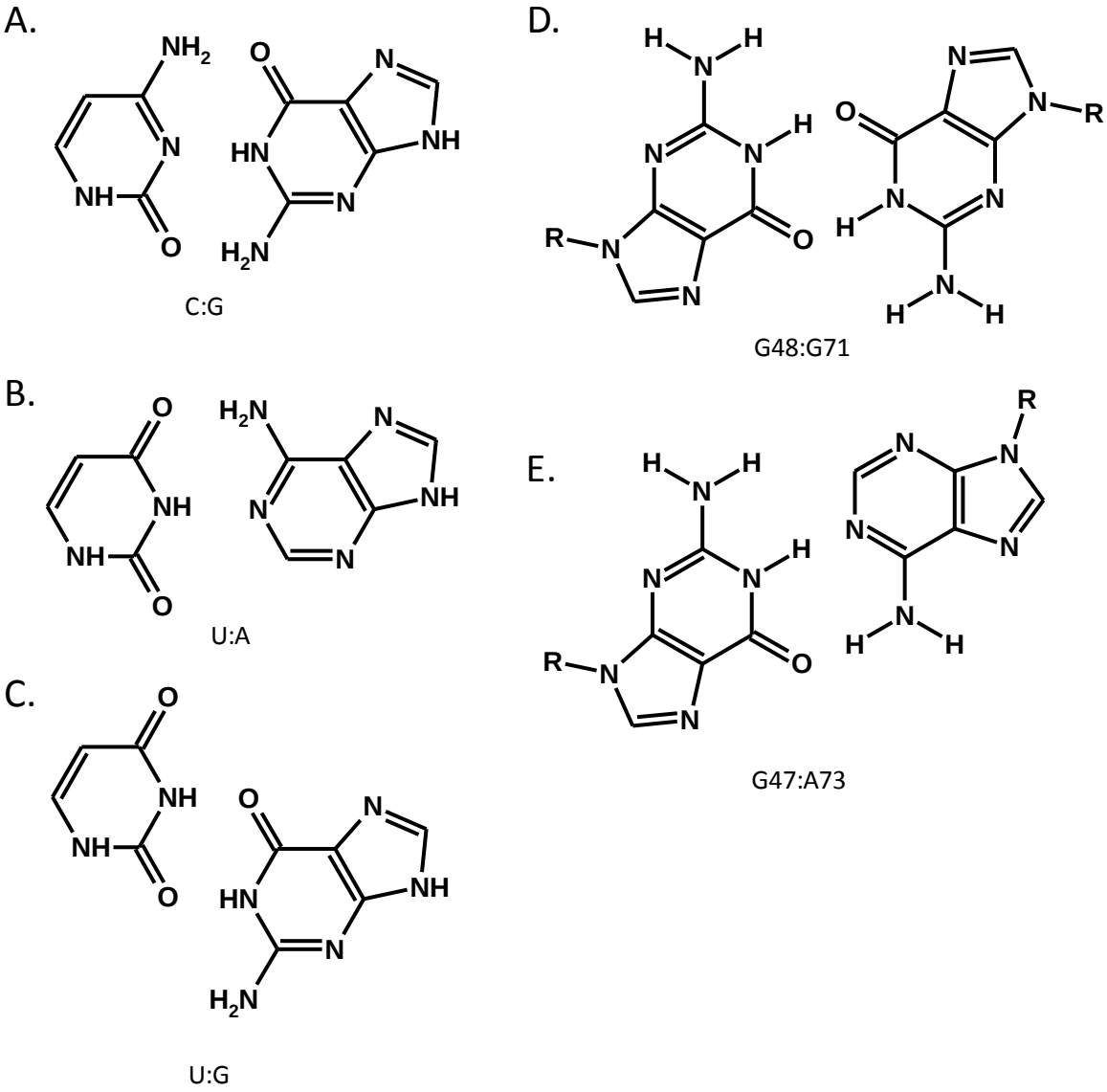


Figure 4. Schematic representation of hydrogen-bonding arrangements of canonical and non-canonical base pairs in RRE IIB. A) C:G. B) U:G. C) U:G. D) G48•G71. E) G474•A73.

2. Arginine-Rich Motifs

The cellular functions of most RNA molecules involve proteins that bind their targets with high affinity and specificity, either in permanent complexes such as ribosomes or in transient association such as the mRNA splicing machinery.

Arginine rich-motifs, initially found in viral and ribosomal proteins, bind their RNA target sites with high affinity and specificity while folding into different conformations independently from their protein context (Lazinski et al., 1989; reviewed in Weiss and Narayana, 1998). The high affinity of arginine-rich peptides are conferred by the positively charged guanidinium group of their arginine content that forms a network of ionic, hydrogen bonding, weakly polar and van der Waals interactions rendered accessible by the long aliphatic side chain (reviewed in Patel, 1999 and Draper, 1999) .

NMR structures of peptide-RNA complexes reveal wide range of structures adopted by ARMs including α helix (Battiste et al., 1996; Ye et al., 1996), bent α helices (Legault et al., 1998; Cai et al., 1998), β hairpin (Puglisi et al., 1995), and extended conformations (Jiang et al., 1999) (figure 5). This conformational flexibility allows a single peptide to adopt different conformations when bound to different RNA partners (Ye et al., 1996 and Ye et al., 1999).

Although arginine side chains are necessary for RNA binding, critical roles are also played by nonpolar and aromatic side chains. The conformational flexibility and wide-sequence range of the peptide partner reflect the diverse biochemical and structural features of the RNA binding surface (Weiss and Narayana, 1998).

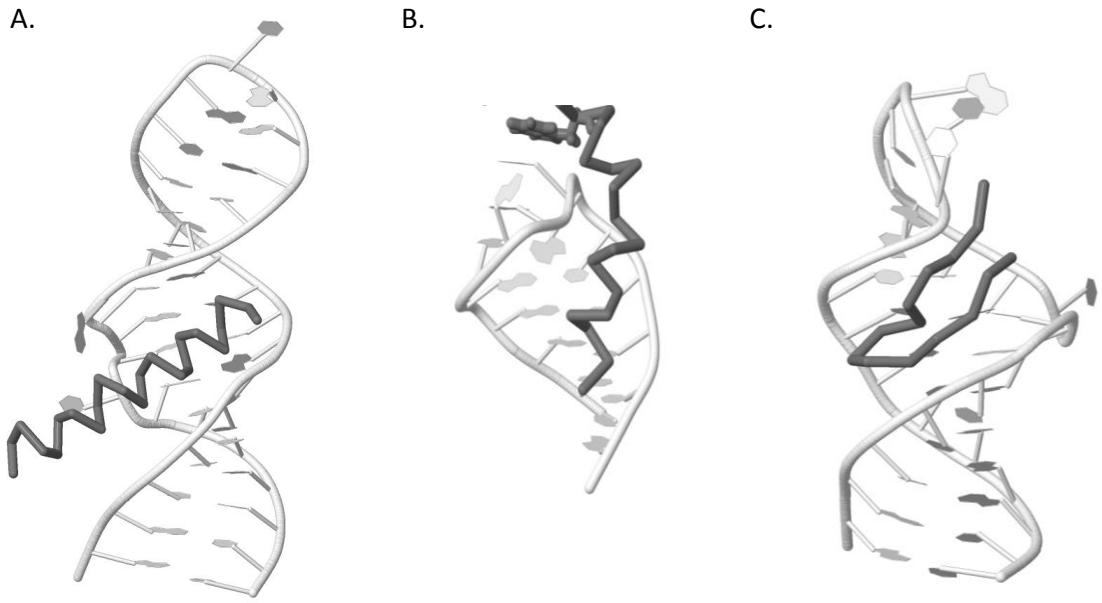


Figure 5. Solution structures of small basic domains bound to their cognate RNA partners. A) The human immunodeficiency virus type 1 (HIV-1) Rev peptide-RRE RNA complex (PDB 1ETf; Battiste et al., 1996). The basic domain of HIV-1Rev forms an α-helix. B) The bacteriophage λ N peptide-boxB RNA complex (PDB 1QFQ; Legault et al., 1998). The N protein of bacteriophage λ forms an irregular, distorted α-helix. C) The bovine immunodeficiency virus (BIV) Tat-Tar peptide-RNA complex (PDB 1mnb; Puglisi et al., 1995). The basic domain of BIV Tat protein forms a β-hairpin upon binding to BIV trans-activation-responsive (TAR) region RNA. The RNA rendered as cartoon and the peptide rendered as backbone. All figures were generated from PDB files using Jmol.

3. Adaptive Recognition

A novel feature of peptide-RNA recognition is the adaptive conformational transitions that both components undergo before they can generate a thermodynamically stable and functional complex (reviewed in Patel, 1999; Weiss and Narayana; 1998 and Williamson, 2000). The RNA structure forms scaffolds discriminating, directing and dictating the binding conformation of the protein partner which adopts a distinct fold on complexing (Varani, 1997).

Although RNA-protein recognition is specific, conformational flexibility allowed by the formation of different hydrogen bonding, hydrophobic and electrostatic interactions makes binding to multiple partners possible. ARMs bind their cognate RNA molecules with high specificity; however, relaxed-specificity interactions are also observed because of the large number of possible electrostatic interactions between the arginine rich peptide and hydrogen bond donors and acceptors of the RNA (Bayer et al., 2005). For example, a single RRE sequence is recognized by multiple peptide partners that adopt distinct conformations and form different nucleobase specific interactions as revealed by structural studies (Iwazaki et al., 2005 and Sugaya et al., 2008).

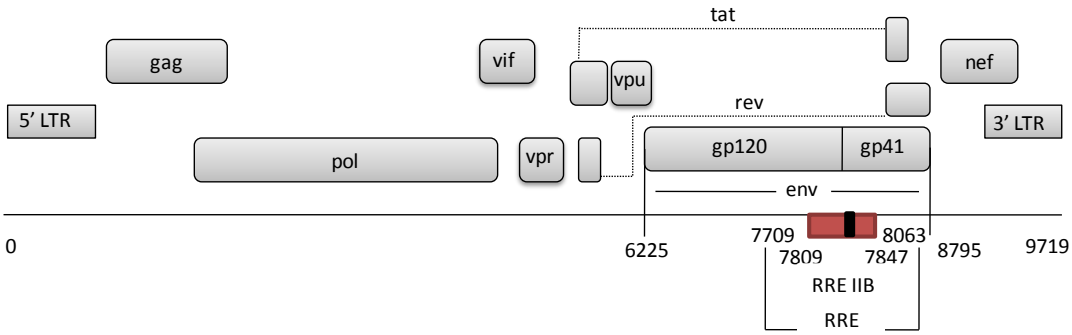
C. HIV-1 Rev-RRE as Model of Recognition

1. HIV-1 Genome and Life Cycle

The human immunodeficiency virus type-1 (HIV-1) Rev-RRE complex has been extensively studied as a model for molecular recognition. HIV-1 is a retrovirus belonging to the lentivirus family (reviewed in Fernandes et al., 2012).HIV has a 9-kb genome whose differential splicing results in 9-kb, 4-kb and 2-kb mRNAs among others (figure 6A). In order to express regulatory and structural proteins as well as genomic RNA, HIV exports unspliced (9-kb), incompletely spliced (4-kb), and multiply spliced transcripts from the nucleus for translation. Lentiviruses encode a regulatory protein, Rev, that mediates essential nuclear export of genomic and mRNAs encoding structural viral proteins (reviewed in Tang et al., 1999). The Rev protein, a 116 amino-acid protein, contains a 17 amino-acid RNA-binding domain (amino-acid residues 34-50) with 10 arginine residues (Lazinski et al., 1989). The arginine-rich motif of Rev binds the Rev Response element (RRE), a ~ 350 nucleotide RNA element in the env gene, at a high affinity site: stem IIB (Malim et al., 1990; Heaphy et al., 1990; Kjems et al., 1991; Tiley et al., 1991).

Rev-RRE interaction is essential for viral replication (figure 6B). In the early phase of HIV life cycle, when the concentration of Rev is below the threshold necessary for function, 2-kb mRNA are constitutively exported to the cytoplasm and used to express Rev and other regulatory proteins. When the level of Rev in the nucleus is high, Rev binds RRE-containing 9-kb and 4-kb mRNAs and mediates their export to the cytoplasm for translation and viral packaging (reviewed in Fernandes et al., 2012).

A.



B.

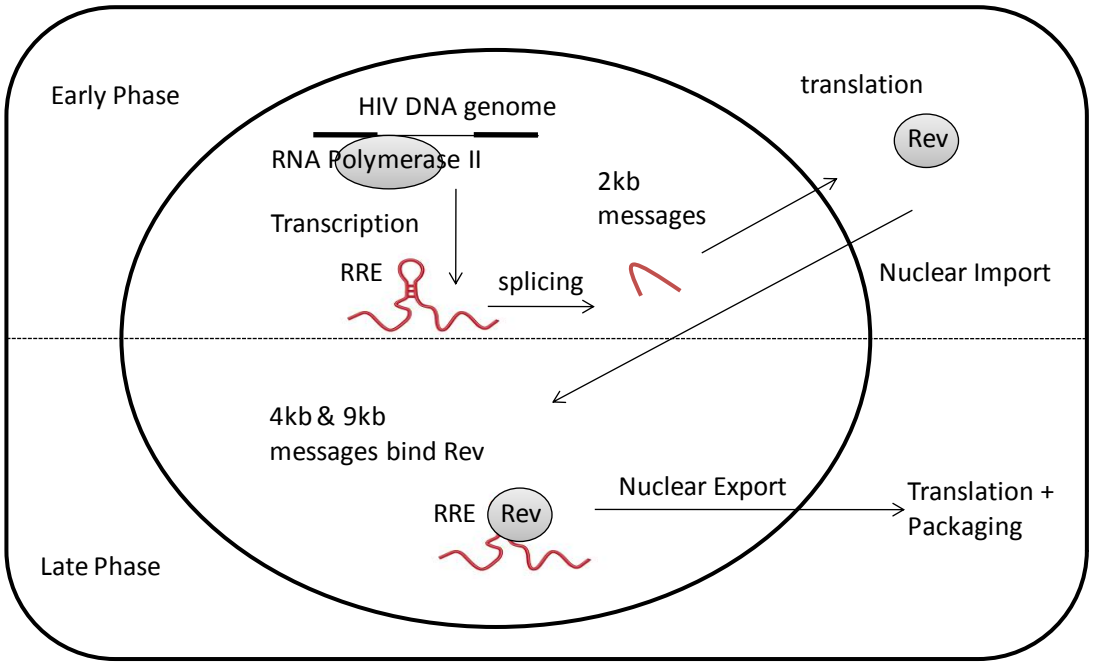


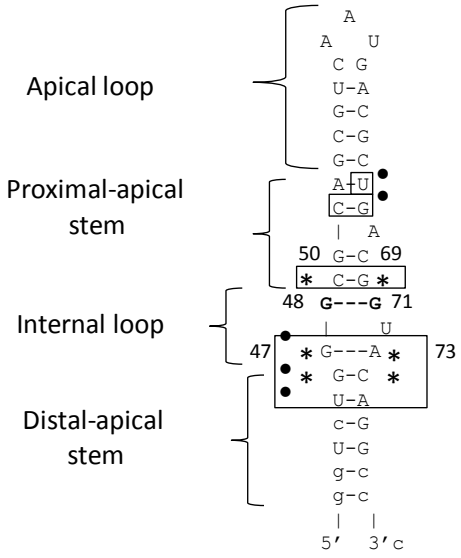
Figure 6. HIV-1 gene regulation and viral replication. A) Genome organization of HIV-1 (Redrawn from HIV Sequence Compendium, 2011). The RRE lies between nucleotides 7709-8063 and RRE stem-loop IIB lies between nucleotides 7809-7847. B) HIV-1 life cycle. In the early phase of the HIV-1 life cycle, splicing generates short RRE free messages that are exported to the cytoplasm where they are translated. Products of translation include the Rev protein which is then imported to the nucleus. In the late phase, unspliced (9Kb) or partially spliced (4Kb) messages containing RRE are transcribed. The binding of Rev to RRE is recognized by host nuclear export factors. In the cytoplasm these massages are either translated into viral proteins or packaged into viral genomes (Fernandes et al., 2012).

2. Rev-RRE Complex

Different experimental approaches have revealed the details of the Rev-RRE interaction and allowed a better understanding of macromolecular recognition (figure 7). Site-directed mutagenesis and RNAase protection experiments revealed stem IIB of RRE as a high affinity binding site for Rev ARM (Kjems et al., 1991; Malim et al., 1990; Heaphy et al., 1990; Holland et al., 1990; Tiley et al., 1991). In vitro selection identified structural features of RRE IIB (Bartel et al., 1991). Seven bases of RRE IIB, highly conserved among the binding variants, and an internal loop containing a non-canonical base pair G48•G7 that can be functionally replaced by an A•A pair, were found to be important for Rev ARM binding (figure 8). Given that the two non-canonical base pairs use different hydrogen bond donors and acceptors, Bartel et al. (1991) proposed that the G48•G71 base pair plays a structural role by distorting the sugar-phosphate backbone. This distortion is a critical determinant of recognition by Rev and the non-canonical A•A base pair maintains the same distinctive backbone structure (Bartel et al., 1991). Structural model of HIV-1 Rev-RRE complex based on nuclear magnetic resonance (NMR) spectroscopy confirmed the structural role of G48•G71 base pair that allows the peptide to access a widened major groove without directly contacting the peptide. The NMR structure showed a α-helical conformation of Rev ARM upon complex formation (Battiste et al., 1996; Ye et al., 1996) as suggested by previous circular dichroism studies (Tan et al., 1993; Tan and Frankel, 1994). It also revealed base-specific contacts; the residues Arg35 and Arg39 interact with nucleotides U66, G67 and G70 and the residues Asn40 and Arg44 interact with nucleotides U45, G46, G47 and A73. These nucleotides were invariant in in vitro selection experiments (Bartel et al. 1991) and all four amino acids are critical for peptide-binding specificity (Tan

et al., 1993) (figure 9). The important nucleotides have functional groups within reasonable hydrogen-bonding distances from the side chains of the four amino acids; however, the precision of the side chain position is insufficient to determine exact hydrogen bond arrangement of these contacts. Nevertheless, Asn40 is coplanar with the conserved purine-purine G47•A73 base pair and makes hydrogen bonds to bases in the major groove (Battiste et al., 1996; Jain and Belasco, 1996**).** The structure formed by the two purine-purine base pairs, G47•A73 and G48•G71 creates a distinctive binding pocket that the peptide uses for specific recognition. Phosphate backbone contacts and Van der Waals contacts also occur in Rev ARM-RRE IIB interaction and they may help positioning the α-helical Rev peptide. Although Rev ARM- RRE IIB is essential for nuclear export, other RRE stem-loops interact with several cellular factors of the nuclear export machinery and are therefore important for transport of mRNA to the cytoplasm (Dillon et al., 1990).

A.



A Watson –Crick interaction is not important for binding at position 50-69 (Bartel et al., 1991)

G48•G71 widens the groove and A48•A71 is allowed (Bartel et al., 1991; Giver et al., 1996)

G47•A73 widens the major groove (Battiste et al., 1996)

* Nucleotides conserved in high affinity aptamers (Giver et al., 1993)
• Phosphates that interfere with Rev protein binding (Kjems et al., 1992)

B.

Figure 7. Minimal HIV-1 Rev ARM-RRE IIB complex. A) Sequence of RRE stem-loop IIB. Boxed nucleotides indicate those that were conserved and predicted to be important for Rev binding as summarized in literature. Capitalized letters indicate wild-type RRE IIB nucleotides, while small letters indicate those that have been modified. The label defined for each region is chosen to make reference to the regions easier. Arg35 is predicted to interact with U66, G67 and G70 while Asn40 is predicted to interact with G47, A73, G46 and U45. B) Amino acid sequence of the RNA-binding domain of the Rev peptide which contains amino acids 34 through 50 of the RNA-binding domain. The large letters indicate aminoacids found to be critical for RRE binding (Tan et al., 1993).

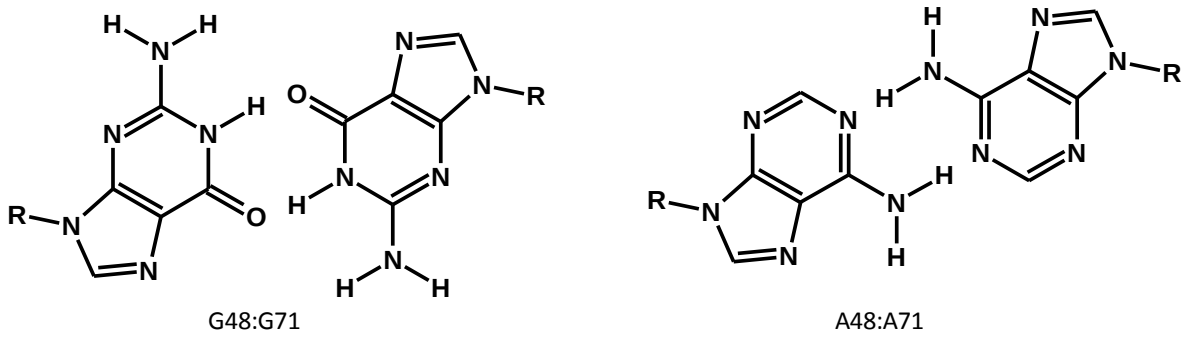


Figure 8. Schematic representation of hydrogen-bonding arrangements of base pairs G48• G71 and the isosteric A48• A71 (Ye et al., 1996).

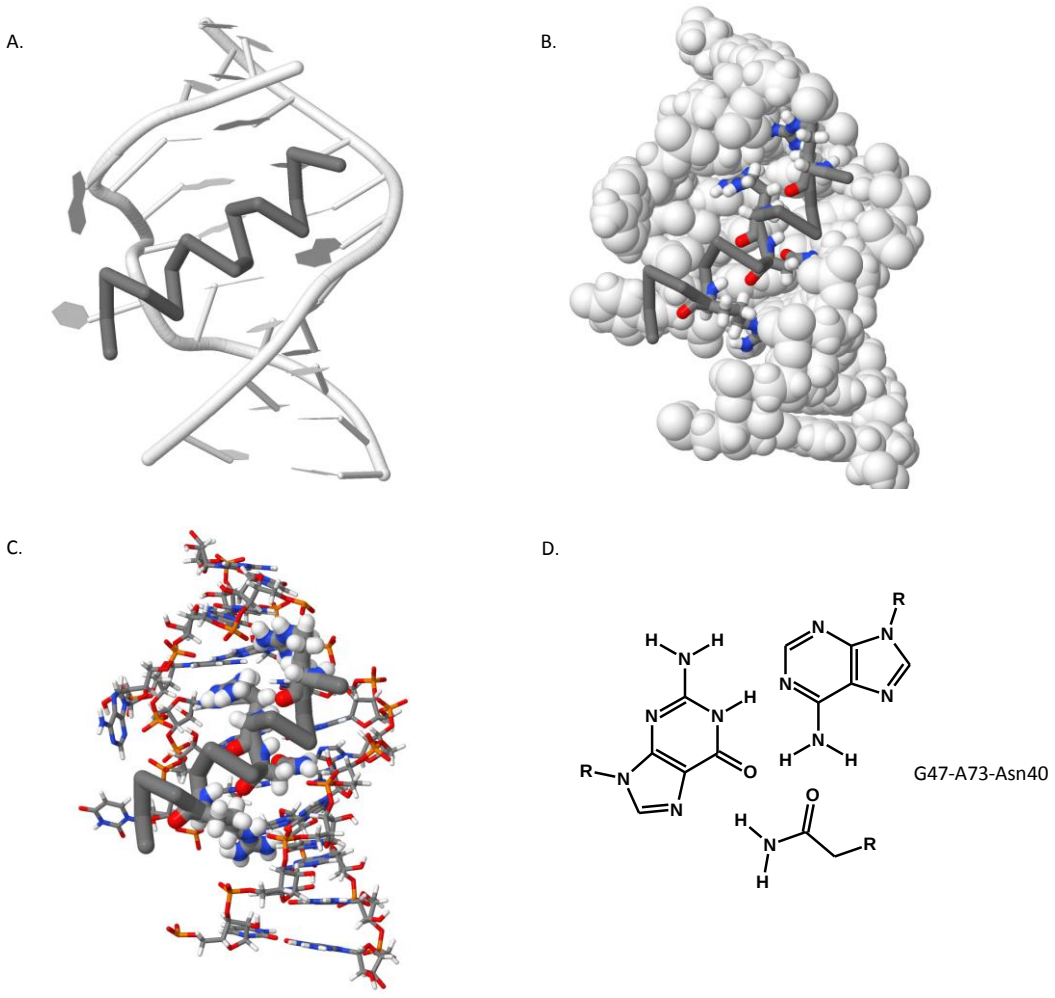


Figure 9. Views of important interactions between restrictive amino acid side chains and nucleotides. A) Solution structure of the human immunodeficiency virus type 1 (HIV-1) Rev ARM-RRE IIB (Battiste et al., 1996). Rev binds deeply within the major groove of RRE stem-loop IIB as a α-helix. RNA rendered as cartoon and peptide rendered as backbone. B) As in A, but RNA rendered as space-filling. The important binding residues Arg35, Arg39, Asn40 and Arg44 are shown. C) As in B, but RNA rendered as sticks. D) Schematic representation of the hydrogen-bonding arrangements of G47-A73-Asn40 interaction. The side chain of asparagine 40 forms specific hydrogen bonds to G47•A73.

3. Sequence and Structure Space of Rev and RRE Binders

The optimization of Rev-RRE interaction has been an interesting approach to understand how proteins and RNA achieve their high specificity and evolve new recognition strategies.

The systematic evolution of ligands by exponential amplification (SELEX) has generated several Rev-binding RNA aptamers with novel sequence and structural motifs. However, all RRE variants obtained exhibit high affinity for Rev only when a G•G or an A•A pair could be formed between the base positions 48 and 71. These results support Rev recognizing structural features associated with a homopurine base pair at the internal loop of RRE (Giver et al., 1993a, Giver el at., 1993b). Ye et al. (1996) found a 35-mer RRE RNA aptamer of class I type with higher binding affinity for Rev ARM (Ye et al., 1996). The complex formation involves the α-helical Rev ARM targeting a widened RNA major groove with non-canonical base pairs G47•A73 and A48•A71 (instead of G48•G71 in wild-type RRE IIB). The same HIV-1 Rev ARM binds an RRE RNA aptamer of class II type that has a very different secondary structure using a distinct binding strategy (Ye et al., 1999).

In addition to selecting RNA aptamers, different peptide aptamers that bind RRE were also selected (Harada et al., 1996; Harada et al., 1997). The nucleotide base requirements for the binding of RRE to Rev aptamers RSG-1.2 and K1peptides were determined and revealed distinct binding strategies (Iwazaki et al., 2005; Sugaya et al., 2008). The nucleotides important for RSG-1.2 binding were localized at the internal loop while nucleotides in the upper stem required for wild-type Rev binding (C51, U66, and G67) were not conserved. The G48•G71 base pair was found to be important for both RSG-

1.2 and K1 peptide binding. The RSG-1.2 peptide was found to bind RRE IIB in an unstructured-turn-helix conformation (Gosser et al., 2001; Zhang et al., 2001) while K1 peptide adopts an α-helical confirmation upon RRE binding (Peled-Zehavi et al., 2003) . The K1 peptide contained a conserved glutamine residue that may bind to the G47•A73 base pair of RRE in a manner analogous to the asparagine residue of the Rev peptide. Another study used randomized-codon libraries of Rev ARM and assayed them for their ability to bind RRE IIB using a two-plasmid lambda antitermination system. Arg39, Asn40 and Arg44 appeared immutable with the RRE IIB reporter while Arg35 tolerated mutations with only moderate loss of function. Interestingly, the activity of N40V mutant was restored by another R35G mutation. The Rev-ARM double mutant, Rev-R35G:N40V binds RRE IIB well, indicating a distinct binding strategy exists (Possik et al., 2013). Diverse natural sequences of Rev and RRE sequences in HIV-1 also exist. The Los Alamos National Laboratory (LANL) HIV database contains nucleotide and protein sequence alignments of HIV genomes sequenced from AIDS patients of different origins. Previous work has shown that minimal conflict appears between experimentally selected variants of Rev and the sequences found in the clinic (Possik et al., 2013), suggesting that Rev-RRE binding reflects Rev viral function. Consistent with the literature, positions R35, R39, N40 and R44 were the most restricted residues in the clinical database. Comparing natural RRE IIB sequences to the experimentally selected sequences will help explain the evolution and diversity of RRE IIB sequence as well as the restrictions imposed on them.

D. Evolution of RNA-Protein Recognition

Recognition of an RNA sequence by different peptides presents a model system that explores how RNA protein recognition strategies can evolve. Kimura's neutral theory of molecular evolution states that genotypes are sufficiently mutable and most mutations maintain phenotype and are therefore neutral (Kimura, 1968). Neutral network theories later suggested molecules evolve by single step mutations without intermediate loss of function and new phenotypes evolve by random genetic drifts (Fontana and Schuster, 1998b, Wilke and Adami, 2003; Ohta, 2002). A neutral network connects all sequences that differ by single mutations yet encode the same phenotype and speciation occurs upon transit from one network to another through an intersection sequence that displays two phenotypes (figure 10). When applied to RNA-protein recognition, a neutral network connects all RNA-protein complexes that adopt the same recognition strategy.

Schultes and Bartel (2000) established the existence of neutral paths between two evolutionary unrelated catalytic RNAs; tracing a stepwise change in sequence at over more than 40 positions while retaining function (Scultes & Bartel, 2000). Iwazaki et al. (2005) found a single nucleotide change in the proximal-apical stem of RRE IIB that switches peptide-binding specificity of the RRE from a bifunctional Rev- and RSG-1.2- binding mode to either Rev-specific or RSG-1.2-specific binding. This observation demonstrates how RNA can evolve alternative binding strategies in discrete steps without intermediate loss of function (Iwazaki et al., 2005).

In all lentiviruses the Rev-RRE interaction is essential for nuclear export of the genomic mRNAs necessary for viral replication (Tang et al., 1999); therefore, conserving a functional Rev-RRE complex is essential. All lentiviruses are believed to have diverged

from a common ancestor (Foley, 2000); therefore, there exists strong evidence that all Rev-RRE complexes of all lentiviruses are related through a smooth mutational path. Using computational strategies Lesnik et al. (2002) revealed that the RREs in each sub-group of lentivirus possess some common structural features and other structural features unique to each sub-group. This observation reveals neutrality in RNA and protein genotype-phenotype maps where structures/functions are much better conserved during evolution than sequences. Therefore, predicting a fitness landscape model for lentiviruses might help understanding how these viruses evolve mutational robustness.

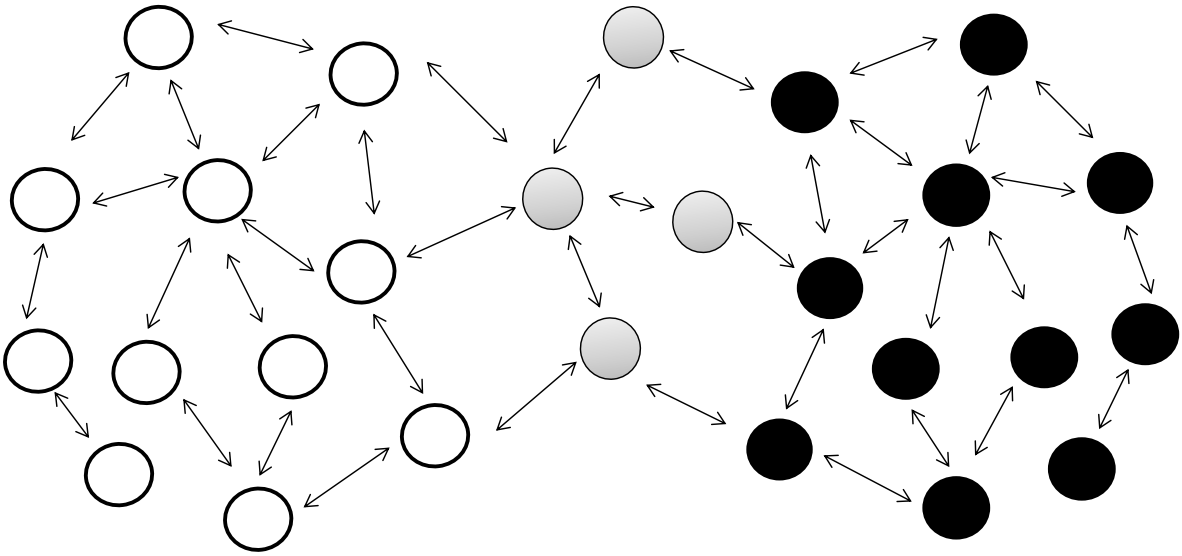


Figure 10. Schematic representation of neutral evolution of macromolecules. Protein or RNA sequences (genotypes) shown as circles are connected by point mutations shown in arrows. The colors white and black represent distinct functions (phenotypes) forming neutral mutational networks. The grey color represents transitional functions of bifunctional intermediates.

CHAPTER II

MATERIALS AND METHODS

A. General

Bacterial medium components were obtained from Oxoid Ltd (United Kingdom). ONPG (ortho-nitro-phenol-galactopyranoside) and IPTG (isopropyl-D-thiogalactoside) were obtained from Acros organics (Belgium). X-Gal (5-bromo-4-chloro-3-indolyl-D-galactoside) was obtained from Amresco (United States). Fine chemicals were obtained from Amresco (United States), Sigma (United States), and Amersham (United Kingdom). Disposable plasticware was obtained from Sarstedt (France).

B. Bacterial Strains and Plasmids

Escherichia coli supporting antitermination, N567 was obtained from Naomi Franklin (University of Utah). One Shot Max Efficiency DH5α-T1 competent cells were obtained from Invitrogen (United States). DNA oligonucleotides were obtained from TIB Mol BIOL (Germany). Plasmids expressing the BIV Tat N fusion (pBRN-BIV Tat) were obtained from Kazou Harada (Tokyo Gakugei University). Plasmids expressing the Rev17 N fusion (pBRN-HIV Rev17) and Rev-R35G:N40V N fusion (pBRN-HIV Rev-R35G:N40V) were available in house (Possik et al., 2013).

C. Library Construction

1. Oligonucleotides

a. Primer Extension

The complementary sequence of the library oligonucleotides were first synthesized by primer extension to produce double stranded DNA sequence using the primer boxAF with the following sequence: 5'-GTC GAC GCT CTT AAA AAT TAA-3'. In 100 µl reactions, 80 pmol of dsDNA libraries were synthesized with the following primer extension reaction: 10 µM template, 25 µM primer, 25 µM dNTP, 25 mM MgCl2, 1 X Taq buffer (500 mM KCl, 100 mM Tris HCl pH 8.3, 15 mM MgCl2, 0.1% gelatin) and Taq polymerase (30U as described in Engelke et al.,1990), and H_2O. The primer, template, $MgCl_2$, Taq buffer, and H_2O were first mixed and incubated at 95 ºC for 5 min. The tubes were placed on ice then dNTP and Taq polymerase were added. Primer extension reactions were carried out with the following thermo cycler program: 5 min at 40 ºC; 5 min at 50 ºC; 5 min at 55 ºC; 5 min at 60 ºC; 5 min at 65 ºC; 5 min at 70 ºC; 20 min at 72 ºC, hold at 4 ºC (Bio-Rad DNA engine, Pelter Thermal Cycler, California, USA).

b. Restriction Enzyme Digestion

Twenty pmol of the primer extended libraries with a PstI cohesive end were cut with restriction enzyme BamHI (Roche, Germany) to allow ligation into pAC plasmid with matching ends made by PstI and BamHI restriction endonucleases as described later. Twenty U BamHI was added to 40 pmol dsDNA. The digested DNA libraries were phenol chloroform extracted, ethanol precipitated and resolved on a 15% acrylamide gel (for

$10cm^3$; 7.5 g urea (Scharlau), 7.5 ml of 40% acrylamide (1:19), 1.5ml 10X TBE, 75 µl ammonium persulfate (APS), 10 µl TEMED) where complete digestion was observed.

c. Phenol Chloroform Extraction/Ethanol Precipitation

Following primer extension, the dsDNA libraries were purified from proteins, salts and other reaction components by phenol/ chloroform extraction. An equal volume of phenol/chloroform isoamyl alcohol (DNase, RNase, Protease free pH 7.8-8.2) (Acros Organics, United States) was added to the DNA solution. The tube was vortexed for 10 sec then centrifuged for 30 sec at 14000 rpm, room temperature. This step was repeated as necessary to get rid of any white interface. The top aqueous phase containing DNA was transferred to a new tube and an equal volume of chloroform (Scharlau, Spain) was added and the tube was vortexed for 10 sec then centrifuged for 30 sec at 14000 rpm, room temperature. The supernatant was transferred to a new tube to which 0.1 volume of 3M NaOAc (PH 5.2) was added. The tube was vortexed for 10 sec and 2.5 volume 100 % ethanol was added; the tube was mixed by inversion, placed on ice for 30 min, and centrifuged for 15 min at 14000 rpm at 4ºC. The supernatant was discarded and 1 ml of 75 % ethanol was added to wash the pellet. The tube was centrifuged for 5 min at 14000 rpm, room temperature. The supernatant was discarded and the pellet was left to dry. The library DNA was resuspended in 40 µl ddH_2O to obtain a DNA concentration of 2 µM. To confirm successful primer extension, 2 µl of the DNA solution was mixed with 5 µl 2x agarose loading buffer and checked on a 4% agarose gel by electrophoresis.

2. Backbone Digestion and Purification

a. Plasmid Restriction Enzyme Digestion

The pAC plasmid extracted was digested with PstI and BamHI to allow insertion of library constructs. 10 µg of plasmid DNA were incubated at 37 ºC for two hours with 200 U Pstl, 1 X Tango buffer (Roche, Germany. Another 10 µg of plasmid DNA were incubated at 37 ºC for two hours with 200 U BamHI, 1 X Tango buffer (Roche, Germany). 10 µl of each digestion were saved at 0 h and 1 h after digestion to run on a gel for comparison. The two digestion mixtures were mixed after two hours, 200 U Bam HI and 200 U PstI were added and re-incubated for an hour at 37 ºC. Complete double digestion was confirmed by agarose gel electrophoresis.

b. Column Purification of Digested Plasmid

To purify the plasmid DNA, the digested plasmid solution was ran on a 1% agarose gel. The DNA band corresponding to the double digested plasmid was cut from the gel and column purified using the GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, United Kingdom). The agarose gel band containing the pAC plasmid was transferred to a 1.5 ml microcentrifuge tube. For each 10 mg gel slice, 10 µl of capture buffer was added and incubated at 60 ºC until the agarose was completely dissolved. The sample was then transferred to a GFX column placed in a collection tube and was incubated at room temperature for 1 min. The sample was centrifuged for 30 sec at 14000 rpm, room temperature and the flow-through was discarded. The GFX column was placed inside the collection tube again and 500 µl of wash buffer (Tris-HCl pH 8.0, 1 mM EDTA, and 80% ethanol) was added. The tube was centrifuged for 30 sec at 14000 rpm and the flow-through

was discarded. The tube was centrifuged again for 30 sec and the GFX column was transferred to a 1.5 ml micro centrifuge tube and 50 µl elution buffer (autoclaved, de-ionized distilled water) was added. The sample was incubated for 1 min at room temperature then centrifuged for 1 min at 14000 rpm to recover purified DNA. The concentration of the purified plasmid backbone was checked by a 1% agarose electrophoresis.

3. Ligation

The digested library constructs were ligated into the pAC plasmid using the following 10 µl ligation reaction: 7 fmol digested library dsDNA, 21 fmol double digested backbone plasmid, 10X T4 DNA ligase buffer (Roche, Germany), 10 mM ATP, 10 mM DTT. The ligation reaction was incubated overnight at room temperature. 20 µg of glycogen (USB Ultrapure 77534) was added to the ligations followed by phenol extraction and ethanol precipitation. The ligations were tested by transforming into competent cells; uncut plasmid was used as a positive control and cut plasmid with no insert was used to estimate background (figure 11).

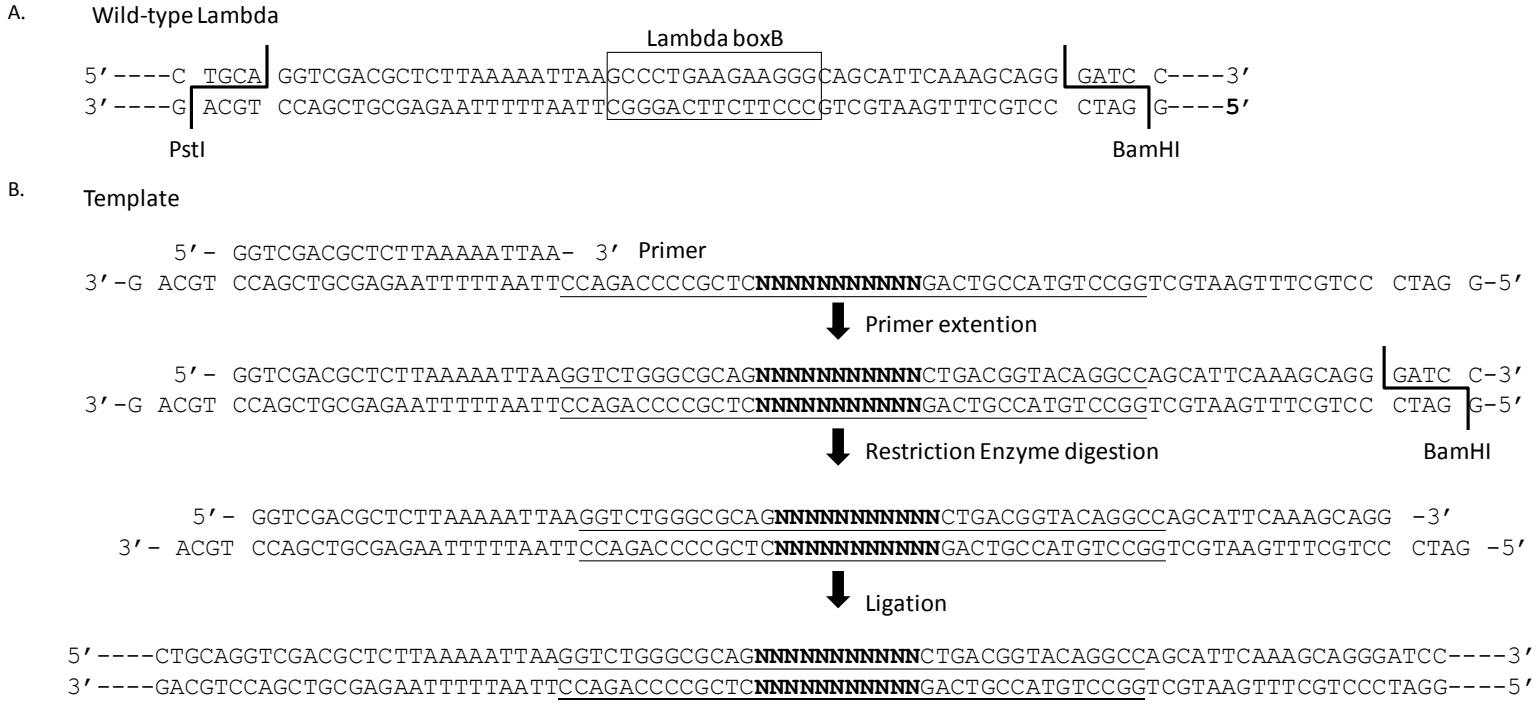


Figure 11. Cloning strategy. A) pAC nut- plasmid's cloning cassette for replacement of lambda boxB with RRE IIB libraries. B) Representative design of degenerate library oligonucleotide (RRE FL11NL) showing synthetic template for primer extention followed by restriction enzyme digestion and ligation into pACnut- plasmid's cloning cassette. The RRE IIB sequence replacing lambda boxB is underlined with the random nucleotides shown in bold.

D. Competent Cell Preparation

Two days before the experiment, LB agar medium (10 g NaCl, 10 g tryptone, 5 g yeast extract, and 15 g bacto-agar per liter) supplemented with ampicillin (100 µg/ml) was streaked with the desired glycerol stock of cells (N567/ pBRN-HIV Rev17, pBRN-HIV Rev-R35G:N40V and pBRN-BIV Tat) and incubated overnight at 34 ºC. The next day a colony was picked and inoculated into 5 ml LB broth with ampicillin and grown overnight at 37 ºC with shaking at 200 rpm. On the day of the experiment, 2 ml of the overnight culture were added to 200 ml of LB medium and grown at 37 ºC with shaking at 200 rpm while frequently measuring the OD_{600}. At OD_{600}= 0.37 the growth was stopped and the cultures were transferred immediately to an ice water bath and swirled for about 5 min to allow rapid cooling. The cultures were poured into pre-chilled 50 ml conical tubes and centrifuged for 15 min at 4500 rpm at 2 ºC. The supernatant was discarded and pellets were re-suspended with 10 ml ice cold Competent Cell Elixir solution (60 mM $CaCl_2$, 15% glycerol, and 10 mM PIPES; piperazine-N, N'-bis-2-ethanesulfonic acid; pH 7). The cells were centrifuged again for 10 min at 4500 rpm at 2 ºC and the supernatant was discarded. While keeping cells constantly on ice, the pellets were re-suspended a second time with 10 ml of ice cold Competent Cell Elixir solution and the suspension was kept on ice for 30 min. The tubes were then spun for 10 min at 4500 rpm at 2 ºC, the supernatant was discarded, and the pellet was re-suspended in 2 ml of ice cold Competent Cell Elixir solution, aliquots of 300 µl were transferred to pre-chilled 1.5 ml tubes and stored at – 80 ºC.

E. DNA Preparation

1. DNA Midiprep

A glycerol stock of DH5α /*E. coli* cells previously transformed with the pAC plasmid was streaked onto Luria Bertani (LB) agar medium (10 g NaCl, 10 g tryptone, 5 g yeast extract, and 15 g bacto-agar per liter) supplemented with chloramphenicol (15µg/ml). The plate was incubated overnight at 37 ºC and a single colony was picked with an autoclaved wooden skewer, inoculated into 200 ml of LB broth (10 g NaCl, 10 g tryptone, and 5 g yeast extract) and grown overnight at 37 ºC. The following day, the culture was transferred to 50 ml conical tubes and centrifuged for 15 min at 4500 rpm at 4 ºC. The supernatant was discarded and the pellet was resuspended in 1.5 ml SET (20 % sucrose, Tris HCl (50 mM), EDTA (50 mM), pH 8.0, autoclaved). The suspension was placed on ice and 3.5 ml of alkaline lysis solution (NaOH (0.2 M), 1% sodium dodecyl sulfate) was added and kept on ice for exactly 5 min after mixing by inversion. To the lysed suspension, 2.5 ml of cold 3 M potassium acetate was added. The suspension was mixed by inversion and kept on ice for 10 min after which the tube was centrifuged for 20 min at 45000 rpm at 4 ºC. The supernatant was then transferred to 40 ml centrifuge tube to which 6 ml of isopropanol were added, mixed by inversion and centrifuged for 10 min at 14000 rpm at 4 ºC. The supernatant was discarded. The pellet was washed with 20 ml 75% ethanol and precipitated by centrifugation for 5 min at 14000 rpm at room temperature. The supernatant was discarded and the pellet was dried and resuspended in 400 µl ddH_2O. The solution was incubated with 10 µl RNase A (Roche, Germany) for an hour at room temperature. Two micro liters of the DNA preparation was ran on a 1% agarose gel to check for RNA digestion. The DNA preparation was then phenol/chloroform extracted and ethanol

precipitated as described above and checked again by agarose electrophoresis for concentration.

Two ml of library plasmids' glycerol stocks in One Shot Max Efficiency DH5α-T1 competent cells were inoculated into 250 ml LB broth supplemented with chloramphenicol (15 µg/ml) and grown at 37 ºC with shaking for 3 hours until an optical density of OD_{600}= 1 was reached. The cultures were transferred to 50 ml conical tubes and plasmid DNA was then prepared following the procedure described for pAC plasmid preparation.

2. DNA Miniprep

The active clones obtained from transforming library containing plasmids (reporter plasmids) into N567 competent cells carrying the supplier plasmids (pBRN Rev17, pBRN Rev-R35G:N40V and pBRN BTat) were picked and pooled. DNA isolation was carried out as described for pAC plasmid preparation. The pAC plasmids containing active RNA sequences were isolated by running 250 µl of the isolated DNA on a 1% gel in a 360 µl well (10 mm height, 18 mm width, 2 mm thick) along with pAC and pBRN plasmid for size comparison. The bands were visualized under UV at wavelength 560 nm and the band corresponding to the pAC plasmid was cut and placed in a pierced 500 µl tube with polyester at the bottom. The 500 µl tube was placed in a 1.5 ml microcentrifuge tube and the setup was spun for 5 sec. The DNA was transformed into N567 competent cells. Cells were scraped off the plates and mixed with 4x glycerol stock solution for storage.

Streaks from glycerol stocks with pool of active clones were made on LB/chloramphenicol plates and incubated overnight at 34 ºC. Single colonies from each

library were individually restreaked on LB agar medium supplemented with chloramphenicol and incubated overnight at 34 ºC.10 ml of LB medium supplemented with chloramphenicol were inoculated with the desired active bacterial colonies in 15 ml culture tubes a day before the experiment and incubated overnight at 37 ºC. On the day of the experiment, the cultures were centrifuged at 4500 rpm for 15 min, at 4 ºC. Supernatants were discarded and pellets were suspended with 225 µl of SET (20% sucrose, 50 mM Tris pH 8.0, and 50 mM EDTA). The suspension was transferred to 2 ml microcentrifuge tubes and kept on ice. 525 µl of freshly made lysis buffer (0.2 M NaOH and 1% SDS) was added and the suspensions were gently mixed by inversions and kept on ice for exactly 5 min. Two hundred seventy five µl of 3 M KOAc (pH 4.8) were then added to the lysed suspension and tubes and left on ice for 10 min. Suspensions were then centrifuged at 14 krpm for 10 min and supernatants were transferred to 2 ml microcentrifuge tubes. Six hundred µl of isopropanol was added to the supernatant and mixed by inversion then centrifuged at 14 krpm for 10 min. Supernatant was discarded, the pellets were dried under the hood and resuspended in 50 µl of autoclaved water mixed with 5 µl of RNase A at 10 mg/ml. One µl of each sample was run on a 1% agarose gel for visualization and concentration estimation.

F. Transformation of Chemically Competent Cells

The transformation efficiency of One Shot Max Efficiency DH5α-T1 competent cells was tested as follows: 25 µl of One Shot Max Efficiency DH5α-T1 competent cells were added to 1 µl of DNA in a 1.5 ml tube, the tube was tapped gently for mixing and incubated at ice for 30 min. Cells were heat-shock for exactly 30 sec at 42 ºC water bath then placed back on ice. One ml of warm SOC medium (20 g tryptone, 5 g yeast extract, 2.5 ml 4M NaCl for 1 liter; 2.5ml 1M $MgCl_2$, 10 ml $MgSO_4$, 20 ml 1M glucose added before use) was added to the transformation and allowed to grow for 2 hrs at 37 ºC. After incubation, 10% of the transformation was plated on dried LB/chloramphenicol plates. The plates were incubated overnight at 34 ºC. In parallel and following the same procedure, 25 µl of chemically competent DH5α cells were transformed with 1 µl of the same DNA to compare the transformation efficiency difference between the two cells.

One µl of the ligation reactions were transformed into One Shot Max Efficiency DH5α-T1 competent cells following the procedure described earlier while plating the whole transformation on LB/chloramphenicol plates. For the ten libraries, ~ 1000 colonies were obtained, scraped off the plates and mixed with 4 x glycerol stocks for storage.

The efficiency of chemically competent cells prepared in the laboratory was tested as follows: chemically competent cells were thawed on ice and 50 µl were added to 1 µl DNA in a 1.5 ml tube. The tubes were incubated on ice for 20 min. Cells were heat-shocked for 3 min in a 37 ºC water bath and were placed immediately on ice afterwards. One ml of LB medium was added to the transformation tubes and incubated at 37 ºC for 1 hour. Immediately after incubation, the whole transformation was plated on

LB/ampicillin/chloramphenicol plates. The plates were incubated overnight at 34 ºC. Ten thousand colonies were obtained from each transformation.

N567 chemically competent cells carrying the supplier plasmids (pBRN Rev17, pBRN Rev-R35G:N40V and pBRN BTat) were transformed with reporter plasmids with the libraries of interest (pAC RRE Libraries) or with the control plasmids (pAC flRRE and pAC BTAR). After 20 min of incubation on ice and 3min of heat shock at 37 ºC, 1ml of LB broth was added to the transformation, and tubes were incubated for 2 hours at 37 ºC. Transformations were plated in percentages that would give around 300 colonies per plate on dried tryptone agar plates (5 g NaCl, 10 g tryptone, and 15 g bacto-agar per liter) supplemented with 100 µg/ml ampicillin, 12 µg/ml chloramphenicol, 80 µg/ml X-Gal and 0.1mM IPTG (isopropyl-D-thiogalactoside). Plates were incubated overnight at 34 ºC and left at room temperature after then. Transformation results were recorded after 24 and 48 hours.

G. Antitermination Assays

1. β-galactosidase Colony Color Assay

Blue colonies showing 4+ (wild-type), 5+ or 3+ activities from the transformations of pAC RRE Libraries with N567 chemically competent cells carrying the supplier plasmids pBRN Rev17and pBRN Rev-R35G:N40V were picked and grown individually in 200 µl LB medium supplemented with chloramphenicol. The cultures were grown overnight at 37 ºC with shaking at 180 rpm. Individual colonies from same libraries were pooled and DNA was extracted as described for library plasmids.

Individual active clones were transformed with 20 µl of diluted competent cells to test for activity and specificity. The whole transformation was plated on dried tryptone agar plates (5 g NaCl, 10 g tryptone, and 15 g bacto-agar per liter) supplemented with 100 µg/ml ampicillin, 12 µg/ml chloramphenicol, 80 µg/ml X-Gal and 0.1mM IPTG (isopropyl-D-thiogalactoside). The plates were incubated overnight at 34 ºC and checked for activity and specificity after 24 and 48 hours. Samples showing 4+ or 3+ activities with Rev17 or Rev-R35G:N40V and 0 activity with BTat were chosen for sequencing (Harada and Frankel, 1991).

2. β-galactosidase Solution Assay

To quantify the activity between RRE IIB and Rev17 or Rev-R35G:N40V reported by the X-gal plate assays, four representative individual colonies from every screen were inoculated in sterile tryptone medium supplemented with chloramphenicol (34 µg/ml), ampicillin (100 µg/ml), and IPTG (50 µM). Cultures were grown overnight for 17-20 hours at 30 ºC with shaking at 210 rpm. On the day of the experiment, in 1.5 ml tubes, 500 µl of

Z 2+ buffer (Z buffer: 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$+ 2% SDS+ β-mercaptoethanol (175 µl for 50 ml Z buffer)) were added to 125, 250 or 500 µl of cells depending on the expected activity of the interaction; strong, intermediate and weak respectively. Forty µl of chloroform were added to the reaction tubes. The total volume in the tube was normalized to 1 ml using sterile tryptone broth as needed. Cells were lysed by vigorous vortexing for 10 sec followed by incubation at 28 ºC for 5 min. Reactions were started after addition of 200 µl of ortho-nitro-phenol-galactopyranoside (ONPG in Z buffer; 4 mg/ml), the substrate of the β-galactosidase enzyme. The reactions were stopped by adding 500 µl of 1 M Na_2CO_3 when a faint yellow color started to develop. Tubes were centrifuged to pellet debris. The start and end times were accurately recorded. The intensity of yellow color reflecting activity was measured by spectrophotometry at OD 420 and the background activity that accounts for presence of cellular debris was subtracted by measuring at OD 550. Cellular growth in culture was measured separately at OD 600. The following formula was used to calculate units of β-galactosidase:

$$1000 \times \frac{(OD420 - 1.75OD550)}{t \times v \times OD600}$$

in units βGal with t and v representing the time of the reaction in minutes and the volume of culture used in the assay respectively (Miller, 1992).

H. Sequencing

Individual active clones were column purified using the GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, United Kingdom). Thirty µl of DNA ready to be purified was mixed with 500 µl of capture buffer and transferred to a GFX column. After

centrifugation for 30 sec at 14 krpm, the elutant was discarded and 500 µl of wash buffer (Tris-HCl pH 8.0, 1 mM EDTA, and 80% ethanol) was added to the column. The tube was centrifuged for 30 sec at 14 krpm, the elutant was discarded and the column was centrifuged another time. The column was transferred to a previously labeled 1.5 ml microcentrifuge tube and 30 µl of elution buffer was added. The setup was incubated for 1 min then centrifuged one last time for 1 min at 14 krpm. One µl of each sample was run on a 1% agarose gel for visualization. Twenty µl of the column purified DNA was transferred to a sequencing plate. PACF2 (100 mg/ml) (AATCACTGCATAATTCGTGTC) was used as a primer for sequencing. Samples were sequenced at Macrogen.

I. Sequence and Structure Visualization Program

Jmol (htto://www.jmol.org/), an open-source Java viewer for chemical structures in 3D, was used to view NMR structures of HIV-1 Rev-RRE (PDB 1ETf; Battiste et al., 1996), λ N-boxB (PDB 1QFQ; Legault et al., 1998), and BIV Tat-Tar (PDB 1mnb; Puglisi et al., 1995).

J. Analysis of RRE Variants from HIV Database

A pre-aligned set of 1848 sequences of the RRE IIB (7809-7847) was extracted from the Los Alamos National Laboratory HIV Sequence Database (http:// http://www.hiv.lanl.gov/).

CHAPTER III

RESULTS

Asn40 is essential and Arg35 is important in Rev-RRE recognition: the NMR model suggests that Asn40 makes hydrogen bonds to G47 and A73 and Arg35 contacts the nucleotides U66, G67 and G70 (Battiste et al., 1996) (figure 9). Surprisingly, Rev-R35G:N40V binds RRE IIB (Possik et al., 2013). Presumably, Rev-R35G:N40V uses a distinct recognition strategy from wild-type Rev. How different the recognition strategy of Rev-R35G:N40V is from wild-type Rev is unclear. The existence of a nearby, distinct recognition strategy is intriguing and supports neutral theories of evolution that posit sufficient incremental mutations are functional such that new phenotypes are accessible by genetic drift. To explore the how well neutral theories describe the mutability of Rev-RRE and Rev-R35G:N40V-RRE interactions, RRE IIB libraries were screened for active clones with Rev-R35G:N40V and the wild-type Rev17 using a two-plasmid reporter system based on lambda N-nut antitermination (figure 2). In the N expressor plasmid, the RNA-binding domain of Rev17 or Rev-R35G:N40V replaces the arginine-rich domain of the lambda N protein to create N-ARM fusion (Harada and Frankel, 1991). RRE IIB randomized libraries replace boxB of the lambda nut site in the RNA reporter plasmid. Upon binding of the N-ARM fusion to boxB site, the antitermination complex assembles and transcribes past terminators to express the LacZ reporter gene. The activity of the LacZ gene reflects the affinity of the interaction. Following screening, active clones were collected.

To test the specificity of the reporter system, an initial screen of wild-type and control interactions were used to verify that the two-plasmid antitermination reporter assay accurately measures Rev ARM-RRE IIB recognition (table 1). A bovine immunodeficiency virus Tat N-ARM fusion expressor (BTat) was used as a heterologous control because RRE IIB does not bind BTat. The activity of each interaction was measured by antitermination solution assay; RRE libraries displayed antitermination activities with the cognate Rev significantly above background. These results show that activities measured with the plasmid reporter assay accurately reflect recognition between RRE IIB and the RNA-binding domain of Rev peptide.

Table 1 X-gal activities of positive and negative controls.

RNA-binding domain	Sequence[c]	X-gal[a](β-galactosidase units)[b] flRRE[d]	nmrRRE[e]	BTAR[f]
Rev17[g]	TRQARRNRRRRWRERQR	4 (383±100)	3	0
Rev-R35G:N40V[h]	T**G**QARR**V**RRRRWRERQR	4 (40±10)	3	0
Rev-R35G[i]	T**G**QARRNRRRRWRERQR	(13±2)	NA	0
Rev-N40V[j]	TRQARR**V**RRRRWRERQR	0	0	0
BTat[k]	RPRGTRGKGRRIRR	0	0	5

[a] X-gal colony color assays were performed as described (Harada and Frankel, 1991). Numbers represent +'s used to score blue color intensity.

b β-galactosidase solution assays were performed as described (Miller, 1992). RRE IIB libraries were transformed into expressor cells. Four replicates of each clone were grown at 30ºC in tryptone medium supplemented with IPTG and assayed for β-galactosidase activity.

[c] The amino-acid sequences of the RNA-binding domain of peptides used with the nucleotides differing from the wild-type Rev shown in bold and underlined.

[d] flRRE has wild-type sequence: GGUCUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGCC

[e] nmrRRE is the same as the RNA used in NMR structural studies. DNA sequence: GGUCUGGGCGCAGCGCAAGCUGACGGUACAGGCC.

[f] BIV TAR is a BIV Tat-binding RNA from bovine immunodeficiency virus that serves as a heterologous control. DNA sequence: GCUCGUGUAGCUCAUUAGCUCCGAGC.

[g] Rev17 contains Rev34-50.

[h] Rev-R35G:N40V is a double mutant that binds RRE II in a distinct strategy from wild-type Rev (Possik et al., 2013).

[i] Rev-R35G is a single mutant that binds weakly to RRE IIB.

[j] Rev-N40V is a single mutant that does not bind to RRE IIB.

[k] BIV Tat contains an arginine-rich beta-turn peptide (figure 5C; Puglisi et al., 1995) from the bovine immunodeficiency virus that binds BIV TAR with high affinity and specificity for use as a heterologous control. Background was measured as antitermination activity of the wild-type flRRE with the BTat protein which has the BTat RNA-binding domain replacing the N peptide in the pBRN plasmid. FlRRE displayed no activity with the negative control BTat.

A. Initial Screens Reveal Different Proportions of RRE Active Clones with Rev17 and Rev-R35G:N40V

Ten libraries were constructed based on degenerate synthetic oligonucleotides and ligated into reporter plasmids (figures 1and 11). The actual complexity of each library was less than theoretical due to the number of colonies obtained after transforming the ligations into competent cells (table 2A). Separate, multiple rounds of screening were anticipated to collect sufficiently enriched pools of active clones with wild-type Rev and Rev-R35G:N40V. Figure 12 illustrates a representative screening and selection process of RRE FL11NL with Rev17. Following the same procedure, other RRE IIB libraries were screened and active clones with Rev17 and Rev-R35G:N40V were selected. The proportions of active clones are summarized (table 2B). The number of verified positives obtained after several rounds of selection is a reflection of the strictness of the RRE-binding requirements.

Most libraries yielded different proportions of active clones with Rev17 and Rev-R35G:N40V (table 2B), with the greatest differences observed in the RRE IIB libraries 43N46-74N77, 50N52-66N69 and fl11NL. The 43N46-74N77and 50N52-66N69 regions are involved in base-specific contacts in wild-type Rev, and the proportions of active clones at each region are consistent with wild-type and double mutant Rev employing different recognition strategies.

Different regions also yielded different proportions of active clones in the ten libraries. The highest proportions of active clones were observed in libraries distal to the internal loop region; at the apical loop library RRE fl11NL and the distal apical stem library RRE 43N46-74N77. Lower proportions of active clones were obtained in libraries

spanning the proximal-apical stem and internal loop region, consistent with the known roles of different bases at these regions in recognition.

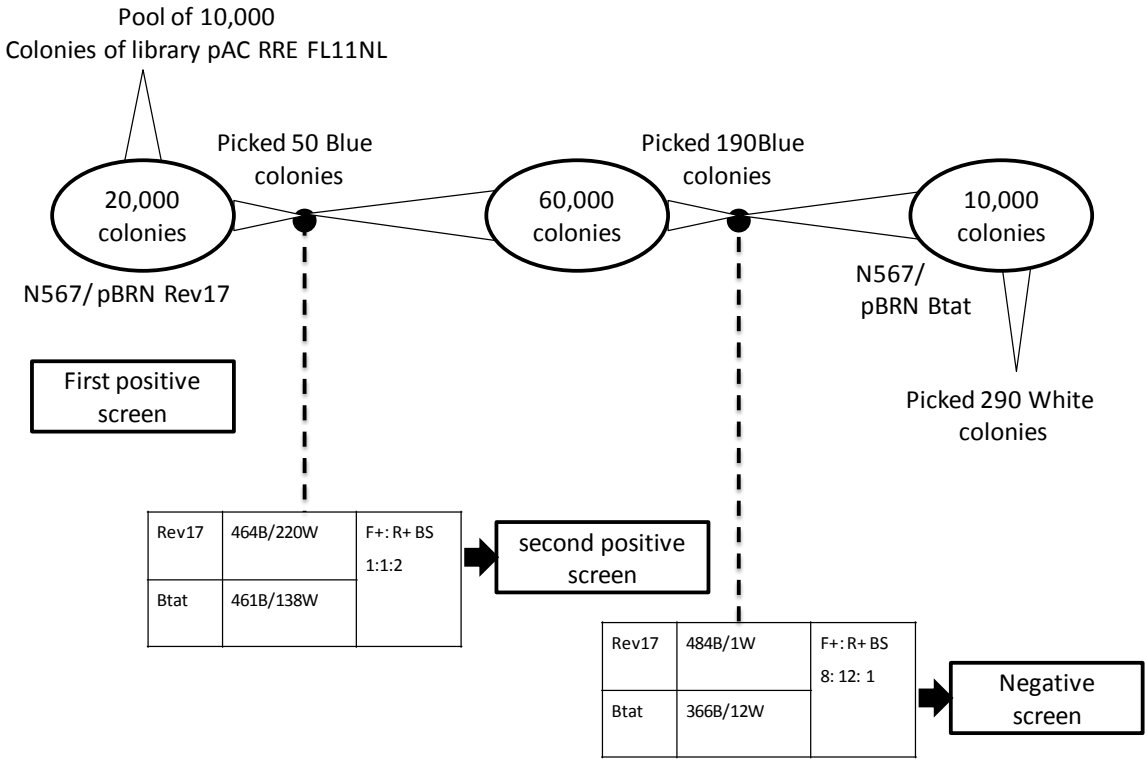


Figure 12. Screening and selection of RRE IIB active clones. A representative scheme of selection of RRE 51N53 active clones screened with Rev17. After ligation into the pAC plasmid and amplification of clones in One Max Shot efficiency DH5α-T1 competent cells, 40,000 colonies of library RRE 51N53 were transformed into pBRN Rev17. Of the 20,000 colonies obtained, 50 blue colonies were picked and positive clones were isolated. The isolated DNA was transformed with Rev17 and BTat, number of blue and white colonies from each screen were recorded and ratio of false+ (F+): real+ (R+): bystanders were calculated. Of the total 50 pool, 50 % were bystanders; therefore, a second positive screen was performed to get rid of the bystanders. Out of 60,000 colonies, 190 blue colonies active with Rev 17 were picked and screened for activity on the BTat to detect false positives. Out of 190 colonies, 38% were false positives; therefore, a negative screen was next performed to get rid of the false positives. Plasmid DNA from second screen was transformed into N567/pBRN BTat and 290 white colonies were picked out of the total 10,000 colonies. The plasmids of the 290 white colonies were isolated, pooled and transformed into N567 competent cells. Glycerol stocks were prepared for later DNA preparation of individual active clones to be screened for activity and specificity as well as for sequencing.

Table 2A Estimated proportions of positive clones obtained in each library.

RRE IIB library	Sequence[a]	Number of random nucleotides	Theoretical complexity[b]	Estimated # of positives[c]	Estimated % of positives
43N46-74N77	GG**NNNN**GGCGCAGCGUCAAUGACGCUGACGGUA**NNNN**CC	8	66,536	120	0.20
45N47-73N75	GGUC**NNN**GCGCAGCGUCAAUGACGCUGACGGU**NNN**GGCC	6	4,096	10	0.15
46N48-71N74	GGUCU**NNN**CGCAGCGUCAAUGACGCUGACG**NNNN**AGGCC	6	16,384	50	0.30
47N49-70N73	GGUCUG**NNN**GCAGCGUCAAUGACGCUGAC**NNNN**CAGGCC	7	16,384	20	0.10
48N50-69N71	GGUCUGG**NNN**CAGCGUCAAUGACGCUGA**NNN**UACAGGCC	6	4,096	50	1.30
49N51-67N70	GGUCUGGG**NNN**AGCGUCAAUGACGCU**NNNN**GUACAGGCC	7	16,384	50	0.30
50N52-66N69	GGUCUGGGC**NNN**GCGUCAAUGACGC**NNNN**GGUACAGGCC	7	16,384	100	0.60
51N53-65N67	GGUCUGGGCG**NNN**CGUCAAUGACG**NNN**ACGGUACAGGCC	6	4,096	300	7.80
52N54-64N66	GGUCUGGGCGC**NNN**GUCAAUGAC**NNN**GACGGUACAGGCC	6	4,096	30	0.80
FL11NL	GGUCUGGGCGCAG**NNNNNNNNNNN**CUGACGGUACAGGCC	11	4,194,304	1,300	0.05

[a] Nucleotide sequence of RRE IIB libraries with nucleotides randomized, N, shown in bold.

[b] The total complexity is 4^n different sequences, with n being the number of nucleotides randomized at a time.

[c] Estimated number of positives to be obtained in each library is calculated based on published data on mutability of RRE IIB summarized in figure 6.

Table 2B Observed proportions of positive clones obtained in each library.

RRE IIB library	Initial screen		Proportion of active clones picked after each round of selection			
	% of active RRE		Primary positive screen[a]		Secondary positive screen	
	Rev17	Rev-R35G:N40V	Rev17	Rev-R35G:N40V	Rev17	Rev-R35G:N40V
43N46-74N77	33.00	6.70	50/20,000	290/20,000	190/40,000	670/30,000
45N47-73N75	0.20	0.30	30/30,000	90/30,000	70/2,000	120/2,000
46N48-71N74	0.10	0.05	180/43,000	180/86,560	1,430/40,000	2,290/86,560
47N49-70N73	0.50	13.30	540/20,000	180/20,000	860/40,000	380/80,000
48N50-69N71	0.30	0.05	410/24,000	170/40,000	3,650/80,000	770/6,000
49N51-67N70	0.50	0.10	130/20,000	180/20,000	380/20,000	290/60,000
50N52-66N69	1.0	0.15	1,310/40,000	660/60,000	2,500/40,000	1,920/40,000
51N53-65N67	12.0	11.50	270/30,000	660/30,000	900/1,000	1,200/2,820
52N54-64N66	3.0	2.0	360/20,000	270/20,000	380/5,000	530/30,000
RRE FL11NL	47.0	13.0	50/20,000	310/20,000	190/60,000	670/40,000

[a] The proportion of the active clones in positive screens is the number of blue colonies picked, with activity similar or close to wild-type over the number of total colonies obtained after transformations

Table 2B Observed proportions of positive clones obtained in each library.

RRE IIB library	Proportion of active clones picked after each round of selection			
	Tertiary positive screen[b]		Negative screen[c,d]	
	Rev17	Rev-R35G:N40V	Rev17	Rev-R35G:N40V
43N46-74N77	190/4,000	580/4,000		
45N47-73N75			10/10,000	190/10,000
46N48-71N74			880/1,460	360/3,193
47N49-70N73	580/4,000	240/5,000	580/10,000	860/21,000
48N50-69N71			600/1,000	670/5,000
49N51-67N70	580/4,000	580/2,500	430/15,000	580/15,000
50N52-66N69			420/1,000	540/4,800
51N53-65N67			720/2,500	900/2,500
52N54-64N66	1,120/3000	580/5,000		
FL11NL		480/5000	288/10,000	

[b] Some libraries had a high number of bystanders after the first two positive screens and therefore required a tertiary positive screen.

[c] The proportion of the active clones in negative screens is the number of white colonies picked, with activity similar or close to wild-type over the number of total colonies obtained after transformations.

[d] The number of bystanders in some libraries remained high after three rounds of positive screen; therefore, negative screens were not performed for these libraries.

B. Rev-R35G:N40V Accepts a Different Panel of RRE Variants than Wild-Type Rev

Characterizing RRE IIB active clones revealed many active clones with both Rev17 and Rev-R35G:N40V and fewer that are functional with one RNA-binding domain. This confirms that the recognition strategies of wild-type Rev and the double mutant Rev are distinct and suggests that Rev-R35G:N40V binds in a similar mode with possible minor conformational adjustments compensating for the absence of Asn40 and Arg35.

RRE IIB residues C51, G69 and U66 have been previously found to be conserved and were shown to form hydrogen-bonding with R35 of the wild-type Rev (Battiste et al., 1996). Herein, these residues were not conserved in variants active with Rev-R35G:N40V consistent with the role of Arg35 in forming hydrogen bonds with these residues (figure13).

RRE IIB residues G50, A68, C69 and U72 have not been previously identified as important for sequence- or structure- specific recognition. Similarly, active clones obtained do not show a requirement for any particular sequence (Bartel et al., 1991). Previously a Watson-Crick base pair at position G50-C69 was not found to be important (Bartel et al., 1991); however, all active clones with Rev-R35G:N40V maintained a Watson-Crick base pair at this position suggesting a structural role of this base-pair in the recognition of the double mutant.

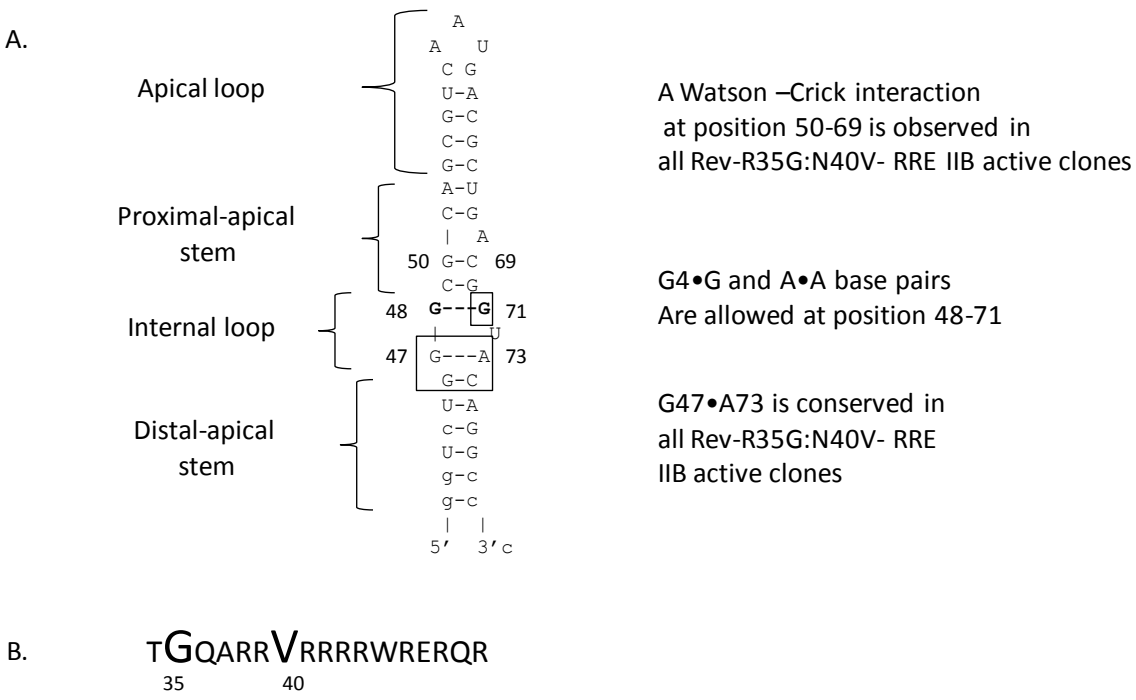


Figure 13. Minimal Rev-R35G:N40V-RRE complex. A) Sequence of RRE IIB. Boxed nucleotides indicate those that are conserved by antitermination screens. B) Amino acid sequence of the RNA-binding domain of Rev-R35G:N40V, which contains amino acids 34 through 50 of the RNA-binding domain. The large letters indicates amino acids differing from wild-type Rev.

All RRE variants active with Rev-R35G:N40V had a G•G or an A•A at position 48-71 and a conserved G47•A73, consistent with the reported structural roles of the two base pairs in widening the major groove (Bartel et al. 1991). The side chain of Asn40 was found to interact with nucleotides G47, A73, G46 and U45 (Battiste et al., 1996), variants active with Rev-R35G:N40V had G47, A73, G46 and C74 conserved but not U45 (figure 13). These findings suggest slightly different interactions between Rev-R35G:N40V and these residues.

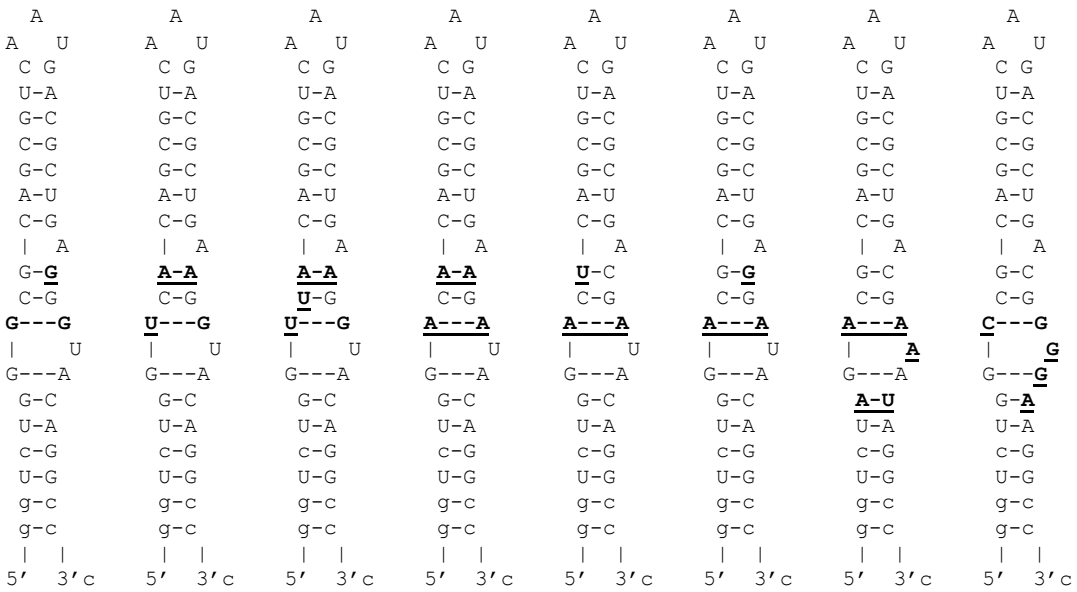
To check for covariation between the RRE IIB and Rev-R35G:N40V at sites of base specific interactions, RRE IB variants obtained were screened for activity with Rev-R35G and Rev-N40V ARMs separately. Unexpectedly, all RRE IIB variants active with Rev- R35G:N40V were also active with Rev-R35G, but with lower activity. Most RRE IIB variants showed no activity with Rev-N40V ARM. Surprisingly, three variants with mutations at the apical loop showed very low activity towards Rev-N40V ARM (table 3).

C. The Structure-Specific Recognition Element G48•G71 Base Pair Can Be Compensated by a Nearby A50•A69 Base Pair with Wild-Type Rev

The internal loop, important for wild-type Rev binding, is conserved in most RRE variants active with both Rev17and Rev-R35:N40V, with A•A or a G•G at position 48-71 and the G47•A73 pair strictly conserved in all active clones. This is consistent with the reported structural roles of the homopurine base pairs in widening the major groove (Bartel et al., 1991). Exceptions are observed in two variants active with Rev17 and with a U-G at position 48-71(figure 14). According to the literature, a U48-G71 is not tolerated (Giver et al., 1993 (A)). The U48-G71 mutation observed however was accompanied by another

mutation, A50•A69. This suggests that the homo-purine base pair at position 50-69 might be compensating for the structural role of G48-G71 in widening the groove to allow binding of Rev at the major groove.

A.



B.

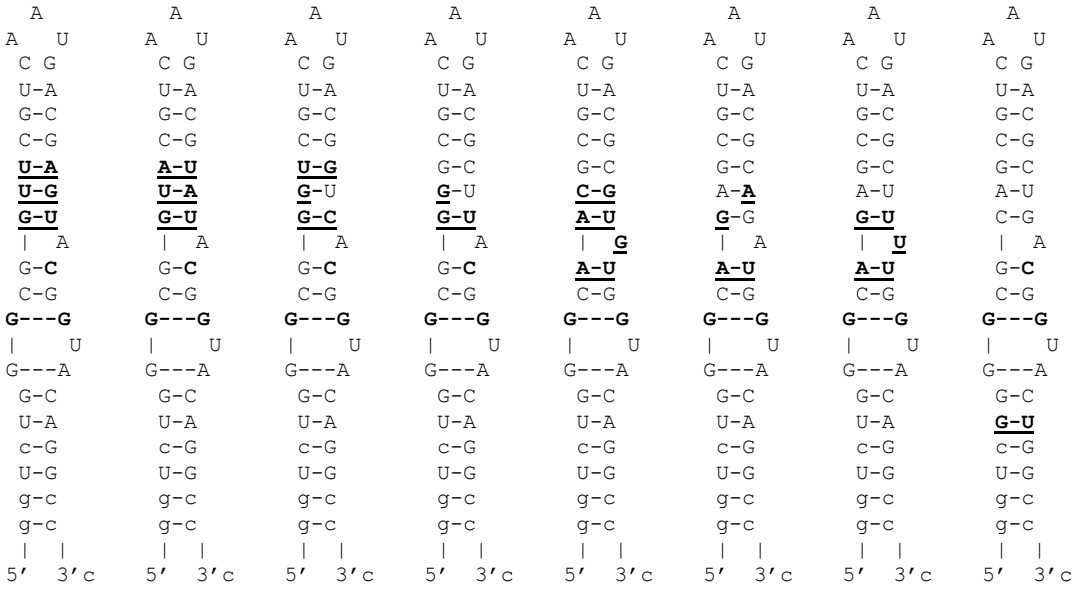


Figure 14. Primary and secondary structure of RRE IIB active clones obtained. A) Rev17-binding RRE IIB active clones. The 2nd and 3rd sequences have a U-G instead of a G•G at position 48-71. B) Rev-R35G:N40V-binding RRE IIB active clones. Nucleotides differing from the wild-type RRE are underlined and shown in bold.

D. As Few as One or Two Substitutions Are Sufficient to Alter the Specificity of RRE IIB toward Wild-Type Rev or Rev-R35G:N40V

In addition to RRE IIB active clones functional with both Rev17 and Rev-R35G:N40V, active clones specific to either RNA-binding domain were observed. All Rev-R35G:N40V-specific RRE IIB variants had mutations between the base pair positions 50-69 and 53-65 and one variant with mutations at position 46-74; these positions contain the sites forming specific interactions with wild-type Rev (figure 14). All Rev17-specific RRE IIB variants had mutations at the internal loop. These observations suggest that different regions confer specificity towards different peptides and suggest that different recognition strategies occur at these regions in the two peptides.

Interestingly, as few as one or two substitutions are sufficient to alter the specificity of RRE IIB toward wild-type Rev or Rev-R35G:N40V (figure 15). A single nucleotide change in wild-type RRE IIB induced the switch in specificity from a relaxed binding to one which is specific to Rev17. Three nucleotide substitutions were able to switch specificity towards Rev-R35G:N40V. These results demonstrate how RNA can evolve alternative binding strategies in discrete steps without loss of function supporting the neutral theory of molecular evolution.

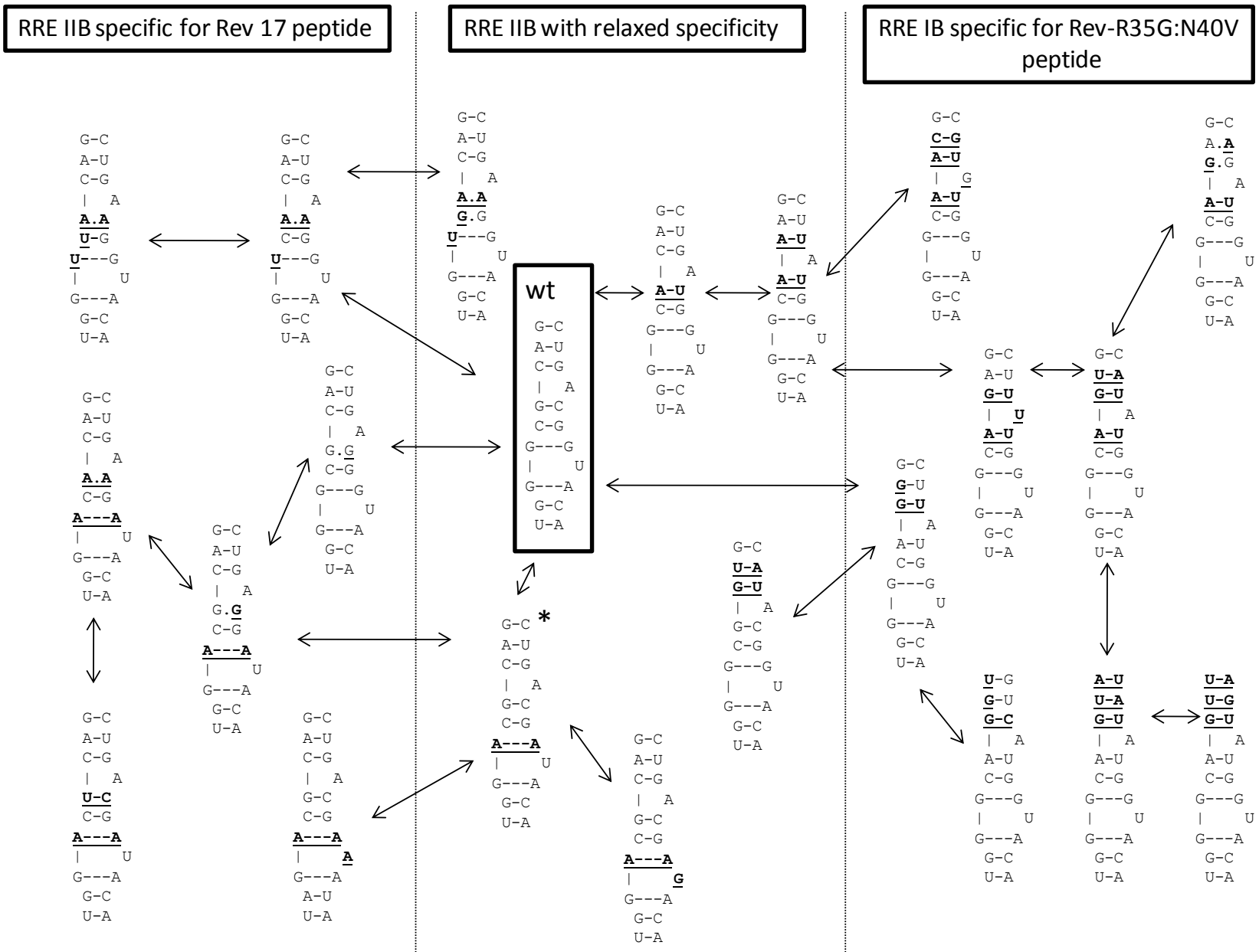

RRE IIB specific for Rev 17 peptide
RRE IIB with relaxed specificity
RRE IB specific for Rev-R35G:N40V peptide
wt
*

Figure 15. Mutational network, as predicted by neutral theories, connecting the RRE IIB active clones with relaxed specificity and those specific to one peptide. Arrows show single and multiple mutational steps, respectively. Wild-type RRE IIB sequence is boxed. The RRE IIB sequence labeled with a star was not obtained in this work, but was made as a construct in a previous study (Possik et al., 2013) and was shown to bind both Rev17 and Rev-R35G:N40V. A single nucleotide or base-pair change in the RRE IIB induced the switch in specificity from a relaxed binding towards one which is specific to Rev17. Multiple substitutions are required to switch specific towards Rev-R35G:N40V.

Chapter IV

DISCUSSION

Regional mutagenesis was used to examine the mutagenic potential of RRE IIB and the mutability of Rev-RRE IIB and Rev-R35G:N40V-RRE IIB interactions. The approach identified functional sequence substitutions that would have been difficult to discover by site-directed mutagenesis alone. This approach allows the detection of how RRE IIB can mutate while maintaining binding to Rev and how specificity can arise by genetic drift.

A. Rev ARM-RRE IIB Mutagenesis Supports Neutral Theories of Molecular Evolution

The potential to evolve a new mode of RNA-ARM interaction in an incremental manner without generating loss-of-function intermediates was suggested in several studies (Cocozaki et al., 2008a; Cocozaki et al., 2008b; Iwazaki et al., 2005; Possik et al., 2013; Smith et al., 1998; Smith et al., 2000; Sugaya et al., 2008). Based on neutral theory, single, incremental mutations are sufficient to change recognition strategy and switch specificity, without any loss-of-function intermediates. In this project, a single nucleotide change in a bifunctional RRE IIB mutant is sufficient to switch its interaction with both Rev17 and Rev-R35G:N40V to a specific one. For example, a Rev17-specific RRE IIB active clone can evolve from a bifunctional wild-type intermediate by a single nucleotide change C69G

(figure 15). Similarly, starting with a bifunctional RRE IIB-G48A:G71A, single substitutions lead to a Rev17-specific RRE IIB. RRE IIB variants specific for Rev-R35G:N40V were also found with few nucleotide substitutions in the proximal-apical stem of wild-type RRE IIB that shift specificity towards the double mutant. Many variants were obtained, enough to trace partial incremental neutral paths between wild-type Rev and Rev-R35G:N40V as predicted by neutral theories demonstrating the possibility to evolve peptide-binding strategy of RNA in discrete steps without intermediate loss of function (figure 15).

All lentiviruses encode a regulatory protein Rev that is essential for nuclear export of unspliced and incompletely spliced mRNAs encoding the structural viral proteins (Tang et al., 1999). The Rev protein binds to the Rev responsive element (RRE) to ensure viral replication. The comparison of RRE sequences and structures reported for different lentivirus sub-groups revealed that the RREs in each sub-group possess some common features, but also possess structural features that are unique to each sub-group (Lesnik et al., 2002). HIV-2 shares many structural and functional features with the HIV-1; the RRE of HIV-2 is similar to that of HIV-1 in structure and location in the env gene although the high-affinity binding site is not located in the same positions as in HIV-1(Le et al., 1990). The nucleotide sequences, secondary structure and location of the RRE regions in different simian immunodeficiency viruses are very similar to those in HIV-2 genomes (Kuiken et al., 1999). Similarities are also seen in viruses branching away from HIV-1. The HIV-1 and Visna Virus (VV) Revs function by similar mechanisms (Lesnik et al., 1999). The bovine immunodeficiency virus (BIV) and Jembrana disease virus (JDV) also encode Rev proteins that have function similar to those in HIV-1 and VV genomes; however, the positions of

Rev high affinity binding site of Rev is not identified (Chadwick et al., 1995). Because all lentiviruses presumably share a common ancestor (Foley BT, 2000), neutral networks of related lentiviral Rev-RREs presumably intersect at common ancestors.

B. Rev-R35G:N40V Recognizes RRE IIB Using a Distinct Strategy

Possik et al. (2013) found that while the Rev residue Asn40 is strictly conserved and residue Arg35 barely tolerates mutations upon binding to wild-type RRE IIB, a Rev-R35G:N40V double mutant is functional.

The nucleotides important for Rev-R35G:N40V recognition are localized in the vicinity of the internal loop (figure 16). Mutations at the proximal-apical stem specific for Rev-R35G:N40V support a different binding strategy for the double mutant. Sequence variation at the proximal-apical stem of RRE IIB is more limited for Rev17 binding. This observation can be explained by the hydrogen bonds between Arg35 of Rev17 and C51, G67 and U65 bases at the proximal apical stem, those interactions are absent in Rev-R35G:N40V. A complex sequence covariation that maintains Watson-Crick base-pairing at the upper stem is observed as well in both peptides suggesting a structural role for those base-pairs. The formation of the G48•G71 and G47•A73 base pairs was suggested to open the major groove and allow binding of Rev peptide (Battiste et al., 1996) and Asn40 forms hydrogen bonds with G47•A73 base pair. RRE IIB variants active with Rev-R35G:N40V have the two purine-purine base pairs conserved, confirming a structural role of these base pairs in widening the major groove and allowing peptide binding. The differences in functional RRE IIB active clones with the two peptides at the internal loop suggest that the

molecular mechanisms of major-groove widening and the specific contacts to the non-canonical elements differ in the two complexes.

Surprisingly, all RRE IIB variants that are active with Rev-R35G:N40V were also active with Rev-R35G but not with Rev-N40V. That the N40V mutation does not allow activity with RRE IIB suggests disruption of important contacts. A model can be proposed in which the R35G mutation removes hydrogen bonds of R35 therefore giving the peptide some conformational flexibility. The R35G nucleotide substitution is likely to shift the location and orientation of hydrogen bond donors and acceptors in the RRE IIB upper stem which has significance for N40V mutation.

Interestingly, Rev17 specific-RRE IIB active clones have their nucleotide substitutions at the internal loop whereas Rev-R35G:N40V-specific-RRE IIB sequences have their nucleotide substitutions at the proximal-apical stem and distal-apical stem (figure 14). These observations show that different regions confer specificity towards different peptides and suggest that different recognition strategies in the two peptides occur at these regions.

C. Mutagenesis Largely Supports Known Structural Motifs of RRE IIB

In the Rev-RRE complex, two purine-purine base pairs widen the major groove, a process facilitated by a distortion of the RNA backbone that results from the formation of the G48•G71 and G47•A73 base pairs and an opening of the major groove by ~ 5 Å (Battiste et al., 1996). Consistent with the importance of the two non-canonical base pairs, initial screens of libraries containing these two base pairs yielded low proportion of active

clones (table 2B). Consistent with the literature (Bartel el al., 1991; Peterson et al., 1994), A48•A71 is tolerated for both Rev17 and Rev-R35G:N40V peptide binding and is filling the role of the non-canonical base-pairs G48•G71 in widening the groove. The G47•A73 base pair was conserved in almost all variants obtained active with Rev17 confirming that Asn40 interacts with both bases of the base pair through formation of hydrogen bonds. According to Giver et al. (1993), the U48-G71 mutant is not expected to work; interestingly, U48-G71 accompanied with an A•A base pair at position 50-69 allows binding of Rev 17. Giver et al. (1993) suggested that interactions between Rev and the context-independent structure-specific recognition element at 48:71 appear to be modulated by base substitutions at other residues. Giver et al. (1993) have shown that although positions 50, 68 and 69 are not critical for sequence-specific recognition, they covary with the identity of the non-Watson-Crick pairings at positions 48 and 71. These observations are consistent with the results of this project, although the active clones obtained and the patterns of covariation in the two studies are different. Ye et al. (1996) show that the adjacent A48•A71 and G47•A73 mismatches contribute to the widening of the major groove at the binding site since the 13.4 Å separation between the C1'-C1' atoms across both mismatch pairs is larger by 3 Å relative to the corresponding distance in Watson-Crick base pairs. The findings therefore predict that A50•A69 widens the major groove and compensates for the structural role of G48•G71 with wild-type Rev.

At the apical loop, all possible mutations are allowed while maintaining Watson-Crick base-pairing with both Rev17 and Rev-R35G:N40V. Unpublished results that assayed the activity of canonically mutated loops of RRE IIB with Rev17 show that alterations in the region apical to the Rev binding site sequence modulate the affinity and

specificity of peptide-RRE IIB interaction. Rev17 does not tolerate tetraloops close to the binding site suggesting that the apical loop stabilizes the binding of Rev17 to RRE IIB. Consistent with these observations; base-pairs at the apical loop in RRE IIB active clones obtained may be required for a stable interaction.

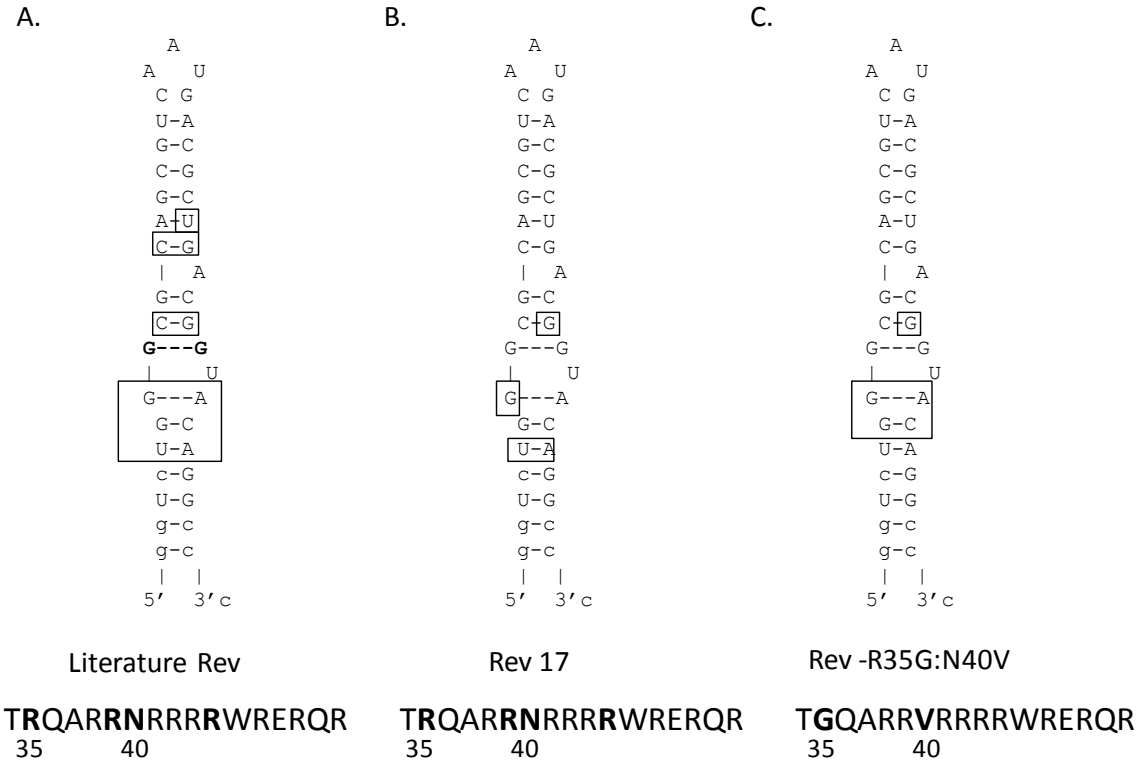


Figure 16. Primary and secondary structural elements important for Rev17 and Rev-R35G:N40V ARMs. A) Boxed nucleotides indicate those that were conserved and predicted to be important for Rev binding as summarized in the literature (Bartel et al., 1991). B) and C) Boxed nucleotides indicate nucleotides that were conserved upon regional mutagenesis of RRE IIB screened with Rev 17 and Rev-R35G:N40V respectively. The G-G base pair shown in bold can be replaced by an A-A base-pair in the three peptides.

D. Clinical Sequences Vary Little from Possible Variations Seen In Vitro

The HIV RRE IIB clinical sequences vary little compared to in vitro screens. All clinical sequences have mutations at sites not known to be important for binding with the most mutations occurring at position 51 and at the apical loop (figure 17). The RRE stem-loop IIB sequence coincides with a large predicted RNA secondary structure present in the transmembrane protein (TMP) coding region just downstream from the boundary of the outer membrane glycoprotein (OMP) gp120 and glycoprotein (TMP) gp41 (Malim et al., 1989). This region was reported to have greater sequence conservation than in the aligned total envelope region suggesting that the role of this sequence as the Rev response element is mediated by secondary structure as well as RNA primary structure. A conservation pressure from both the RNA structure and the overlapping env protein reading frame set limitations on the evolution of this sequence and mutations at the apical loop presumably do not interfere with the folding of this secondary structure. Interestingly, similar series of unusual folding regions immediately 3' to the cleavage site of the outer membrane protein(OMP) and transmembrane protein (TMP) were detected in the envelope gene RNA of the human immunodeficiency virus type 2 (HIV-2) and simian immunodeficiency virus (SIV) (Le et al., 1990; Lesnik et al., 2002). This suggests that the conserved primary sequence is necessary for folding of a defined secondary structure which plays a common biological function in HIV-1, HIV-2 and SIV.

Many functional RRE IIB variants obtained had activities higher than that of wild-type which may interfere with the in vivo regulation of HIV gene expression. The higher sequence variability of RRE IIB observed in vitro suggests that the virus has much unexplored mutagenic potential.

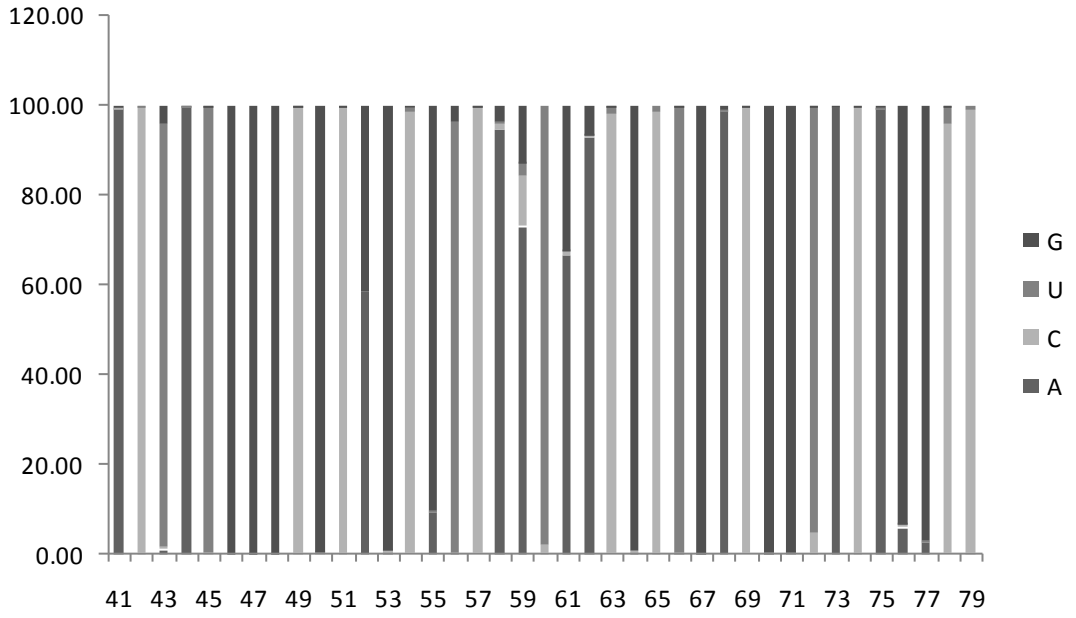


Figure 17. Histogram plotting the percentage conservation of clinical RRE stem-loop IIB sequences. The greatest variations are observed at position 51 and at the apical loop.

E. Adaptive Recognition Facilitates Multiple Partners

Neutral evolution of macromolecules gives access to new phenotypes and can thus play important role in adaptive recognition. A set of different RRE IIB active clones obtained are able to bind specifically to Rev17 and Rev- R35G:N40V (figure 15). Although structural analysis is not available to tell how different the interactions are between the two complexes; a number of RRE-binding peptides have been identified from random peptide libraries suggesting that the flexibility and adaptability of arginine-rich peptides generates multiple strategies for recognition of the same RNA (Harada et al., 1996; Frankel and Young, 1998; Hermann and Patel, 2000; Peled-Zehavi et al., 2003). RSG-1.2 adopts a recognition strategy distinct from that of HIV-1 Rev-RRE (Harada et al., 2008; Zhang et al., 2001). Importantly, a single nucleotide change was found to switch the peptide-binding specificity of RRE from a bifunctional Rev- and RSG-1.2 binding mode to either a Rev-specific or a RSG-1.2-specific mode, demonstrating how RNA can evolve alternative binding strategies in discrete steps without intermediate loss of function (Iwazaki et al., 2005). Based on findings of this work and work on other RRE-binding aptamers, it is tempting to speculate that there exist paths of single mutation steps that permit speciation of binding strategies between wild-type Rev-RRE interaction and non-natural Rev-RREs.

APPENDIX

Table 3 Sequences and activities of selected RRE IIB active clones.

Clone[c]	RRE IIB library of origin	Clone sequence[d]	X-gal[a] (β-galactosidase units)[b]				
			Rev17	Rev-R35G:N40V	Rev-R35G	Rev-N40V	BTat
XIV-220-1	flRRE	GGUCUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGCC	4 (383±100)	4 (40±10)			0
XIV-537-4	RRE FL 11NL	GGUCUGGGCGCAG**UCA**C**GUAUUG**GCUGACGGUACAGGCC	4	3	3	1	0
XIV-537-4	RRE FL 11NL	GGUCUGGGCGCAG**AAGA**AA**GAGUU**CUGACGGUACAGGCC	5	4	3	2	0
XIV-537-59	RRE FL 11NL	GGUCUGGGCGCAG**UAGA**AA**ACCUA**CUGACGGUACAGGCC	5	4	3	1	0
XIV-537-67	RRE FL 11NL	GGUCUGGGCGCAG**G**GU**G**A**CGC**AC**U**CUGACGGUACAGGCC	4	4	3	1	0
XIV-537-69	RRE FL 11NL	GGUCUGGGCGCAG**AAAA**AAUG**C**C**A**CUGACGGUACAGGCC	4	4	NA[e]	NA	0
XIV-537-52	RRE FL 11NL	GGUCUGGGCGCAGC**AAAU**A**GA**A**U**GCUGACGGUACAGGCC	4	4	3	1	0
XIV-537-51	RRE FL 11NL	GGUCUGGGCGCAG**AA**U **GG**UGAC**A**CUGACGGUACAGGCC	4	4	3	2	0
XIV-549-29	RRE FL 11NL	GGUCUGGGCGCAG**GUCU**A**UG**GAC**A**CUGACGGUACAGGCC	5	4	3	1	0
XIV-549-49	RRE FL 11NL	GGUCUGGGCGCAG**A**GU**A**A**U**U**U**A**UU**CUGACGGUACAGGCC	4	4	4	1	0
XIV-549-52	RRE FL 11NL	GGUCUGGGCGCAG**AU**U AA**G**G**GAU**CUGACGGUACAGGCC	4	4	3	1	0
XIV-549-33	RRE FL 11NL	GGUCUGGGCGCAG**UACGU**A**C**G**GGA**CUGACGGUACAGGCC	4	4	3	2	0
XIV-549-37	RRE FL 11NL	GGUCUGGGCGCAG**UUAAUC**U**UUAA**CUGACGGUACAGGCC	4	4	3	2	0
XIV-549-23	RRE FL 11NL	GGUCUGGGCGCAG**U**G**GA**A**U**UG**C**CGCUGACGGUACAGGCC	4	4	3	2	0
XIV-537-19[f]	52N54-64N66	GGUCUGGGCGCA**AG**GUCAAUGAC**UU**UGACGGUACAGGCC	4	4	2	1	0
XIV-537-21	52N54-64N66	GGUCUGGGCGCA**UU**GUCAAUGACG**G**UGACGGUACAGGCC	4	4	3	2	0
XIV-350-1	51N53-65N67	GGUCUGGGCG**AUU**CGUCAAUGACG**AA**GACGGUACAGGCC	4	4	3	0	0
XIV-503-21	51N53-65N67	GGUCUGGGCG**AUU**CGUCAAUGACG**GA**GACGGUACAGGCC	4	3	2	0	0
XIV-350-4[g]	51N53-65N67	GGUCUGGGCG**AUA**CGUCAAUGACG**UA**GACGGUACAGGCC	4	4	3	1	0
XIV-526-17	51N53-65N67	GGUCUGGGCG**GUU**CGUCAAUGACG**AGU**ACGGUACAGGCC	0	4	3	0	0
XIV-526-20	51N53-65N67	GGUCUGGGCG**GUA**CGUCAAUGACG**UAU**ACGGUACAGGCC	1	3	3	0	0
XIV-350-9	51N53-65N67	GGUCUGGGCG**GGU**CGUCAAUGACG**G**U**C**ACGGUACAGGCC	0	4	3	0	0
XIV-503-63	51N53-65N67	GGUCUGGGCG**GG**GCGUCAAUGACGCU**U**ACGGUACAGGCC	1	4	2	0	0
XIV-350-10	51N53-65N67	GGUCUGGGCG**GUA**CGUCAAUGACG**UAA**ACGGUACAGGCC	3	4	2	1	0
XIV-350-8	51N53-65N67	GGUCUGGGCG**GU**GCGUCAAUGACGC**AU**ACGGUACAGGCC	4	4	3	0	0
XIV-568-119	50N52-66N69	GGUCUGGG**UU**CAGCGUCAAUGACGCU**AUU**GGUACAGGCC	5	4	1	0	0

Table 3 Sequences and activities of selected RRE IIB active clones.

Clone	Library	Clone sequence	X-gal (β-galactosidase units) Rev17	Rev-R35G:N40V	Rev-R35G	Rev-N40V	BTat
XIV-568-98	50N52-66N69	GGUCUGGGC**AAC**GCGUCAAUGACGC**GUGU**GGUACAGGCC	0	4	2	0	0
XIV-568-97	50N52-66N69	GGUCUGGGC**AG**UGCGUCAAUGACGC**A**GA**U**GGUACAGGCC	4	3	3	1	0
XIV-549-7	49N51-67N70	GGUCUGGGC**AU**AGCGUCAAUGACGCU**A**AA**A**GGUACAGGCC	5	3	3	1	0
XIV-568-93	48N50-69N71	GGUCUGGGCGCAGCGUCAAUGACGCUGA**G**GGUACAGGCC	4	0	0	0	0
XIV-503-38[h]	48N50-69N71	GGUCUGGGC**A**CAGCGUCAAUGACGCUGA**U**GGUACAGGCC	4	4	1	0	0
XIV-503-16	48N50-69N71	GGUCUGG**UGA**CAGCGUCAAUGACGCUGA**U**GGUACAGGCC	4	4	2	0	0
XIV-503-15	48N50-69N71	GGUCUGG**UUA**CAGCGUCAAUGACGCUGA**A**GGUACAGGCC	3	0	NA	NA	0
XIV-568-82	48N50-69N71	GGUCUGG**U**C**A**CAGCGUCAAUGACGCUGA**A**GGUACAGGCC	5	0	3	0	0
XIV-568-65	48N50-69N71	GGUCUGG**A**C**A**CAGCGUCAAUGACGCUGA**A**G**A**UACAGGCC	4	1	1	0	0
XIV-568-67	48N50-69N71	GGUCUGG**A**C**U**CAGCGUCAAUGACGCUGACG**A**UACAGGCC	5	0	0	0	0
XIV-568-79	48N50-69N71	GGUCUGG**A**CGCAGCGUCAAUGACGCUGA**G**G**A**UACAGGCC	4	1	1	0	0
XIV-321-1[i]	46N48-71N74	GGUCUGG**A**CGCAGCGUCAAUGACGCUGACG**AG**ACAGGCC	4	3	1	0	0
XIV-321-5	46N48-71N74	GGUCUGG**C**CGCAGCGUCAAUGACGCUGACGG**GGA**AGGCC	3	0	0	0	0
XIV-549-54[j]	45N47-73N75	GGUC**G**GGGCGCAGCGUCAAUGACGCUGACGGUAC**U**GGCC	0 ()	5 ()	4	0	0
XIV-537-1k	43N46-74N77	GG**AU**GGGGCGCAGCGUCAAUGACGCUGACGGUACA**AA**CC	5	3	NA	NA	0

[a] X-gal colony color assays were performed as described (Harada and Frankel, 1991). Numbers represent +'s used to score blue color intensity.

[b]β-galactosidase solution assays were performed as described (Miller, 1992). RRE IIB libraries were transformed into expressor cells. Four replicates of each clone were grown at 30ºC in tryptone medium supplemented with IPTG and assayed for β-galactosidase activity.

[c] Laboratory stock number of RRE active clones.

[d]The nucleotide sequence of RRE IIB active clones with the nucleotides differing from the wild-type RRE shown in bold.

[e]Not availablef Laboratory stock number of minipreps with the same sequence: XIV-537-22

g Laboratory stock number of minipreps with the same sequence: XIV-350-7

[h] Laboratory stock number of minipreps with the same sequence: XIV-503-40, XIV-526-15, XIV-503-21, XIV-503-53, XIV-503-78, XIV-503-81, XIV-503-84, XIV-330-1, XIV-330-4, XIV-330-5

[i] Laboratory stock number of minipreps with the same sequence: XI-325-3, XIV-330-2, XIV-330-3

[j] Laboratory stock number of minipreps with the same sequence: XIV-549-60, XIV-549-61

[k] Laboratory stock number of minipreps with the same sequence: XIV-537-2, XIV-537-5, XIV-537-7, XIV-537-11

Table 4 Synthetic DNA oligonucleotide sequences.

Insert reference name	Insert DNA sequence	Cloning restriction	Cloning plasmid
RRE43N46-74N77	5'CCAGGATCCCTGCTTTGAATGCTGGNNNNTACCGTCAGCGTCATTGACGCTGCGCCNNNNCCTTAATTTTTAAGAGCGTCGACTGCA-3	PstI-BamHI	pACnutTAT13
RRE45N47-73N75	5'CCAGGATCCCTGCTTTGAATGCTGGCCNNNACCGTCAGCGTCATTGACGCTGCGCNNNGACCTTAATTTTTAAGAGCGTCGACTGCA-3	PstI-BamHI	pACnutTAT13
RRE46N48-71N74	5'CCAGGATCCCTGCTTTGAATGCTGGCCTNNNNCGTCAGCGTCATTGACGCTGCGNNNAGACCTTAATTTTTAAGAGCGTCGACTGCA-3	PstI-BamHI	pACnutTAT13
RRE47N49-70N73	5'CCAGGATCCCTGCTTTGAATGCTGGCCTGNNNNGTCAGCGTCATTGACGCTGCNNNCAGACCTTAATTTTTAAGAGCGTCGACTGCA-3	PstI-BamHI	pACnutTAT13
RRE48N50-69N71	5'CCAGGATCCCTGCTTTGAATGCTGGCCTGTANNNTCAGCGTCATTGACGCTGNNNCCAGACCTTAATTTTTAAGAGCGTCGACTGCA-3	PstI-BamHI	pACnutTAT13
RRE49N51-67N70	5'CCAGGATCCCTGCTTTGAATGCTGGCCTGTACNNNNAGCGTCATTGACGCTNNNCCCAGACCTTAATTTTTAAGAGCGTCGACTGCA-3	PstI-BamHI	pACnutTAT13
RRE50N52-66N69	5'CCAGGATCCCTGCTTTGAATGCTGGCCTGTACCNNNNGCGTCATTGACGCNNNGCCCAGACCTTAATTTTTAAGAGCGTCGACTGCA-3	PstI-BamHI	pACnutTAT13
RRE51N53-65N67	5'CCAGGATCCCTGCTTTGAATGCTGGCCTGTACCGTNNNCGTCATTGACGNNNCGCCCAGACCTTAATTTTTAAGAGCGTCGACTGCA-3	PstI-BamHI	pACnutTAT13
RRE52N54-64N66	5'CCAGGATCCCTGCTTTGAATGCTGGCCTGTACCGTCNNNGTCATTGACNNNGCGCCCAGACCTTAATTTTTAAGAGCGTCGACTGCA-3	PstI-BamHI	pACnutTAT13
RREFL11NL	5'CCAGGATCCCTGCTTTGAATGCTGGCCTGTACCGTCAGNNNNNNNNNNNCTGCGCCCAGACCTTAATTTTTAAGAGCGTCGACTGCA-3	PstI-BamHI	pACnutTAT13

Table 4 Synthetic DNA oligonucleotide sequences.

Insert reference name	Insert DNA sequence	Cloning restriction	Cloning plasmid
46FLRREF	5'CCAGGATCCCTGCTTTGAATGCTGGTCTGGGCGCAGCGTCAATGACGCTGACGGTACAGGCC TTAATTTTTAAGAGCGTCGACTGCA-3	PstI-BamHI	pACnutTAT13
nmrRRE	5'CCAGGATCCCTGCTTTGAATGCTGGTCTGGGCGCAGCGCAAGCTGACGGTACAGGCC TTAATTTTTAAGAGCGTCGACTGCA-3	PstI-BamHI	pACnutTAT13
BIVTAR	5'CCAGGATCCCTGCTTTGAATGCTGCTCGTGTAGCTCATTAGCTCCGAGCTTAATTTTTAAGAGCGTCGACTGCA-3	PstI-BamHI	pACnutTAT13
Rev17NF	5'-CATGGCAACCCGCCAGGCCCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAATG-3'	NcoI-BsmI	pBR-ptac-N*λ
BIVTATN	5'-CCATGGGTCGTCCTCGTGGTACCCGCGGTAAAGGTCGCCGTATTCGCCGTGGTGGCGGGAATG-3'	NcoI-BsmI	pBR-ptac-N*λ

PstI-BamHI fragment cloned into pACnutTAT13 replacing the lambda left nut site such that BoxB is replaced with RRE IIB libraries, full-length RRE IIB and BIV TAR. Rev17N cloned into pBR-ptac-N*λ as NcoI-BsmI fragment replacing the aminoacid terminus of lambda N. The BIVTat N fusion plasmid was obtained from Kazuo Harada (Harada et al., 1996) (Tokyo Gakugei University). Sequences that match flRRE, RRE IIB libraries, RAI and BIVTAR are underlined.

REFERENCES

Bartel DP, Zapp ML, Green MR, Szostak JW. HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell*. 1991 Nov 1; 67(3):529–536.

Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, Frankel AD, Williamson JR. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science*. 1996 Sep 13; 273(5281):1547-51.

Bayer TS, Booth LN, Knudsen SM, Ellington AD. Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA*. 2005 Dec; 11(12):1848-57.

Cai Z, Gorin A, Frederick R, Ye X, Hu W, Majumdar A, Kettani A, Patel DJ. Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex. *Nat Struct Biol*. 1998 Mar; 5(3):203-12.

Chadwick BJ, Coelen RJ, Wilcox GE, Sammels LM, Kertayadnya G. Nucleotide sequence analysis of Jembrana disease virus: A bovine lentirus associated with an acute disease syndrome. *J Gen Virol*. 1995; 76: 1637-1650.

Chen Y, Varani G. Protein families and RNA recognition. *FEBS J.* 2005 May; 272(9):2088-97.

Cocozaki AI, Ghattas IR, Smith CA, The RNA-Binding Domain of Bacteriophage P22 N Protein Is Highly Mutable and a Single Mutation Relaxes Specificity Toward λ. *J Bacteriol.* 2008a Sep 26.

Cocozaki AI, Ghattas IR, Smith CA, Bacteriophage P22 antitermination boxB sequence requirements are complex and overlap with those of lambda. *J Bacteriol.* 2008b Jun;190(12):4263-71.

Dillon PJ, Nelbock P, Perkins A, Rosen CA. Function of the human immunodeficiency virus types 1 and 2 Rev proteins is dependent on their ability to interact with a structured region present in env gene mRNA. *J Virol*. 1990 Sep; 64(9):4428-37.

Draper DE. Themes in RNA-protein recognition. *J Mol Biol.* 1999 Oct 22; 293(2):255-70.

Engelke DR, Krikos A, Bruck ME, Ginsburgt D. Purification of Thermus aquaticus DNA Polymerase Expressed in Escherichia coli. *Analytical Biochemistry*. 1990 July; 191: 396-400.

Fernandes JD, Jayarman B, Frankel AD. The HIV-1 Rev Response Element. 2012. *RNA Biology* January 9; 1: 1-6.

Foley BT. An overview of the molecular phylogeny of lentiviruses. *Los Alamos National Laboratory, Los Alamos.* 2000.

Fontana W, Schuster P. Continuity in evolution: on the nature of transitions. *Science.* 1989a May; 280(5368):1451-5.

Fontana W, Schuster P. Shaping Space: the Possible and the Attainable in RNA Genotype‑phenotype Mapping**.** *Journal of Theoretical Biology.* 1989b Oct; 194 (4): 491‑515.

Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem.* 1998; 67:1-25.

Gesteland R, Cech T, Atkins J. The RNA World. Third edition.1993.© by Cold Spring Harbor Laboratory Press.

Giver L, Bartel D, Zapp M, Pawul A. Green M, Ellington AD. Selective optimization of the Rev-binding element of HIV-1. *Nucleic Acids Res.* 1993a Nov 25; 21(23):5509-16.

Giver L, Bartel DP, Zapp ML, Green MR, Ellington AD. Selection and design of high-affinity RNA ligands for HIV-1 Rev. *Gene*. 1993b Aug 137: 19-24.

Gosser Y, Hermann T, Majumdar A, Hu W, Frederick R, Jiang F, Xu W, Patel DJ. Peptide-triggered conformational switch in HIV-1 RRE RNA complexes. *Nat Struct Biol.* 2001 Feb; 8(2):146-50.

Harada K, Frankel AD. Screening RNA-binding libraries using a bacterial transcription antitermination assay. From *Methods in Molecular Biology*. 1991 118: RNA-Protein Interaction Protocols.

Harada K, Martin SS, Frankel AD. Selection of RNA-binding peptides in vivo. *Nature*. 1996 Mar 14; 380(6570):175-9.

Harada K, Martin SS, Tan R, Frankel AD. Molding a peptide into an RNA site by in vivo peptide evolution. *Proc Natl Acad Sci U S A.* 1997 Oct 28; 94(22):11887-92.

Heaphy S, Dingwall C, Ernberg I, Gait M J, Green S M, Karn J, Lowe A D, Singh M, Skinner M A. HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the Rev response element. *Cell*. 1990; 60:685–693.

Hermann T, Patel DJ. Adaptive recognition by nucleic acid aptamers. *Science*. 2000 Feb 4; 287(5454):820-5.

Hermann T, Westhof E. Non-Watson-Crick base pairs in RNA-protein recognition. *Chem Biol*. 1999 Dec; 6(12):R335-43.

Holland S.M., Ahmad N., Maitra R.K., Wingfield P., Venkatesan S. Human immunodeficiency virus rev protein recognizes a target sequence in rev-responsive element RNA within the context of RNA secondary structure. *J. Virol*. 1990; 64:5966–5975.

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

Jain C, Belasco JG. A structural model for the HIV-1 Rev-RRE complex deduced from altered-specificity rev variants isolated by a rapid genetic strategy. *Cell*. 1996 Oct 4; 87(1):115-25.

Jiang F, Gorin A, Hu W, Majumdar A, Baskekerville S, Xu W, Ellington A, Patel DJ. Anchoring an extended HTLV-1 Rex peptide within an RNA major groove containing junctional base triples. *Structure*.1999; 7: 1461-1472.

Kimura M. Evolutionary rate at the molecular level. *Nature*. 1968 Feb 17; 217(5129):624-6.

Kimura M. The neutral theory of molecular evolution: a review of recent evidence. *Jpn J Genet.* 1991 Aug; 66(4):367-86.

Kjems J, Brown M, Chang DD, Sharp PA. Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proc Natl Acad Sci U S A.* 1991 Feb 1; 88(3):683-7.

Kuiken CL, Foley B, Hahn B, Marx P, McCutchan F, Mellors JW, Mullins J, Wolinsky S, Korber B. Human Retroviruses and AIDS 1999: A compilation of nucleic acid and amino acid sequences. Los Alamos, NM: Theoretical Biology and Biophysics Group, Los Alamos Natl Lab; 1999. P. iv.

Lazinski D, Grzadzielska E, Das A. Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. *Cell.* 1989 Oct 6; 59(1):207-18.

Le SY, Malim MH, Cullen BR Maizel JV. A highly conserved RNA folding region coincident with the Rev response element of primate immunodeficiency viruses. *Nucleic Acids Res*. 1990; 18: 1613- 1623.

Legault P, Li J, Mogridge J, Kay LE, Greenblatt J. NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell.* 1998 Apr 17; 93(2):289-99.

Lesnik EA, Sampath R, Ecker DJ. Rev response elements (RRE) in lentiviruses: an RNA Motif algorithm-based strategy for RRE prediction. *Med Res Rev*. 2002 Nov; 22(6):617-36.

Leulliot N, Varani G. Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry*. 2001 40(27):7947-56.

Malim MH, Hauber J, Le S-Y, Maizel JV, Cullen BR. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature.* 1989; 338: 254-257.

Malim MH, Tiley LS, McCarn DF, Rusche JR, Hauber J, Cullen BR. HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell*. 1990 Feb 23; 60(4):675–683.

Miller JH. A short course in bacterial genetics: a laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 1992.

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. Epub 2002 Dec 2.

Patel DJ. Adaptive recognition in RNA complexes with peptides and protein modules. *Curr Opin Struct Biol.* 1999 Feb; 9(1):74-87.

Peled-Zehavi H, Horiya S, Das C, Harada K, Frankel AD. Selection of RRE RNA binding peptides using a kanamycin antitermination assay. *RNA*. 2003 Feb; 9(2):252-61.

Possik E, Bou Sleiman M, Ghattas I, Smith C. 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*. 2013 Jun; 26(6):286-96.

Puglisi JD, Chen L, Blanchard S, Frankel AD. Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science*. 1995 Nov 17; 270(5239):1200-3.

Schultes EA, Bartel DP. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science*. 2000; 289 (5478): 448-52.

Smith CA, Calabro V, Frankel AD. An RNA-binding chameleon. *Mol Cell*. 2000 Nov;6(5):1067-76.

Smith CA, Crotty S, Harada Y, Frankel AD. Altering the context of an RNA bulge switches the binding specificities of two viral Tat proteins. *Biochemistry*. 1998 Jul 28;37(30):10808-14.

Sugaya M, Nishimura F, Katoh A, Harada K. Tailoring the peptide-binding specificity of an RNA by combinations of specificity-altering mutations. *Nucleosides, Nucleotides and Nucleic Acids.* 2008 27: 534-545.

Tan R, Chen L, Buettner JA, Hudson D, Frankel AD. RNA recognition by an isolated alpha helix. *Cell*. 1993 Jun 4; 73(5):1031-40.

Tan R, Frankel AD. Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. *Biochemistry*. 1994 Dec 6; 33(48):14579-85.

Tang H, Kuhen KL, Wong-Staal F. Lentivirus replication and regulation. *Annu Rev Genet.* 1999; 33:133-70.

Tiley LS, Malim MH, Cullen BR. Conserved functional organization of the human immunodeficiency virus type 1 and visna virus Rev proteins. *J Virol.* 1991 Jul; 65(7):3877-81.

Varani G. RNA-protein intermolecular recognition. *Acc. Chem Res*.1997 (30):189-95.

Weiss MA, Narayana N. RNA recognition by arginine-rich peptide motifs. *Biopolymers*. 1998; 48(2-3):167-80.

Williamson JR. Induced fit in RNA-protein recognition. *Nat Struct Biol*. 2000 Oct; 7(10):834-7.

Wilke CO, Adami C. Evolution of mutational robustness. *Mutation Research* 2003; 522: 3-11.

Ye X, Gorin A, Ellington AD, Patel DJ. Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex. *Nat Struct Biol.* 1996 Dec; 3(12):1026-33

Ye X, Gorin A, Frederick R, Hu W, Majumdar A, Xu W, McLendon G, Ellington A, Patel DJ. RNA architecture dictates the conformations of a bound peptide. *Chem Biol.* 1999 Sep; 6(9):657-69.

Zhang Q, Harada K, Cho S, Frankel AD, Wemmer DE. Structural characterization of the complex of the Rev response element RNA with a selected peptide. Chem. Biol.2001; 8: 511-520.