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

ARTICLE *in* JOURNAL OF MOLECULAR RECOGNITION · JUNE 2013

Impact Factor: 2.15 · DOI: 10.1002/jmr.2272 · Source: PubMed

4 AUTHORS, INCLUDING:

Elite Possik
McGill University
7 PUBLICATIONS **46** CITATIONS
SEE PROFILE

Maroun S Bou Sleiman
École Polytechnique Fédérale de Lausanne
6 PUBLICATIONS **14** CITATIONS
SEE PROFILE

Colin Smith
American University of Beirut
36 PUBLICATIONS **912** CITATIONS
SEE PROFILE

Research Article

Journal of Molecular Recognition

Received: 12 January 2013, Revised: 9 February 2013, Accepted: 10 February 2013, Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jmr.2272

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

Elite J. Possik[a], Maroun S. Bou Sleiman[b], Ingrid R. Ghattas[c] and Colin A. Smith[c]*

The binding of the arginine-rich motif (ARM) of HIV Rev protein to its high-affinity site in stem IIB in the Rev response element (RRE) initiates assembly of a ribonucleoprotein complex that mediates the export of essential, incompletely spliced viral transcripts. Many biochemical, genetic, and structural studies of Rev–RRE IIB have been published, yet the roles of many peptide residues in Rev ARM are unconfirmed by mutagenesis. Rev aptamer I (RAI) is an optimized RRE IIB that binds Rev with higher affinity and for which mutational data are sparse. Randomized-codon libraries of Rev ARM were assayed for their ability to bind RRE IIB and RAI using a bacterial reporter system based on bacteriophage λ N-nut antitermination. Most Rev ARM residues tolerated substitutions without strong loss of binding to RRE IIB, and all except arginine 39 tolerated substitution without strong loss of binding to RAI. The pattern of critical Rev residues is not the same for RRE IIB and RAI, suggesting important differences between the interactions. The results support and aid the interpretation of existing structural models. Observed clinical variation is consistent with additional constraints on Rev mutation. By chance, we found double mutants of two highly critical residues, arginine 35 (to glycine) and asparagine 40 (to valine or lysine), that bind RRE IIB well, but not RAI. That an apparently distinct binding mode occurs with only two mutations highlights the ability of ARMs to evolve new recognition strategies and supports the application of neutral theories of evolution to protein–RNA recognition.

Supporting information may be found in the online version of this paper

Keywords: HIV rev; Rev-response element; Rev aptamer I; Protein–RNA recognition; arginine-rich motif; neutral evolution

INTRODUCTION

The HIV regulates the export of its singly and unspliced mRNAs using the interaction between the virally encoded trans-activator protein, Rev, and a *cis*-acting, ~350 nucleotide Rev response element (RRE), located in the *env* gene (Malim *et al*., 1989; Zapp and Green, 1989; Heaphy *et al*., 1990). The 116 amino acid Rev protein includes an arginine-rich motif (ARM) that binds to a high affinity site in stem IIB of RRE (Figure 1). A 17-mer Rev peptide comprising residues 34–50 adopts an α-helical conformation and binds to RRE IIB with high affinity and specificity *in vitro* (Kjems *et al*., 1992; Tan *et al*., 1993; Battiste *et al*., 1996). The Rev ARM binding site in RRE IIB includes an asymmetric internal loop with two purine : purine base pairs (Hope *et al*., 1990; Bartel *et al*., 1991; Kjems *et al*., 1992). Substitutions of Arg35, Arg39, Asn40, and Arg44 to alanine and lysine (Tan *et al*., 1993; Tan *et al*., 1994) support NMR-based structural models of Rev–RRE IIB (Battiste *et al*., 1996) and Rev–Rev aptamer I (RAI) (Ye *et al*., 1996), in which these residues make base-specific contacts in the major groove (Figure 2). Although Trp45 is not important for binding *in vitro*, it is required for essential Rev oligomerization *in vivo* (Zapp *et al*., 1991; Kjems *et al*., 1992). Despite intense investigation of the Rev–RRE IIB interaction, mutational data are limited, and precise roles of specific ARM residues are uncertain.

Rev aptamer I (Figure 1) was selected from a random RNA library on the basis of a minimal RRE IIB (Giver *et al*., 1993). RAI shares substantial identity with RRE IIB and adopts a very similar structure (Figure 2). The structural models show that the Rev peptide binds RAI similarly to RRE IIB, although it has unique aspects including a U:A:U base triple in the Rev binding site (Ye *et al*., 1996). RAI can replace RRE IIB functionally *in vivo* (Symensma *et al*., 1996). The Rev–RAI interaction has been examined far less than that of Rev–RRE IIB, and its tolerance to Rev–ARM substitutions is unknown.

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

a *E. J. Possik*
McGill Cancer Centre and Department of Biochemistry, McGill University, Montreal, Quebec, Canada

b *M. S. Bou Sleiman*
Doctoral Program in Molecular Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

c *I. R. Ghattas, C. A. Smith*
Department of Biology, American University of Beirut, Beirut, Lebanon

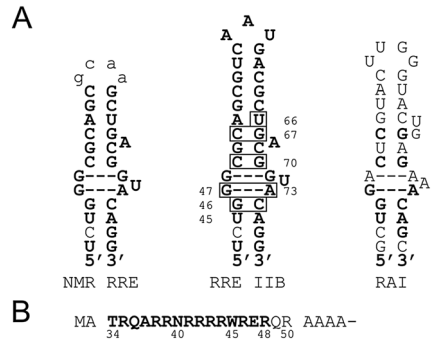

Figure 1. Rev-binding RNAs and Rev peptide. (A) Sequence and secondary structures of the high-affinity site in RRE used in structural studies (NMR RRE), the longer site used in this study (RRE IIB), and the Rev aptamer I used in structural studies (RAI). Wild-type nucleotides are bold, and the added GNRA tetraloop in NMR RRE is in lowercase. Boxed residues in RRE IIB have been identified as important by biochemical and mutational studies. Numbered RRE IIB bases contact Rev. Numbers are standard to RRE studies and refer to placement in the entire RRE element (Malim *et al*., 1989). (B) The sequence of the Rev peptide fusion to the N activation domain is shown with HIV Rev residues numbered. Randomized codon libraries were made of residues 34–48, shown in bold.

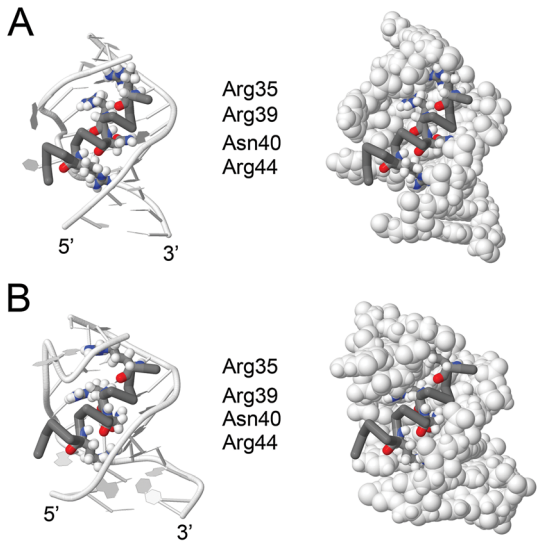

Figure 2. Structural models of Rev–RRE IIB and Rev–RAI. NMR models of (A) HIV Rev–RRE IIB (1ETF) and (B) Rev–RAI (1ULL) are shown to the right with RNA white cartoon, peptide backbone gray wireframe, and side chains of critical residues Arg35, Arg39, Asn40, and Arg44 colored by atom type. Views and renderings to the right are identical, except RNAs are rendered at van der Waals radii. Critical side-chain labels are placed level to their locations.

Arginine-rich motifs binding small RNA structures occur in important biological systems, including gene expression and its regulation (Chen and Varani, 2005). Because they are small, structurally diverse, and occur in important systems, they serve as model systems to understand RNA–protein recognition. (Draper, 1999). In addition to HIV Rev, examples of α-helical ARMs include bacteriophage λ N (Legault *et al*., 1998; Schärpf *et al*., 2000) and its relatives P22 N (Cai *et al*., 1998) and ϕ21 N (Cilley and Williamson, 2003) that recognize small, boxB stem-loop RNAs during the regulation of transcriptional antitermination. Important non-α-helical examples include lentiviral Tat proteins that bind their small, stem-loop TAR RNAs (transactivating response elements) during the regulation of transcriptional elongation (Ott *et al*., 2011). HIV Tat binds its target in an extended conformation (Long and Crothers, 1995), and its relative bovine immunodeficiency virus Tat (BIV Tat) binds its target in a β-turn conformation (Puglisi *et al*., 1995).

In addition to RRE IIB and its variant RAI, unrelated Rev-binding RNAs (Xu and Ellington, 1996; Ye *et al*., 1999) and RRE IIB-binding peptides (Harada *et al*., 1996; Harada *et al*., 1997; Peled-Zehavi *et al*., 2003; Tan and Frankel, 1998) have been isolated by artificial selection. The occurrence of diverse and specific recognition strategies raises questions as to how new recognition strategies evolve: what mutations can occur without loss of function, how many different specific interactions might exist, and what is the nature of evolutionary intermediates? Interestingly, the conformational variability of proteins has been recently argued to facilitate evolvability (Tokuriki and Tawfik, 2009).

Neutral theories of evolution (Kimura, 1991; Ohta, 2002) assert that most observed genetic change is the result of the random fixation of selectively neutral mutations, rather than the result of positive selection. When applied to molecular recognition, these theories predict that sequences are sufficiently mutable that new recognition strategies can evolve by random genetic drift and relaxed- and altered-specificity intermediates. Studies of arginine-rich peptide–RNA interactions, including Rev–RRE relatives (Iwazaki *et al*., 2005; Sugaya *et al*., 2008a, Sugaya *et al*., 2008b), Tat–TAR (Smith *et al*., 1998, Smith *et al*., 2000), and P22N–boxB (Cocozaki *et al*., 2008a; Cocozaki *et al*., 2008b), support these predictions by finding altered-specificity variants. Neutral theories would predict that Rev ARM is sufficiently mutable such that distinct recognition strategies are accessible by incremental changes without loss-of-fitness intermediates.

To improve understanding of the Rev–RRE IIB and Rev–RAI interactions and to explore how well neutral theories apply to RNA–protein interactions, we constructed randomized-codon libraries of the Rev ARM and assayed them for their ability to bind RRE IIB and RAI using a bacterial reporter system (Figure 3) that recapitulates bacteriophage λ antitermination in *Escherichia coli* (Franklin, 1993). Arg39, Asn40, and Arg44 appear immutable with the RRE IIB reporter. Arg35 can be substituted with the conservative basic amino acid lysine with only moderate loss of activity. Most residues are mutable, reflecting their less important role in recognition. In contrast to RRE IIB, only Arg39 appears immutable with the RAI reporter, suggesting differences between RRE IIB and RAI recognition strategies. The existing structural models are interpreted in light of results. Clinical variation was compared with results and is consistent with the existence of additional constraints on Rev mutation *in vivo*. Intriguingly, we found double mutants at two highly critical residues, Arg35 (to glycine) and Asn40 (to valine or lysine), that bind RRE IIB well, but not RAI. That an apparently distinct binding mode occurs with only two mutations highlights the ability of ARMs to evolve new recognition strategies and supports the application of neutral theories of molecular evolution to small peptide–RNA recognition.

MATERIALS AND METHODS

General

Restriction enzymes and T4 DNA ligase were obtained from Roche (Mannheim, Germany). Bacterial media components were

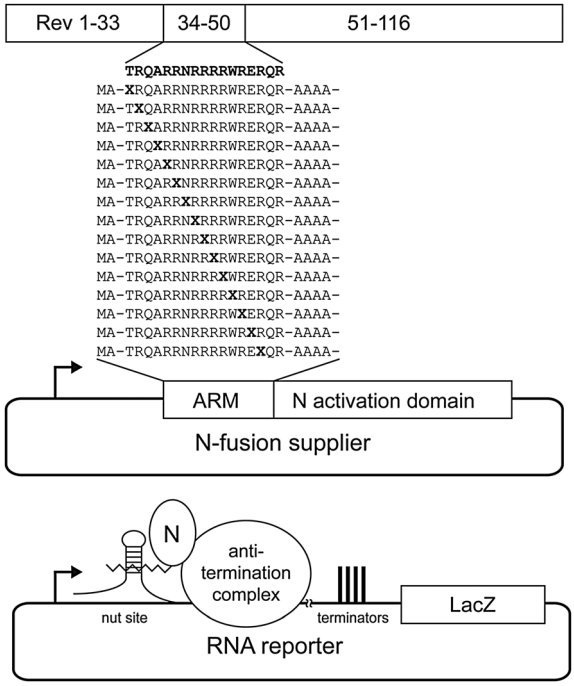

Figure 3. Experimental approach. The ARM of HIV Rev (residues 34–50) replaces the ARM of λN such that Rev–ARM is fused to the activation domain of λN and flanked by methionine–alanine at the amino terminus and 4 alanines at the fusion junction. Oligonucleotide libraries, coding for Rev–ARM sequences in which codons of residues 34–48 are individually randomized, were inserted into N-fusion supplier plasmids and screened for functional members by transformation into *E. coli* hosting RNA reporter plasmids. When the N-fusion protein containing a Rev–ARM mutant binds RRE IIB or RAI in the *nut* site of the reporter transcript, the N activation domain recruits host factors to assemble an anti-termination complex causing downstream transcriptional terminators to be ignored and expression of β-galactosidase from the LacZ reporter gene. β-Galactosidase activity can be monitored colorimetrically in colonies with X-gal plates and quantified with *ortho*-nitrophenol-β-D-galactoside in extracts of cultures. The BIV Tat–TAR interaction placed in this reporter system serves as a heterologous control for specificity.

obtained from Oxoid (Basingstoke, Hampshire, UK). Fine chemicals were obtained from Amersham (Little Chalfont, Buckinghamshire, UK), Sigma (St. Louis, MO, USA), and Amresco (Solon, OH, USA). Laboratory chemicals were obtained from Acros (Geel, Belgium). Disposable plasticware was obtained from Sarstedt (Numbrecht, Germany).

Strains and plasmids

Escherichia coli supporting antitermination, N567 (Franklin and Doelling, 1989), pACnutTAT13, and pBRptac-N* (Franklin, 1993), were obtained from Naomi Franklin (University of Utah). Plasmids expressing the Rev14 N fusion (pBRN–HIVRev14), the BIV Tat N fusion (pBRN–BIVTat), and the NMR RRE reporter (pAC–HIV–RRE) (Harada *et al*., 1996) were obtained from Kazuo Harada (Tokyo Gakugei University).

Construction of libraries and mutants

Plasmids expressing the Rev17 N fusion and reporting on RRE IIB binding, RAI binding, and BIV TAR binding were constructed in-house. Rev17 and its libraries were cloned into pBRptac-N* as NcoI-BsmI fragments replacing the amino terminus of λ N (residues 1–19). Rev17 and library fragments were generated by mutually priming oligonucleotides followed by NcoI-BsmI digestion using oligonucleotide pairs based on the Rev17 coding strand 5′-CC ATG GCA ACC CGC CAG GCC CGT CGT AAC CGT AGA CGT CGT TGG CGT GAG CGT CAG CGT GCA GCT GCG GCG AAT GCA-3′. RRE IIB, RAI, and BIV TAR reporters were cloned into pACnutTAT13 replacing the λ left *nut* site oligonucleotides based on the coding strand sequences flanked by PstI and BamHI sites: 5′-CTG CAG TCG ACG CTC TTA AAA ATT AAG GTC TGG GCG CAG CGT CAA TGA CGC TGA CGG TAC AGG CCA GCA TTC AAA GCA GGG ATC C-3′ for RRE IIB, 5′-CTG CAG TCG ACG CTC TTA AAA ATT AAG CTC GTG TAG CTC ATT AGC TCC GAG CAG CAT TCA AAG CAG GGA TCC-3′ for RAI, and 5′-CTG CAG TCG ACG CTC TTA AAA ATT AAG CTC GTG TAG CTC ATT AGC TCC GAG CAG CAT TCA AAG CAG GGA TCC-3′ for BIV TAR. Standard molecular biology procedures were used, and all clones were tested by function. Entire synthetic inserts were confirmed by sequencing using PBRNR2 (5′-GGC TTG CTG TAC CAT GTG-3′) for N-fusion constructs and PACF2 (5′-AAT CAC TGC ATA ATT CGT GTC-3′) for RNA-reporter constructs.

Library screening with X-gal

Competent N567 host cells carrying reporter plasmids were transformed with the library or clones of interest and control plasmids. Approximately 10–100 ng of plasmid per 100 μl of competent cells were transformed by heat shock and plated on tryptone plates containing 50 μg/ml ampicillin and 15 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 and photographed after 1 day at 34°C and after 2 days at 24°C.

ARM–RNA recognition assays with *ortho*-nitrophenol-β-D-galactoside

For each interaction, from four to six representative colonies were picked from X-gal plates for solution assays. For measurement of ARM–RNA recognition via N-mediated antitermination, cultures were grown overnight at 30°C with aeration in tryptone with 50 μg/ml ampicillin and 15 μg/ml chloramphenicol as antibiotics, and with 0.05 mM IPTG. The cells were then permeabilized, assayed for β-galactosidase activity using *ortho*-nitrophenol-β-D-galactoside), and units of β-galactosidase were calculated following Miller (1992). Percent activities are reported normalized to the activity of Rev17 for RRE and RAI reporters and BIV Tat for the BIV TAR reporter of that day's experiment.

Structure visualization

Jmol (http://www.jmol.org/), an open-source Java viewer for chemical structures in 3D, was used to view NMR structures of Rev–RRE (Protein Data Bank 1ETF and 1ETG; Battiste *et al*., 1996) and Rev–RAI (Protein Data Bank 1ULL; Ye *et al*., 1996).

Journal of Molecular Recognition

RESULTS

Anti-termination assay reflects Rev–RRE IIB and Rev–RAI binding

A protein–RNA reporter system (Franklin, 1993) based on bacteriophage λ N-nut antitermination has been previously exploited for screening and assay of heterologous interactions, including Rev–RRE IIB and relatives (Harada *et al*., 1996; Harada *et al*., 1997; Iwazaki *et al*., 2005; Peled-Zehavi *et al*., 2003; Sugaya *et al*., 2008a. Sugaya *et al*., 2008b). The system consists of two plasmids transformed into *E. coli* (Figure 3). One plasmid is an N-fusion supplier plasmid in which the RNA-binding peptide of interest is fused to amino terminus of λ N protein, replacing that of N. The other plasmid is a boxB reporter plasmid, in which the RNA of interest replaces the boxB hairpin in the λ *nut* (N-utilization) site upstream from transcriptional terminators. Expression of the downstream β-galactosidase gene reflects binding of the peptide–RNA interaction by N-nut mediated antitermination. To choose the context for library screening, we first compared the activity of two different Rev sequences, Rev14 and Rev17 (Table 1), expressed as fusions with the activation domain of N, and three reporters, in which the Rev-binding RNA replaces boxB in the reporter transcript (Figure 1A). Rev14 is Rev 34–47, used previously (Harada *et al*., 1996; Peled-Zehavi *et al*., 2003). Rev17 is Rev34–50, also used previously (Battiste *et al*., 1996; Tan *et al*., 1993; Tan and Frankel, 1994). NMR RRE is the high affinity site in RRE IIB (C44-G53:C65-G76) truncated and capped by a tetraloop, previously used for structural studies and in previous screens (Harada *et al*., 1996, Harada *et al*. 1997). RRE IIB is RRE C44-G76 with the apical 9 nt of the wild-type stem loop intact. RAI is Rev Aptamer I, used for NMR studies (Ye *et al*., 1996). An N supplier and a nut reporter for the heterologous BIV Tat–TAR interaction served as a specificity control (Harada *et al*., 1996). On the basis of solution assays indicating increased activity with Rev17 and RRE IIB constructs without loss of specificity, they were chosen as the basis for library construction and screening, and they were used for all further experiments (Table 1).

Rev arginine-rich domain is robust to substitutions

A set of 15 plasmid libraries, spanning Rev Thr34 to Arg48, each expressing a Rev17-N fusion with one Rev ARM codon randomized, was constructed via cassette mutagenesis and transformed into reporter cells on plates containing the chromogenic substrate X-gal (Figure 3). The proportion of each library transformation appearing to have wild-type activity by deposition of blue pigment was estimated by eye (Figure 4). Computer-assisted quantification of RRE IIB plate images produced similar results to those estimated by eye (data not shown). The proportion of positives ranged from less than 1% for Asn40 to 50% for Arg36, Arg 41, and Glu47. The existing structural model of Rev–RRE posits essential roles for base-specific contacts by four residues: Arg35, Arg39, Asn40, and Arg44 (Battiste *et al*., 1996), and the results of Figure 2A accord with these roles, although the Arg35 library screen has more positives than the 6/64 (~9%) that would be expected for only arginine codons. Some libraries had more positives than was expected from published mutational analysis, including Thr34, Arg38, and Arg46 (Tan *et al*., 1993; Tan and Frankel, 1994). Mutation of Ala37 had been neglected previously, yet its library contained relatively few positives.

Screening the libraries with the RAI reporter revealed a different pattern, in which all libraries contain more positives than with the RRE IIB reporter, and only Arg39 appeared to be poorly tolerant of substitution. Notably, the two most restrictive Rev positions when screened against RRE IIB, Asn40 and Arg44, have far higher proportions of positives when screened against RAI, suggesting differences between Rev–RRE IIB and Rev–RAI interactions. No library had positives when screened against BIV TAR reporter cells.

Activities of single mutations support and refine existing structural models

To determine the effect of specific mutations, colonies displaying wild-type activity were selected by eye from each screen, grown individually, and pooled for plasmid preparations. The N-supplier plasmid was separated from the reporter plasmid by gel

Table 1. Antitermination activities of ARM–RNA interactions

		β-Gal units[c]				Percent activation[d]			
N-fusion[a]	Sequences[b]	NMR RRE[e]	RRE IIB[f]	RAI[g]	BTAR[h]	NMR RRE[e]	RRE IIB[f]	RAI[g]	BTAR[h]
Rev14[i]	-TRQARRNRRRRWRR-	46 ± 8	130 ± 30	470 ± 90	2 ± 0.4	46 ± 8	57 ± 13	101 ± 20	4.2 ± 0.7
Rev17[j]	-TRQARRNRRRRWRERQR-	100 ± 15	230 ± 30	460 ± 70	2 ± 0.6	100 ± 15	100 ± 14	100 ± 16	4 ± 1.1
BTat[k]	-GRPRGTRGKGRRIRR-	1 ± 0.07	2 ± 0.06	2 ± 0.4	60 ± 10	1.37 ± 0.07	0.7 ± 0.03	0.51 ± 0.09	100 ± 18

[a]Laboratory names of N-fusions expressed by pBRN plasmids.
[b]The sequences of ARM fused to N. See footnotes i, j, and k.
[c]Plasmids were transformed into reporter cells with the named RNA in place of λ *nut* site boxB. 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 *ortho*-nitrophenol-β-D-galactoside.
[d]Percent activation represents antitermination activities of N-fusions normalized to Rev17–NMR RRE, Rev17–RRE IIB Rev17–RAI, or BIV Tat–TAR.
[e]NMR RRE = 5′-GGUCUGGGCGCAGCGCAAGCUGACGGUACAGGCC-3′.
[f]RRE IIB = 5′-GGUCUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGCC-3′.
[g]RAI = 5′-GGCUGGACUCGGACUUCGGUACUGGAGAAACAGCC-3′.
[h]BTAR = 5′-GCUCGUGUAGCUCAUUAGCUCCGAGC.
[i]Rev14 = MA TRQARRNRRRRWRR AAAA.
[j]Rev17 = MA TRQARRNRRRRWRERQR AAAA.
[k]BTat = MG RPRGTRGKGRRIRR GGG.

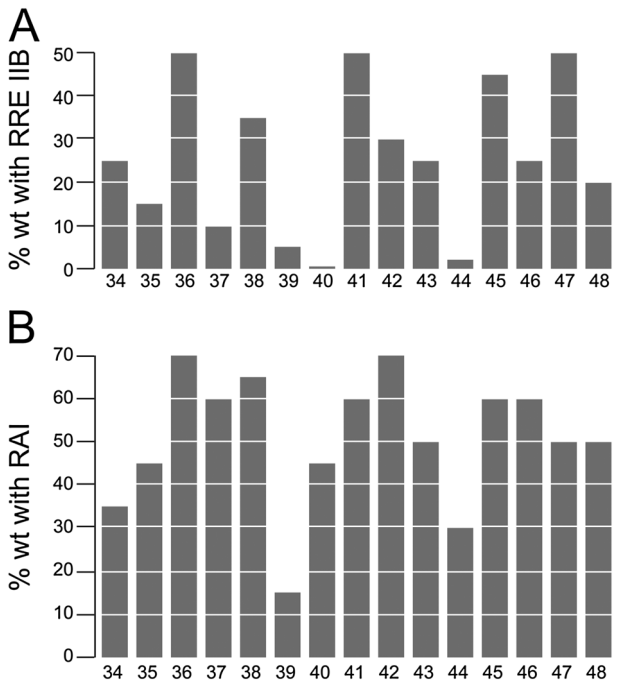

Figure 4. Rev library activities against RRE IIB and RAI. Estimated proportions of Rev library members displaying wild-type activity in X-gal plate screens with RRE IIB (A) and RAI (B) reporters are shown as histograms, by Rev residue.

electrophoresis and transformed into *E. coli* cells. Colonies were selected randomly and restreaked, and plasmids were prepared and individually tested against reporters. Clones were sequenced and measured for activity by solution assay against Rev, RAI, and BIV TAR reporters (Table 2).

As expected from the base-specific contacts observed in the Rev–RRE model (Battiste *et al*., 1996), published data on mutants (Tan *et al*., 1993; Tan and Frankel, 1994), and the proportion of active library members (Figure 4), no functional mutants of Arg39, Asn40, and Arg44 were found, and only lysine could replace Arg35, with moderately reduced activity (Table 2). The average Rev–RAI structural model has a similar core structure to Rev–RRE. Using mutational data of Rev–RRE to support the Rev–RAI structural model, Ye *et al*. (1996) posit base-specific contacts by Arg35, Arg39, Asn40, and Arg44. While screening the Rev libraries against RAI, a markedly higher proportion of active members was noted. The pattern of restrictive residues appears distinct: Arg39 is immutable with RAI, Arg44 intermediate, and Asn40 tolerant to substitution.

In the average Rev–RRE model, Arg39 is deep in the major groove within hydrogen-bonding distance of G70 (Figure 2).The Rev–RAI model shows hydrogen bonds between the Arg39 ε- and η-nitrogens to the equivalent of G70 O6 and G70 N7, respectively, and structure 3 of the Rev–RRE PDB file shows a similar orientation. We recovered only arginine from RRE IIB and RAI screens. Although the proportion of positives in the Arg39 library was consistent with no mutation being functional, we prepared the R39K construct to determine whether the positive charge or the guanidium group of Arg39 was important. The poor activity of R39K with both RRE IIB and RAI reporters (Table 2) strongly supports the proposed archetypal arginine–guanine interaction between Arg39 and G70 being responsible for the critical role of Arg39.

In the Rev–RRE model, Asn40 is within hydrogen-bonding distance and nearly co-planar with the G47:A73 base pair (Figure 2). The Rev–RAI average structure shows Asn40's carboxamide donating a hydrogen bond from its nitrogen to the O6 of the G47 equivalent and accepting a hydrogen bond from its oxygen to N6 of the A73 equivalent. The Asn40 library had the lowest proportion of active mutants, and no active single substitution was found in screens against the RRE IIB reporter. Substitutions with arginine, leucine, valine, and glycine had only background activity. Glutamine was not recovered, despite sequencing 16 clones, strongly supporting the critical role of Asn40 hydrogen bonds to G47:A73 in the Rev–RRE interaction.

The Rev–RRE model positions Arg44 in the major groove close to U45 and G46, although the side chain is not well resolved. The Rev–RAI model is similarly resolved and suggests hydrogen bonds to the equivalent U45 O4 and G46, although the published analysis only discusses the interaction with the U45 (Ye *et al*., 1996). Individual structures of both interactions suggest hydrogen bonding to either or both U45 O4 and G46 O6, although the contact to U45 O4 is most frequent. We observed a low proportion of active library members and recovered only arginine from Arg44-RRE IIB screens and found R44K to have moderate activity with RRE IIB. The results support the critical role of Arg44 as donating hydrogen bonds to U45.

The Rev–RRE model has Arg35 within hydrogen bonding distance of U66 and G67. Individual Rev–RRE structures and the average Rev–RAI structure suggest η-nitrogens of Arg35 donate hydrogen bonds to O6 and N7 of the equivalent to G67. Individual Rev–RRE structures infrequently suggest an Arg35 hydrogen bond to O4 of U66. The RAI equivalent of U66 is a cytosine lacking an O4, preventing the Rev–RAI interaction from supporting the Arg35–U66 interaction. The observation that more than 10% of library members were active in RRE IIB screens suggested functional substitutions existed, and indeed, R35K was recovered from screens with RRE IIB, although with moderately reduced activity. Interestingly, R35S showed activity well above background with the RRE IIB reporter, consistent with a critical role for the positive charge of Arg35 and the donation of at least one hydrogen bond to G67.

In contrast to literature reports that Thr34 was important (Tan *et al*., 1993) and the Rev–RRE and Rev–RAI models suggesting important contact between the Thr34 hydroxyl to the 5′-phosphate of G47, we observed a proportion of the library active with RRE IIB and moderate to high activity of smaller amino acids, including glycine and proline (Table 2). Interestingly, most individual Rev–RAI structures and some Rev–RRE structures suggest the backbone nitrogen of the terminal Thr34 making the phosphate contact. The recovery of T34N and T34G from screening and the relatively high activity of these mutants, including T34P, conflicts with published gel shift data showing T34A had low affinity (Tan *et al*., 1993). The T34A mutant was constructed in hopes of resolving this discrepancy, but it too showed more than 50% activity. Our data suggest that the hydroxyl contact is not important, although the Thr34 amide could be, and that steric occlusion of the side chain limits Thr34 mutation to smaller residues. Why the activity of the T34A mutant might differ between our system and a gel shift assay is not clear. The effect of chemical modification of the G47 phosphate (Kjems *et al*., 1992) could be a reflection of steric crowding rather than a hydrogen bond with Thr34. The activity of several small residues unable to make productive contacts with phosphate, especially T34G, suggests that either the threonine side chain is not making a hydrogen bond to G47 phosphate or it contributes little to binding.

Journal of Molecular Recognition

Table 2. Activities of single Rev substitutions

Clone[a]	Rev17 N-fusion[b]	Percent activation[c] RRE IIB	RAI	BTAR
II820-2	T34S	114 ± 15	148 ± 26	2.9 ± 0.4
V635-1	T34N	68 ± 7	129 ± 28	2.5 ± 0.4
V658-2	T34G	93 ± 9	158 ± 39	1.8 ± 0.4
II752-1	T34P	66 ± 7	134 ± 30	3.1 ± 0.4
V635-2	T34L	38 ± 2	117 ± 25	2.2 ± 0.5
II922-3[d]	T34A	66 ± 9	141 ± 16	2 ± 0.5
V635-3	R35K	49 ± 10	33 ± 6	1.8 ± 0.3
V715-1	R35S	17 ± 5	70 ± 13	5.4 ± 0.8
V696-1	R35G	5.5 ± 0.5	16 ± 4	5 ± 0.4
V635-5	Q36S	308 ± 42	172 ± 21	1.7 ± 0.11
II750-2	Q36L	70 ± 4	124 ± 23	3.9 ± 0.4
V658-6	A37S	48 ± 7	163 ± 15	1.8 ± 0.5
II778-2	A37C	2.3 ± 0.1	74 ± 6	2.0 ± 0.5
II778-1	A37L	2.0 ± 0.2	64 ± 9	2.8 ± 0.4
II778-5	A37V	5.3 ± 0.4	43 ± 8	2.2 ± 0.6
II752-10	R38K	74 ± 15	208 ± 35	2.7 ± 0.5
V658-10	R38L	52 ± 20	157 ± 24	1.0 ± 0.2
II752-9	R38E	29 ± 2	127 ± 27	2.5 ± 0.4
II751-2	R38G	4.1 ± 0.3	90 ± 8	2.0 ± 0.2
II922-1[d]	R39K	2.2 ± 0.4	0.73 ± 0.25	2 ± 0.2
V701-1	N40L	0.8 ± 0.3	119 ± 21	2.2 ± 0.2
V701-7	N40R	1.2 ± 0.2	58 ± 13	6.5 ± 0.6
V696-10	N40G	0.68 ± 0.2	41 ± 6	5 ± 1.0
II945-1[d]	N40V	0.72 ± 0.14	12 ± 3	1.9 ± 0.2
II751-4	R41L	46 ± 5	146 ± 40	1.5 ± 0.3
752-14	R41I	24 ± 4	147 ± 25	1.4 ± 0.4
V635-15	R41S	28 ± 2	205 ± 68	1.6 ± 0.3
II909-2[d]	R41K	82 ± 6	135 ± 36	2 ± 0.2
V635-18	R42L	80 ± 13	189 ± 29	1.6 ± 0.2
II747-18	R42M	64 ± 6	187 ± 12	2.3 ± 0.06
V635-17	R42A	60 ± 7	199 ± 55	1.6 ± 0.11
V658-17	R42F	40 ± 6	128 ± 36	1.7 ± 0.11
II752-16	R42S	11 ± 6	150 ± 22	1.7 ± 0.3
II747-20	R43K	55 ± 10	92 ± 25	2.7 ± 0.9
V635-20	R43S	63 ± 5	137 ± 25	2.0 ± 0.1
V696-13	R43T	14 ± 2	92 ± 12	3.0 ± 2
V715-5	R43L	3 ± 0.7	42 ± 4	3.0 ± 0.9
II755-5	R44K	0.88 ± 0.11	44 ± 13	1.5 ± 0.4
V635-23	W45Q	117 ± 22	134 ± 26	2.9 ± 0.6
II755-6	W45L	103 ± 20	108 ± 28	2.3 ± 0.6
II747-24	W45N	127 ± 31	118 ± 9	1.6 ± 0.14
II752-20	W45H	80 ± 6	115 ± 27	2.3 ± 1
V658-23	W45A	101 ± 11	117 ± 17	1.6 ± 0.2
II752-19	W45T	81 ± 13	103 ± 13	2.6 ± 0.5
V635-24	W45S	71 ± 7	107 ± 15	2.3 ± 0.6
V635-25	R46K	24 ± 2	50 ± 9	3.2 ± 0.9
V635-26	R46L	10 ± 1	117 ± 20	2.5 ± 0.9
II751-6	R46A	5 ± 1	109 ± 9	2.9 ± 0.8
II755-7	R46S	4 ± 1	91 ± 22	1.9 ± 0.1
V696-15	E47V	136 ± 15	101 ± 17	1 ± 0.1
V635-28	E47M	142 ± 37	96 ± 24	2.8 ± 0.5
V696-14	E47R	150 ± 22	107 ± 16	4.0 ± 1.5
V696-19	E47C	134 ± 10	95 ± 21	2.1 ± 0.2
V635-27	E47N	63 ± 13	102 ± 26	2.5 ± 0.7
II752-24	E47S	120 ± 27	117 ± 14	3.1 ± 1.1
II752-23	E47I	78 ± 7	58 ± 7	2.7 ± 0.6
V715-12	E47A	121 ± 31	56 ± 5	3.3 ± 1.5
V635-29	R48K	75 ± 12	77 ± 11	2.5 ± 1.3
V751-1	R48Q	21 ± 4	39 ± 12	1.9 ± 0.2
II755-9	R48H	50 ± 11	98 ± 16	2.1 ± 0.3
II747-30	R48A	35 ± 5	87 ± 9	2.1 ± 0.4

[a]Laboratory stock numbers of single mutant clones in Rev17 N-fusion.
[b]Rev mutants are indicated by single-letter codes for wild-type and mutant residues separated by the numerical position in Rev.
[c]Library plasmids were transformed into RRE IIB and BTAR reporter cells. At least four replicates of each clone were grown at 30°C overnight in tryptone medium supplemented with IPTG and assayed for β-galactosidase activity with *ortho*-nitrophenol-β-D-galactoside. Percent activation represents antitermination activities of Rev mutants normalized to Rev17–RRE, Rev17–RAI, or BIV Tat–TAR.
[d]Synthetic constructs.

Arg46 is suggested to stabilize the concentration of three phosphates near RRE U72 (Battiste *et al*., 1996). As a result of the alterations to the equivalent region of RAI caused by the formation of the A:U:A base triple, Arg46 is near only two phosphates. That our library had approximately a quarter of its members active with RRE IIB suggested that Arg46 may not be absolutely required, but we found no strongly active substitutions. R46K, which would be capable of favorable electrostatic interactions with phosphates, had only a quarter of the activity of wild-type Rev with RRE IIB. Arg46 appears relatively important, supporting its role in stabilizing the charge of three phosphates in the Rev–RRE IIB interaction, although it appears to contribute to recognition less than that of the residues contacting bases.

The Rev–RRE model shows Arg38 close to the phosphates 5′ to U66 and G67 and proposed to donate hydrogen bonds to them. From RRE IIB screens, we recovered R38K and R38L with only moderately reduced activity (Table 2). Why R38A and R38K severely decreased gel shift affinity (Tan *et al*., 1993) and reporter activation (Tan and Frankel, 1994), respectively, is unclear. Similar to Thr34, the difference in context may be responsible, and the varying activity of multiple mutants is consistent with it having a relatively minor role in RRE recognition

In the Rev–RRE model, Arg41, Arg42, Arg43, and Arg48 are positioned to interact with backbone phosphates. Arg48 is seen with Gln49 contacting the non-ARM Asn26 in a Rev crystal structure (Daugherty *et al*., 2010). Screening Arg41, Arg42, Arg43, and Arg48 libraries with RRE IIB and RAI reporters indicated active mutants, and a variety of moderately active substitutions were recovered. Our data suggest that these arginines make minor contributions to affinity consistent with roles contacting backbone phosphates, and they are individually dispensable for Rev–ARM binding to RRE IIB. The moderate to high activities of R38L, R41L, and R42L and their adjacent locations in the complex suggest favorable van der Waals interactions between the arginine aliphatic chains.

Overlooked in previous mutational studies, Ala37 appears to be restricted in the Rev–RRE interaction to smaller amino acids. The models suggest it is sterically confined between the peptide and RNA, making van der Waals contacts to a backbone ribose. Consistent with steric confinement, the proportion of positives in the Ala37 library was low, and only A37S had even moderate activity with RRE IIB. Although Ala37 may not be making significant enthalpic contributions, it could have a role in stabilizing the bound structure conformationally, and it is likely that few substitutions would be allowed.

Gln36 is in the major groove, and the Rev–RRE model suggests it could make van der Waals contacts to G47 and G48 (Battiste *et al*., 1996). Gel shift data indicate that it can be substituted with alanine (Tan *et al*., 1993). As expected, the Gln36 library had a large proportion of positives, and active substitutions were recovered. Its position in the major groove could allow its carboxamide nitrogen to donate a hydrogen bond to the N7 of G48. Interestingly, only one mutant in all libraries was found to have significantly higher than wild-type activity with RRE IIB, Q36S. It is conceivable that the high activity of Q36S with RRE IIB and RAI is the result of a new favorable contact between the serine hydroxyl and the N7 of G48 and its equivalent in RAI.

As expected from the distance between the Glu47 and the RNA, its negative charge, and gel shift affinity unaffected by substitution (Tan *et al*., 1993), the Glu47 library had a large proportion of positives. Many different substitutions had activities similar to or higher than the wild-type sequence. Glu47 does not appear to have a role in recognition. Trp45 has an essential role in Rev function *in vivo* (Zapp *et al*., 1991). Although it could make van der Waals contact to A68, mutational data *in vitro* and the Rev–RRE model indicate it has little role in RRE IIB recognition. As expected, a high proportion of its library was active in screens with both RRE IIB and RAI reporters, and a wide variety of substitutions were found to have near wild-type activity.

Compensatory mutations of critical residues suggest a distinct recognition strategy nearby

By chance, the first eight active clones isolated from the Asn40 library screen with RRE IIB included four clones representing double substitutions, three of R35G–N40V and one of R35G–N40K. The Arg35 codon in the Asn40 library should be CGC, and the double mutants had glycine GGC codons. Presumably, active single substitutions in the Asn40 library were so rare that active double substitutions arising from single nucleotide errors in the Asn40 library oligonucleotide synthesis were relatively common. Both double mutants displayed moderate activity with RRE IIB (Table 3). We prepared the single substitution N40V construct and assayed it for activity, which was as expected minimal, as were single substitutions R35G and N40L (Table 2).

To pursue these observations, a library randomized at both Arg35 and Asn40 was constructed and screened with RRE IIB. The double substitution library displayed a much higher proportion than did the Asn40 library (~4% vs 0.4%, see Supporting Information, Figures S1 and S2). Nine sequences were isolated, of which five were R35G–N40V, two were R35G–N40L, and two were R35G–N40A. No wild-type sequences were recovered. The double substitutions were assayed for activity (Table 3). That functional double mutants of two critical residues making base-specific contacts restore activity when replaced by the absence of a side chain and aliphatic side chains argues that the double mutant binds RRE using an alternate recognition strategy dependent on a glycine at position 35 and independent of hydrogen bonding side chains at position 40. Interestingly, R35G–N40L maintained a much greater activity on RRE IIB compared with wild type compared with RAI, which was not observed for any single substitution, demonstrating significant differences between RRE IIB and RAI specificities.

Clinical variation suggests constraints on rev genetic drift other than binding

Extraction of HIV Rev basic domain (Thr34–Arg48) sequences from the 2009 Los Alamos HIV Compendium (Kuiken *et al*., 2009) followed by filtering of incomplete and stop-codon-containing variants yields 1379 entries comprising 129 distinct sequences. With regard to only the composition of sequences, the number of substitutions relative to our basis sequence, HXB2, indicates several interesting features (Table 4). Very few substitutions are seen at most positions. Trp45, required for multimerization, appears least frequently mutated. Critical residues that make base-specific contacts, Arg35, Asn40, and Arg44, have the least number of substitutions. The critical residue Arg39 has a large number of R39K substitutions, but otherwise, it has as few substitutions as other critical, base-contacting residues. Slightly more substitutions are seen

Table 3. Activity of Rev double mutants

Clone[a]	Mutants[b]	Sequences[c]	Percent activation[d] RRE IIB	RAI	BTAR
V795-4	R35G; N40V	TGQARRVRRRRWRER	46 ± 1.0	10 ± 0.95	1.9 ± 0.2
V696-11	R35G; N40K	TGQARRKRRRRWRER	40 ± 4.0	8.9 ± 2.1	1.9 ± 0.2
V795-1	R35G; N40L	TGQARRLRRRRWRER	31 ± 1.4	3.1 ± 0.35	2.0 ± 0.5
II1026-1[e]	R39K; E47A	TRQARKNRRRRWRAR	9.3 ± 1.4	11 ± 3.4	2.2 ± 0.4

[a]Laboratory stock numbers of Rev17 N-fusions.
[b]Rev mutants are indicated by single-letter codes for wild-type and mutant residues separated by the numbered position.
[c]The sequences of Rev mutants with mutations indicated by underlining.
[d]Plasmid clones were transformed into RRE IIB, RAI, and BTAR reporter cells. At least four replicates of each clone were grown at 30°C overnight in tryptone medium supplemented with IPTG and assayed for β-galactosidase activity with *ortho*-nitrophenol-β-D-galactoside. Percent activation represents antitermination activities of Rev mutants normalized to Rev17–RRE, Rev17–RAI, or BIV Tat–TAR.
[e]Synthetic construct.

Table 4. Composition of Rev 34–48 in clinical HIV sequences

	Clinical variation of HIV-1 compared with HXB2 Rev 35–48[a]														
Codon[b]	T34	R35	Q36	A37	R38	R39	N40	R41	R42	R43	R44	W45	R46	E47	R48
Ala	2		2	**1308**					1					987[d]	
Arg		**1369**	10		**1313**	**856**	1	**1357**	**1369**	**1362**	**1373**	2	**1364**	31	**1350**
Asn	2		14				**1370**						1		
Asp							1							1	
Cys															
Gln		3	**1328**		56	1	1	14						41	
Glu											1			**285**	
Gly	1				1			2	1	3	1		3	2	5
His							1								
Ile							1						1		
Leu					1							1			
Lys		2	12		1	515	2	2	4	13	4		9	19	23
Met															
Phe															
Pro	4	4	1		1	1	1	4							
Ser	142		10	8	1	1	1			1			1	1	1
Thr	**1228**	1	2	63		5			4					12	
Trp					5							**1376**			
Tyr							1								
Val															
Total[c]	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379

[a]The Los Alamos HIV sequence database (Kuiken *et al*., 2009) was queried for Rev34–48 as nucleotides aligning to 8402–8447 of the reference genome HXB2. Incomplete, artifact, and stop-codon containing sequences were removed, and the 1379 remaining sequences were analyzed. The total number of each amino acid at each position, indicated by single-letter code for amino acid and location in Rev protein, was counted, and the number referring to the identity of HXB2 at that position is in bold.
[b]The 20 possible amino acids are represented by standard three-letter codes.
[c]The total number of sequences analyzed is shown as a checksum.
[d]E47A is more common than the reference HXB2 sequence.

at arginines that do not contact bases: Arg38, Arg41, Arg42, Arg43, Arg46, Arg48. More substitutions are seen at non-arginines that we found mutable: Thr34, Gln36, Ala37, Glu47. Many frequent substitutions appear quite conservative, such as T34S, A37T, R38Q, R39K, R41Q, R43K, R44K, R46K, and R48K. Our basis sequence for Rev–ARM, HXB2, is not the most common: E47A constitutes more than half of residue 47 entries.

The frequency of R39K was surprising, because our data show that this mutant has very low activity. Further analysis showed that of the 515 entries with R39K, none were single mutations. Indeed, excluding those of Glu47, there were only 18 single substitutions in the 34–48 region: The remainder were multiple substitutions. Of the 515 R39K entries, 498 are also E47A. This frequency suggested covariation, in which the replacement of the negatively charged Glu37 with an uncharged residue compensated for the mutation of Arg39 to lysine. We prepared and assayed the R39K–E47A construct and found it to be weakly active, yet well above the background activity of R39K (Table 3).

DISCUSSION

Noting incomplete mutational data in support of the Rev–RRE and Rev–RAI structural models, we constructed 15 libraries containing all single substitutions in Rev35–48 and screened them for activity using a well-established protein–RNA reporter system (Cocozaki *et al*., 2008b; Franklin, 1993; Harada *et al*., 1996; Iwazaki *et al*., 2005). Our results show that Rev basic domain residues range from invariant to highly mutable when binding RRE IIB. The data support the Rev–RRE structural model with minor refinements, consistent with critical base-specific contacts by Arg35, Arg39, Asn40, and Arg44. The Rev–RAI interaction appears far more mutable than that of Rev–RRE and is absolutely dependent on only one critical residue, Arg39, suggesting differences between Rev–RRE IIB and Rev–RAI recognition. New findings include the activity of R35K with RRE IIB, the unexpected mutability of Thr34, the restricted mutability of Ala37, and the fortuitous discovery of active double mutants of Arg35–Asn40. Comparison with variation in clinical sequences is consistent with Rev mutation being limited by factors other than binding RRE IIB. The mutability of Rev with different partners and the activity of the double mutant support the application of neutral theories of evolution to small peptide–RNA recognition.

Rev–RRE and Rev–RAI structural models

Supported by published gel shifts of alanine substitutions (Tan *et al*., 1993) and heterologous reporter assays of arginine-to-lysine substitutions (Tan and Frankel, 1994), the NMR-based structural models of Rev–RRE (Battiste *et al*., 1996) and Rev–RAI (Ye *et al*., 1996) present four residues making important base-

specific contacts, Arg35, Arg39, Asn40, and Arg44, while the backbone amide or hydroxyl of Thr34 and remaining arginines contact phosphates. Gln36, Ala37, and Trp45 are shown making van der Waals contacts, and Glu47 does not contribute to binding. It is tempting to imagine well-defined interactions in which each of the critical residues makes ideal hydrogen bonds in plane to their base pair partners. The mutability of Rev seen here, the conformational flexibility of Rev arginines observed in Rev–RRE IIB and Rev–RAI complexes (Wilkinson *et al*., 2004), and the adaptive binding of partners in these complexes (Battiste *et al*., 1994, Ye *et al*., 1996) suggest that less perfect interactions and even an ensemble of minor variations is likely.

The similarity of the Rev–RRE and Rev–RAI structural models promotes the assumption that they are identical in all important aspects. How similar are the two interactions, and how much can we infer about one interaction from the other? We found the Rev ARM to be highly mutable with the RAI reporter, consistent with Rev–RAI having more intermolecular contacts, perhaps unsurprisingly as it is the product of artificial selection for affinity, whereas the natural Rev–RRE interaction may have evolved under selection for biologically relevant rates of binding and unbinding in addition to specificity. The structures have important differences. The U:A:U triple in RAI introduces a large aberration in the backbone that repositions phosphates and more encloses the peptide, possibly facilitating more or better contacts. RAI, lacking the equivalent of RRE U66, cannot form a hydrogen bond between U66 O4 and Arg35, which was found to be important for Rev binding by conservation in selection experiments (Bartel *et al*., 1991) and footprinting and modification-interference experiments (Kjems *et al*., 1992). Counter-intuitively, arginines in Rev ARM are more dynamic in the higher affinity Rev–RAI interaction than in Rev–RRE IIB (Wilkinson *et al*., 2004).

The two interactions' determinants of specificity, while overlapping, are not identical. Only Arg39 of the critical Rev–RRE residues Arg35, Arg39, Asn40, and Arg44 is absolutely required for RAI binding, obscuring the defining pattern of the Rev–RRE recognition strategy. Rev substitutions throughout the peptide decreased binding to RRE IIB at least 10-fold more than RAI: A37L, R38G, N40L, R42S, R43L, R44K, and R46S (Table 2). Although no Rev single substitution was found to increase binding to RRE IIB markedly more than RAI, the Rev double mutant R35G–N40L maintained 10-fold more activity to RRE than to RAI. Thus, our results do not cast doubt on Arg35, Asn40, and Arg44 making base-specific contacts to RAI, but they do show that the recognition strategy of Rev–RAI, although structurally similar, has a distinct specificity.

Clinical sequence variation

To what extent does the mutability of the Rev basic domain seen in our binding assay reflect its potential mutability in the wild? Although our reporter system has been found to accord with affinity measurements *in vitro* (Harada *et al*., 1996; Harada *et al*., 1997; Iwazaki *et al*., 2005; Sugaya *et al*., 2008b), N-mediated antitermination relies on bacterial host factors unrelated to the host factors involved in Rev-mediated nuclear export in human cells. Compared with our results, the clinical data have very little variability, although the profile is similar in that critical residues have the lowest frequency of mutations and unimportant residues have the highest, with the exception of Arg39 (Table 4). The very low levels of substitution and our isolation of few mutations conferring higher-than-wild-type activity are consistent with purifying selection operating against weaker binders. In agreement with this, Glu47 is the most variable residue in the clinic, it has no clear RNA-binding role being negatively charged and projecting away from the RNA, and we isolated many functional mutants. In the clinic, the conservative mutation R39K is very common (Table 4), although we found the R39K mutant had only background activity with both RRE IIB and RAI reporters (Table 2). This specific discrepancy between clinical sequences and our assays is ameliorated by the observation that the vast majority of R39K occur in conjunction with E47A and the moderately restored activity of our R39K–E47A construct (Table 3). Interestingly, neither R35G nor N40V were found in the database (Table 4). Thus, Rev ARM does display some diversity of recognition strategy *in vivo*, although much less than found in our reporter system. In addition to maintaining affinity to RRE IIB, there are likely strong constraints on Rev ARM evolution *in vivo*, including the overlapping open reading frames of *env* and *tat* and its function as a nuclear localization signal.

Evolution of recognition

How different is the recognition strategy of R35G–N40V? How are the loss of critical, base-specific contacts made by Arg35 and Asn40 compensated? Energetically, there should be new contacts to compensate for the loss of hydrogen bonds. The replacement of the charged guanidinium group of Arg35 with a proton suggests deeper penetration of the peptide and conformational flexibility. The replacement of the carboxamide of Asn40 with an aliphatic group suggests burial of hydrophobic surface by extensive van der Waals interactions. We argue that specificity to RRE in preference to RAI upon mutation of two critical residues is most consistent with the double mutant binding using a distinct recognition strategy, although more physical data or computational modeling will be needed to understand how structurally similar the interaction is to Rev–RRE.

Proteins are observed to be quite tolerant of random substitution (Guo *et al*., 2004; Bloom *et al*., 2005). Neutral theories of evolution (Kimura, 1991; Ohta, 2002) assert that most observed genetic change is neutral or nearly neutral, such that genetic drift can lead to new phenotypes without positive selection. RNA has been used as a model to study neutral theories, its sequence being its genotype and its folding being its phenotype (Fontana, 2002). Intersection sequences (Schultes and Bartel, 2000), RNAs able to fold into different structures, would allow transitions to new phenotypes. Similar to intersection sequences, chameleon peptides (Minor and Kim, 1996) are able to adopt different conformations in different contexts, possibly facilitating evolutionary transitions and modeling how alternate structural conformers facilitate protein evolution (Tokuriki and Tawfik, 2009). When applied to small peptide–RNA recognition such as the interaction between Rev ARM and RRE IIB, neutral theories would explain the evolution of new recognition strategies by incremental genetic drift along paths of active mutations of one strategy until paths originating from different strategy are encountered, at which point transition can occur. Thus, neutral theories predict that peptide and RNA partners will be robust (Wilke and Adami, 2003), robustness being the ability to incur mutations without loss of activity, and that there are other recognition strategies nearby.

We see a range of alterations to the Rev–RRE recognition strategy: the minor variation of R39K–E47A, a somewhat different strategy employed by Rev–RAI, and an apparently

distinct strategy employed by the Rev R35G–N40V double mutant when binding RRE IIB. Many ARM–RNA variations have been isolated from random libraries (Harada *et al*., 1996; Xu and Ellington, 1996; Harada *et al*., 1997; Tan and Frankel, 1998; Ye *et al*., 1999; Das and Frankel, 2003; Peled-Zehavi *et al*., 2003; Bayer *et al*., 2005). These variations illustrate the potential evolution of recognition strategies. The R39K–E47A variation may represent a simple energetic compensation in which the decreased affinity caused by the disruption of a guanidinium hydrogen bond is restored by the removal of a distal, destabilizing negative charge. Rev–RAI may represent a variation in which Rev makes sufficient additional contacts that the otherwise critical residues Arg35, Asn40, and Arg44 are individually dispensable. The robustness of Rev–RAI suggests that the recognition strategy of Rev–RAI, possibly very similar to Rev–RRE in structure, is poised between being largely identical and being able to access new recognition strategies with very few mutations. One can imagine new recognition strategies easily arising by incremental mutations that repeatedly acquire new contacts and dispense of old contacts, thus maintaining binding while specificity is altered.

Our findings support neutral theories, both by the robustness of Rev ARM to substitutions and by the discovery of a distinct recognition strategy with only two Rev mutations, R35G and N40V. Are there examples from related systems supporting the possibility that Rev R35G–N40V binds in a distinct conformation? In an example of chameleonism, JDV Tat adopts distinct strategies to bind both BIV TAR and HIV-1 TAR, in β-turn and extended conformations, respectively (Smith *et al*., 2000). Hybrid boxB RNAs are able to bind two normally type-specific N proteins in bacteriophages λ and P22 (Cocozaki *et al*., 2008a), and a single substitution relaxed P22N ARM specificity such that it complemented both N-deficient λ and P22 viruses *in vivo* (Cocozaki *et al*., 2008b).

CONCLUSION

Our findings that Rev can access increasingly variant recognition strategies resonates with other studies on the diversity and adaptability of the Rev and RRE and viral ARM–RNA interactions (Williamson, 2000). Importantly, Rev binds other sites on RRE than stem IIB (Daugherty *et al*., 2008), and Rev binds a non-RAI aptamer (Xu and Ellington, 1996) in an extended, non-helical conformation (Ye *et al*., 1999). Using the N-boxB antitermination system applied here, structurally diverse peptides have been isolated that bind RRE IIB (Harada *et al*., 1996; Peled-Zehavi *et al*., 2003). Mutational analyses and further selections have shown how few mutations in ARMs or RRE can increase affinity and alter specificity (Harada *et al*., 1997, Iwazaki *et al*., 2005; Sugaya *et al*., 2008a; Sugaya *et al*., 2008b). Thus, there are many different specificities accessible, often with few structural changes. The adaptive binding (Battiste *et al*., 1994; Patel, 1999; Ye *et al*., 1996) and the conformational flexibility of Rev arginines (Wilkinson *et al*., 2004) observed in Rev–RRE and Rev–RAI complexes may reflect requirements of specificity (Leulliot and Varani, 2001), but they may also cooperate with the mutability of Rev reported herein to allow multiple partners and binding modes that facilitate evolutionary transitions.

Acknowledgements

We thank Naomi Franklin and Kazuo Harada for materials. This work was supported by the American University of Beirut Research Board.

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**: 529–536.

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**: 1547–1551.

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**:2741–2747.

Bayer TS, Booth LN, Knudsen SM, Ellington AD. 2005. Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA* **11**: 1848–1857.

Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, Arnold FH. 2005. Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 606–611. DOI: 10.1073/pnas.0406744102.

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**: 203–212.

Cilley CD, Williamson JR. 2003. Structural mimicry in the phage phi21 N peptide–boxB RNA complex. *RNA* **9**: 663–676.

Chen Y, Varani G. 2005. Protein families and RNA recognition. *FEBS J.* **272**: 2088–2097.

Cocozaki AI, Ghattas IR, Smith CA. 2008a. Bacteriophage P22 antitermination boxB sequence requirements are complex and overlap with those of lambda. *J. Bacteriol.* **190**: 4263–4271. 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** :7699–7708. DOI: 10.1128/JB.00997-08.

Das C, Frankel AD. 2003. Sequence and structure space of RNA-binding peptides. *Biopolymers* **70**: 80–85.

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. DOI: 10.1016/j.molcel.2008.07.016.

Daugherty MD, Liu B, Frankel AD. 2010. Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. *Nat. Struct. Mol. Biol.* **17**: 1337–1342. DOI: 10.1038/nsmb.1902.

Draper DE. 1999. Themes in RNA–protein recognition. *J. Mol. Biol.* **293**: 255–270.

Fontana W. 2002. Modelling 'evo-devo' with RNA. *Bioessays* **24**: 1164–1177.

Franklin NC. 1993. Clustered arginine residues of bacteriophage lambda N protein are essential to antitermination of transcription, but their locale cannot compensate for boxB loop defects. *J. Mol. Biol.* **231**: 343–360

Franklin NC, Doelling JH. 1989. Overexpression of N antitermination proteins of bacteriophages lambda, 21, and P22: loss of N protein specificity. *J. Bacteriol.* **171**: 2513–2522.

Giver L, Bartel D, Zapp M, Pawul A, Green M, Ellington AD. 1993. Selective optimization of the Rev-binding element of HIV-1. *Nucleic Acids Res.* **21**: 5509–5516.

Guo HH, Choe J, Loeb LA. 2004. Protein tolerance to random amino acid change. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 9205–9210. DOI: 10.1073/pnas.0403255101.

Harada K, Martin SS, Frankel AD. 1996. Selection of RNA-binding peptides in vivo. *Nature* **380**: 175–179.

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**: 11887–11892.

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.

Hope TJ, McDonald D, Huang XJ, Low J, Parslow TG. 1990. Mutational analysis of the human immunodeficiency virus type 1 Rev transactivator: essential residues near the amino terminus. *J. Virol.* **64**: 5360–5366.

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

Kimura M. 1991. The neutral theory of molecular evolution: a review of recent evidence. *Jpn J Genet* **66**: 367–386.

Kjems J, Calnan BJ, Frankel AD, Sharp PA. 1992. Specific binding of a basic peptide from HIV-1 Rev. *EMBO J.* **11**: 1119–1129.

Kuiken C, Leitner T, Foley B, Hahn B, Marx P, McCutchan F, Wolinsky S, and Korber B, Eds. 2009. HIV Sequence Compendium. Theoretical Biology and Biophysics Group: Los Alamos National Laboratory, NM.

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**: 289–299.

Leulliot N, Varani G. 2001. Current topics in RNA–protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry* **40** :7947–7956.

Long KS, Crothers DM. 1995. Interaction of human immunodeficiency virus type 1 Tat-derived peptides with TAR RNA. *Biochemistry* **34**: 8885–8895.

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**: 254–257.

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**: 730–734.

Ohta T. 2002. Near-neutrality in evolution of genes and gene regulation. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 16134–16137.

Ott M, Geyer M, Zhou Q. 2011. The control of HIV transcription: keeping RNA polymerase II on track. *Cell Host Microbe* **10**: 426–435. DOI: 10.1016/j.chom.2011.11.002.

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**: 252–261.

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

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**: 2397–2408.

Schultes EA, Bartel DP. 2000. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* **289**: 448–452.

Smith CA, Calabro V, Frankel AD. 2000. An RNA-binding chameleon. *Mol. Cell* **6**: 1067–1076.

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**: 10808–10814.

Sugaya M, Nishimura F, Katoh A, Harada K. 2008a. Tailoring the peptide-binding specificity of an RNA by combinations of specificity-altering mutations. *Nucleosides Nucleotides Nucleic Acids* **27**: 534–545. DOI: 10.1080/15257770801944493.

Sugaya M, Nishino N, Katoh A, Harada K. 2008b. 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**: 924–935. DOI: 10.1002/psc.1027.

Symensma TL, Giver L, Zapp M, Takle GB, Ellington AD. 1996. RNA aptamers selected to bind human immunodeficiency virus type 1 Rev in vitro are Rev responsive in vivo. *J. Virol.* **70**: 179–187.

Tan R, Chen L, Buettner JA, Hudson D, Frankel AD. 1993. RNA recognition by an isolated α helix. *Cell* **73**: 1031–1040.

Tan R, Frankel AD. 1994. Costabilization of peptide and RNA structure in an HIV Rev peptide–RRE complex. *Biochemistry* **33**: 14579–14585.

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**: 4247–4252.

Tokuriki N, Tawfik DS. 2009. Protein dynamism and evolvability. *Science* **324**: 203–207. DOI: 10.1126/science.1169375.

Wilke CO, Adami C. 2003. Evolution of mutational robustness. *Mutat. Res.* **522**: 3–11.

Wilkinson TA, Zhu L, Hu W, Chen Y. 2004. Retention of conformational flexibility in HIV-1 Rev–RNA complexes. *Biochemistry* **43**: 16153–16160.

Williamson JR. 2000. Induced fit in RNA–protein recognition. *Nat. Struct. Biol.* **7**: 834–837.

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**: 7475–7480.

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**: 1026–1033

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**: 657–669.

Zapp ML, Green MR. 1989. Sequence-specific RNA binding by the HIV-1 Rev protein. *Nature* **342**: 714–716.

Zapp M L, Hope T J, Parslow T G, Green M R. 1991. Oligomerization and RNA binding domains of the type 1 human immunodeficiency virus Rev protein: a dual function for an arginine-rich binding motif. *Proc. Natl. Acad. Sci. U. S. A.* **88**: 7734–7738.