Journal of Molecular Recognition
Online Journal:
=Free Color
=More Color
=Faster Publication

Diverse Mutants of HIV RRE IIB Recognize Wild-Type Rev ARM or Rev ARM R35G-N40V

Journal:	*Journal of Molecular Recognition*
Manuscript ID:	JMR-15-0004.R2
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Abdallah, Emane; American University of Beirut, Biology Smith, Colin; American University of Beirut, Biology
Keywords:	HIV, Rev-Response Element, Protein-RNA recognition, Arginine-rich motif, Neutral evolution

Diverse Mutants of HIV RRE IIB Recognize Wild-Type Rev ARM or Rev ARM R35G-N40V

Emane Y. Abdallah[a] and Colin A. Smith[a]*
Department of Biology
American University of Beirut
Beirut, Lebanon

* Correspondence to: Colin A. Smith, Department of Biology, American University of Beirut, PO Box 11-0236, Riad El Solh 1107 2020, Beirut, Lebanon.
Email: colin.smith@aub.edu.lb

[a] E. Y. Abdallah, C. A. Smith
Department of Biology, American University of Beirut, Beirut, Lebanon.

Short title: **Diverse Mutants of RRE Recognize Rev ARM or Rev R35G-N40V**

ABSTRACT

The binding of human immunodeficiency virus Rev protein via its arginine-rich motif (ARM) to an internal loop in the Rev Response Element region IIB (RRE IIB) is necessary for viral replication. Many variant RNAs and ARMs that bind Rev and RRE IIB have been found. Despite the essential role of Rev asparagine 40 in recognition, the Rev ARM double mutant R35G-N40V functions well in a Rev-RRE IIB reporter assay, indicating R35G-N40V uses a distinct recognition strategy. To examine how RRE IIB may evolve specificity to wild-type Rev ARM and R35G-N40V, ten RRE IIB libraries, each completely randomized in overlapping regions, were screened with wild-type Rev ARM and R35G-N40V using a reporter system based on bacteriophage λ N-*nut* antitermination. Consistent with previous studies, a core element of RRE IIB did not vary and substitutions occurred at conserved residues only in the presence of other substitutions. Notably, the groove-widening, non-canonical base pair G48:G71 was mutable to U48:G71 without strong loss of binding to wild-type Rev ARM, suggesting U48:G71 performs the same role by adopting the nearly isosteric, reverse-wobble base pair. Originating from RRE IIB, as few as one or two substitutions are sufficient to confer specificity to wild-type Rev or Rev R35G-N40. The diversity of RRE IIB mutants that maintain binding to wild-type Rev ARM and R35G-N40V support neutral theories of evolution and illustrate paths by which viral RNA-protein interactions can evolve new specificities. Rev-RRE offers an excellent model with which to study fine structure of how specificity evolves.

Keywords: HIV; Rev-Response Element; Protein-RNA recognition; Arginine-rich motif; Neutral evolution.

INTRODUCTION

The recognition of the Rev Response Element (RRE) of the human immunodeficiency virus (HIV) by HIV Rev protein initiates recruitment of host factors that mediate nuclear export of unspliced and singly spliced viral transcripts, including the essential genomic and structural mRNAs (Malim *et al.*, 1989; Zapp and Green, 1989; Heaphy *et al.*, 1990). Many studies have illuminated details of the Rev ARM-RRE IIB interaction (Figure 1), including extensive mutagenesis (Bartel *et al.*, 1991; Possik *et al.*, 2013), biochemical and biophysical experiments (Kjems *et al.*, 1992; Tan *et al.*, 1993; Tan *et al.*, 1994), an NMR structural model (Battiste *et al.*, 1996), and a crystal structure (Jayaraman *et al.*, 2014). The initial, critical event is the binding of the short, α-helical, arginine-rich motif (ARM) of Rev to an internal loop in region IIB of RRE with high specificity. The core recognition element of RRE IIB is an asymmetric, internal loop of purine:purine base pairs that widens the major groove to allow ARM binding deep in the major groove and provides the critical

contact of Rev Asn40 to RRE G47-A73 (NMR model; Battiste *et al.*, 1996) or Rev Asn40 to RRE G47 and G71 (crystal structure; Jayaraman *et al*., 2014).

Selection experiments have found many variant peptides that bind RRE IIB and many RNAs that bind Rev ARM (Bartel *et al.,* 1991; Giver *et al.*, 1993b; Harada *et al.*, 1996; Xu and Ellington 1996; Harada *et al.*, 1996; Peled-Zehavi *et al.*, 2003; Bayer *et al.*, 2005). The ability of the Rev-RRE interaction to accept variant partners combined with available mutagenesis and structural models makes it an attractive system to study how recognition strategies and specificity can evolve.

Some natural RNAs and ARMs recognize both their cognate and a variant partner using different recognition strategies (Harada *et al.*, 1996; Xu and Ellington, 1996; Tan and Frankel, 1998; Smith *et al.*, 2000; Peled-Zehavi *et al.*, 2003; Bayer *et al.*, 2005). Protein sequences able to adopt multiple conformations have been called chameleons (Minor and Kim, 1996; Smith *et al.*, 2000; Bayer *et al.*, 2005; Abroggio and Kuhlman, 2006). Similar to chameleon sequences, intersection sequences are those RNAs able to adopt different structures (Schuster *et al.*, 1994; Schultes and Bartel, 2000). The ability of short ARMs and small, structured RNAs to recognize multiple partners with high specificity may explain their prevalence in viral regulatory interactions. Questions emerge as to how many distinct recognition strategies exist and how easily new strategies and specificities can evolve. An interesting feature of RNA-ARM interactions is induced fit, conformational change occurring upon binding (Frankel and Smith, 1998; Patel, 1999; Williamson, 2000), which may permit the recognition of multiple partners and facilitate evolutionary transitions (Smith *et al*, 1998; Smith *et al*. 2000; Das and Frankel, 2003).

Observing that most genetic change is the result of random mutation, neutral theories of evolution posit that for every phenotype, there are sufficient genotypes that any two phenotypes are linked by paths of incremental mutation (Kimura, 1991; Ohta, 2002). Computational simulations of RNA evolution, in which RNA folds represent phenotype and intersection sequences bridge phenotypes, have lent much evidence in support of applying neutral theories to small RNAs (Schuster *et al.*, 1994; Schuster, 2001; Wagner, 2008). Applied to ARM-RNA recognition, neutral theories predict genetic drift alone is sufficient to lead to new recognition strategies. In addition to genetic drift, other mechanisms contributing to the diversity of ARM-RNA recognition could include recombination, duplication followed by neofunctionalization, transition through low-activity intermediates, or simultaneous mutation at multiple sites, mechanisms that would not rely upon nucleotide-by-nucleotide, incremental paths in which every RNA intermediate functions.

Previously, a screen of single-mutation Rev ARMs identified serendipitously an active double-mutant Rev ARM, R35G-N40V (Figure 1B; Possik *et al.*, 2013). We presume that the simultaneous substitution of the important Arg35 with glycine and the critical Asn40 with a hydrophobic residue creates a possibly structurally related, yet distinct and unknown, recognition strategy. The NMR model and crystal structure are very similar, with some important details resolved in the crystal structure, including base-specific hydrogen bonds between Arg35 and G67. The crystal structure is that of a Rev dimer bound to a hybrid RRE fragment comprising the high-affinity site at RRE IIB and an adjacent, lower affinity site (Daugherty *et al*., 2008; Jayaraman *et al*., 2014). Though differences from the NMR model likely reflects the greater precision of the crystallography data, they could reflect minor adjustments in during the cooperative assembly of the second Rev. In the NMR-based structural model of Rev ARM-RRE IIB, Asn40 forms critical, base-specific hydrogen bonds to G47-A73 (Battiste *et al.*, 1996). In the crystal structure, Asn40 is twisted out of the plane of base pairs and makes base-specific hydrogen bonds to G47 and G71 across adjacent base pairs of the internal loop (Jayaraman *et al*., 2014). These base-specific interactions of Arg35 and Asn40 cannot exist with R35G-N40V. Thus, one might expect bases contacted by Arg35 and Asn40 may not be as important in binding R35G-N40V, though they may have other roles, such as that of G47-A73 and G48-G71 to widen the major groove for ARM access. Other RRE IIB bases have important roles binding Rev. NMR model, U45 and G46 are within hydrogen-bonding distance of critical residue Arg44, though the crystal structure shows Arg44 hydrogen bonding G46. C74 and A75 pair with G46 and U45, respectively. Because G48-G71 can be replaced with the isosteric G48A-G71A, its role has been assumed to widen the major groove, though the crystal structure has Asn40 hydrogen bonding to G71. G70 forms a hydrogen bond to critical residue Arg39. C49 and C51 pair to G70 and G67, respectively, which have direct roles. The bulged nucleotides A68 and U72 position phosphates to interact with Arg42, Arg43, Arg46, and 43.

With an interest in how recognition strategies evolve and assuming R35G-N40V binds using a distinct recognition strategy, we wondered how RRE IIB might be free to mutate and whether RRE IIB mutants might display specificities for wild-type Rev ARM or R35G-N40V. Neutral theories predict that sufficient functional RRE mutants should exist to form incremental paths of mutation that link the two distinct phenotypes of binding wild-type Rev ARM and binding R35G-N40V. Alternatively, it is conceivable that the absence of Arg35 and Asn40 are compensated by new or improved interactions with precisely the same RRE IIB contacts used by wild-type Rev ARM. If so,

no specific RRE IIB mutants would be found, and RRE IIB would not be acting as an intersection sequence between specific phenotypes.

Selections using a phage λ N-*nut* (N-utilization) antitermination reporter system adapted for heterologous interactions have found variant and specific RRE IIBs to Rev and the non-natural partner RSG1.2, and the activities of RNA-peptide interactions in this system agree well with affinities measured *in vitro* (Harada *et al.*, 1996; Harada et al., 1997; Iwazaki *et al.*, 2005). This *in vivo* approach promotes the finding of variants that have the specificity to function in living systems, though lacking the massive screening capabilities of SELEX (Mayer, 2009).

To find mutant RRE IIBs displaying specificity to wild-type Rev or Rev R35G-N40V, we constructed 10 libraries in the phage λ N-*nut* antitermination reporter plasmid in which regions of three base pairs or more were completely randomized. These libraries were screened for active clones with wild-type Rev ARM and R35G-N40V through three or more rounds of selection, and active clones were sequenced and assayed for activity and specificity. We report the activity of RRE mutants isolated from these selections, including those specific for wild-type Rev ARM and those specific for R35G-N40V. We find RRE mutants that function with wild-type Rev, including G48U, that suggest Rev-RRE is more mutable than expected. Originating from RRE IIB, as few as one or two substitutions are sufficient to confer specificity to wild-type Rev or Rev R35G-N40, respectively, supporting the existence of neutral evolutionary paths between the distinct recognition strategies. The few substitutions between RRE IIB and mutants with distinct specificities to wild-type Rev ARM and R35G-N40V support the predictions of neutral theories of evolution and illustrate paths by which viral RNA-protein interactions can evolve new specificities.

MATERIALS AND METHODS

General

Restriction enzymes and T4 DNA ligase were obtained from Roche (Mannheim, Germany). *Thermus aquaticus* DNA polymerase I was prepared in house (Engelke *et al.*, 1990). ONPG (*ortho*-nitrophenol-β-D-galactopyranoside) and IPTG (isopropyl-D-thiogalactoside) were obtained from Acros (Geel, Belgium). X-gal (5-bromo-4-chloro-3-indolyl-D-galactoside) was obtained from Amresco (Solon, OH, USA). Single-stranded oligonucleotides were obtained from Sigma (St. Louis, MO, USA). Fine and laboratory chemicals were obtained from Acros, Amresco, and Sigma.

Bacterial media components were obtained from Oxoid Ltd (Basingstoke, Hampshire, UK). Disposable plasticware was obtained from Sarstedt (Numbrecht, Germany).

Strains and plasmids

Escherichia coli supporting N-*nut* antitermination, N567 (Franklin and Doelling, 1989), was obtained from Naomi Franklin (University of Utah). *One Shot Max Efficiency DH5α-T1* competent cells were obtained from Invitrogen (Waltham, MA, USA). Plasmids expressing the bovine immunodeficiency virus (BIV) Tat N fusion (Harada *et al.*, 1996), the RSG1.2 ARM N fusion (Harada *et al.*, 1997), and the NMR RRE reporter (Harada *et al.*, 1996), were obtained from Kazou Harada (Tokyo Gakugei University, Tokyo, Japan). Plasmids expressing the wild-type Rev ARM N fusion, the R35G-N40V N fusion, the RRE IIB reporter (Figure 2), and the BIV TAR reporter were available in house (Possik *et al.*, 2013). All clones, including received, isolated, and constructed, were verified by sequencing.

Construction of RRE IIB reporter libraries and constructs

Double-stranded DNA of the RRE IIB library oligonucleotides that span the entire λ left *nut* site with RRE IIB in place of boxB (Table 2) were prepared in 100 µl reactions, 80 pmole primer extension reactions using primer boxAF (5'-GTC GAC GCT CTT AAA AAT TAA-3') on RRE IIB library templates based on the RRE IIB sequence complement (5'-CCA GGA TCC CTG CTT TGA ATG CTG GCC TGT ACC GTC AGC GTC ATT GAC GCT GCG CCC AGA CCT TAA TTT TTA AGA GCG TCG ACT GCA-3'). The resulting double-stranded DNA libraries contained a PstI cohesive end, and 40 pmoles of each library were digested with BamHI to allow ligation into a reporter plasmid digested by PstI and BamHI. Specific clones were constructed by annealing complementary oligonucleotides to form inserts with PstI and BamHI cohesive ends. For initial transformations, 1 µl of the ligation reaction representing 20 femtomole of plasmid reporter were transformed into 25 µl *One Shot Max Efficiency DH5α-T1* competent cells three times for each RRE library. The transformations were plated on Luria Bertani media with 12 µg/ml chloramphenicol as antibiotic. The plates were incubated overnight at 34 ºC. For each library, ~ 10, 000 colonies were obtained, scraped off the plates, and stored in 50% glycerol at -70 °C.

Library screening, and X-gal assays on solid media

Transformations for screening or assessing N-*nut* antitermination activity by solid media assays used competent N567 host cells carrying N-fusion supplier plasmids (Figure 2). Approximately 10-

100 ng of reporter plasmid per 100 µl of competent cells were transformed by heat shock and plated on tryptone plates containing 100 µg/ml ampicillin and 12 µg/ml chloramphenicol as antibiotics, 80 µg/ml X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) as the chromogenic substrate of the β-galactosidase reporter protein, and 0.05 mM IPTG (isopropyl β-D-thiogalactoside) to induce the tac promoters expressing N protein and the reporter transcript. The plates were scored after 24 hours at 34 °C and after a second 24-hour incubation at 24 °C. The intensity of the blue colonies was used to score the antitermination activity for selections and the preliminary assessment, by comparison: the wild-type Rev ARM-RRE IIB interaction is scored as 4+, the wild-type Rev ARM-NMR RRE interaction is 3+, and non-cognate interactions are 0 (background white). For library selections, colonies showing 3+ or greater activity with wild-type Rev ARM or R35G-N40V on positive screens and 0 activity with BIV Tat on negative screens were picked and grown overnight at 37 °C with shaking as individual 200 µl cultures in LB medium supplemented with chloramphenicol. Like cultures were pooled, plasmid DNA extracted, and reporter plasmid separated from supplier plasmid by agarose gel electrophoresis.

ONPG solution assay of antitermination

For each interaction, four to six representative colonies were picked from X-gal plates for solution assays. For measurement of ARM–RNA recognition via N-*nut* antitermination, cultures were grown overnight for 16 hours at 30 ºC with aeration in 3 ml of tryptone containing 100 µg/ml ampicillin and 12 µg/ml chloramphenicol as antibiotics, and with 0.05 mM IPTG. The cells were then permeabilized and assayed for β-galactosidase activity using ONPG. Units of β-galactosidase were calculated following Miller (1992). Percent activities are reported normalized to the activity of wild-type Rev ARM or R35G-N40V with the RRE IIB reporter and BIV Tat with the BIV TAR reporter of that day's experiment.

Structure visualization

Jmol (http://www.jmol.org/ [18 January 2015]), an open-source Java viewer for chemical structures in 3D, was used to view and measure distances in NMR structures of HIV-1 Rev-RRE (Protein Data Bank 1ETF and 1ETG; Battiste *et al.*, 1996), a crystal structure of Rev-RRE (Protein Data Bank 4PMI; Jayaraman *et al.*, 2014), and U:G reverse wobble base pairs in an engineered glutamine tRNA (Protein Data Bank 1EXD; Bullock *et al.*, 2000) and domain IIID of the hepatitis C virus IRES (Protein Data Bank 1FQZ; Klinck *et al.*, 2000), which were located via "Database of non-canonical base pairs found in known RNA structures" (http://prion.bchs.uh.edu/bp_type/ [18 January 2015]; Nagaswamy *et al.*, 2000).

RESULTS

We first confirmed that the N-*nut* antitermination reported assay functioned as expected when the RNA reporter was transformed into ARM-supplier hosting competent cells (Table 1). RSG1.2 (Harada *et al.*, 1997), a non-natural ARM structurally unrelated to Rev, served as a RRE-binding control, and the bovine immunodeficiency virus (BIV) Tat-TAR interaction (Puglisi *et al.*, 1995) served as a heterologous control for specificity. Consistent with previous results, the activity of full-length RRE IIB was higher than its apically truncated variant, NMR RRE, which has been used in structural studies (Battiste *et al.*, 1996) and previous selections (Harada *et al.*, 1996; Harada *et al.*, 1997; Tan and Frankel, 1998; Peled-Zehavi *et al.*, 2003). It should be noted that this reporter system depends on the proper folding of RNA and assembly of a cellular complex that does not directly depend on affinities *in vitro*, although previous studies with Rev-RRE in this system predict close agreements between activities *in vivo* and affinities *in vitro*. We chose the longer, full-length RRE IIB for its higher activity relative to background and because it is the RNA with which R35G-N40V had been isolated (Possik *et al.*, 2013).

Library design and selection for active RRE IIB mutants

Reasoning that library approaches to finding RRE IIB mutants active with wild-type Rev ARM and R35G-N40V could be more efficient than testing of individually designed constructs, we chose to construct and screen a series of overlapping, regionally randomized libraries covering the entire RRE IIB from basal stem base pair U43:G77 to the apical loop (Table 2). Randomized regions of three adjacent base pairs would contain all single substitutions of individual nucleotides and base pairs as well as all combinations of mutant base pairs and adjacent base pairs that could act as compensatory mutations using base stacking interactions or the shifting of bulged nucleotides. Assuming that bases in the conserved regions of RRE IIB (Figure 1A) cannot vary except with the known isosteric replacement of G48A:G71A and that flanking base pairs must be Watson-Crick or wobble base pairs, these libraries contain at most 6 conserved positions, and would be expected conservatively to have a recoverable proportion of at least 1/4096 of sequences meeting the wild-type Rev ARM-binding consensus. This frequency is within the numbers manageable by hand screening using the N-*nut* a reporter system (Figure 2). This approach would not find compensatory mutations more than two base pairs distant, but randomizing larger regions would dilute active mutants and require substantially larger screening.

Libraries transformed into wild-type Rev ARM or R35G-N40V host cells yielded between 30 and 1300 colonies appearing active of 20,000 to 90,000 transformants per screen (Table 2). Because false positives sometimes occur in this screening system as frequently as 1/200 (Harada *et al.*, 1996), the initial active colonies were selected and pooled, used to prepare plasmid, and their RNA reporter plasmid separated from the N-supplier plasmid by gel electrophoresis. The resulting RNA reporter plasmid pools were retransformed for at least one additional round of positive selection to remove bystander plasmids and usually a final negative selection with the heterologous BIV Tat ARM supplier plasmid to reduce the proportion of false positives (Table 2). Isolated clones were chosen for DNA preparation, and individual clones of reporter plasmid activities were confirmed by plate assays and chosen for sequencing. From the twenty selections, a total of 40 unique sequences were collected, some of which were isolated in multiple selections (Table S1). In three R35G-N40V selections, no individuals were tested or sequenced because the proportion of false positives remained high. Three other selections yielded no confirmed, active sequences.

Diverse RRE IIB mutations were isolated

Of the 40 sequences isolated from selections, eight particularly interesting sequences were reconstructed from synthetic oligonucleotides to confirm their activity. Three additional mutant sequences were constructed to assess the role of specific base pairs in a wild-type RRE IIB context. All resulting 43 RRE IIB mutant reporters were assayed for activity with wild-type Rev ARM, R35G-N40V, R35G, N40V, and BIV Tat ARM on solid media with X-gal (Table S1). Of the sequenced isolates, many had strong function on both wild-type Rev ARM and R35G-N40V and some demonstrated high specificity. The discovery of mutant RREs specific to either wild-type Rev ARM or R35G-N40V corroborates that R35G-N40V uses a distinct recognition strategy.

We chose the three constructed mutants and 21 representative library-derived sequences and measured their activities with wild-type Rev ARM, R35G-N40V, and BIV Tat ARM by solution assay (Table 3). N-*nut* antitermination solution assays largely agree with visually scored plate assays and offer an objective and larger range measurement of antitermination activity (Harada *et al.*, 1996). Several observations emerge, including the RRE IIB requirements shared between wild-type Rev ARM and R35G-N40V, that RRE mutant G48U (Table 3, mutant #8) is active with wild-type Rev ARM, that the two ARMs tolerate different mutation to bases flanking the internal loop, and that tolerance to mutation in the apical stem is complex, though a preference for Watson-Crick and wobble base pairs is clear.

Consistent with previous studies (Bartel *et al.*, 1991; Giver *et al.*, 1993b; Iwazaki *et al.*, 2005), we found the wild-type Rev ARM core binding site, an internal loop flanked by base pairs (comprising U45:A75, G46:C74, G47:A73, either G48:G71 or the isosteric G48A:G71A, and C49:G70), was nearly invariant with the exception of G48U (Table 3, mutants #8, #10). Interestingly, R35G-N40V appears to have a very similar core binding site, but it accepts variation at U45:A75 and does not tolerate G48U. That R35G-N40V also recognizes the internal loop is consistent with the possibility it shares many interactions found in wild-type Rev ARM-RRE IIB, even if it lacks the critical Asn40 interaction with G47:A73.

The G48:G71 base pair

Unlike the G47:A73 base pair that makes contacts to wild-type Rev ARM, the role of G48:G71 appears to be to widen the major groove to accommodate the Rev ARM α helix (Battiste *et al.*, 1996). G48:G71 adopts a symmetric homopurine pair, Leontis-Westhof classification G•G *trans* W.C./W.C. (Leontis and Westhof, 2001; pair III of Table 6-1 in Saenger, 1984). The primary evidence for its structural role in widening the groove is that it can be replaced with the isosteric symmetric homopurine pair A:A, Leontis-Westhof classification A•A *trans* W.C./W.C. (Leontis and Westhof, 2001; pair I of Table 6-1 in Saenger, 1984), yet does not make hydrogen contacts to Rev ARM (Bartel *et al.*, 1991; Battiste *et al.*, 1996). In our assay system, G48A:G71A (Table 3, mutant #6) leads to no loss of activity with wild-type Rev ARM and moderately reduced in activity with R35G-N40V, suggesting that it forms a similar structural role without specific contacts in both ARM-RRE interactions.

Intriguingly, a G48U mutation simultaneous with G50A:C69A was found to have strong activity with wild-type Rev ARM, though little with R35G-N40V (Table 3, mutant#10). This suggested that wild-type Rev ARM tolerates G48U, an observation we did not find noted previously. To pursue this, we constructed RRE IIB reporters with G48U alone (Table 3, mutant #8) and G50A:C69A alone (Table 3, mutant #15). Unexpectedly, G48U had strong activity with wild-type Rev ARM, yet only background activity with R35G-N40V.

Some RRE IIB mutants display high specificity

Wild-type RRE IIB has relaxed specificity toward wild-type Rev ARM and R35G-N40V. Several mutant RREs showed high specificity, in which the activity with one partner was reduced by a factor of 10 relative to RRE, but the activity with the other partner was reduced less than 2-fold (Table 3). In some cases, specificity arises by only one or two substitutions of RRE IIB. How does

specificity arise? Both wild-type Rev ARM and R35G-N40V appear to share a core binding site of the internal loop formed by the two purine:purine base pairs and their flanking base pairs: G46:C74, G47:A73, either G48:G71 or the isosteric G48A:G71A, and less stringently, C49:G70. That both wild-type Rev ARM and R35G-N40V recognize the same core site suggests these bases contribute to the structure of the binding site regardless of base-specific contacts to ARM. Specificity could arise by violation of one partner's requirements while maintaining the other's. In the internal loop, specificity arises by mutation of the symmetric purine:purine base pair by G48U, which is tolerated by wild-type Rev ARM but not by R35G-N40V (Table 3, mutants #8, #10). Immediately basal to the core binding site, wild-type Rev ARM prefers U45:A75, and R35G-N40V does not (Table 3, mutant #2). In contrast, wild-type Rev ARM is little affected by distal mutations in the basal stem (Table 3, mutant #1) compared to R35G-N40V.

Our data are consistent with wild-type Rev ARM and R35G-N40V both recognizing the internal loop and their distinct specificities arising from differences in their tolerance to G48U and dependence on proximal flanking regions. A consensus of wild-type Rev ARM preferences apical to the core binding site is defined by published data: no requirement at 50:69, and preference for C51:G67, U66, and base pairing in the apical stable stem (Bartel *et al.*, 1991; Giver *et al.*, 1993b; Iwazaki *et al.*, 2005). Our data support the importance of these bases for wild-type Rev ARM recognition, but we note mutations are sometimes tolerated in the presence of other mutations (Table 3, mutants #11, #14, #17, and #18), which suggests complex compensatory mechanisms. In contrast, R35G-N40V tolerates only some mutations at G50:C69 and tolerates many substitutions at C51:G69 and U66, and library isolates most often have base-paired apical stems. No strict R35G-N40V consensus outside the internal loop is apparent in our data. Thus, specificity to wild-type Rev ARM can arise from mutation of G50:C69 (Table 3, mutants #4 and #15), and specificity to R35G-N40V can arise from mutation of C51:G67 and U66 (Table 3, mutants #12, #19, #20, and #21).

Apical stem

Many sequences isolated from library 10, in which the terminal 11 nucleotides of the apical stem are randomized, appear to have a predominance of Watson-Crick or wobble-base pairing consistent with a stable A-form stem. This suggests apical stem stability or continuation of the A-form helix is important for binding both ARMs. Several non-exclusive possibilities arise: 1) base pairing acts as a simple clamp to stabilize the basal binding site; 2) base pairing favors bound conformations via stacking interactions; 3) base pairs in the apical stem prevent alternate, incompatible secondary structures. None of these explanations are easily distinguishable without physical studies.

Reflecting on the diminished activity of NMR RNA, which is truncated and capped with a stable GNRA tetraloop (Heus and Pardi, 1991), we noted that its activities are about one-third of RRE IIB with wild-type Rev ARM, R35G-N40V, and RSG1.2 (Table 1). Curious as to this effect of truncation and capping, we constructed 4 truncated RRE IIB mutants (Table 4) in which the stem immediately apical to the consensus wild-type Rev ARM binding site was replaced with 4 different tetraoligonucleotides representing classes of stable tetraloops: GCAA (Heus and Pardi, 1991), UUCG (Molinaro and Tinoco, 1995), CUUG (Jucker and Pardi, 1995), and UUUU (Proctor *et al.*, 2004). These tetraloops adopt specific and stable structures, or in the case of UUUU, no stable structure, which could affect ARM binding by altering the stability or conformation of the adjacent binding site. Interestingly, activities with wild-type Rev ARM were reduced to nearly background levels, 10-fold less than with NMR RRE, though activities with R35G-N40V were similar to NMR RRE, and activities with RSG1.2 activities, whose binding site does not extend apically from the internal loop beyond G50:C69 (Gosser *et al.*, 2001; Zhang *et al.*, 2001; Iwazaki *et al.*, 2005), were enhanced relative to NMR RRE. This suggests that the RRE IIB apical stem stabilizes the wild-type Rev ARM binding site without distortion and perhaps allows necessary conformational adjustments upon binding. That R35G-N40V activity is not altered much between NMR RRE and the tetraloop-bearing truncated RREs is consistent with our mutagenesis data that its binding site does not extend apically past G50:C59.

DISCUSSION

Curious as to how the change in recognition strategy engendered by the R35G-N40V mutation in Rev ARM could alter the ability of RRE IIB to evolve, we selected active mutants from an overlapping series of RRE IIB libraries with randomized regions. We find many RRE mutants active with wild-type Rev ARM and many active with R35G-N40V. Some mutants have increased specificity toward wild-type Rev ARM, and some have increased specificity toward R35G-N40V, a finding that corroborates the use of a different recognition strategy by Rev R35G-N40V. We find the core binding site comprising the internal loop and flanking base pairs G46:C74 and C49:G70 is nearly invariant in selections against both ARMs. We find that the 48:71 pair can be U:G with wild-type Rev ARM, but not with R35G-N40V. Wild-type Rev ARM appears more dependent on base identities and pairing further from the core binding site. Unlike wild-type Rev ARM, R35G-N40V is sensitive to the identity and pairing of G50:C69 yet tolerant of mutation to C51:G67 and U66. The influence of the distal apical stem is complex and affects activities of the two ARMs differently. Compensatory interactions may allow variation at conserved bases, and specific mutants arise when one strategy is thwarted.

Approach

Our strategy of regional mutagenesis was designed to be unbiased and to randomize at least three base pairs at a time using the N-*nut* antitermination reporter. The reporter system has been successfully applied previously to similar questions (Iwazaki *et al.*, 2005), preserves conditions *in vivo*, and though no affinity measurements were made *in vitro*, other studies with Rev-RRE in this system report agreement between activities *in vivo* and affinities *in vitro* (Harada *et al.*, 1996; Harada *et al.*, 1997; Iwazaki et al., 2005). The regional randomization facilitated recovery of multiple substitutions that suggest complex compensatory mutations occur. Active mutants with substitutions at more than three nearby positions (Table 3, mutants #3, #9, #11, #12, #13, #14, #17, #18, #21, and #22) would be difficult to find by strategies of designed constructs, doping, or complete randomization without the screening and sequencing of many more clones.

Structural insights

Though our screens were limited, we find many variants of RRE IIB are active with wild-type Rev ARM, as expected from earlier studies (Bartel *et al.*, 1991; Giver *et al.*, 1993b; Iwazaki *et al.*, 2005). Despite the internal loop of the core binding site being nearly invariant, several RRE mutants with alterations to conserved C51:G67, U66, and even C49:G70, are active with wild-type Rev ARM. These mutations were found only in the presence of other mutations, suggesting that consensus features flanking the internal loop can be compensated by nearby mutations. Importantly, wild-type Rev ARM and R35G-N40V appear to recognize the same internal loop and impose similar constraints on its mutability. To what extent this conserved core is a simple reflection of a shared binding site or reflects similarity in the recognition strategies is difficult to disentangle without structural studies. On one hand, R35G-N40V shares 15 of 17 amino acids with wild-type Rev ARM and was isolated from a screen with RRE IIB. Thus, parsimony would predict that many aspects of the recognition strategy remain intact despite the absence of the important Arg35 and critical Asn40, whose replacements in R35G-N40V do not have hydrogen-bonding side chains. On the other hand, some RRE IIB-binding peptides unrelated to Rev bind the internal loop using very different recognition strategies (Harada *et al.*, 1996; Harada *et al.*, 1997; Gosser *et al.*, 2001; Zhang *et al.*, 2001). That all these selected Rev-RRE variants use a structurally related internal loop suggests the RRE IIB internal loop is a particularly "good" binding site: a set of structural elements that are accessible to stable peptide structures and display sufficient hydrogen bond donors and acceptors for specificity.

What aspects of the recognition strategy might remain without Arg35 and Asn40, how might their lack be compensated, and what light might our data shed? The lack of Arg35 and Asn40 contacts could conceivably be compensated by new or improved interactions accompanied by some adjustment to the ARM-RNA interaction. Arg35 interacts with the G67 base. Thus, tolerance of R35G-N40V to mutation in this region is unsurprising. Noting that a previous study found wild-type Rev ARM tolerates U45C:A75G (Bartel *et al.*, 1991), the specific tolerance of R35G-N40V to U45G:A75U could be a manifestation of altered contacts near Val40. The sensitivity of R35G-N40V to mutation of C50:C69 could be the result of new or improved interactions of Arg38, Arg41, Arg42, Arg43, Arg46, and Arg48, which interact with backbone phosphates.

One means to new or improved contacts could be deeper penetration of the major groove by R35G-N40V (Possik *et al.*, 2013). The non-canonical G48:G71 pair (Leontis-Westhof classification G•G *trans* W.C./W.C., Leontis and Westhof, 2001; pair III of Table 6-1 in Saenger, 1984) is remarkable for being an important element of the RRE core binding site without making critical contacts to Rev. It is replaceable with the isosteric A:A pair (Leontis-Westhof classification A•A *trans* W.C./W.C., Leontis and Westhof, 2001; pair I of Table 6-1 in Saenger, 1984). Based on similarity (Westhof, 2014), its role has been explained as groove-widening, as the canonical A-form RNA double helix has a major groove too narrow to accommodate an α helix. Intriguingly, the crystal structure indicates a hydrogen bond between Asn40 and G71 in RRE IIB that would not exist in the G48A:G71A mutant. The G48:G71 base pair contributes to forming the distorted backbone spanning G70 - A73, which includes several phosphates making important electrostatic contacts to Arg46 (Battiste *et al.*, 1996). The activity of G48U with wild-type Rev ARM had not been noted previously. Interestingly, Giver *et al.* (1993a; 1993b) find G48C:G71A active, which they explain by the formation of the asymmetric reverse wobble C:A base pair (Leontis-Westhof classification A•C *trans* W.C./W.C., Leontis and Westhof, 2001; pair XXVI of Table 6-1 in Saenger, 1984) whose glycosyl carbons are nearly 1 Å narrower than the G48:G71 pair. Giver *et al.* (1993a; 1993b) did not test U48:G71, but it does appear without comment in other selections (clone R41 of Table 1 of Iwazaki *et al.*, 2005; clone 8 of Figure 3 of Bartel *et al.*, 1991). We note that a reverse wobble U:G base pair (Leontis-Westhof classification G•U *trans* W.C./W.C., Leontis and Westhof, 2001; pair XXVII of Table 6-1 in Saenger, 1984) isostructural to the reverse wobble C:A base pair has been observed in RNA structures (Bullock *et al.*, 2000; Klinck *et al.*, 2000). Thus, we propose the U48 mutation may pair with G71 in a reverse wobble configuration and narrow the major groove compared to G48:G71. This reverse wobble pair would widen the groove to accept wild-type Rev ARM binding, yet be too narrow for R35G-N40V. Importantly, it could preserve the Asn40

hydrogen bond to G71 observed in the crystal structure. Reviewing the literature in detail, we note G48U (in the context of C69U) having undiminished activity with wild-type Rev ARM in the same antitermination assay used here (clone R41 Table 1 of Iwazaki *et al.*, 2005). Werstuck *et al.*, (1996) list a clone with G48U in the context of other mutations that has only 4-fold increase in its dissociation constant. That R35G-N40V functions with A48:A71, but not with U48, suggest R35G-N40V is more dependent on a widened RRE major groove than wild-type Rev ARM. This could reflect deeper penetration of the major groove that would bury the hydrophobic surface of N40V and could accommodate improved arginine-phosphate contacts. We note that the R35G mutation could facilitate this by reducing peptide bulk at the binding interface, possibly explaining why the N40V mutation is active only with R35G.

Flanking sequences affect recognition in interactions unrelated to Rev-RRE IIB. BIV TAR does not bind BIV Tat ARM when an adjacent clamping base pair is mutated (Smith *et al.*, 1998). Consistent with simple stabilization by clamping base pairs, all active RRE mutants have a Watson-Crick or wobble base pair at 52:65, and there is a predominance of Watson-Crick and wobble pairing in the apical stems of active RRE mutants isolated from library 10 (Table S1, mutants #32 to #43). The results of Table 4, in which a variety of canonical tetraloops appended immediately apical to the Rev-binding site are moderately active with R35G-N40V and RSG1.2, yet not active with wild-type Rev ARM, are consistent with the truncations disrupting the interaction of Arg35 with G67. These results are also consistent with the binding site of R35G-N40V being similar to that of RSG1.2 (Iwazaki *et al.*, 2005), which does not extend apical past G50:C69. R35G-N40V would be then less dependent on an extended α helix, and the absence of Arg35 would be compensated by improved contacts to the internal loop and immediate vicinity.

Neutral evolution and origins of specificity

Neutral theories of evolution assert that for any one phenotype, there are sufficient genotypes to create incremental mutational paths that intersect paths originating from distinct phenotypes (Kimura, 1991; Ohta, 2002). Thus, neutral theories apply better to ARM-RNA recognition when many recognition strategies exist, when ARMs and RNAs have many functional mutants, and when ARMs and RNAs can participate in multiple recognition strategies. The Rev ARM-RRE IIB presents a particularly apt candidate for applying neutral theories to ARM-RNA interactions. In addition to the wild-type Rev ARM recognition strategy and that of R35G-N40V, there exist variant ARMs and RNAs that use related and unrelated strategies, including well-characterized examples. RSG1.2 (Harada *et al.*, 1997) binds as a partial α helix to the same internal loop as wild-type Rev

ARM (Gosser *et al.*, 2001; Zhang *et al.*, 2001). Glutamine-containing ARMS R_6QR_7 (Tan and Frankel, 1998), K1 (Peled-Zehavi *et al.*, 2003), and DLA (Sugaya *et al.*, 2008) may be α helices with very similar RRE IIB recognition strategies, though their specificities are not identical. Rev aptamer I binds RRE in a similar manner to wild-type Rev ARM, but relies on different contacts (Giver *et al.*, 1993b; Ye *et al.*, 1996; Possik *et al.*, 2013). Rev aptamer II binds RRE with lower affinity and in an extended conformation (Xu and Ellington, 1996; Ye *et al.*, 1999). Structural models exist for RSG1.2-RRE IIB (Zhang *et al.*, 2001; Gosser *et al.*, 2001), Rev ARM-RAI (Ye *et al.*, 1996;), and Rev ARM-RAII. (Ye *et al.*, 1999). Thus, if sufficient functional and incremental mutants exist, RRE IIB may serve as an intersection sequence linking distinct recognition strategies.

The wild-type Rev ARM-RRE IIB interaction tolerates many mutations. In this reporter system, Rev ARM tolerates some substitutions without two-fold loss of activity at all positions other than Arg39, Arg40, Arg44, and Arg46 (Possik *et al.*, 2013). Though RRE IIB tolerates few substitutions to its internal loop, flanking sequences are mutable. Viewing RRE IIB as a relaxed-specificity intermediate that recognizes both wild-type Rev ARM and R35G-N40V, we find nearly incremental, neutral paths between distinct specificity RRE mutants (Figure 3), where one substitution connects RRE IIB to mutant #8, which is specific to wild-type Rev ARM, and two and three substitutions connect RRE IIB to mutants #2 and # 20, respectively, which are specific to R35G-N40V. Similarly, mutant #16 connects wild-type Rev ARM-specific mutant #15 to R35G-N40V-specific mutant #12. Indeed, the analysis of the structural basis for specificity above would predict that exhaustive screening would likely find sufficient RRE mutants to describe a neutral path with increments of single substitutions. Some RRE mutants are active with wild-type Rev ARM and single mutant Rev ARM R35G (Table S1, mutants #6, #10, #14, #18, #22, #23), thus providing neutral paths via mutation of the RNA and ARM sequentially.

Similar to how RRE IIB bridges the specificity of wild-type Rev ARM and R35G-N40V, structurally unrelated ARM-RNA interactions have been found to undergo transitions to new specificities via intermediate ARMs and RNAs that bind multiple partners. A hybrid HIV-BIV TAR functions with both HIV and BIV Tat proteins by having flanking base pairs necessary for BIV Tat binding and an apical loop that recruits the host factor necessary for HIV Tat activity (Smith *et al.*, 1998). The ARMs of bacteriophage λ and P22 N proteins bind as α helices to their boxB hairpin loops. Specificity arises from the ARMs recognizing their boxB GNRA-like pentaloops adopting either a 4-out (λ) or 3-out (P22) conformations. (Legault *et al.*, 1998; Cai *et al.*, 1998; Schärpf *et al.*, 2000) As an example of intersection sequences, several boxB mutants are active with both λ and

P22 N ARMs, apparently by being able to adopt both conformations (Cocozaki *et al.*, 2008a). Conversely, a mutant P22 N ARM, R30W, is able to recognize both P22 and λ boxBs by having an essential aromatic side chain required for stacking on the λ boxB 4-out conformation (Cocozaki *et al.*, 2008b).

The key features of these model ARM-RNA interactions, whether the internal loop of RRE IIB, the bulge of TARs, or the hairpin loop of boxBs, include alternate conformations of RNAs, roles for base pairs flanking a common binding site, and ARMs that adopt different conformations: RRE IIB, like TARs and boxBs, displays induced-fit recognition by undergoing conformational changes upon binding wild-type Rev ARM (Battiste *et al.*, 1994; Tan and Frankel, 1994) and RSG1.2 (Gosser *et al.*, 2001; Zhang *et al.*, 2001). Concordantly, Rev ARM, similarly to Tat and N ARMs, becomes more structured upon binding (Tan and Frankel, 1994). One can imagine that by accommodating α helices, presenting energetically important arrangements of phosphates, and having major groove bases with specific hydrogen bond donors and acceptors, the RRE IIB internal loop acts as a scaffold for many variations of binding themes and complements the apparent rich variations of α-helical ARMs.

More broadly, by allowing diverse and plastic structural variations within an ARM-RNA theme, induced fit may facilitate neutral evolution by allowing either partner to adapt to the conformation necessary to bind in related and distinct recognition strategies (Bayer *et al.*, 2005). Thus, it is unsurprising to find ARM-RNA interactions employed by viruses, which with small genomes and simplistic regulatory systems, have need for small interactions of high specificity (Frankel, 2000).

CONCLUSION

Our findings corroborate that R35G-N40V recognizes RRE IIB using a strategy distinct from wild-type Rev ARM and support the existence of neutral evolutionary paths. The data herein will support structural studies needed to understand R35G-N40V-RRE IIB interaction better. The core binding site of RRE IIB, the internal loop, accepts diverse and unrelated binding modes, and the flanking sequence allow many specific variants. RRE may be unusually compatible with neutral evolution of recognition strategies, in that the internal loop allows different peptide secondary structures and base specific contacts, and flanking sequence allows modulation of specificity: it may be an oasis in a neutral network.

Our findings align with studies of other ARM-RNA recognition, in which distinct specificities exist and transitions to new specificities occur via multifunctional intermediates (Smith *et al.*, 1998; Smith *et al.*, 2000; Iwazaki *et al.*, 2005; Cocozaki *et al.*, 2008a; Cocozaki *et al.*, 2008b; Possik *et al.*, 2013), sometimes with single residue changes (Iwazaki *et al.*, 2005; Cocozaki *et al.*, 2008a; Cocozaki *et al.*, 2008b). Internal loop and α-helical ARMs may be particularly suited to evolutionary transitions by having many variations and distinct specificities accessible by few mutations, and Rev-RRE offers a model with which to study fine structure of evolution of specificity.

ACKNOWLEDGEMENTS

We thank Naomi Franklin and Kazuo Harada for materials and Ingrid Ghattas for advice. This work was supported by the Research Board and the Kamal A. Shair Central Research Science Laboratory of the American University of Beirut.

REFERENCES

Bartel DP, Zapp ML, Green MR, Szostak JW. 1991. HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell.* **67** :529-36.

Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, Frankel AD, Williamson JR. 1996. α helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science.* **273** :1547-51.

Battiste JL, Tan R, Frankel AD, Williamson JR. 1994. Binding of an HIV Rev peptide to Rev responsive element RNA induces formation of purine-purine base pairs. *Biochemistry.* **33** :2741-7.

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

Bullock TL1, Sherlin LD, Perona JJ. 2000. Tertiary core rearrangements in a tight binding transfer RNA aptamer. *Nat Struct Biol.* **7** :497-504.

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

Cocozaki AI, Ghattas IR, Smith CA. 2008a. Bacteriophage P22 antitermination boxB sequence requirements are complex and overlap with those of lambda. *J Bacteriol.* **190** :4263-71. DOI: 10.1128/JB.00059-08.

Cocozaki AI, Ghattas IR, Smith CA. 2008b. The RNA-Binding Domain of Bacteriophage P22 N Protein Is Highly Mutable and a Single Mutation Relaxes Specificity Toward λ. *J Bacteriol.* **190** :7699-708. DOI: 10.1128/JB.00997-08.

Das C, Frankel AD. 2003. Sequence and structure space of RNA-binding peptides. *Biopolymers.* **70** :80-5.

Daugherty MD, D'Orso I, Frankel AD. 2008. A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA. *Mol Cell.* **31** :824-834.

Engelke DR, Krikos A, Bruck ME, Ginsburg D. 1990. Purification of *Thermus aquaticus* DNA polymerase expressed in *Escherichia coli*. *Anal Biochem.* **191** :396-400.
Frankel AD. 2000. Fitting peptides into the RNA world. *Curr Opin Struct Biol.* **10** :332-40.
Frankel AD, Smith CA. 1998. Induced folding in RNA-protein recognition: more than a simple molecular handshake. *Cell.* **92** :149-51.
Franklin NC, Doelling JH. 1989. Overexpression of N antitermination proteins of bacteriophages lambda, 21, and P22: loss of N protein specificity. *J Bacteriol.* **171** :2513-22.
Giver L, Bartel DP, Zapp ML, Green MR, Ellington AD. 1993a. Selection and design of high-affinity RNA ligands for HIV-1 Rev. *Gene.* **137** :19-24.
Giver L, Bartel D, Zapp M, Pawul A, Green M, Ellington AD. 1993b. Selective optimization of the Rev-binding element of HIV-1. *Nucleic Acids Res.* **21** :5509-16.
Gosser Y, Hermann T, Majumdar A, Hu W, Frederick R, Jiang F, Xu W, Patel DJ. 2001. Peptide-triggered conformational switch in HIV-1 RRE RNA complexes. *Nat Struct Biol.* **8** :146-50.
Harada K, Martin SS, Frankel AD. 1996. Selection of RNA-binding peptides in vivo. *Nature.* **380** :175-9.
Harada K, Martin SS, Tan R, Frankel AD. 1997. Molding a peptide into an RNA site by in vivo peptide evolution. *Proc Natl Acad Sci U S A.* **94**:11887-92.
Heaphy S, Dingwall C, Ernberg I, Gait MJ, Green SM, Karn J, Lowe AD, Singh M, Skinner MA. 1990. HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the Rev response element. *Cell.* **60**:685–693.
Heus HA, Pardi A. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science.* **253** :191-4.
Iwazaki T, Li X, Harada K. 2005. Evolvability of the mode of peptide binding by an RNA. *RNA.* **11** :1364-73.
Jayaraman B, Crosby DC, Homer C, Ribeiro I, Mavor D, Frankel AD. 2014. RNA-directed remodeling of the HIV-1 Rev protein orchestrates assembly of the Rev-Rev response element complex. eLife 3:e04120. doi: 10.7554/eLife.04120
Jucker FM, Pardi A. 1995. Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry.* **34** :14416-27.
Kimura, M. 1991. Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proc Natl Acad Sci U S A.* **88** :5969-5973.
Kjems J, Calnan BJ, Frankel AD, Sharp PA. 1992. Specific binding of a basic peptide from HIV-1 Rev. *EMBO J.* **11** :1119-29.
Klinck R, Westhof E, Walker S, Afshar M, Collier A, Aboul-Ela F. 2000. A potential RNA drug target in the hepatitis C virus internal ribosomal entry site. *RNA.* **6** :1423-31.
Legault P, Li J, Mogridge J, Kay LE, Greenblatt J. 1998. NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell.* **93** :289-99.
Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA.* **7** :499-512.
Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR. 1989. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature.* **338** :254-7.
Mayer G. 2009. The chemical biology of aptamers. *Angew Chem Int Ed Engl.* **48** :2672-89
Miller JH. 1992. *A short course in bacterial genetics: a laboratory manual and handbook for* Escherichia coli *and related bacteria.* Cold Spring Harbor Laboratory Press, Plainview, NY.
Minor DL Jr, Kim PS. 1996. Context-dependent secondary structure formation of a designed protein sequence. *Nature.* **380** :730-4.
Molinaro M, Tinoco I Jr. 1995. Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: thermodynamic and spectroscopic applications. *Nucleic Acids Res.* **23** :3056-63

Nagaswamy U, Voss N, Zhang Z, Fox GE. 2000. Database of non-canonical base pairs found in known RNA structures. *Nucleic Acid Res.* **28** :375-376.
Ohta T. 2002. Near-neutrality in evolution of genes and gene regulation. *Proc Natl Acad Sci U S A.* **99** :16134-7.
Patel DJ. 1999. Adaptive recognition in RNA complexes with peptides and protein modules. *Curr Opin Struct Biol.* **9** :74-87.
Peled-Zehavi H, Horiya S, Das C, Harada K, Frankel AD. 2003. Selection of RRE RNA binding peptides using a kanamycin antitermination assay. *RNA.* **9**:252-61.
Possik EJ, Bou Sleiman MS, Ghattas IR, Smith CA. 2013. Randomized codon mutagenesis reveals that the HIV Rev arginine-rich motif is robust to substitutions and that double substitution of two critical residues alters specificity. *J Mol Recognit.* **26**:286-96. doi: 10.1002/jmr.2272.
Proctor DJ, Ma H, Kierzek E, Kierzek R, Gruebele M, Bevilacqua PC. 2004. Folding thermodynamics and kinetics of YNMG RNA hairpins: specific incorporation of 8-bromoguanosine leads to stabilization by enhancement of the folding rate. *Biochemistry.* **43** :14004-14.
Puglisi JD, Chen L, Blanchard S, Frankel AD. 1995. Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science.* **270** :1200-3.
Puglisi JD, Chen L, Frankel AD, Williamson JR. 1993. Role of RNA structure in arginine recognition of TAR RNA. *Proc Natl Acad Sci U S A.* **90** :3680-4.
Saenger W. 1984. *Principles of Nucleic Acid Structure.* Springer-Verlag: New York.
Schärpf M, Sticht H, Schweimer K, Boehm M, Hoffmann S, Rösch P. 2000. Antitermination in bacteriophage lambda. The structure of the N36 peptide-boxB RNA complex. *Eur J Biochem.* **267** :2397-408.
Schultes EA, Bartel DP. 2000. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science.* **289** :448-52.
Schuster P. 2001. Evolution in silico and in vitro: the RNA model. *Biol Chem.* **382** :1301-14.
Schuster P, Fontana W, Stadler PF, Hofacker IL. 1994. From sequences to shapes and back: a case study in RNA secondary structures. *Proc Biol Sci.* **255** :279-84.
Smith CA, Calabro V, Frankel AD. 2000. An RNA-binding chameleon. *Mol Cell.* **6** :1067-76.
Smith CA, Crotty S, Harada Y, Frankel AD. 1998. Altering the context of an RNA bulge switches the binding specificities of two viral Tat proteins. *Biochemistry.* **37** :10808-14.
Sugaya M, Nishino N, Katoh A, Harada K. 2008. Amino acid requirement for the high affinity binding of a selected arginine-rich peptide with the HIV Rev-response element RNA. *J Pept Sci.* **14** :924-35. DOI: 10.1002/psc.1027.
Tan R, Chen L, Buettner JA, Hudson D, Frankel AD. 1993. RNA recognition by an isolated α helix. *Cell.* **73** :1031-40.
Tan R, Frankel AD. 1994. Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. *Biochemistry.* **33** :14579-85.
Tan R, Frankel AD. 1998. A novel glutamine-RNA interaction identified by screening libraries in mammalian cells. *Proc Natl Acad Sci U S A.* **95** :4247-52.
Wagner A. 2008. Robustness and evolvability: a paradox resolved. *Proc Biol Sci.* **275** :91-100.
Werstuck G1, Zapp ML, Green MR. 1996. A non-canonical base pair within the human immunodeficiency virus rev-responsive element is involved in both rev and small molecule recognition. *Chem Biol.* **3** :129-37.
Westhof E. 2014. Isostericity and tautomerism of base pairs in nucleic acids. *FEBS Letters*. **588** :2464-2469.
Williamson JR. 2000. Induced fit in RNA-protein recognition. *Nat Struct Biol.* **7** :834-7.
Xu W, Ellington AD. 1996. Anti-peptide aptamers recognize amino acid sequence and bind a protein epitope. *Proc Natl Acad Sci U S A.* **93** :7475-80.

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

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

Zapp ML, Green MR. 1989. Sequence-specific RNA binding by the HIV-1 Rev protein. *Nature.* **342** :714-6.

Zhang Q, Harada K, Cho HS, Frankel AD, Wemmer DE. 2001. Structural characterization of the complex of the Rev response element RNA with a selected peptide. *Chem Biol.* **8** :511-20.

FIGURE LEGENDS

Figure 1. RREs and Rev ARMs: A) Left, the average HIV-1 Rev ARM-RRE IIB NMR structure of Battiste *et al.* (1996; Protein Data Bank 1ETF) in which the Rev ARM backbone is gray and the RNA is a light gray cartoon. Arg35 and Asn40 residues are shown as CPK-colored sticks. Important nucleotides are shown colored salmon, and groove-widening G48:G71 pair is olive. Middle, the secondary structure of the apically truncated stem IIB of RRE with a GNRA tetraloop that was used in NMR structural studies, with nucleotides numbered according to Battiste *et al.* (1996). Nucleotides found important in previous studies are bold. Mutant bases are shown as small letters. Right, the secondary structure of the RRE IIB used in this study is shown. B) The wild-type Rev ARM (amino acids 34-50) fused to the activation domain of λ N is shown above with residues found important for RRE binding by mutagenesis bold (Tan *et al.*, 1993; Tan and Frankel, 1994; Possik *et al.*, 2013) and the R35G-N40V mutant ARM sequence below. The numbering is that of HIV-1 Rev protein. The amino-terminus and carboxy linker used in the N-fusion are shown separated from the ARM with spaces.

Figure 2. ARM-RNA reporter system. Above, HIV-1 Rev protein ARM, amino acids 34 to 50, is expressed as a fusion to the activation domain of λ N by the wild-type Rev ARM-N supplier plasmid. Plasmids with RSG1.2 and BIV Tat ARM are analogous. Below, the RNA reporter plasmid expresses a transcript containing a λ *nut* (N-utilization) site in which the boxB RNA hairpin is replaced with RRE IIB, libraries, mutants, or other RNAs such as BIV TAR. The RNA reporter expresses β-galactosidase (LacZ) downstream of intrinsic transcriptional terminators. LacZ expression is dependent on the N-fusion binding the RNA structure at boxB and recruiting host factors that cause the transcription complex to become antiterminating by ignoring downstream terminators. Competent cells hosting ARM-N suppliers are prepared and transformed with RNA reporter plasmids. LacZ expression in colonies can be monitored by X-gal on solid media and by ONPG in cell extracts. The BIV Tat-TAR interaction served as a heterologous control for specificity.

Figure 3. Paths between specific RRE mutants. RRE IIB and RRE mutants are shown from basal U45:A75 to apical G53:C65 with activities on wild-type Rev ARM (wt) and R35G-N40V (GV) as percent activity relative to RRE IIB. Lines between RNAs represent one, two, or more substitutions by number of segments. One substitution links RRE IIB to wild-type Rev ARM-specific mutant #8, and two substitutions link RRE IIB to R35G-N40V-specific mutant #2.

TABLES

Table 1. Activities of RNA-ARM interactions

	β-galactosidase units[a]			
RNA	Wild-type Rev ARM[b]	R35G-N40V[c]	RSG1.2[d]	BIV Tat[e]
RRE IIB[f]	80 ± 12	40 ± 11	180 ± 60	3.3 ± 0.18
NMR RRE[g]	31 ± 5	13 ± 2	73 ± 8	3.8 ± 0.3
BIV TAR[h]	1.4 ± 0.19	1.1 ± .3	3.5 ± 0.4	100 ± 10

[a] RNA reporter plasmids were transformed into N-ARM supplier cells. At least four replicates of each clone were grown at 30 °C overnight in tryptone medium supplemented with 50 μM IPTG and assayed for β-galactosidase activity with ONPG.
[b] Wild-type Rev ARM contains Rev34-50: TRQARRNRRRRWRERQR
[c] R35G-N40V is a mutant Rev ARM (Possik *et al.*, 2013): TGQARRVRRRRWRERQR.
[d] RSG1.2 contains a non-natural, high-affinity RRE binder (Harada *et al.*, 1997): DRRRRGSRPSGAERRRRRAAAA.
[e] BIV Tat ARM contains an arginine-rich motif from the bovine immunodeficiency virus: MG RPRGTRGKGRRIRR GGGNAAN.
[f] RRE IIB has a wild-type sequence from U43 to G77, except for A44C and a two base pair clamp at the base of the stem: 5'-GGUCUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGCC-3'
[g] NMR RRE: 5'-GGUCUGGGCGCAGCGCAAGCUGACGGUACAGGCC-3'.
[h] BIV TAR (BTAR) is a BIV Tat-binding RNA from bovine immunodeficiency virus that serves as a heterologous control with BIV Tat ARM. RNA sequence: GCUCGUGUAGCUCAUUAGCUCCGAGC.

Table 2. Library selections

RRE IIB library[a]	Sequence[b]	Transformed complexity[c]	Initial WT actives/titer[d]	WT screens[e]	Test:good: seq:unique WT actives[f]	Initial GV actives/titer[d]	GV screens[e]	Test:good: seq:unique GV actives[f]
1-43N46-74N77	gg**NNNN**G–GCG–CAGCGUCA ||||||| ||| |||||| A cc**NNNN**AUGGCAGUCGCAGU	9000	50/20k	+++	18:18:6:1	290/20k	+++	0:0:0:0
2-45N47-73N75	ggUc**NNN**–GCG–CAGCGUCA ||||||| ||| |||||| A ccGG**NNN**UGGCAGUCGCAGU	5000	30/30k	++–	43:1:1:0	90/30k	++–	60:16:7:1
3-46N48-71N74	ggUcU**NN**–**N**CG–CAGCGUCA ||||||| ||| |||||| A ccGGA**NNNN**GCAGUCGCAGU	9000	180/40k	++–	29:27:9:3	180/90k	++–	40:21:11:2
4-47N49-70N73	ggUcUG**N**–**NN**G–CAGCGUCA ||||||| ||| |||||| A ccGGAC**NNNN**CAGUCGCAGU	9000	540/20k	++–	14:13:4:0	180/20k	+++	28:17:3:0
5-48N50-69N71	ggUcUGG–**NNN**–CAGCGUCA ||||||| ||| |||||| A ccGGACAU**NNN**AGUCGCAGU	12500	410/20k	++–	65:52:15:5	170/40k	++–	70:18:12:1
6-49N51-67N70	ggUcUGG–G**NN**–**N**AGCGUCA ||||||| ||| |||||| A ccGGACAUG**NNNN**UCGCAGU	9000	130/20k	+++–	16:16:3:1	180/20k	+++–	0:0:0:0
7-50N52-66N69	ggUcUGG–GC**N**–**NN**GCGUCA ||||||| ||| |||||| A ccGGACAUGG**NNNN**CGCAGU	12000	1310/40k	++–	29:29:13:4	660/60k	++–	34:33:17:4
8-51N53-65N67	ggUcUGG–GCG–**NNN**CGUCA ||||||| ||| |||||| A ccGGACAUGGCA**NNN**GCAGU	12400	270/30k	++–	14:13:7:5	660/30k	++–	66:65:11:7
9-52N54-64N66	ggUcUGG–GCG–C**NNN**GUCA ||||||| ||| |||||| A ccGGACAUGGCAG**NNN**CAGU	9000	360/20k	+++	19:18:4:2	270/20k	+++	0:0:0:0
10-FL11NL	ggUcUGG–GCG–CAG**NNNNN** ||||||| ||| |||||| **N** ccGGACAUGGCAGUC**NNNNN**	9000	50/20k	+++–	16:8:8:8	310/20k	+++–	18:8:6:4

[a] Each library was assigned a number used to indicate the library of origin for every sequence, and the ranges of randomized bases are indicated.
[b] Nucleotide sequence of RRE IIB libraries with nucleotides randomized, N, shown in bold.
[c] Transformed complexity is the total number of colonies pooled from the ligation of each library and used to prepare the plasmid libraries.
[d] Number of colonies appearing active (blue colonies on X-gal media) in the first screen of the plasmid library with wild-type Rev ARM (WT) or R35G-N40V (GV) over the number of total colonies obtained after transformation.
[e] The order of positive (+) screens with wild-type Rev ARM or R35G-N40V and negative screens (–) with BIV Tat.

[f] The number of individual reporter plasmid clones tested (test) to those whose activity was confirmed on controls (good) to those that were sequenced (seq) to those that had unique sequences (unique) in that selection with wild-type Rev ARM (WT) or R35G-N40V (GV).

Table 3. Solution antitermination activities of RNA reporters

Name[a]	Clone sequence[b]	Origin[c]	Percent activation[d]		
			Wild-type Rev ARM	R35G-N40V	BIV Tat
s-RRE IIB	ggUcU**GG**–**GC**G–**C**AGCGUCA ||||||| ||| |||||| A ccGGA**CA**U**GG**CA**GU**CGCAGU	control	**100±14**	**100±30**	3.4 ± 0.18
s-BIV TAR	ggGCUCG**UG**UA**G**CUCA ||||||||| | | ccCGAGC–**C**–U**CG**AUU	control	1.7 ± 0.2	2.7 ± 0.6	**100 ± 10**
1-i-43A44U76A77A	gg**AU**UGG–GCG–CAGCGUCA ||||||| ||| |||||| A cc**AA**ACAUGGCAGUCGCAGU	WT	90 ± 3	32 ± 4	7 ± 1.2
2-r-45G75U	ggUc**G**GG–GCG–CAGCGUCA ||||||| ||| |||||| A ccGG**U**CAUGGCAGUCGCAGU	GV	1.6 ± 0.5	120 ± 24	2.8 ± 0.9
3-i-48A50A69A71A	ggUcUGG–**A**C**A**–CAGCGUCA ||||||| ||| |||||| A ccGGACAU**A**G**A**AGUCGCAGU	WT	60 ± 19	20 ± 3	3.0 ± 0.3
4-r-48A50U71A	ggUcUGG–**A**C**U**–CAGCGUCA ||||||| ||| |||||| A ccGGACAU**A**GCAGUCGCAGU	WT	90 ± 28	1.8 ± 0.18	2.8 ± 0.3
5-i-48A69G71A	ggUcUGG–**A**CG–CAGCGUCA ||||||| ||| |||||| A ccGGACAU**A**G**G**AGUCGCAGU	WT	80 ± 10	19 ± 1.3	2.6 ± 0.3
6-s-48A71A	ggUcUGG–**A**CG–CAGCGUCA ||||||| ||| |||||| A ccGGACAU**A**GCAGUCGCAGU	Synth	90 ± 19	50 ± 19	2.5 ± 0.5
7-i-48A71A72G	ggUcUGG–**A**CG–CAGCGUCA ||||||| ||| |||||| A ccGGACA**GA**GCAGUCGCAGU	WT&GV	110 ± 19	31 ± 7	2.9 ± 0.3
8-s-48U	ggUcUGG–**U**CG–CAGCGUCA ||||||| ||| |||||| A ccGGACAUGGCAGUCGCAGU	Synth	65 ± 9	2.1 ± 0.5	2.4 ± 0.6
9-i-48U49G50A69U	ggUcUGG–**UGA**–CAGCGUCA ||||||| ||| |||||| A ccGGACAUGG**U**AGUCGCAGU	WT	28 ± 6	60 ± 3	3.1 ± 0.2

10-r-48U50A69A	ggUcUGG–**U**C**A**–CAGCGUCA ||||||| ||| ||||||| A ccGGACAUGG**A**AGUCGCAGU	WT	110 ± 27	5 ± 3	2.8 ± 0.3
11-i-49U50U67A68U69U	ggUcUGG–G**UU**–CAGCGUCA ||||||| ||| ||||||| A ccGGACAUGG**UUA**UCGCAGU	GV	80 ± 28	80 ± 10	2.7 ± 0.3
12-i-50A51G66A69U	ggUcUGG–GC**A**–**G**AGCGUCA ||||||| ||| ||||||| A ccGGACAUGG**U**AG**A**CGCAGU	GV	5 ± 0.6	110 ± 16	3.1 ± 0.5
13-r-50A51G67U68U69U	ggUcUGG–GC**A**–**G**AGCGUCA ||||||| ||| ||||||| A ccGGACAUGG**UUU**UCGCAGU	GV	19 ± 5	50 ± 6	2.2 ± 0.4
14-i-50A51U67A69A	ggUcUGG–GC**A**–**U**AGCGUCA ||||||| ||| ||||||| A ccGGACAUGG**A**A**A**UCGCAGU	WT	68 ± 5	21 ± 6	7 ± 1.0
15-s-50A69A	ggUcUGG–GC**A**–CAGCGUCA ||||||| ||| ||||||| A ccGGACAUGG**A**AGUCGCAGU	Synth	140 ± 37	10 ± 1.0	3.1 ± 0.3
16-i-50A69U	ggUcUGG–GC**A**–CAGCGUCA ||||||| ||| ||||||| A ccGGACAUGG**U**AGUCGCAGU	WT&GV	61 ± 7	83 ± 3	3.0 ± 0.4
17-i-51A52U53A65G66A	ggUcUGG–GCG–**AUU**CGUCA ||||||| ||| ||||||| A ccGGACAUGGCAG**AG**GCAGU	WT	50 ± 14	78 ± 6	2.4 ± 1.1
18-i-51A52U53A65U66A	ggUcUGG–GCG–**AUA**CGUCA ||||||| ||| ||||||| A ccGGACAUGGCAG**AU**GCAGU	WT&GV	39 ± 6	70 ± 13	2.8 ± 0.5
19-i-51G52G53U65G67C	ggUcUGG–GCG–**GGU**CGUCA ||||||| ||| ||||||| A ccGGACAUGGCA**C**U**G**GCAGU	GV	1.7 ± 0.4	82 ± 6	2 ± 1.0
20-r-51G52G67U	ggUcUGG–GCG–**GG**GCGUCA ||||||| ||| ||||||| A ccGGACAUGGCA**U**UCGCAGU	GV	1.8 ± 0.19	78 ± 4	2.8 ± 0.17
21-i-51G52U53U 65A66G67U	ggUcUGG–GCG–**GUU**CGUCA ||||||| ||| ||||||| A ccGGACAUGGCA**UGA**GCAGU	GV	1.8 ± 0.7	110 ± 17	3.0 ± 0.5

22-i-51G52U66A67U	ggUcUGG–GCG–**GU**GCGUCA ||||||| ||| ||||||| A ccGGACAUGGCA**UA**CGCAGU	GV	34 ± 3	67 ± 6	2.8 ± 0.7
23-r-53U54U65G	ggUcUGG–GCG–CA**UU**GUCA ||||||| ||| ||||||| A ccGGACAUGGCAGU**G**GCAGU	WT	70 ± 10	80 ± 12	2.7 ± 0.6
24-r-69G	ggUcUGG–GCG–CAGCGUCA ||||||| ||| ||||||| A ccGGACAUGG**G**AGUCGCAGU	WT	23 ± 1.8	1.9 ± 0.3	2.8 ± 0.3

[a] Control RNA reporters are named as in Table 1. Mutant RRE reporters are numbered; all reporters were given a prefix indicating origin, i- (isolated from a library), r- (reconstructed from the sequence of an isolated clone), or s- (synthetic); and RRE mutants described by the numerical position and identity of the mutations relative to RRE IIB.

[b] The nucleotide sequences and assumed secondary structures of RNAs are shown. Important positions in RRE IIB and BIV TAR are bold and underlined. RRE mutants have nucleotides with those differing from RRE IIB bold and underlined.

[c] The origin of sequences are described by the selections from which RRE mutants were isolated: either wild-type Rev ARM (WT) or R35G-N40V (GV). Synthetic constructs are indicated (Synth).

[d] RRE IIB libraries were transformed into ARM-N supplier cells. At least four replicates of each clone were grown at 30 °C overnight in tryptone medium supplemented with 50 μM IPTG and assayed for β-galactosidase activity with ONPG. Percent activation represents antitermination activities of Rev mutants normalized to wild-type Rev ARM–RRE, R35G-N40V–RRE, or BIV Tat–TAR assayed the same day.

Table 4. Activities of apically truncated constructs

Name[a]	RNA sequence[b]	Percent activation[c] wild-type Rev ARM	R35G-N40V	RSG1.2	BIV Tat
RRE IIB	ggUCUGG-GCG-CA**GCGUCA** ||||||| ||| |||||| **A** ccGGACAUGGCAGU**CGCAGU**	**100 ± 1.4**	**100 ± 30**	100 ± 30	3.4 ± 0.2
NMR RRE	ggUCUGG-GCG-CA**GCgc** ||||||| ||| |||| | ccGGACAUGGCAGU**CGaa**	40 ± 5	33 ± 5	40 ± 4	3.9 ± 0.3
RRE-GCAA	ggUCUGG-GCG-CA**gc** ||||||| ||| || | ccGGACAUGGCAGU**aa**	4 ± 2	30 ± 12	90 ± 20	3.8 ± 0.3
RRE-UUCG	ggUCUGG-GCG-CA**uu** ||||||| ||| || | ccGGACAUGGCAGU**gc**	3 ± 1.0	40 ± 10	110 ± 14	3.6 ± 0.2
RRE-CUUG	ggUCUGG-GCG-CA**cu** ||||||| ||| || | ccGGACAUGGCAGU**gu**	1.9 ± 0.4	30 ± 7	80 ± 30	3.5 ± 0.2
RRE-UUUU	ggUCUGG-GCG-CA**uu** ||||||| ||| || | ccGGACAUGGCAGU**uu**	3 ± 1.0	20 ± 5	60 ± 17	3.4 ± 0.3
BIV TAR	**ggGCUCGUGUAGCUCA** ||||||| | ||| | **ccCGAGC-C-UCGAUU**	1.7 ± 0.2	2.7 ± 0.6	2.0 ± 0.2	**100 ± 10**

[a] Control RNA reporters are named as in Table 1. Truncated RRE reporters are named by the sequence of the loop.
[b] The nucleotide sequences and assumed secondary structures of RNAs are shown. Sequences not common to all are bold.
[c] RRE IIB libraries were transformed into ARM-N supplier cells. At least four replicates of each clone were grown at 30 °C overnight in tryptone medium supplemented with 50 μM IPTG and assayed for β-galactosidase activity with ONPG. Percent activation represents antitermination activities of Rev mutants normalized to wild-type Rev ARM–RRE, R35G-N40V–RRE, RSG1.2, or BIV Tat–TAR assayed the same day.

SUPPORTING MATERIAL

Table S1. Solid media antitermination activities of all RNA reporters

Name[a]	Clone sequence[b]	Origin[c]	Activity by X-gal scoring[d]				
			Wild-type Rev ARM	R35G-N40V	R35G	N40V	BIV Tat
s-flRRE	GGUcUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGCC	control	4	4	1	0	0
s-NMR RRE	GGUcUGGGCGCAGCG caa GCUGACGGUACAGGCC	control	3	3	1	0	0
s-BIVTAR	ggGCUCG**UG**UA**G**CUCAUUA**GC**U**C**CGAGCcc	control	0	0	0	0	4
1-i-43A44U76A77A	GG**AU**UGGGCGCAGCGUCAAUGACGCUGACGGUACA**AA**CC	WT	5	3	0	0	0
2-r-45G75U	GGUC**G**GGGCGCAGCGUCAAUGACGCUGACGGUAC**U**GGCC	GV	0	5	4	0	0
3-i-48A50A69A71A	GGUCUGG**A**C**A**CAGCGUCAAUGACGCUGA**A**G**A**UACAGGCC	WT	4	1	1	0	0
4-r-48A50U71A	GGUCUGG**A**C**U**CAGCGUCAAUGACGCUGACG**A**UACAGGCC	WT	5	0	0	0	0
5-r-48A69G71A	GGUCUGG**A**CGCAGCGUCAAUGACGCUGA**G**G**A**UACAGGCC	WT	4	1	1	0	0
6-s-48A71A	GGUcUGG**AC**GCAGCGUCAAUGACGCUGAC**GA**UACAGGCC	Synth	4	5	3	3	0
7-i-48A71A72G	GGUCUGG**A**CGCAGCGUCAAUGACGCUGACG**AG**ACAGGCC	WT&GV	4	3	1	0	0
8-s-48U	GGUcUGG**U**CGCAGCGUCAAUGACGCUGACGGUACAGGCC	Synth	4	0	0	0	0
9-i-48U49G50A69U	GGUCUGG**UGA**CAGCGUCAAUGACGCUGA**A**GGUACAGGCC	WT	4	4	2	0	0
10-r-48U50A69A	GGUcUGG**U**C**A**CAGCGUCAAUGACGCUGA**A**GGUACAGGCC	WT	5	0	3	0	0
11-i-49U50U67A68U69U	GGUCUGGG**UU**CAGCGUCAAUGACGCU**AUU**GGUACAGGCC	GV	5	4	1	0	0
12-i-50A51G66A69U	GGUCUGGGC**AGU**GCGUCAAUGACGC**A**GA**U**GGUACAGGCC	GV	1	4	3	1	0
13-r-50A51G67U68U69U	GGUcUGGGC**AG**AGCGUCAAUGACGCU**UUU**GGUACAGGCC	GV	3	3	0	0	0
14-i-50A51U67A69A	GGUCUGGGC**AU**AGCGUCAAUGACGCU**A**A**A**GGUACAGGCC	WT	5	3	3	1	0
15-s-50A69A	GGUcUGGGC**AC**AGCGUCAAUGACGCUGA**A**GGUACAGGCC	Synth	4	0	1	1	0
16-i-50A69U	GGUCUGGGC**A**CAGCGUCAAUGACGCUGA**U**GGUACAGGCC	WT&GV	4	4	1	0	0
17-i-51A52U53A65G66A	GGUCUGGGCG**AUU**CGUCAAUGACG**GA**GACGGUACAGGCC	WT	4	3	2	0	0
18-i-51A52U53A65U66A	GGUCUGGGCG**AUA**CGUCAAUGACG**UA**GACGGUACAGGCC	WT&GV	4	4	3	1	0
19-i-51G52G53U65G67C	GGUCUGGGCG**GGU**CGUCAAUGACG**G**U**C**ACGGUACAGGCC	GV	0	4	3	0	0
20-r-51G52G67U	GGUCUGGGCG**GG**GCGUCAAUGACGCU**U**ACGGUACAGGCC	GV	1	4	2	0	0
21-i-51G52U53U 65A66G67U	GGUCUGGGCG**GUU**CGUCAAUGACG**AGU**ACGGUACAGGCC	GV	0	4	3	0	0
22-i-51G52U66A67U	GGUCUGGGCG**GU**GCGUCAAUGACGC**AU**ACGGUACAGGCC	GV	4	4	3	0	0

23-r-53U54U65G	GGUCUGGGCGCA**UU**GUCAAUGACG**G**UGACGGUACAGGCC	WT	4	4	3	2	0
24-r-69G	GGUCUGGGCGCAGCGUCAAUGACGCUGA**G**GGUACAGGCC	WT	3	0	0	0	0
25-48C72G73G74A	GGUCUGG**C**CGCAGCGUCAAUGACGCUGACGG**GGA**AGGCC	WT	3	0	0	0	0
26-48U49U50A69A	GGUCUGG**UUA**CAGCGUCAAUGACGCUGA**A**GGUACAGGCC	WT	3	0	0	0	0
27-50A51A52C66G67U68U69U	GGUCUGGGC**AAC**GCGUCAAUGACGC**GUGU**GGUACAGGCC	GV	0	4	2	0	0
28-51A52U53U65A66A	GGUCUGGGCG**AUU**CGUCAAUGACG**AA**GACGGUACAGGCC	WT	4	4	3	0	0
29-51G52U53A65U66A67A	GGUCUGGGCG**GUA**CGUCAAUGACG**UAA**ACGGUACAGGCC	WT	3	4	2	1	0
30-51G52U53A65U66A67U	GGUCUGGGCG**GUA**CGUCAAUGACG**UAU**ACGGUACAGGCC	GV	1	3	3	0	0
31-53A54G64U65U	GGUCUGGGCGCA**AG**GUCAAUGAC**UU**UGACGGUACAGGCC	WT	4	4	2	1	0
32-54A55A56A57A62C64A	GGUCUGGGCGCAG**AAAA**AAUG**C**C**A**CUGACGGUACAGGCC	WT	4	4	2	1	0
33-54A55A56G57A60G61A62G63U64U	GGUCUGGGCGCAG**AAGA**AA**GAGUU**CUGACGGUACAGGCC	WT	5	4	3	2	0
34-54A55A57G58G59Δ64A	GGUCUGGGCGCAG**AA**U**GG**_UGAC**A**CUGACGGUACAGGCC	WT	4	4	3	2	0
35-54A55U57Δ60G62G63A64U	GGUCUGGGCGCAG**AU**U_AA**G**G**GAU**CUGACGGUACAGGCC	GV	4	4	3	1	0
36-54A57A59U61U63U64U	GGUCUGGGCGCAG**A**GU**A**A**U**U**A**U**U**CUGACGGUACAGGCC	GV	4	4	4	1	0
37-54G55U56C57U59U60G64A	GGUCUGGGCGCAG**GUCU**A**UG**GAC**A**CUGACGGUACAGGCC	WT	5	4	3	1	0
38-54G57G59C60G61C64U	GGUCUGGGCGCAG**G**GU**G**A**CGC**AC**U**CUGACGGUACAGGCC	WT	4	4	3	1	0
39-54U55A56G57A60A61C62C63U64A	GGUCUGGGCGCAG**UAGA**AA**ACCUA**CUGACGGUACAGGCC	WT	5	4	3	1	0
40-54U55A56C57G58U60C62G63G64G	GGUCUGGGCGCAG**UACGU**A**C**G**GGA**CUGACGGUACAGGCC	WT	4	4	3	2	0
41-54U55C56A58G59U60A61U62U63G	GGUCUGGGCGCAG**UCA**C**GUAUUG**GCUGACGGUACAGGCC	WT	4	3	3	1	0
42-54U55U56A67A58U59C61U62U63A64A	GGUCUGGGCGCAG**UUAAUC**U**UUAA**CUGACGGUACAGGCC	GV	4	4	3	2	0
43-55A56A57A58U60G61A63U	GGUCUGGGCGCAGC**AAAU**A**GA**A**U**GCUGACGGUACAGGCC	GV	4	4	3	1	0

[a] RRE mutants are indicated by a unique sequential number, the prefix i-, r-, or s- indicating the clone's origin as isolated, reconstructed, or synthetic, respectively; and a description of mutant residues by numerical position in RRE and base identity or absence (Δ).

[b] The nucleotide sequences of RNAs are shown. Gaps in NMR RRE are to align the internal loop to RRE IIB. RRE mutants have nucleotides with those differing from RRE IIB bold and underlined.

[c] The origin of sequences are described by the selections from which RRE mutants were isolated: either wild-type Rev ARM (WT) or R35G-N40V (GV). Synthetic constructs are indicated (Synth).

[d] RRE-ARM X-gal colony color assays were performed as described in material and methods. Numbers represent the intensity of the blue in colonies by comparison: the wild-type Rev ARM-RRE IIB interaction is scored as 4+, the wild-type Rev ARM-NMR RRE interaction is 3+, and non-cognate interactions are 0 (background white).

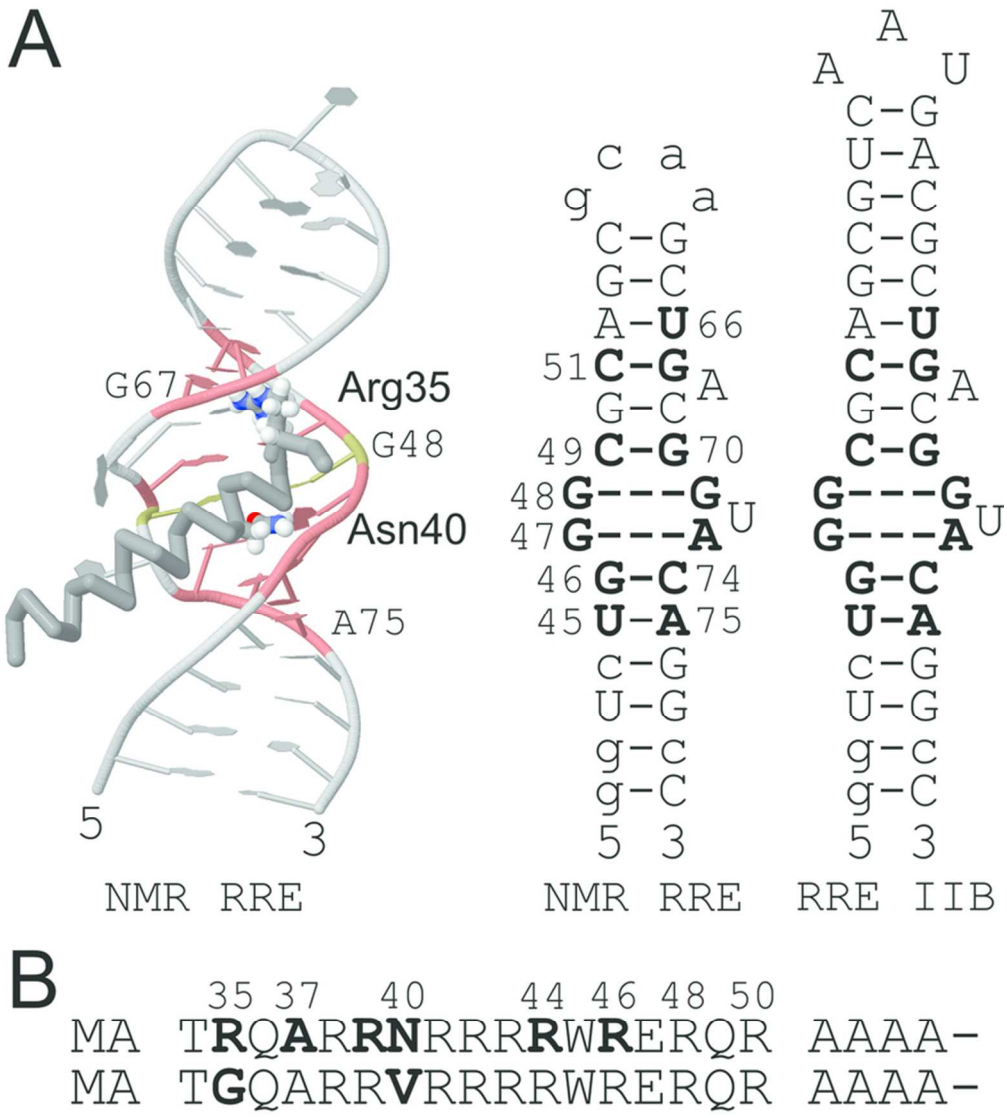

Figure 1. RREs and Rev ARMs: A) Left, the average HIV-1 Rev ARM-RRE IIB NMR structure of Battiste et al. (1996; Protein Data Bank 1ETF) in which the Rev ARM backbone is gray and the RNA is a light gray cartoon. Arg35 and Asn40 residues are shown as CPK-colored sticks. Important nucleotides are shown colored salmon, and groove-widening G48:G71 pair is olive. Middle, the secondary structure of the apically truncated stem IIB of RRE with a GNRA tetraloop that was used in NMR structural studies, with nucleotides numbered according to Battiste et al. (1996). Nucleotides found important in previous studies are bold. Mutant bases are shown as small letters. Right, the secondary structure of the RRE IIB used in this study is shown. B) The wild-type Rev ARM (amino acids 34-50) fused to the activation domain of λ N is shown above with residues found important for RRE binding by mutagenesis bold (Tan et al., 1993; Tan and Frankel, 1994; Possik et al., 2013) and the R35G-N40V mutant ARM sequence below. The numbering is that of HIV-1 Rev protein. The amino-terminus and carboxy linker used in the N-fusion are shown separated from the ARM with spaces.
74x82mm (300 x 300 DPI)

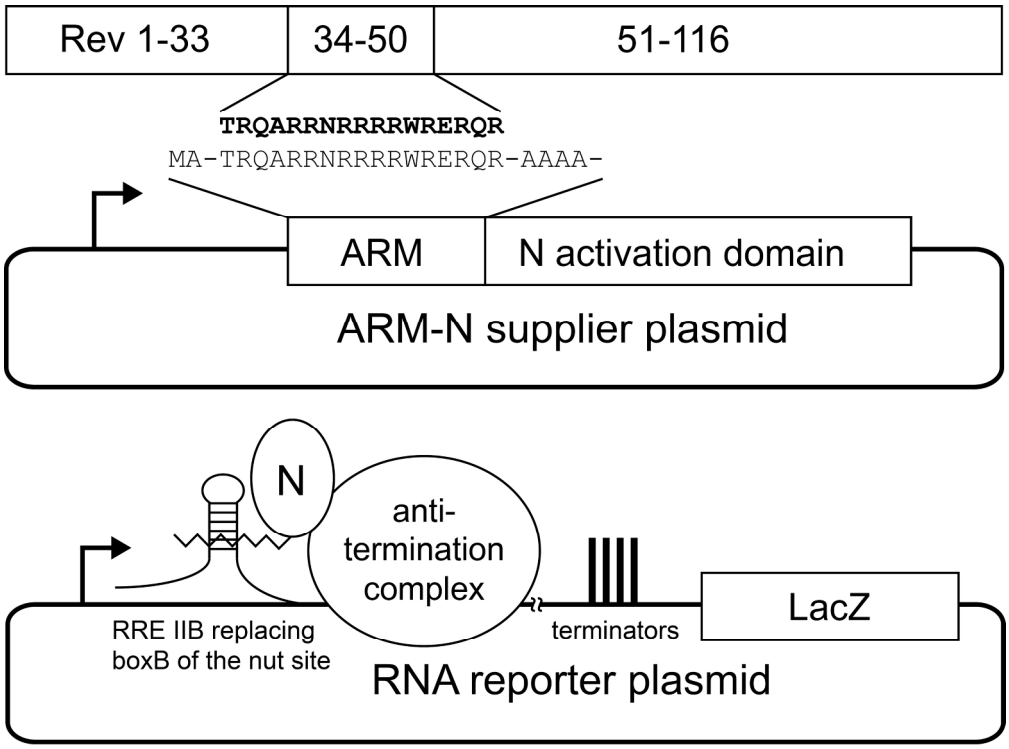

Figure 2. ARM-RNA reporter system. Above, HIV-1 Rev protein ARM, amino acids 34 to 50, is expressed as a fusion to the activation domain of λ N by the wild-type Rev ARM-N supplier plasmid. Plasmids with RSG1.2 and BIV Tat ARM are analogous. Below, the RNA reporter plasmid expresses a transcript containing a λ nut (N-utilization) site in which the boxB RNA hairpin is replaced with RRE IIB, libraries, mutants, or other RNAs such as BIV TAR. The RNA reporter expresses β-galactosidase (LacZ) downstream of intrinsic transcriptional terminators. LacZ expression is dependent on the N-fusion binding the RNA structure at boxB and recruiting host factors that cause the transcription complex to become antiterminating by ignoring downstream terminators. Competent cells hosting ARM-N suppliers are prepared and transformed with RNA reporter plasmids. LacZ expression in colonies can be monitored by X-gal on solid media and by ONPG in cell extracts. The BIV Tat-TAR interaction served as a heterologous control for specificity.
112x82mm (600 x 600 DPI)

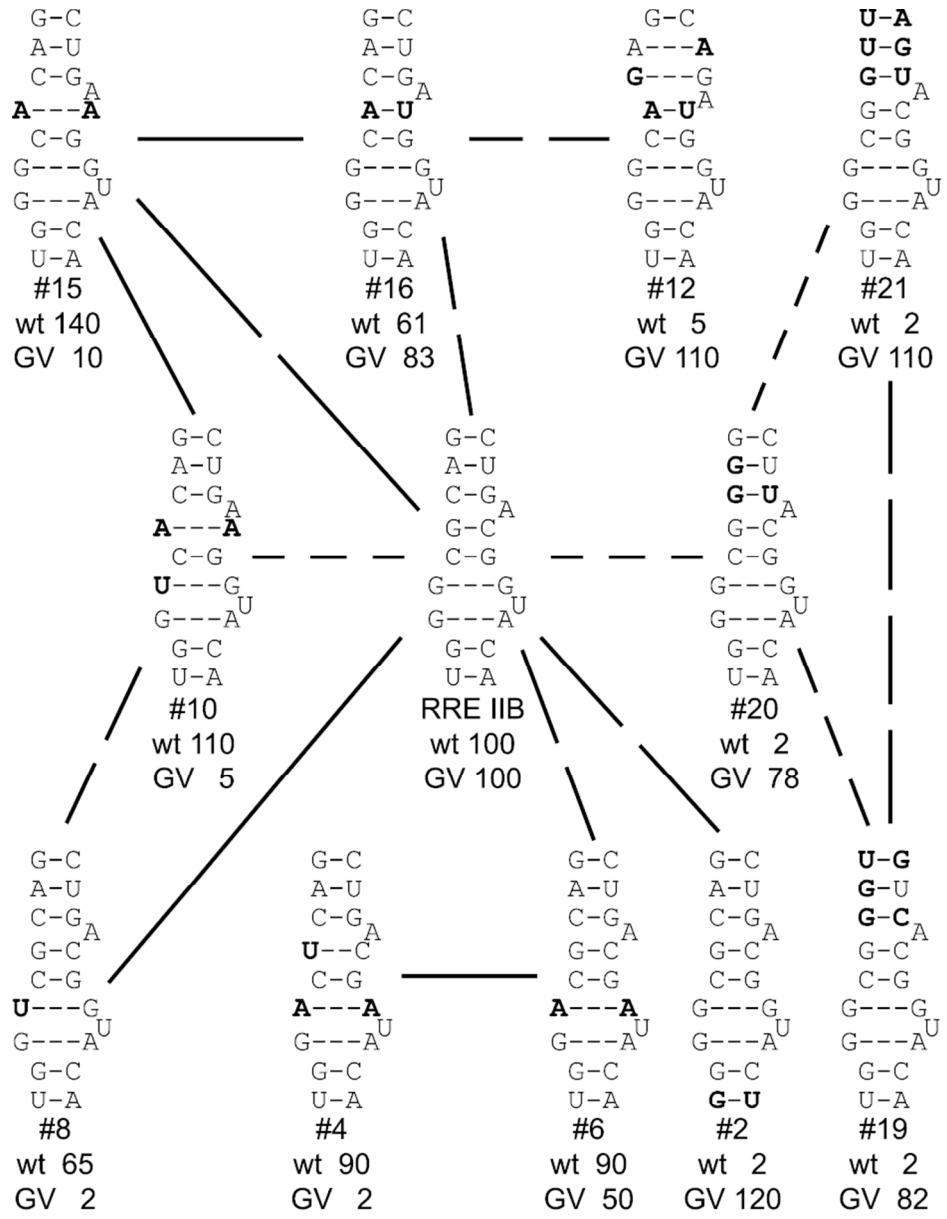

Figure 3. Paths between specific RRE mutants. RRE IIB and RRE mutants are shown from basal U45:A75 to apical G53:C65 with activities on wild-type Rev ARM (wt) and R35G-N40V (GV) as percent activity relative to RRE IIB. Lines between RNAs represent one, two, or more substitutions by number of segments. One substitution links RRE IIB to wild-type Rev ARM-specific mutant #8, and two substitutions link RRE IIB to R35G-N40V-specific mutant #2.
79x102mm (300 x 300 DPI)