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

Altered-Specificity Mutants of the HIV Rev Arginine-Rich Motif-RRE IIB Interaction

Journal:	*Journal of Molecular Recognition*
Manuscript ID	JMR-19-0080.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a

Complete List of Authors:	Raad, Nicole; American University of Beirut, Biology Ghattas, Ingrid; American University of Beirut, Biology Amano, Ryo; Chiba Institute of Technology, Department of Life Science Watanabe, Natsuki; Chiba Institute of Technology, Department of Life Science Sakamoto, Taiichi; Chiba Institute of Technology, Department of Life Science Smith, Colin; American University of Beirut, Biology

Keywords:	HIV Rev, RNA-protein interaction, arginine-rich motif, altered specificity, neutral evolution, disordered protein, fitness landscape

SCHOLARONE™
Manuscripts

Altered-Specificity Mutants of the HIV Rev Arginine-Rich Motif-RRE IIB Interaction

Short running title: **Altered-Specificity Mutants of Rev-RRE**

Nicole G. Raad[a], Ingrid R. Ghattas[a], Ryo Amano[b], Natsuki Watanabe[b], Taiichi Sakamoto[b], and Colin A. Smith[a]*

*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] N. G. Raad, I. R. Ghattas, C. A. Smith

Department of Biology, American University of Beirut, Beirut, Lebanon.

[b] R. Amano, N. Watanabe, T. Sakamoto

Department of Life Science, Chiba Institute of Technology, Chiba Japan

For Peer Review

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

1 **ABSTRACT**

2 Arginine-rich motifs (ARMs) bind RNA structures with high affinity and specificity, and the
3 human immunodeficiency virus (HIV) exploits ARM-RNA interactions to regulate its lifecycle.
4 The expression of HIV structural genes relies on recognition between the ARM of its Rev
5 protein and its primary binding site, an internal loop in the viral RNA, the Rev-response element
6 region IIB (IIB). Many functional variants of the Rev ARM-IIB interaction have been
7 discovered, yet how easily it can evolve new specificities is poorly explored. A double mutant of
8 Rev ARM, R35G-N40V, uses an unknown strategy to recognize IIB. Here, isothermal titration
9 calorimetry and gel shift assays show that the R35G-N40V-IIB interaction has high affinity and
10 specificity *in vitro* and a larger unfavorable entropy change upon binding than that of wild-type
11 Rev ARM-IIB. In stark contrast with the critical dependence of wild-type Rev on Arg35, Arg39,
12 Asn40, and Arg44, mutational profiling shows R35G-N40V is highly mutable at positions 40
13 and 44, and dependent on Gly35, Arg38, Arg39, Arg42, and Arg43. Affinity measurements *in*
14 *vitro* and reporter assay measurements *in vivo* are consistent with the wild-type Rev ARM and
15 R35G-N40V maintaining their recognition strategies when binding IIB mutants specific to wild-
16 type Rev ARM and R35G-N40V, respectively. Some single amino acid mutants of wild-type
17 Rev ARM and R35G-N40V have enhanced specificity, recognizing mutant IIBs yet not wild-
18 type IIB. These results provide another example of viral ARM-RNA interactions evolving new
19 specificities with few mutations, consistent with neutral theories of evolution.

20

21 **Keywords**: HIV Rev; RNA-protein interaction; arginine-rich motif; altered specificity; neutral
22 evolution, disordered protein, fitness landscape

23

For Peer Review

INTRODUCTION

RNA-protein interactions comprise diverse structures exploited for many biological processes, including multiple, essential steps in gene expression.[1] Arginine-rich motifs (ARMs) binding small, structured RNAs occur in important viral regulatory processes, including well-studied examples with structural models: human immunodeficiency virus (HIV) Rev-RRE[2] and Tat-TAR,[3] bovine immunodeficiency virus (BIV) Tat-TAR,[4,5] and lambdoid bacteriophage N-boxBs.[6-9] Because viral ARM-RNA interactions are small, exhibit structural diversity, have high affinities and specificities, and display complex characteristics such as induced fit,[10] recognition of multiple partners,[11] and structural plasticity,[12] they are attractive models with which to understand RNA-protein recognition and how specificities alter and evolve new strategies.

HIV Rev is a small, 114 amino acid, essential regulatory protein that mediates the nuclear export of incompletely spliced viral transcripts that contain a large, structured RNA, the Rev-response element (RRE).[13] Rev initially binds to a high-affinity site in RRE region IIB (IIB) via its ARM that comprises residues 34-50, and then Rev binds to secondary sites with multimerization and further conformational changes followed by recruitment of host factors that mediate export.[14,15]

Many findings, including NMR[16] and x-ray crystallography structures[2] and biochemical and genetic results,[17,18] support a model in which the Rev ARM-IIB interaction relies on its α-helical ARM and 4 amino acids, Arg35, Arg39, Asn40, and Arg44, that make contacts to 5 guanine bases in the internal loop formed by non-canonical purine-purine base pairs G47:A73 and G48:G71 (Figure 1A). Arg35 contacts G67, Arg39 contacts G70, Asn40 contacts G47 and G71, and Arg44 contacts G46. Other arginines form ionic interactions with the phosphates of the backbone, and Trp45 is not important for binding isolated IIB, but required for multimeric binding of Rev to larger regions of RRE.[19] Importantly, the mutually inducted fit of Rev ARM and IIB binding[20-22] appears to extend through complex and cooperative binding of multiple Rev proteins to other binding sites in RRE.[2,23-25]

Although Rev ARM and RRE IIB have limited sequence variation in clinical databases,[26,27] diverse synthetic variants have been discovered, including a Rev-aptamer in which Rev ARM binds in an extended conformation[28] and a peptide selected from a random library that binds in a

For Peer Review

partially α-helical conformation unlike wild-type Rev.[29] The sequence and structural variety raises questions as to the richness and diversity of recognition strategies accessible to ARM-RNA interactions generally, and to Rev-RRE specifically.

Rev ARM R35G-N40V (GV) (Figure 1B) binds IIB using an uncharacterized recognition strategy. It was serendipitously discovered by screening Rev ARM libraries[18] in a reporter system[30] based on λ phage N-boxB antitermination. Despite GV lacking two of the four amino acids that wild-type Rev ARM (WT) uses to contact IIB bases, it appears to bind with similar specificity.[18] WT and GV have similar requirements for the internal loop binding site of IIB, both relying on the same core internal loop formed by non-canonical base pairs G47:A73 and G48:G71, yet they recognize different sets of RNAs (Figure 1C).[31]

The activity of GV with IIB and discovery of IIB mutants that bind specifically to WT or GV prompted closer examination of GV. Here, isothermal titration calorimetry and gel shift assays show that the GV-IIB interaction has high affinity and specificity *in vitro* and a larger unfavorable entropy change upon binding than WT-IIB. In stark contrast with the critical dependence of WT on Arg35, Arg39, Asn40, and Arg44, mutational profiling using reporter assays in bacteria shows GV is highly mutable at positions 40 and 44, and depends on Gly35, Arg38, Arg39, Arg42, and Arg43. Affinity measurements *in vitro* and reporter assay measurements *in vivo* are consistent with WT and GV maintaining their recognition strategies when binding their specific RRE mutants. Single amino acid mutants of WT and GV have been found with enhanced specificity, recognizing mutant RRE IIBs yet not wild-type RRE IIB. These results show that viral ARM-RNA interactions can evolve new specificities with few mutations, consistent with neutral theories of evolution.[32]

MATERIALS AND METHODS

Peptides and DNA oligonucleotides

Peptides with the wild-type HIV-1 Rev_{34-50} (WT), maTRQARRNRRRRWRERQRaaaa, and its R35G-N40V mutant (GV), maTGQARRVRRRRWRERQRaaaa, were obtained from Hokkaido Science Systems (Sapporo, Japan), and included flanking amino acids (lower case) to match the sequence used in the bacterial reporter assay. No provisions were made to remove residual

trifluoroacetic acid from deprotection, and peptides were resuspended in water. ^{19}F NMR experiments indicate 50 and 40 molar equivalents of TFA for WT and GV peptides, respectively (data not shown). Peptide concentrations were estimated from tryptophan absorbance at 280 nm and an extinction coefficient of 5,500 M^{-1} cm^{-1}. DNA oligonucleotides used for transcription templates and cloning were obtained from Hokkaido Science Systems, Sigma-Aldrich (Darmstadt, Germany), and TIB MolBiol (Berlin, Germany).

RNA synthesis

All RNAs were transcribed by T7 RNA polymerase using synthetic oligonucleotide templates and standard methods.[33] Briefly, transcriptions were conducted in T7 buffer (80 mM HEPES-KOH, pH 8.1; 50 mM dithiothreitol; 10 mM spermidine; 0.01% Triton X-100) with 80 mg/ml PEG-8000, 8 mM GTP, 8 mM UTP, 4 mM ATP, 4 mM CTP, 42 mM $MgCl_2$, and 50 μg/ml T7 RNA polymerase on 300 nM templates annealed to T7 promoter. After 2 to 4 hours at 37 °C, transcription reactions were centrifuged to remove precipitated magnesium pyrophosphate, and the supernatant processed by ethanol precipitation before purification by gel electrophoresis. RNAs for gel shift assays were based on the 39-nt RRE IIB sequence (IIB) used in reporter assays: 5'-gGUcUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGCcGUUCCCCUGCAGUGCA-3', in which mutated bases are lower case and the 16 underlined 3'-nucleotides serve as a hybridization site for a biotinylated DNA oligonucleotide probe. RNAs for isothermal titration calorimetry were based on the 34-nt RRE IIB sequence used for NMR studies:[16] 5'-gGUcUGGGCGCAGCgcaaGCUGACGGUACAGGCc-3'. RNAs were purified on 10% polyacrylamide/7 M urea gels, soaked from crushed excised bands, serially centrifuged to remove gel fragments, and precipitated with ethanol. RNA concentrations were determined by 260 nm absorbance and estimated extinction coefficients.

Gel shift assays

Purified RNAs were resuspended in water and renatured from 95 °C with 1.5-fold ratio of biotinylated DNA oligonucleotide complementary to the 3'-end of the RNA and diluted to 20 nM in gel shift buffer (10 mM HEPES, pH 7.5, 100 mM KCl; 1 mM $MgCl_2$; 0.5mM EDTA; 10% glycerol; 50 μg/ml yeast tRNA) and mixed with equal volumes of serial, two-fold dilutions of

peptides ranging from 20 nM to 5120 nM in gel shift buffer. After incubation on ice for 30 minutes, RNA-peptide mixtures were applied to pre-run, uncooled, room temperature, native, 0.5 X TBE (45 mM Tris base; 45 mM boric acid; 1 mM EDTA), 10% (37.5:1 mono:bis) polyacrylamide gels at 300 volts for 90 minutes. The RNAs and RNA-peptide complexes were transferred via semi-dry transfer to Hybond-N+ (GE Healthcare Life Sciences) membrane blots, and RNAs were crosslinked to blots by 254 nm irradiation from a transilluminator for 2-5 minutes. Blots were blocked with 1% non-fat milk in TBS-T (20 mM Tris-HCl, pH 7.4; 150 mM NaCl), washed, and bound to horseradish peroxidase-streptavidin conjugate (Invitrogen) for 120 minutes. Blots were washed with TBS-T, and the RNAs were visualized by incubation with ECL reagents (GE Healthcare Life Sciences). Images were collected with a Chemidoc XRS (Bio-Rad) and their contrasts were adjusted to enhance band recognition using QuantityOne software (Bio-Rad).

Isothermal Titration Calorimetry

RNAs for isothermal titration calorimetry were filtered using 0.8 nm filters. RNAs were placed in calorimetry buffer (20 mM sodium phosphate, pH 7.0; 50 mM sodium chloride) using Vivaspin 2 centrifugal filter units (GE Healthcare Life Sciences). Isothermal calorimetry titrations were performed with a MicroCal iTC$_{200}$ (Malvern Instruments) at 25 °C. RNAs, 10 µM, were titrated with WT and GV peptides in calorimetry buffer, 100 µM, two or three times each in a 200 µl sample cell stirred at 1000 rpm over 19 injections with 150 s between injections. The individual titration peaks are integrated by the instrument software Origin and presented in Wiseman plots where ΔH values are plotted against molar ratio. After subtracting the ΔH of non-specific binding estimated from the heat observed during the late, flat portion of the titration plot, the resulting data points were fitted to 'One set of sites' binding model provided by the instrument software Origin and yielded the binding enthalpy ΔH, the dissociation constant (K_{d}), and the stoichiometry n. The change in Gibbs free energy, ΔG, and entropy, ΔS, were calculated on the basis of K_d and ΔH. Each thermodynamic parameter is represented by the mean ± standard error of two or three independent measurements.

Library screening

For Peer Review

A reporter system measures heterologous RNA-protein binding using two plasmids that reconstitute bacteriophage λ N-boxB transcriptional antitermination in bacteria:[30] an ARM-N supplier in which the ARM (residues 1−19) of λ N protein is replaced with an NcoI-BsmI fragment coding for wild-type Rev ARM (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') or variants, and an RNA reporter in which λ left nut boxB is replaced with the PstI-BamHI fragment expressing RRE IIB (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') or variants. Following published methods,[34] Rev R35G-N40V ARM and its libraries were cloned as NcoI-BsmI fragments generated by mutually priming oligonucleotides followed by NcoI-BsmI digestion. WT libraries were available in house.[18] WT and GV libraries identities and diversities were assessed by blind sampling to yield at least 5 readable sequences of each library (see Table S1). *Escherichia coli* supporting λ N-nut antitermination, N567,[35] were transformed with reporter plasmids, made competent, and transformed with library plasmids. For selections, colonies showing activity comparable to the standard ARM-RNA interaction by plate assay were chosen and grown as individual cultures, plasmid DNA was extracted, and supplier plasmid was separated from reporter plasmid by agarose gel electrophoresis. Supplier plasmids were retransformed, prepared, retested, and sequenced.

ARM-RNA reporter assays

For plate assays, approximately 5−50 ng of N-fusion plasmid per 50 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 5-bromo-4-chloro-3-indolyl-$_{D}$-galactoside (X-gal) as the chromogenic substrate of the β-galactosidase reporter protein, and 50 µM isopropyl-D-thiogalactoside (IPTG) to induce the tac promoters expressing N-protein and the reporter transcript. The plates were scored after 16 h at 34 °C and after a second 24 h 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 WT-IIB interaction is scored as 4+, and non-cognate interactions are 0+ (background white). The BIV Tat-TAR interaction in which mgRPRGTRGKGRRIRRggg (flanking amino acids are in lower case) is fused to the λ N

For Peer Review

activation domain and 5'-GCUCGUGUAGCUCAUUAGCUCCGAGC-3' replaces boxB, is scored as 3+ and served as a control for specificity.

For solution assays, cultures of four to six representative colonies were picked from X-gal plates and grown overnight for 16 h at 30 °C with aeration in 3 ml of tryptone containing 100 μg/ml ampicillin and 12 μg/ml chloramphenicol as antibiotics, and with 50 μM IPTG. The cultures were assayed for β-galactosidase activity using *ortho*-nitrophenol-β-$_D$-galactopyranoside (ONPG), and units of β-galactosidase were calculated following Miller.[36] Observed β-galactosidase units with standard deviations are reported of representative experiments.

Structure visualization

Jmol (http://www.jmol.org/ [18 January 2015]), an open-source Java viewer for chemical structures in 3D, was used to view structures and create images of the crystal structure of Rev–RRE (PDB ID: 4PMI)[2] and to view the minimized, average NMR structure of HIV-1 Rev–RRE (PDB ID: 1ETF).[16]

RESULTS

Gel shift assays corroborate high affinity and isothermal titration calorimetry reveals a large unfavorable entropic contribution to GV-RRE IIB binding

An important characteristic of HIV Rev-RRE is that the isolated ARM-IIB interaction displays high specificity *in vitro*. Consistent with N-boxB reporter assay results,[18,31] gel shift assays in the presence of non-specific tRNA competitor (Figure 2A) are consistent with GV-IIB having a similar affinity and specificity as WT-IIB, with dissociation constants between 80 and 160 nM in the conditions used. Competitor RNA avoids measuring non-specific affinity of positively charged peptides with negatively charged RNAs. Published affinities of similar WT-IIB complexes measured by gel shift assays range from 25 nM[37] to 160 nM[11] and presumably depend on precise conditions used, especially Mg^{++}, ionic strength, temperature, and competitor RNA. The GV peptide, despite having one less positive charge than WT peptide, appears to cause a slightly greater separation in its IIB complex, suggesting a gross structural change consistent with a different binding mode (Figure 2A).

For Peer Review

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

Isothermal titration calorimetry was used to compare the interaction of WT-IIB and GV-IIB in the absence of competitor tRNA (Figure 2B). As observed for the gel shift assays and as expected from reporter assays, the affinities of WT-IIB and GV-IIB are similar, yet the relative contribution of enthalpy and entropy differ substantially between WT-IIB and GV-IIB, with a larger unfavorable contribution of entropy in the GV-IIB interaction than the WT-IIB (Table 1). Importantly, isothermal titration calorimetry of the inactive single mutants WT R35G and WT N40V show no specific binding (Figure S2). Nonetheless, with the methods used, it is difficult to discriminate quantitatively between specific and non-specific binding. The relative contributions of enthalpy and entropy of WT-IIB are roughly similar to what has been reported for a similar interaction in detailed thermodynamic study[23] characterizing a Rev ARM peptide binding to IIB and IA RRE RNAs. Jayaraman et al.[23] report dissociation constants ranging from 0.02 nM at 283 K and 100 mM KCl to 16.3 nM at 303 K and 300 mM KCl, approaching the dissociation constant of 22 ± 8 nM reported here for WT-IIB. Meaningful comparison to the values reported here is limited by differences in the order in which molecules are titrated, the presence of Mg^{++}, the use of different monovalent cations and their concentrations, the pH of the buffers, and the method of data analysis.[23] Favorable changes in enthalpy upon binding are usually ascribed to increased non-covalent interactions, such as hydrogen and ionic bonding. Entropy changes are usually ascribed to the balance of unfavorable contributions from decreases in conformational freedom and favorable contributions from decreases in exposed hydrophobic surface area with concomitant release of water.[38] Both Rev ARM and RRE IIB are known to undergo conformational changes upon binding,[20] and the origin of the relatively larger unfavorable entropy of GV-IIB binding could be a greater conformation change of GV upon binding associated with the conformational flexibility of Gly35, a larger conformational change of IIB RNA, or complex contributions from less burial of hydrophobic surface area. Because preliminary ^{1}H-NMR data are consistent with the free WT and GV peptides being unstructured (data not shown), it seems unlikely that differences in free peptide conformations affect the thermodynamics of binding. It must be noted that because of the presence of residual trifluoroacetic acid in the peptides, its influence on peptide conformation and isothermal calorimetry results cannot be formally excluded. Although the data are insufficient to ascertain the origin of the thermodynamic differences, they are consistent with gross structural differences between recognition strategies.

For Peer Review

Mutagenesis reveals amino acid requirements of GV Rev ARM

To reveal which amino acids of GV are likely important in binding IIB, we followed the approach used in a previous study of WT-IIB[18] and constructed 15 plasmid libraries spanning Thr34 to Arg48, each expressing GV with one amino acid position targeted for randomization in the ARM-N supplier plasmid. If an amino acid in the ARM is required for function, such as by making an essential contact to the RNA partner, it should be immutable and a small proportion of the corresponding library will be active with the RNA reporter. If an amino acid has no specific role in binding, it should be mutable and a relatively large proportion of the corresponding library will be active. Although the representation of amino acids is substantially skewed by the genetic code and by errors and non-random incorporation of nucleotides during oligonucleotide synthesis, codon randomization provides diversity beyond alanine scanning and is an inexpensive alternative to the use of triplet phosphoramidites.[39] Here, sampling of individual WT and GV libraries without selection found targeted ARM mutations with distinct triplets in at least one-third of 5 to 8 readable sequences (Table S1). Using a reporter assay in *E. coli* based on bacteriophage λ N-boxB antitermination, library plasmids were screened for ARM-IIB recognition in which ARM-RNA binding is reflected in the expression of β-galactosidase and intensity of blue pigment in colonies on X-gal solid media.[30] The proportions of each library displaying colonies with activity similar to GV-IIB were estimated by visual inspection and counting (Figure 3). Importantly, the profile obtained of GV-IIB differs strongly from that of WT-IIB[18] at multiple positions, consistent with GV-IIB interaction employing a distinct recognition strategy. To corroborate the mutagenic profiling, at least 24 active colonies were selected and their plasmids were prepared, retested, and sequenced from each assay plate (Figure 3). To suggest functional roles of interesting positions, some alanine, lysine, and glutamine mutants were constructed and assayed (Table 2).

The immutability of Gly35 is consistent with previously published selections from a WT R35X-N40X library in which only glycines were recovered at position 35.[18] That Gly35 of GV cannot be substituted, even with alanine, suggests steric restraints imposed by a close approach of the peptide to the RNA or by the peptide backbone conformation. The immutability of Arg38 and Arg39 and low activity of their lysine mutants suggest critical roles in specific recognition of

For Peer Review

bases via hydrogen bonding.[40] That GV R38K and R39K have very weak activity is consistent with base-specific hydrogen bonds, as lysine can often substitute for arginine ionic contacts to phosphates. Of the three arginines that contact guanines in WT-IIB[2] (Figure 1A), WT R35K has moderate activity, and WT R39K and WT R44K have background levels of activity.[18] Thus, the restricted mutability of Arg42 and Arg43, the recovery of only arginines and lysines, and the moderate activity of lysine substitutions, suggest either hydrogen bonding to specific bases or ionic contacts to phosphates. Intermediate proportions and limited diversity and preferences suggest Thr34, Gln36, and Trp45 may have specific roles or restrictions on size, shape, and hydrophobicity. Other positions appear tolerant of mutation and without obvious restrictions.

Wild-type Rev ARM and R35G-N40V display high affinity and specificity to mutant RREs *in vitro*

Using representative, specific RRE IIB mutants,[31] we characterized the binding of WT and GV to WT-specific RRE mutants G50A-C69A (#15) and G48U (#8), and GV-specific mutants U45G-A75U (#2), and C51G-A52U-G53U-C65A-U66G-G67U (#21) (Figure 1C). As expected, the affinity of WT for RRE IIB mutants #15 and #8 and GV for #2 and #21 observed in the reporter system[31] was corroborated by gel shift assays (Figure 4) and isothermal titration calorimetry experiments (Figure 5 and Table 3). Interestingly, by gel shift assays, the binding of WT-#15 and GV-#2 appear stronger than that of WT-IIB and GV-IIB. The relative specificities measured by reporter and gel shift assays agree better than the relative affinities measured by calorimetry. This highlights the possible effects of conditions and context on ARM-RNA measurements.

Recognition strategies of WT and GV maintained while binding specific RRE mutants

We next asked if WT and GV bind mutant RREs using the same recognition strategies as they bind IIB. Mutational profiling of WT libraries with #15 and #8 resulted in profiles (Figure 6A, B) that are similar to that of WT-IIB,[18] in which Arg39, Asn40, and Arg44 have very low proportions active library members. Likewise, mutational profiling of GV libraries with #2 and #21 (Figure 6C, D) are similar to that of GV-IIB (Figure 2), in which Gly35, Arg38, Arg39, Arg42, and Arg43, yet not Asn40 or Arg44, are low. These observations are consistent with the

For Peer Review

existence of merely two recognition strategies, possibly with minor variations: WT-like and GV-like.

Enhanced specificity mutants of WT and GV

Considering that RNA change in an ARM-RNA complex could affect the function of some ARM mutants, we reasoned that there may be enhanced-specificity mutants of WT and GV that do not recognize IIB yet do recognize mutant RNAs. Thus, we used the X-gal plate assays to select four pools comprising at least 24 active members each from WT libraries[18] for binding to #15 and #8 and for active members of the GV libraries for binding to #2 and #21. After isolating and selecting clones that were active with the specific RRE mutant and inactive with IIB, modest numbers of clones were sequenced, reconstructed, and retested. Selected altered-specificity mutants and other available mutants were assayed for activity with IIB, #15, #8, #2, #21, and BIV TAR (Table 4). Remarkably, several mutants of WT and GV display altered specificity. GV R42A, GV R43A, and GV W45A have much higher activity with #2 than IIB, and WT R35P, WT R38G, and WT R41G have much higher activity on #15 than IIB.

GV and Specific RRE Mutants are not found in clinical data

Searches via the National Center for Biotechnology Information BLAST server[41] restricted to HIV-1 found no occurrences of the GV sequence or specific mutant IIB sequences #15, #8, #2, and #21. This was anticipated in light of a previous analysis that found neither R35G-N40V, R35G, nor N40V,[18] and more recent publications examining clinical variation of Rev-RRE,[26,27,42,43] which describe relatively little variation and none related to this study.

DISCUSSION

Structural insights

How different could the GV-IIB structure be from WT-IIB? The structural diversity of ARM-RNA complexes, which include extended conformations[3,28] and β turns,[4,5,44] as well as α helices, provides no obvious limits. More relevant could be selections from arginine-rich libraries for sequences that bind IIB, which have found diverse amino acid identities embedded in arginine-rich contexts,[30,37,45-47] all of which appear partially or completely α helical, including one for which a structural model is available, RSG-1.2.[45] RSG-1.2, binds IIB in a partially α-helical

conformation inserted deep in the major groove and nearly perpendicular to the RNA axis.[29] Interestingly, RSG-1.2 contains a glycine in the portion binding IIB and its RNA requirements[48] are similar to that of GV,[31] although they do not share all of each other's important residues.

How similar could GV-IIB be to WT-IIB? Despite sharing a binding site and the identity of 13 of 17 amino acids, GV-IIB cannot be a minor variant of the WT-IIB binding mode. GV tolerates mutations flanking the core RNA-binding site that WT does not, including mutation of G67, U66, and base pair U45:A75. Conversely, unlike WT, GV requires base pair G50:C69. Although specificities of ARM-RNA interactions can differ even with very similar architecture,[7] of the four WT amino acids making base-specific hydrogen bonds to IIB, only Arg39 is critical in GV and in common with WT. Additionally, the greater shift of GV-IIB than WT-IIB and the relative contributions of enthalpy and entropy are consistent with gross structural differences. All these data support a model in which GV binds IIB using a distinct recognition strategy than WT.

Evolutionary Potential

The diverging specificities from WT-IIB to GV-IIB continuing to WT-#15 and GV-#2 and further to WT R38G-#15 and GV W45A-#2 describe paths from the original WT-IIB interaction to orthogonal ARM-RNA interactions (Table 4): WT R38G-#15 and GV W45A-#2 are similar in activity to GV-IIB, yet the inverse interactions WT R38G-#2 and GV W45A-#15 have background activity. Orthogonal interactions, in which cognate partners do not recognize non-cognate partners, allow for viral regulation to be specific in complex cellular environments. How many orthogonal recognition strategies are possible with ARM-RNA interactions? It would seem from the great variety of sequences and strategies employed by RNA-binding proteins and peptides[1,49] and the permutations possible with specific amino acid recognition of bases and base pairs found in RNA motifs,[40,50,51] that the number could be very large.

A fitness landscape[52] that describes the relationship between the ARM-RNA sequences (genotypes) and functional recognition (fitness) is useful in considering how far evolution paths might extend. The question becomes whether regions of fitness are isolated like mountain peaks or connected like ridges. Neutral and nearly neutral theories of evolution predict landscapes of ridges: that for any phenotype, sufficient functional genotypes exist such that incremental paths

For Peer Review

connect distinct phenotypes without any loss-of-function intermediates.[32] Relaxed-specificity ARMs and RNAs have been found that transit recognition strategies between distinct lentiviral Tat-TAR interactions and between distinct lambdoid N-boxB interactions.[12,53-55] Like Rev-RRE,[15,20,21] Tat-TAR and N-boxB interactions are characterized by disordered proteins binding RNAs sites with conformational flexibility, a feature of RNA-protein interactions relating to their role in complex, fine-tuned regulation.[56,57]

Synthetic and natural ARM-RNA interactions include such structural diversity that it is difficult to imagine that neutral paths extend to very different interactions. Yet the conformational flexibility,[21] adaptive binding,[20] recognition of multiple partners, and the existence of structurally distinct, orthogonal interactions found close in sequence space adds support to the application of neutral theories to ARM-RNA interactions. The role of neutral evolution in HIV is complex:[58] HIV genetic diversity is created by error-prone reverse transcription, host editing factors, and retroviral recombination,[59] and over evolutionary timescales, complex retroviruses have found diverse solutions to Rev-RRE binding.[60] The variation in WT-IIB observed in the virus[26,27,42,43] is very limited relative to the mutational potential seen in the laboratory, yet there are many reasons GV may not appear in clinical isolates: it is uncertain how well GV binds to secondary RRE binding sites, whether GV cooperates with Rev multimerization, and how GV might affect overlapping viral reading frames.

CONCLUSION

Although there is insufficient information to propose a structural model, these mutagenesis data will assist interpretation of a future structural model. The orthogonal WT-#15 and GV-#2 interactions illustrate how viral sequences and recognition strategies can evolve and how change in one partner permits change in the other. The architecture of WT-IIB with its genetic and structural plasticity may be unusually capable of functional variations, ideal for achieving and altering specificity in a small size and complex cellular environment, attractive for viruses with small genomes.

ACKNOWLEDGMENTS

For Peer Review

We thank Kazuo Harada, Tokyo Gakugei University for hosting C. A. S. in his laboratory, advice, and critical comments. We thank Core Research for Evolutional Science and Technology (CREST) at Tokyo Gakugei University for access and assistance with the Chemidoc XRS. This work was supported by the University Research Board of the American University of Beirut.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

AUTHORS' CONTRIBUTIONS

N.G.R., I.R.G., R.A., N.W., T.S., and C.A.S. designed and conducted experiments. C.A.S. wrote the paper.

ORCID

Colin Smith https://orcid.org/0000-0001-5652-0855

REFERENCES

1. Hentze MW, Castello, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol*. 2018; 19(5): 327-341. doi: 10.1038/nrm.2017.130.
2. Jayaraman B, Crosby DC, Homer C, Ribeiro I, Mavor D, Frankel AD. RNA-directed remodeling of the HIV-1 Rev protein orchestrates assembly of the Rev-Rev response element complex. *eLife*. 2014; 3:e04120. doi: 10.7554/eLife.04120.
3. Schulze-Gahmen U, Hurley JH. Structural mechanism for HIV-1 TAR loop recognition by Tat and the super elongation complex. *Proc Natl Acad Sci U S A*. 2018; 115(51):12973-12978. doi: 10.1073/pnas.1806438115. doi: 10.1126/science.270.5239.1200.
4. Puglisi JD, Chen L, Blanchard S, Frankel AD. Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science.* 1995; 270(5239):1200-1203. doi: 10.1126/science.270.5239.1200.
5. Ye X, Kumar RA, Patel DJ. Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem. Biol*. 1995; 2(12):827-840. doi: 10.1016/1074-5521(95)90089-6.

For Peer Review

6. Cai Z, Gorin A, Frederick R, et al. Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex. *Nat Struct Biol.* 1998; 5(3):203-12. doi: 10.1038/nsb0398-203.
7. Cilley CD, Williamson JR. Structural mimicry in the phage phi21 N peptide-boxB RNA complex. *RNA*. 2003; 9(6):663-676. doi: 10.1261/rna.2189203.
8. Legault P, Li J, Mogridge J, Kay LE, Greenblatt J. NMR structure of the bacteriophage λ N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell.* 1998; 93(2):289-299. doi: 10.1016/s0092-8674(00)81579-2.
9. Schärpf M, Sticht H, Schweimer K, Boehm M, Hoffmann S, Rösch P. Antitermination in bacteriophage λ. The structure of the N36 peptide-boxB RNA complex. *Eur. J. Biochem*. 2000; 267(8):2397-2408. doi: 10.1046/j.1432-1327.2000.01251.x.
10. Patel DJ. Adaptive recognition in RNA complexes with peptides and protein modules. *Curr Opin Struct Biol.* 1999; 9(1):74-87. doi: 10.1016/S0959-440X(99)80010-4.
11. Bayer TS, Booth LN, Knudsen SM, Ellington AD. Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA.* 2005; 11(12):1848-1857. doi: 10.1261/rna.2167605.
12. Smith CA, Calabro V, Frankel AD. An RNA-binding chameleon. *Mol Cell.* 2000; 6(5):1067-1076. doi: 10.1016/s1097-2765(00)00105-2.
13. Pollard VW, Malim MH. The HIV-1 Rev protein. *Annu Rev Microbiol*. 1998; 52:491-532. doi: 10.1146/annurev.micro.52.1.491.
14. Rausch JW, Le Grice SF. HIV Rev Assembly on the Rev Response Element (RRE): A Structural Perspective. *Viruses*. 2015; 7(6):3053-3075. doi: 10.3390/v7062760.
15. Fernandes JD, Booth DS, Frankel AD. A structurally plastic ribonucleoprotein complex mediates post-transcriptional gene regulation in HIV-1. *Wiley Interdiscip Rev RNA*. 2016; 7(4): 470-486. doi: 10.1002/wrna.1342.
16. Battiste JL, Mao H, Rao NS, et al. α helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science.* 1996; 273(5281):1547-1451. doi: 10.1126/science.273.5281.1547.
17. Tan R, Chen L, Buettner JA, Hudson D, Frankel AD. RNA recognition by an isolated α helix. *Cell.* 1993; 73(5):1031-1040. doi: 10.1016/0092-8674(93)90280-4.

For Peer Review

18. Possik EJ, Bou Sleiman MS, Ghattas IR, Smith CA. 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.* 2013; 26(6):286-296. doi: 10.1002/jmr.2272.
19. Daugherty MD, D'Orso I, Frankel AD. A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA. *Mol Cell*. 2008; 31(6):824-834. doi: 10.1016/j.molcel.2008.07.016.
20. Casu F, Duggan BM, Hennig M. The arginine-rich RNA-binding motif of HIV-1 Rev is intrinsically disordered and folds upon RRE binding. *Biophys J.* 2013; 105(4):1004-1017. doi:10.1016/j.bpj.2013.07.022.
21. Chu CC, Plangger R, Kreutz C, Al-Hashimi HM. Dynamic ensemble of HIV-1 RRE stem IIB reveals non-native conformations that disrupt the Rev-binding site. *Nucleic Acids Res*. 2019; 47(13):7105-7117. doi: 10.1093/nar/gkz498.
22. Tan R, Frankel AD. Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. *Biochemistry.* 1994; 33(48):14579-14585. doi: 10.1021/bi00252a025.
23. Jayaraman B, Mavor D, Gross JD, Frankel AD. Thermodynamics of Rev-RNA interactions in HIV-1 Rev-RRE assembly. *Biochemistry*. 2015; 54(42):6545-6554. doi: 10.1021/acs.biochem.5b00876
24. Tanamura S, Terakado H, Harada K. Cooperative dimerization of a stably folded protein directed by a flexible RNA in the assembly of the HIV Rev dimer-RRE stem II complex. *J Mol Recognit*. 2016; 29(5):199-209. doi: 10.1002/jmr.2518.
25. Bai Y, Tambe A, Zhou K, Doudna JA. RNA-guided assembly of Rev-RRE nuclear export complexes. *Elife*. 2014; 3:e03656. doi: 10.7554/eLife.03656
26. Jackson PE, Tebit DM, Rekosh D, Hammarskjold ML. Rev-RRE Functional Activity Differs Substantially Among Primary HIV-1 Isolates. *AIDS Res Hum Retroviruses*. 2016; 32(9):923-34. doi: 10.1089/AID.2016.0047.
27. Sherpa C, Jackson PEH, Gray LR, et al. Evolution of the HIV-1 Rev Response Element during Natural Infection Reveals Nucleotide Changes That Correlate with Altered Structure and Increased Activity over Time. *J. Virol*. 2019; 93(11): e02102-e02118. doi: 10.1128/JVI.02102-18.

28. Ye X, Gorin A, Frederick R, et al. RNA architecture dictates the conformations of a bound peptide. *Chem Biol.* 1999; 6(9):657-669. doi: /10.1016/S1074-5521(99)80117-3.

29. Gosser Y, Hermann T, Majumdar A, et al. Peptide-triggered conformational switch in HIV-1 RRE RNA complexes. *Nat Struct Biol.* 2001; 8(2);146-50. doi: DOI: 10.1038/84138.

30. Harada K, Martin SS, Frankel AD. Selection of RNA-binding peptides in vivo. *Nature.* 1996; 380(6570):175-179. doi: 10.1038/380175a0.

31. Abdallah EY, Smith CA. Diverse mutants of HIV RRE IIB recognize wild-type Rev ARM or Rev ARM R35G-N40V. *J Mol. Recognit.* 2015; 28(12):710-721. doi: 10.1002/jmr.2485

32. Kimura M. Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proc Natl Acad Sci U S A.* 1991; 88(15):5969-5973. doi: 10.1073/pnas.88.14.5969.

33 Smith CA, Chen L, Frankel AD. Using peptides as models of RNA–protein interactions. *Methods Enzymol*. 2000; 318:423-438. doi: 10.1016/s0076-6879(00)18067-x.

34. Peled-Zehavi H, Smith CA, Harada K, Frankel AD. Screening RNA-binding libraries by transcriptional antitermination in bacteria. *Methods Enzymol*. 2000; 318:297-308. doi: 10.1016/s0076-6879(00)18059-0.

35. Doelling JH, Franklin NC. Effects of all single base substitutions in the loop of boxB on antitermination of transcription by bacteriophage λ's N protein. *Nucleic Acids Res*. 1989; 17(14): 5565–5577. doi: 10.1093/nar/17.14.5565.

36. Miller JH. *A short course in bacterial genetics: a laboratory manual and handbook for* Escherichia coli *and related bacteria.* Plainview, NY: Cold Spring Harbor Laboratory Press; 1997. doi: 10.1002/jobm.3620330412.

37. Sugaya M, Nishino N, Katoh A, Harada K. Amino acid requirement for the high affinity binding of a selected arginine-rich peptide with the HIV Rev-response element RNA. *J Pept Sci.* 2008; 14(8):924-935. doi: 10.1002/psc.1027.

38. Amano R, Furukawa T, Sakamoto T. ITC Measurement for High-Affinity Aptamers Binding to Their Target Proteins. *Methods Mol Biol*. 2019; 1964:119-128. doi: 10.1007/978-1-4939-9179-2_9.

39 Virnekäs B, Ge L, Plückthun A, Schneider KC, Wellnhofer G, Moroney SE. Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res*. 1994 Dec 25;22(25):5600-7. doi: 10.1093/nar/22.25.5600.

For Peer Review

40. Cheng AC, Chen WW, Fuhrmann CN, Frankel AD. Recognition of nucleic acid bases and base-pairs by hydrogen bonding to amino acid side-chains. *J Mol Biol*. 2003; 327(4):781-796. doi: 10.1016/s0022-2836(03)00091-3.

41. Altschul, SF, Gish W, Miller W, Myers EW Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215(3):403-410. doi: 10.1016/S0022-2836(05)80360-2.

42. Cunyat F, Beerens N, García E, Clotet B, Kjems J, Cabrera C. Functional analyses reveal extensive RRE plasticity in primary HIV-1 sequences selected under selective pressure. *PLoS One*. 2014; 9(8): e106299. doi: 10.1371/journal.pone.0106299.

43. Sloan EA, Kearney MF, Gray LR, et al. Limited nucleotide changes in the Rev response element (RRE) during HIV-1 infection alter overall Rev-RRE activity and Rev multimerization. *J. Virol*. 2013; 87(20):11173-11186. doi: 10.1128/JVI.01392-13.

44. Calabro V, Daugherty MD, Frankel AD. A single intermolecular contact mediates intramolecular stabilization of both RNA and protein. *Proc. Natl. Acad. Sci. U. S. A.* 2005; 102(19):6849-6854. doi: 10.1073/pnas.0409282102.

45. Harada K, Martin SS, Tan R, Frankel AD. Molding a peptide into an RNA site by in vivo peptide evolution. *Proc Natl Acad Sci U S A.* 1997; 94(22):11887-11892. doi: 10.1073/pnas.94.22.11887.

46. Tan R, Frankel AD. A novel glutamine-RNA interaction identified by screening libraries in mammalian cells. *Proc Natl Acad Sci U S A.* 1998; 95(8):4247-4252. doi: 10.1073/pnas.95.8.4247.

47. Peled-Zehavi H, Horiya S, Das C, Harada K, Frankel AD. Selection of RRE RNA binding peptides using a kanamycin antitermination assay. *RNA.* 2003; 9(2):252-261. doi: 10.1261/rna.2152303.

48. Iwazaki T, Li X, Harada K. Evolvability of the mode of peptide binding by an RNA. *RNA.* 2005; 11(9):1364-1373. doi: 10.1261/rna.2560905.

49. Das C, Frankel AD. Sequence and structure space of RNA-binding peptides. *Biopolymers.* 2003; 70(1):80-85. doi: 10.1002/bip.10429.

50. Leontis NB, Westhof E. Analysis of RNA motifs. *Curr Opin Struct Biol*. 2003; 13(3):300-308. doi: 10.1016/S0959-440X(03)00076-9.

51. Hermann T, Westhof E. Non-Watson-Crick base pairs in RNA-protein recognition. *Chem Biol*. 1999; 6(120):R335-R343. doi: 10.1016/S1074-5521(00)80003-4.

For Peer Review

52. Fragata I, Blanckaert A, Dias Louro MA, Liberles DA, Bank C. Evolution in the light of fitness landscape theory. *Trends Ecol Evol*. 2019; 34(1);69-82. doi: 10.1016/j.tree.2018.10.009.

53. Cocozaki AI, Ghattas IR, Smith CA. Bacteriophage P22 antitermination boxB sequence requirements are complex and overlap with those of λ. *J Bacteriol*. 2008; 190(12):4263-4271. doi: 10.1128/JB.00059-08.

54. Smith CA, Crotty S, Harada Y, Frankel AD. Altering the context of an RNA bulge switches the binding specificities of two viral Tat proteins. *Biochemistry.* 1998; 37(30):10808-10814. doi: 10.1021/bi980382+.

55. Cocozaki AI, Ghattas IR, Smith CA. The RNA-Binding Domain of Bacteriophage P22 N Protein Is Highly Mutable and a Single Mutation Relaxes Specificity Toward λ. *J Bacteriol*. 2008; 190(23):7699-708. doi: 10.1128/JB.00997-08.

56. Balcerak A, Trebinska-Stryjewska A, Konopinski R, Wakula M, Grzybowska EA. RNA-protein interactions: disorder, moonlighting and junk contribute to eukaryotic complexity. *Open Biol*. 2019; 9(6):190096. doi: 10.1098/rsob.190096

57. Basu S, Bahadur RP. A structural perspective of RNA recognition by intrinsically disordered proteins. *Cell Mol Life Sci*. 2016; 73(21):4075-4084. doi: 10.1007/s00018-016-2283-1

58. Leitner T. The Puzzle of HIV Neutral and Selective Evolution. *Mol. Biol. Evol*. 2018; 35(6):1355-1358. doi: 10.1093/molbev/msy089.

59. Smyth RP, Davenport MP, Mak J. The origin of genetic diversity in HIV-1. *Virus Res*. 2012; 169(2): 415-429 doi: 10.1016/j.virusres.2012.06.015

60. Cavallari I, Rende F, D'Agostino DM, Ciminale V. Converging strategies in expression of human complex retroviruses. *Viruses*. 2011; 3(8):1395-414. doi: 10.3390/v3081395.

For Peer Review

TABLES

TABLE 1 Isothermal titration calorimetry of WT-IIB and GV-IIB interactions[a]

titrand RNA[b]	titrant peptide[c]	stoichiometry	K_d (nM)	ΔH (kcal/mol)	ΔG (kcal/mol)	$T\Delta S$ (kcal/mol)
IIB	WT	1.02 ± 0.08	22 ± 8	-15.8 ± 0.5	-10.5 ± 0.2	-5.3 ± 0.5
IIB	GV	1.01 ± 0.03	39 ± 2	-24.6 ± 0.5	-10.1 ± 0.1	-14.3 ± 0.5

[a]Experiments were performed at 25℃ in 20 mM sodium phosphate buffer (pH 7.0), 50 mM NaCl. Each thermodynamic parameter is represented by the mean ± standard error of two or three independent measurements.
[b]IIB is 5'-GGUCUGGGCGCAGCGCAAGCUGACGGUACAGGCC-3'.
[c]WT is MATRQARRNRRRRWRERQRAAAA, GV is MATGQARRVRRRRWRERQRAAAA.

For Peer Review

TABLE 2 ARM-RNA solution assays of selected ARM-IIB and ARM-BIV TAR[a]

N-fusion[b]	Sequence[c]	IIB[d]	BTAR[e]
WT[f]	-T**R**QARR**N**RRRRWRER-	230±50	3.1±0.7
GV[g]	-T**G**QARR**V**RRRRWRER-	100±22	2.7±0.3
BIV Tat[h]	-GRPRGTRGKGRRIRR-	1.2±0.6	90±30
WT R35G/GV V40N	-T**G**QARR**N**RRRRWRER-	15.8±1.1	3.0±0.4
WT N40V/GV G35R	-T**R**QARR**V**RRRRWRER-	0.0±0.6	2.6±0.4
GV G35A	-T**A**QARRVRRRRWRER-	2.6±0.5	3.1±0.2
GV Q36A	-TG**A**ARRVRRRRWRER-	55±8	2.2±0.3
GV A37Q	-TGQ**Q**RRVRRRRWRER-	150±40	2.2±0.5
GV R38K	-TGQA**K**RVRRRRWRER-	0.3±0.5	2.3±0.4
GV R39K	-TGQAR**K**VRRRRWRER-	0.4±0.3	2.19±0.13
GV V40A	-TGQARR**A**RRRRWRER-	45±5	2.24±0.19
GV R42A	-TGQARRVR**A**RRWRER-	4.1±0.6	2.49±0.07
GV R42K	-TGQARRVR**K**RRWRER-	40±4	2.5±0.3
GV R43A	-TGQARRVRR**A**RWRER-	8±2	1.7±1.0
GV R43K	-TGQARRVRR**K**RWRER-	24.8±1.6	2.4±0.2
GV W45A	-TGQARRVRRRR**A**RER-	11.2±1.8	2.0±0.3

[a]Plasmids were transformed into reporter cells with the named RNA in place of λ nut site boxB. At least three 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. Values represent β-galactosidase units and standard deviations.
[b]Laboratory names of N-fusions expressed by pBRN plasmids. All plasmid inserts were reconstructed from synthetic oligonucleotides and sequenced.
[c]The sequences of ARM fused to λ N.
[d]The IIB sequence replacing boxB expressed by reporter IIB is 5'-GGUCUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGCC-3'.
[e]BTAR is the heterologous control, bovine immunodeficiency virus TAR, 5'-GCUCGUGUAGCUCAUUAGCUCCGAGC-3'.
[f]WT is MA TRQARRNRRRRWRERQR AAAA.
[g]WT is MA TGQARRVRRRRWRERQR AAAA.
[h]BTat, the heterologous control is MG RPRGTRGKGRRIRR GGG.

TABLE 3 Isothermal titration calorimetry of WT-#15, WT-#8, GV-#2, and GV-#21[a]

titrand RNA[b]	titrant peptide[c]	stoichiometry	K_d (nM)	ΔH (kcal/mol)	ΔG (kcal/mol)	$T\Delta S$ (kcal/mol)
#15	WT	1.1 ± 0.2	30 ± 5	-22 ± 2	-10.3 ± 0.1	-12 ± 2
#8	WT	1.1 ± 0.2	8.5 ± 4	-16 ± 1	-11.1 ± 0.3	-4.7 ± 0.9
#2	GV	1.3 ± 0.2	44 ± 3	-25 ± 3	-10.05 ± 0.04	-15 ± 3
#21	GV	0.96 ± 0.07	75 ± 11	-40 ± 1	-9.81 ± 0.01	-30 ± 1

[a]Experiments were performed at 25℃ in 20 mM sodium phosphate buffer (pH 7.0), 50 mM NaCl. Each thermodynamic parameter is represented by the mean ± standard error of two or three independent measurements.

[b]IIB is 5'-GGUCUGGGCGCAGCGCAAGCUGACGGUACAGGCC-3', #15 is IIB-G50A-C69A is 5'-GGUCUGGGCACAGCGCAAGCUGAAGGUACAGGCC-3', #8 is IIB-G48U is 5'-GGUCUGGUCGCAGCGCAAGCUGACGGUACAGGCC-3', #2 is IIB-U45G-A75U is 5'-GGUCGGGGCGCAGCGCAAGCUGACGGUACUGGCC-3', #21 is IIB-C51G-A52U-G53U-C65A-U66G-G67U is 5'-GGUCUGGGCGGUUCGCAAGAGUACGGUACAGGCC-3'.

[c]WT is MATRQARRNRRRRWRERQRAAAA, GV is MATGQARRVRRRRWRERQRAAAA.

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

1
2 TABLE 4 ARM-RNA reporter assays[a]
3

lab ref[b]	N-fusion[c]	#15	#8	IIB	#2	#21	BTAR
ii728-1	WT	4+ (320±70)	3+ (86±11)	4+ (230±50)	0+ (2±0.5)	1+ (5.1±0.7)	0+ (3.1±0.7)
xv26-3	GV	1+ (5.7±0.4)	0+ (2.5±0.3)	3+ (100±22)	3+ (180±30)	3+ (110±18)	0+ (2.7±0.3)
xv363-1	BIV Tat	0+ (1.8±0.5)	0+ (2.14±0.32)	0+ (1.2±0.6)	0+ (1.0±0.2)	0+ (2.1±0.3)	3+ (90±30)
xv26-9	WT R35G/GV V40N	1+ (47±17)	0+ (3.17±0.15)	3+ (15.8±1.1)	3+ (65±16)	3+ (27±4)	0+ (3.0±0.4)
xv26-17	WT N40V/GV G35R	0+ (2.3±0.2)	0+ (2.26±0.16)	0+ (0.0±0.6)	0+ (1.1±0.2)	0+ (1.69±0.12)	0+ (2.6±0.4)
xv188-36	GV G35A	0+ (2.11±0.09)	0+ (1.7±0.3)	1+ (2.6±0.5)	0+ (1.9±0.8)	0+ (3.4±0.4)	0+ (3.1±0.2)
xv220-33	GV Q36A	0+ (4.4±0.3)	0+ (2.6±0.3)	3+ (55±8)	4+ (130±9)	4+ (70±10)	0+ (2.2±0.3)
xv444-18	GV Q36Y	0+ (2.11±0.18)	0+ (2.46±0.19)	0+ (0.3±0.3)	2+ (10±4)	0+ (2.3±0.2)	0+ (2.2±0.3)
xv382-11	GV A37Q	3+ (21±3)	0+ (2.5±0.3)	3+ (150±40)	4+ (120±14)	4+ (11.3±1.3)	0+ (2.2±0.5)
xv220-34	GV R38K	0+ (2.07±0.16)	0+ (2.3±0.2)	0+ (0.3±0.5)	1+ (5.9±1.2)	0+ (2.3±0.2)	0+ (2.3±0.4)
xv220-35	GV R39K	0+ (2.10±0.13)	0+ (2.4±0.5)	0+ (0.4±0.3)	0+ (1.0±0.5)	0+ (2.5±0.3)	0+ (2.19±0.13)
xv220-37	GV V40A	0+ (3.4±0.2)	0+ (2.39±0.16)	3+ (45±5)	4+ (80±20)	3+ (53±7)	0+ (2.24±0.19)
xv444-21	GV R41G	0+ (2.3±0.3)	0+ (2.6±0.5)	0+ (0.3±0.4)	1+ (7.2±0.8)	1+ (3.4±1.5)	0+ (1.8±0.3)
ii2554-8	GV R42A	0+ (1.8±0.3)	0+ (2.16±0.15)	0+ (4.1±0.6)	2+ (38±6)	0+ (2.9±0.2)	0+ (2.49±0.07)
xv382-16	GV R42I	0+	0+	0+	0+	0+	0+
ii2554-31	GV R42K	0+ (3.1±0.5)	0+ (1.9±0.6)	2+ (40±4)	2+ (114±18)	2+ (27±5)	0+ (2.5±0.3)
ii2554-37	GV R43A	0+ (1.8±0.4)	0+ (1.8±0.3)	1+ (8±2)	2+ (77±6)	1+ (19.1±1.6)	0+ (1.7±1.0)
ii2554-22	GV R43K	0+ (2.2±0.3)	0+ (1.83±0.14)	1+ (24.8±1.6)	2+ (70±9)	2+ (50±6)	0+ (2.4±0.2)
xv220-39	GV W45A	0+ (2.56±0.08)	0+ (2.3±0.2)	2+ (11.2±1.8)	2+ (120±20)	2+ (23.1±1.2)	0+ (2.0±0.3)
xv395-21	WT R35P	4+ (100±20)	1+ (3.5±0.4)	2+ (8±3)	0+ (1.5±0.3)	1+ (3.6±0.6)	0+ (2.8±0.7)
xv382-22	WT R38G	4+ (127±8)	1+ (2.5±0.3)	2+ (3.3±0.6)	0+ (1.1±0.2	0+ (2.0±0.3)	0+ (2.4±0.5)
xv444-2	WT R38S	3+	2+	2+	0+	0+	0+
ii922-1	WT R39K	0+	0+	0+	0+	0+	0+
xv382-28	WT R41G	3+ (50±9)	1+ (4.3±0.3)	2+ (2.3±0.7)	0+ (0.9±0.2)	0+ (1.83±0.08)	0+ (1.8±0.5)
ii909-2	WT R41K	4+	3+	4+	0+	1+	0+
xv382-33	WT R42K	4+	3+	4+	1+	1+	0+
xv382-38	WT R43K	4+	3+	4+	1+	2+	0+

4

[a]Plasmids were transformed into reporter cells with the named RNA in place of λ nut site boxB. Number of plusses indicates intensity of blue pigment deposited in colonies of X-gal plate assays containing 50 uM IPTG. Values in parentheses represent β-galactosidase units and standard deviations as in Table 2: at least three replicates of each clone were grown at 30°C overnight in tryptone medium supplemented with 50 μM IPTG and assayed with ONPG.
[b]Laboratory reference of plasmids.
[c]Laboratory names of N-fusions expressed by pBRN plasmids. All plasmid inserts were reconstructed from synthetic oligonucleotides and sequenced.

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FIGURE LEGENDS

FIGURE 1 Rev ARM-IIB structure and variants. (A) Structural model of the high-affinity binding site of HIV Rev-RRE, Rev_{34-50} and RRE nucleotides 45-53 and 65-75 from the X-ray crystal structure[3] (PDB ID: 4PMI). Note that the sequences of the displayed region of the crystal structure differs from those of the NMR model[16] (PDB ID: 1ETF) on which the sequences studied here originate: the crystal structure Rev ARM is E47A and the RRE is G53C-C66G. Left, RNA rendered as white wireframe, with bases making contacts to Rev larger and the Rev peptide backbone gray and with side chains of amino acids that contact bases. Right, the same structure rotated 80 degrees vertically and looking at the helix axis from the amino terminus and rendered the same without the RNA wireframe. (B) The sequences of wild-type Rev ARM_{34-50} (WT) and its R35G-N40V mutant (GV) in the peptide and λ N-fusion contexts used in this study. (C) RRE IIB and its WT- and GV-specific RNAs used in this study. Gel shift, isothermal titration calorimetry, and reporter contexts all share the core binding site shown. Guanines with specific WT amino acid contacts are bold in the IIB structure, and mutations are bold in mutant RNAs.

FIGURE 2 WT-IIB and GV-IIB affinity and specificity *in vitro*. (A) RRE IIB RNA-binding gel shift assays with WT and GV peptides. IIB was mixed with increasing concentrations of peptide, from 10 nM to 2560 nM, in buffer (10 mM HEPES, pH 7.5; 100 mM KCl; 1 mM $MgCl_2$; 0.5mM EDTA; 10% glycerol; 50 µg/ml yeast tRNA). The concentration of peptide required to shift half the RNA is the apparent dissociation constant in nM. Paired bars at fixed separation highlight the difference in shifts. (B) Isothermal titration calorimetry plots in which peptide is titrated into a solution of RNA at 25 °C in 20 mM sodium phosphate buffer (pH 7.0), 50 mM NaCl. Top panels show raw calorimetry traces, and bottom panels show plots of integrated heat values before subtracting the heat of non-specific binding. See Table 1 for calculated values and Figure S1 for plots of integrated heat values after subtraction of the estimated heat of non-specific binding.

FIGURE 3 Mutational profiling of GV-IIB. Bar height represents the percentage of GV library members displaying activity similar to GV in X-gal plate assays with RRE IIB reporter.

For Peer Review

Above the bars, the identities of mutants from randomly selected active colonies are shown with the number as subscript. Below the bars, letters describe the amino acid identity and position in GV. The number of clones with the starting amino acid is shown as the subscript below. Important positions are in bold type.

FIGURE 4 WT-IIB and GV-IIB affinity and specificity *in vitro*. As in Figure 2A, peptide-RNA-binding gel shift assays with WT and GV peptides and RNAs, as labeled. (A) Peptides with #15 (G50A-C69A), left: WT and right: GV. (B) Peptides with #8 (G48U), left: WT and right: GV. (C) Peptides with #2 (U45G-A75U), left: WT and right: GV. (D) Peptides with #21 (C51G-A52U-G53U-C65A-U66G-G67U), left: WT and right: GV. RNAs were mixed with increasing concentrations of peptide described in nM in buffer containing competitor tRNA.

FIGURE 5 Isothermal titration calorimetry of WT and GV with mutant RNAs. As in Figure 2B, raw isothermal titration calorimetry traces and plots of peptides and RNAs, as labeled. (A) Peptides with #15 (G50A-C69A), left: WT and right: GV. (B) Peptides with #8 (G48U), left: WT and right: GV. (C) Peptides with #2 (U45G-A75U), left: WT and right: GV. (D) Peptides with #21 (C51G-A52U-G53U-C65A-U66G-G67U), left: WT and right: GV. See Table 3 for calculated values and Figure S1 for plots of integrated heat values after subtraction of the heat of non-specific binding.

FIGURE 6 Mutational profiling of WT and GV libraries with mutant RNAs. As in Figure 3, bar height represents the percentage of library members displaying activity similar to WT-IIB or GV-IIB in X-gal plate assays with mutant RNA reporters (see Figure 1). (A) WT-#15 (IIB G50A-C69A). (B) WT-#8 (IIB G48U). (C) GV-#2 (IIB U45G-A75U). (D) GV-#21 (IIB C51G-A52U-G53U-C65A-U66G-G67U).

Graphical Abstract

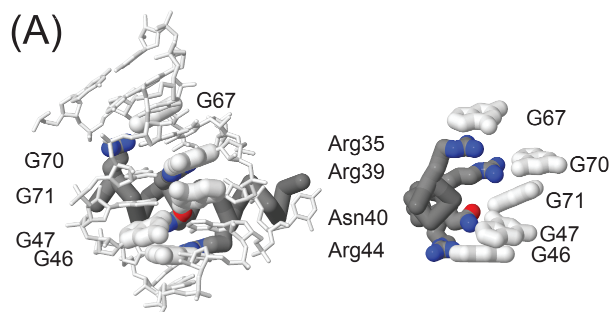
Altered-Specificity Mutants of the HIV Rev Arginine-Rich Motif-RRE IIB Interaction

Nicole G. Raad, Ingrid R. Ghattas, Ryo Amano, Natsuki Watanabe, Taiichi Sakamoto, and Colin A. Smith*

For Peer Review

Isothermal titration calorimetry and gel shift assays show that the interaction of HIV Rev R35G-N40V and RRE IIB has high affinity and specificity *in vitro*. In stark contrast with the critical dependence of wild-type Rev on Arg35, Arg39, Asn40, and Arg44, R35G-N40V is mutable at positions 40 and 44, and dependent on Gly35, Arg38, Arg39, Arg42, and Arg43. Some single amino acid mutants of wild-type Rev ARM and R35G-N40V have enhanced specificity, recognizing mutant IIBs yet not wild-type IIB.

For Peer Review



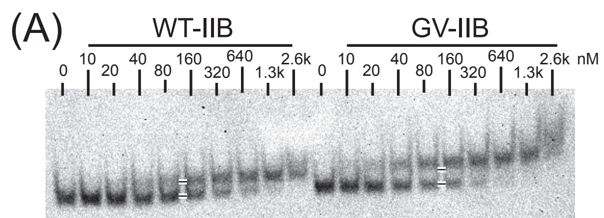
(B)

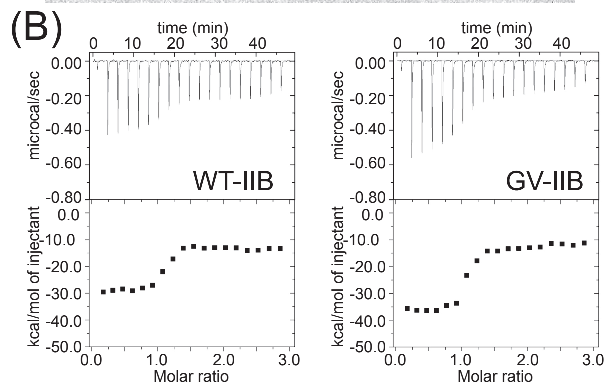
```
wild type (WT)     MA TRQARRNRRRRWRERQR AAAA
                         35      40    44
R35G-N40V (GV)     MA TGQARRVRRRRWRERQR AAAA
```

(C) WT-binding GV-binding

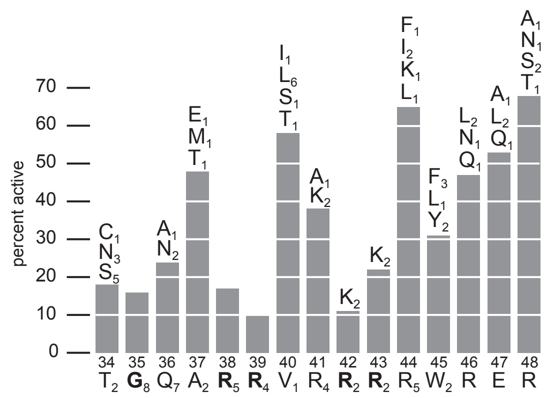
```
 G-C        G-C        G-C          G-C        U-A
 A-U        A-U        A-U          A-U        U-G
 C-G A      C-G        C-G 67       C-G        G-U
A--A        G-C A      G-C A        G-C A      G-C A
 C-G        C-G        C-G 70       C-G        C-G
G---G U     U---G U    G---G 71     G---G U    G---G U
G---A       G---A    47G---A U      G---A      G---A
 G-C        G-C      46G-C          G-C        G-C
 U-C        U-C        U-C          G-U        U-C
 50A69A     48U        5' 3'        45G75U     51A52U53U
                                               65A66G67U

#15         #8         IIB          #2         #21
```

For Peer Review



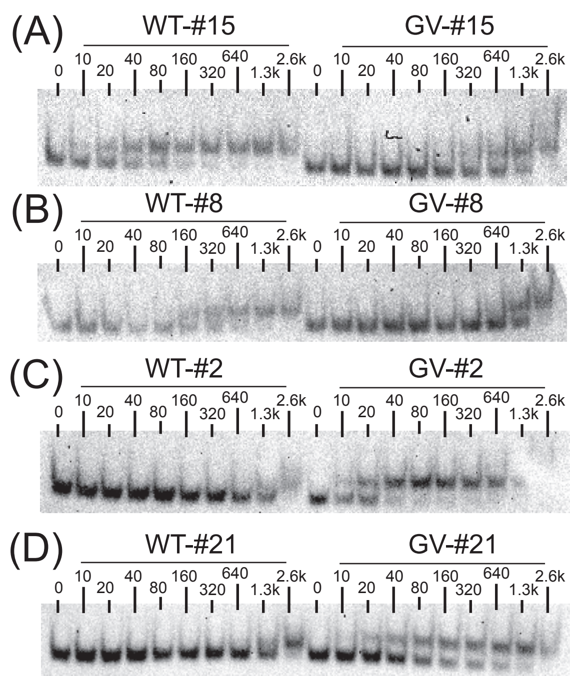


For Peer Review



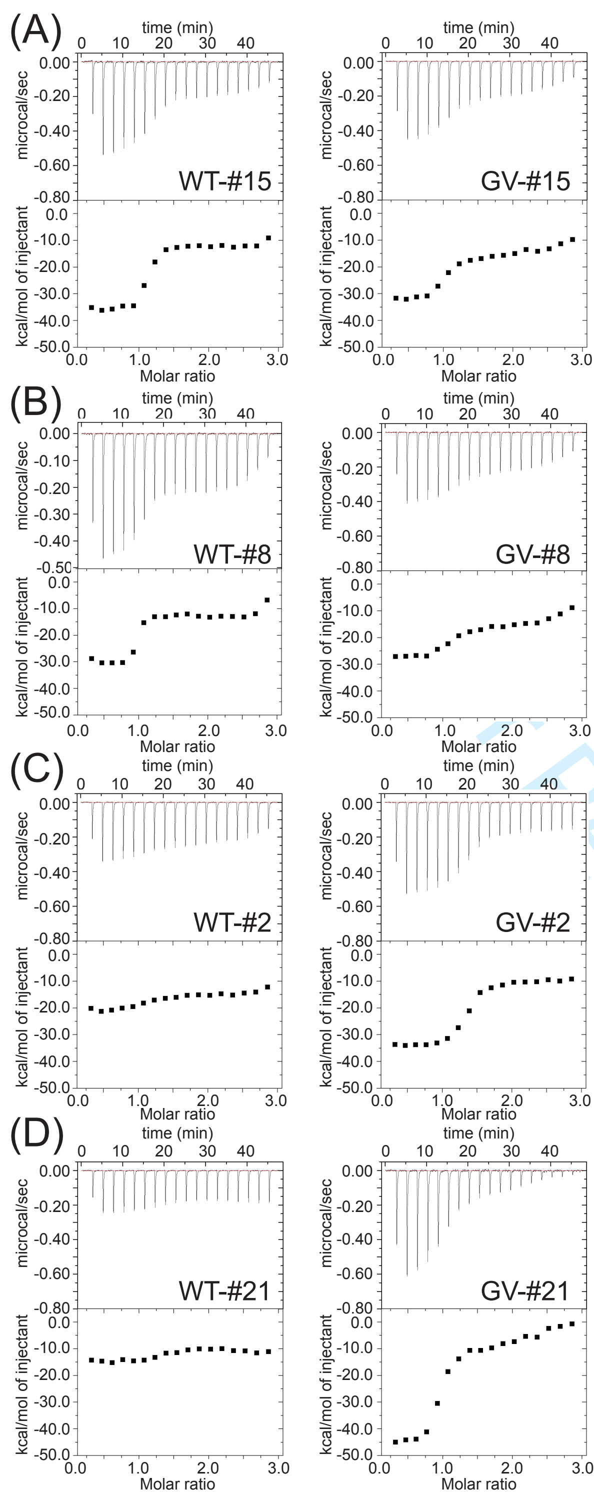
For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



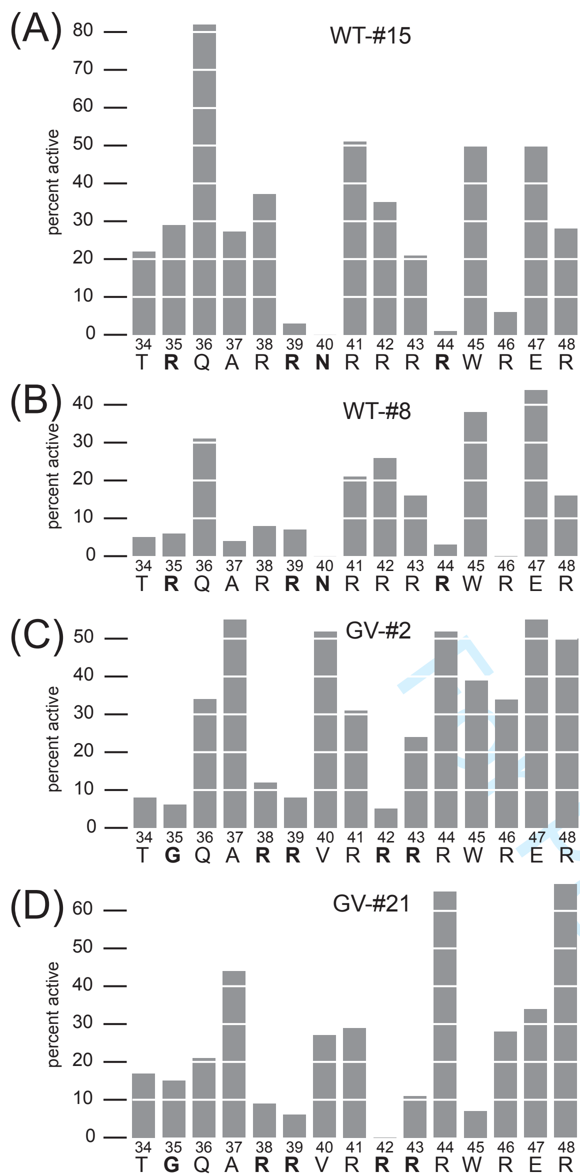
For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



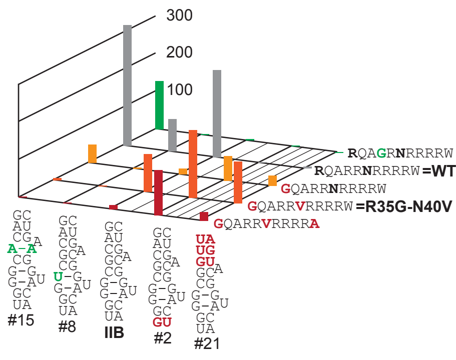
For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



For Peer Review

TABLE S1 Sequencing of library samples without selection[a]

reference[b]	identity[c]	origin[d]	codon[e]	amino acid[f]	NNN[g]	sequence[h]
WT	WT	WT	WT	WT	WT	CCATGGCAACC**CGC**CAGGCCCGTCGT**AAC**CGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
GV	GV	GV	GV	GV	GV	CCATGGCAACC**GGA**CAGGCCCGTCGT**GTC**CGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-1	mutated	WT T34X				CCATGG**G**CA**CCG**CGCCCAGGCCCGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCG
II2614-2	WT T34R	WT T34X	AGG	Arg	AGG	CCATGGCA**AGG**CGCCAGGCCCGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-3	WT T34G	WT T34X	GGC	Gly	GGC	CCATGGCA**GGC**CGCCAGGCCCGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-4	backbone	WT T34X				CCATGGCCTGACTGACTGACTGACGAATGCAGCAAATCCCCTGTTGGTTGGGGTAAGCGCAAAA
II2614-5	WT T34G	WT T34X	GGA	Gly	GGA	CCATGGCA**GGA**CGCCAGGCCCGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-6	mutated	WT T34X				CCATGGCA**ACG**CGCC**C**AGGCCCGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGCGAA
II2614-7	mutated	WT R35X				CCATGGCAACC**ACA**CAGGCCCGTCGTAACCGTAGACGTCGTTGGCGTGAGCG**-**CAGCGTGCAGCTGCGGCGAA
II2614-8	WT R35W	WT R35X	TGG	Trp	TGG	CCATGGCAACC**TGG**CAGGCCCGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-9	WT R35S	WT R35X	AGT	Ser	AGT	CCATGGCAACC**AGT**CAGGCCCGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-10	WT R35C	WT R35X	TGT	Cys	TGT	CCATGGCAACC**TGT**CAGGCCCGTCGTAACCGTAGACGTTGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-11	WT R35V	WT R35X	GTT	Val	GTT	CCATGGCAACC**GTT**CAGGCCCGTCATAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-12	WT R35G	WT R35X	GGA	Gly	GGA	CCATGGCAACC**GGA**CAGGCCCGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-13	WT Q36L	WT Q36X	TTG	Leu	TTG	CCATGGCAACCCGC**TTG**GCCCGTCGTAACCGTAGACGTCGTTGGCGTAAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-14	WT Q36A	WT Q36X	GCT	Ala	GCT	CCATGGCAACCCGC**GCT**GCCCGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-15	WT Q36Y	WT Q36X	TAC	Tyr	TAC	CCATGGCAACCCGC**TAC**GCCCGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-16	WT Q36R	WT Q36X	CGT	Arg	CGT	CCATGGCAACCCGC**CGT**GCCCGTCGTAACCGTAGACGTCGTTGGCCTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-17	WT Q36P	WT Q36X	CCC	Pro	CCC	CCATGGCAACCCGC**CCC**GCCCGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-18	WT Q36D	WT Q36X	GAC	Asp	GAC	CCATGGCAACCCGC**GAC**GCCCGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCCAA
II2614-19	WT A37V	WT A37X	GTT	Val	GTT	CCATGGCAACCCGCCAG**GTT**CGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-20	WT A37V	WT A37X	GTT	Val	GTT	CCATGGCAACCCGCCAG**GTT**CGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-21	WT A37W	WT A37X	TGG	Trp	TGG	CCATGGCAACCCGCCAG**TGG**CGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-22	WT A37G	WT A37X	GGA	Gly	GGA	CCATGGCAACCCGCCAG**GGA**CGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-23	WT A37A	WT A37X	GCT	Ala	GCT	CCATGGCAACCCGCCAG**GCT**CGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-24	WT A37C	WT A37X	TGT	Cys	TGT	CCATGGCAACCCGCCAG**TGT**CGTCGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-25	WT R38P	WT R38X	CCG	Pro	CCG	CCATGGCAACCCGCCAGGCC**CCG**CGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-26	WT R38R	WT R38X	AGG	Arg	AGG	CCATGGCAACCCGCCAGGCC**AGG**CGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-27	WT R38Y	WT R38X	TAT	Tyr	TAT	CCATGGCAACCCGCCAGGCC**TAT**CGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-28	WT R38C	WT R38X	TGT	Cys	TGT	CCATGGCAACCCGCCAGGCC**TGT**CGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-29	mutated	WT R38X				CCATGGCA**-**CCCGCCAGGCC**CTT**CGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-30	WT R38C	WT R38X	TGC	Cys	TGC	CCATGGCAACCCGCCAGGCC**TGC**CGTAACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTACAGCTGCGGCGAA
II2614-31	WT R39M	WT R39X	ATG	Met	ATG	CCATGGCAACCCGCCAGGCCCGT**ATG**AACCGTAGACGTCGTTGGCGTGAGCGTCCGCGTGCAGCTGCGGCGAA

For Peer Review

II2614-32	backbone	WT R39X				CCATGGCCTGACTGACTGACTGACGAATGCAGCAAATCCCCTGTTGGTTGGGGTAAGCGCAA
II2614-33	mutated	WT R39X				CCATG**G**GCAACCCGCCAGGCCCGT**CGG**AACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGA
II2614-34	WT R39L	WT R39X	CTG	Leu	CTG	CCATGGCAACCCGCCAGGCCCGT**CTG**AACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-35	mutated	WT R39X				CCATG**G**GCAACCCGCCAGGCCCGT**GGC**AACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGA
II2614-36	WT R39I	WT R39X	ATT	Ile	ATT	CCATGGCAACCCGCCAGGCCCGT**ATT**AACCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-85	WT N40C	WT N40X	TGC	Cys	TGC	CCATGGCAACCCGCCAGGCCCGTCGT**TGC**CGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-86		WT N40X				no readable sequence
II2614-87	WT N40C	WT N40X	TGC	Cys	TGC	CCATGGCAACCCGCCAGGCCCGTCGT**TGC**CGTAGACGTCGTTGGCGTGAGCGTCAGGGTGCAGCTGCGGCGAA
II2614-88	WT N40C	WT N40X	TGT	Cys	TGT	CCATGGCAACCCGCCAGGCCCGTCGT**TGT**CGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-89	WT N40L	WT N40X	TTG	Leu	TTG	CCATGGCAACCCGCCAGGCCCGTCGT**TTG**CGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-90	WT N40L	WT N40X	TTG	Leu	TTG	CCATGGCAACCCGCCAGGCCCGTCGT**TTG**CGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-37	WT R41T	WT R41X	ACG	Thr	ACG	CCATGGCAACCCGCCAGGCCCGTCGTAAC**ACG**CGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-38	WT R41L	WT R41X	TTA	Leu	TTA	CCATGGCAACCCGCCAGGCCCGTCGTAAC**TTA**AGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-39	WT R41C	WT R41X	TGT	Cys	TGT	CCATGGCAACCCGCCAGGCCCGTCGTAAC**TGT**AGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-40	WT R41Y	WT R41X	TAT	Tyr	TAT	CCATGGCAACCCGCCAGGCCCGTCGTAAC**TAT**AGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-41	WT R41A	WT R41X	GCG	Ala	GCG	CCATGGCAACCCGCCAGGCCCGTCGTAAC**GCG**AGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-42	mutated	WT R41X				CCATGGCAACCCGCCAGGCCCGTC**A**TAAC**AGT**AGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-43	mutated	WT R42X				CCATGGCAACCCGCCAGGCCCGTCGTAAC**T**GT**GTT**CGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-44	WT R42C	WT R42X	TGT	Cys	TGT	CCATGGCAACCCGCCAGGCCCGTCGTAACCGT**TGT**CGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-45	WT R42L	WT R42X	CTG	Leu	CTG	CCATGGCAACCCGCCAGGCCCGTCGTAACCGT**CTG**CGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-46	mutated	WT R42X				CCATGGCAACCCGCCAGGCCCGTCGTAACCGT**CAC**CGT**T**GTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-47	WT R42L	WT R42X	TTG	Leu	TTG	CCATGGCAACCCGCCAGGCCCGTCGTAACCGT**TTG**CGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-48	WT R42H	WT R42X	CAT	His	CAT	CCATGGCAACCCGCCAGGCCCGTCGTAACCGT**CAT**CGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-49	WT R43E	WT R43X	GAA	Glu	GAA	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGA**GAA**CGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-50	WT R43*	WT R43X	TAA	Stop	TAA	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGA**TAA**CGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-51	WT R43L	WT R43X	CTG	Leu	CTG	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGA**CTG**CGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-52	mutated	WT R43X				CCATGGCAACCCGCCAGG**G**CC**T**GTCGTAACCGTAGA**ACA**CGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGA
II2614-53	WT R43E	WT R43X	GAA	Glu	GAA	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGA**GAA**CGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-54	WT R43N	WT R43X	AAC	Asn	AAC	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGA**AAC**CGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-55	mutated	WT R44X				CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGT**CGT**TGGCGTGAGCGTCAGCGTGCAGCTGCGG**G**CGA
II2614-56	mutated	WT R44X				CCATGGCAACCCGCCAGGCCCGTCGTAACCGTA**A**ACGT**GCC**TGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-57	WT R44T	WT R44X	ACA	Thr	*TGT*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGT**ACA**TGGCGTGAGCGTCAGCGTACAGCTGCGGCGAA
II2614-58	unknown	WT R44X				unknown vector-related sequence
II2614-59	mutated	WT R44X				CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGT**CCA**TGGC**AACCCGCCAGGCCCGTCGTAACCGTAG**
II2614-60	WT R44S	WT R44X	AGC	Ser	*GCT*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGT**AGC**TGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-61	WT W45T	WT W45X	ACA	Thr	*TGT*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGT**ACA**CGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-62	mutated	WT W45X				CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGT**AGA**CGTGAGCGTCA**TTTACGACGTTACGGTTA**
II2614-63	WT W45P	WT W45X	CCT	Pro	*AGG*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGT**CCT**CGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-64	WT W45V	WT W45X	GTT	Val	*AAC*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGT**GTT**CGTGAGCGTCAGCGTGCAGCTGCGGCGAA

For Peer Review

II2614-65	WT W45P	WT W45X	CCC	Pro	*GGG*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGT**CCC**CGTGAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-66	mutated	WT W45X				CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGT**ACA**CGTGAGCGTCAG**-------**CTGCGGCGAA
II2614-67	WT R46P	WT R46X	CCA	Pro	*TGG*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGTTGG**CCA**GAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-68	WT R46K	WT R46X	AAA	Lys	*TTT*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGTTGG**AAA**GAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-69	WT R46T	WT R46X	ACA	Thr	*TGT*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGTTGG**ACA**GAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-70	mutated	WT R46X				CCATGGCAACCCGCCAGGCCCGTCGTAACC**-**TAGACGTCGTTGG**TCA**GAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-71	WT R46*	WT R46X	TAA	Stop	*TTA*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAAACGTCGTTGG**TAA**GAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-72	WT R46C	WT R46X	CCC	Pro	*GGG*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGTTGG**CCC**GAGCGTCAGCGTGCAGCTGCGGCGAA
II2614-73	WT E47V	WT E47X	GTT	Val	*AAC*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGTTGGCGT**GTT**CGTCAGCGTGCAGCTGCGGCGAA
II2614-74	WT E47*	WT E47X	TAA	Stop	*TTA*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGTTGGCGT**TAA**CGTCAGCGTGCAGCTGCGGCGAA
II2614-75	WT E47*	WT E47X	TAA	Stop	*TTA*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGTTGGCGT**TAA**CGTCAGCGTGCAGCTGCGGCGAA
II2614-76	mutated	WT E47X				CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGT**T**GTTGGCGT**TAA**CGTCAGCGTGCAGCTGCGGCGAA
II2614-77	WT E47R	WT E47X	AGA	Arg	*TCT*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGTTGGCGT**AGA**CGTCAGCGTGCAGCTGCGGCGAA
II2614-78	WT E47*	WT E47X	TAA	Stop	*TTA*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGTTGGCGT**TAA**CGTCAGCGTGCAGCTGCGGCGAA
II2614-79	mutated	WT R48X				CCATGGCAACCCGCCAGGCCCGTCGTAACCG**-**AGACGTCGTTGGCGTGAG**CCA**CAGCGTGCAGCTGCGGCGAA
II2614-80	WT R48T	WT R48X	ACC	Thr	*GGT*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGTTGGCGTGAG**ACC**CCGCGTGCAGCTGCGGCGAA
II2614-81	mutated	WT R48X				CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGTTGGCGTGAG**AAT**C**A**GCGTGCAGC**C**TGCGGCCG
II2614-82	mutated	WT R48X				CCATGGCAACCCGCCAGGCC**-**GTCGTAACCGTAGACGTCGTTGGCGTGAGA**CAC**AGCGTGCAGCTGCGGCGAA
II2614-83	WT R48S	WT R48X	TCC	Ser	*GGA*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGTTGGCGTGAG**TCC**CAGCGTGCAGCTGCGGCGAA
II2614-84	WT R48T	WT R48X	ACA	Thr	*TGT*	CCATGGCAACCCGCCAGGCCCGTCGTAACCGTAGACGTCGTTGGCGTGAG**ACA**CAGCGTGCAGCTGCGGCGAA
ii2617-1	GV R34S	GV T34X	TCG	Ser	TCG	CCATGGCA**TCG**GGACAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-2	GV R34V	GV T34X	GTA	Val	GTA	CCATGGCA**GTA**GGACAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-3	GV R34T	GV T34X	ACA	Thr	ACA	CCATGGCA**ACA**GGACAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-4		GV T34X				no readable sequence
ii2617-5	GV R34N	GV T34X	AAC	Asn	AAC	CCATGGCA**AAC**GGACAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-6		GV T34X				CCATGGCA**-AC**GGGACAGGCCCGTCGTGTCCGAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2620-4	GV R35G	GV G35X	GGC	Gly	GGC	CCATGGCAACC**GGC**CAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2620-5	GV R35T	GV G35X	ACC	Thr	ACC	CCATGGCAACC**ACC**CAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2620-6	GV R35R	GV G35X	CGT	Arg	CGT	CCATGGCAACC**CGT**CAGGCCCGTCGTGTCCGTAGACGTCGCTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2620-7		GV G35X				no readable sequence
ii2620-8		GV G35X				no readable sequence
ii2620-9	mix	GV G35X				not readable: mix of target sequences
ii2620-10	GV R35C	GV G35X	TGC	Cys	TGC	CCATGGCAACC**TGC**CAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2620-11	mutated	GV G35X				CCATGGCAACC**CGA**CAGGCCCGTCGTGTCCGTA**-**ACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2620-12	mutated	GV G35X				CCATGGCAACC**GGTG**CAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGA
ii2620-13	mutated	GV G35X				CCATGGCAACC**GGG**CAGGCC**T**GTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2620-14	GV R35	GV G35X	CAC	His	CAC	CCATGGCAACC**CAC**CAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2620-15		GV G35X				no readable sequence
ii2617-7	mutated	GV Q36X				CCATGGCAACCGGA**TAA**GCCC**C**GTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGA

For Peer Review

ii2617-8	mutated	GV Q36X				CCATGGCAACCGGA**CCG**GCCCGTCGTGTCCGT**-**GACGCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-9	GV R36M	GV Q36X	ATG	Met	ATG	CCATGGCAACCGGA**ATG**GCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-10	GV R36L	GV Q36X	CTG	Leu	CTG	CCATGGCAACCGGA**CTG**GCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-11	mix	GV Q36X				not readable: mix of target sequences
ii2617-12	GV R36L	GV Q36X	TTG	Leu	TTG	CCATGGCAACCGGA**TTG**GCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-13	backbone	GV A37X				CCATGGGTCGTCCTCGTGGTACCCGCGGTAAAGGTCGCCGTATTCGCCGTGGTGGCGGGAA
ii2617-14	GV A37S	GV A37X	TCG	Ser	TCG	CCATGGCAACCGGACAG**TCG**CGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-15	GV A37V	GV A37X	GTG	Val	GTG	CCATGGCAACCGGACAG**GTG**CGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-16	GV A37S	GV A37X	TCT	Ser	TCT	CCATGGCAACCGGACAG**TCT**CGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-17	GV A37R	GV A37X	AGG	Arg	AGG	CCATGGCAACCGGACAG**AGG**CGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-18	GV A37R	GV A37X	AGA	Arg	AGA	CCATGGCAACCGGACAG**AGA**CGTCGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-19	GV A38S	GV R38X	AGT	Ser	AGT	CCATGGCAACCGGACAGGCC**AGT**CGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-20	GV R38K	GV R38X	AAG	Lys	AAG	CCATGGCAACCGGACAGGCC**AAG**CGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-21	GV R38M	GV R38X	ATG	Met	ATG	CCATGGCAATCGGACAGGCC**ATG**CGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-22	GV R38L	GV R38X	TTG	Leu	TTG	CCATGGCAACCGGACAGGCC**TTG**CGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-23	GV R38L	GV R38X	CTG	Leu	CTG	CCATGGCAACCGGACAGGCC**CTG**CGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-24	GV R38R	GV R38X	CGT	Arg	CGT	CCATGGCAACCGGACAGGCC**CGT**CGTGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-25	mutated	GV R39X				CCATGG**G**CAACCGGACAGGCCCGTTCGGTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGGAA
ii2617-26	GV R39Y	GV R39X	TAT	Tyr	TAT	CCATGGCAACCGGACAGGCCCGT**TAT**GTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-27	GV R39Q	GV R39X	CAG	Gln	CAG	CCATGGCAACCGGACAGGCCCGT**CAG**GTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-28	mutated	GV R39X				CCATGG**G**CAACCGGACAGGCCCGT**GGT**GTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGA
ii2617-29	mutated	GV R39X				CCATGG**G**CAA**A**CCGGACAGGCCC**G**GT**GCT**GTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGC
ii2617-30	GV R39A	GV R39X	GCC	Ala	GCC	CCATGGCAACCGGACAGGCCCGT**GCC**GTCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-31	mutated	GV V40X				CCATGGCAACCGGACAGGCCCGTCGT**ATG**CGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAACTGCGGCGAA
ii2617-32	mutated	GV V40X				CCATGGCAACCGGACAGGCC**-**GTCGTCCCGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-33	GV R40L	GV V40X	CTT	Leu	CTT	CCATGGCAACCGGACAGGCCCGTCGT**CTT**CGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-34	GV R40A	GV V40X	GCG	Ala	GCG	CCATGGCAACCGGACAGGCCCGTCGT**GCG**CGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-35	GV R40L	GV V40X	TTA	Leu	TTA	CCATGGCAACCGGACAGGCCCGTCGT**TTA**CGTAGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-36		GV V40X				no readable sequence
ii2617-37	GV R41S	GV R41X	TCG	Ser	TCG	CCATGGCAACCGGACAGGCCCGTCGTGTC**TCG**AGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-38	GV R41G	GV R41X	GGT	Gly	GGT	CCATGGCAACCGGACAGGCCCGTCGTGTC**GGT**AGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-39	mutated	GV R41X				CCATGGCAACCGGACAGGCCCGTCGTGTCGGAAGA**-**GTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-40	GV R41*	GV R41X	TAG	Stop	TAG	CCATGGCAACCGGACAGGCCCGTCGTGTC**TAG**AGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-41	GV R41L	GV R41X	TTA	Leu	TTA	CCATGGCAACCGGACAGGCCCGTCGTGTC**TTA**AGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-42	GV R41S	GV R41X	TCC	Ser	TCC	CCATGGCAACCGGACAGGCCCGTCGTGTC**TCC**AGACGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-43	GV R42F	GV R42X	TTT	Phe	TTT	CCATGGCAACCGGACAGGCCCGTCGTGTCCGT**TTT**CGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-44	backbone	GV R42X				CCATGGGTCGTCCTCGTGGTACCCGCGGTAAAGGTCGCCGTATTCGCCGTGGTGGCGGGAA
ii2617-45	GV R42A	GV R42X	GCC	Ala	GCC	CCATGGCAACCGGACAGGCCCGTCGTGTCCGT**GCC**CGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-46	GV R42T	GV R42X	ACC	Thr	ACC	CCATGGCAACCGGACAGGCCCGTCGTGTCCGT**ACC**CGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA

For Peer Review

ii2617-47	mutated	GV R42X				CCATGGCAACCGGACAGGCCCGTCGTGTCCGTATCCTCCGTGGCGTGAACGGCCGCGTGCAGCTGCCGCGAA
ii2617-48	GV R42R	GV R42X	AGA	Arg	AGA	CCATGGCAACCGGACAGGCCCGTCGTGTCCGT**AGA**CGTCGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-49	GV R43*	GV R43X	TAA	Stop	TAA	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGA**TAA**CGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-50	backbone	GV R43X				CCATGGGTCGTCCTCGTGGTACCCGCGGTAAAGGTCGCCGTATTCGCCGTGGTGGCGGGAA
ii2617-51	mutated	GV R43X				CCATGGG**AACCGGACCGGGCCCGCCTGGCCCGAAAACGCCTTTGCCGTAACCGCCACGGGGCAGTTCGGGGGA**
ii2617-52	GV R43S	GV R43X	TCG	Ser	TCG	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGA**TCG**CGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-53	GV R43P	GV R43X	CCT	Pro	CCT	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGA**CCT**CGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-54	GV R43K	GV R43X	AAG	Lys	AAG	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGA**AAG**CGTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-55	mutated	GV R44X				CCATGGCAACCGGACAGGGCCCGTCGTGTCCGTAGACGTTCTTGGCGTGAGCGTCAGCGTGCAGCTGCGGCGA
ii2617-56	GV R44A	GV R44X	GCT	Ala	*AGC*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGT**GCT**TGGCGTGAGCGTCAGCGTGCAGCTGCGGC**A**AA
ii2617-57	backbone	GV R44X				CCATGGGTCGTCCTCGTGGTACCCGCGGTAAAGGTCGCCGTATTCGCCGTGGTGGCGGGAATGCAGCAAA
ii2617-58	GV R44N	GV R44X	AAT	Asn	*ATT*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGT**AAT**TGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-59	GV R44R	GV R44X	CGC	Arg	*GCG*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGT**CGC**TGGCGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-60	GV R44N	GV R44X	AAT	Asn	*ATT*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGT**AAT**TGGCGTGAGCGTCAGCGTGCAGCTGGGGCGAA
ii2617-61	GV R45C	GV W45X	TGC	Cys	*GCA*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGT**TGC**CGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-62	mutated	GV W45X				CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGT**CGA**CGTGAGCGTCAGCGTGCAGCTGC**C**GGCGA
ii2617-63	GV R45F	GV W45X	TTC	Phe	*GAA*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGT**TTC**CGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-64	GV R45V	GV W45X	GTA	Val	*TAC*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGT**GTA**CGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-65	mutated	GV W45X				CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGT**ATG**CGTGAGCGTCAGCGTGCAGCTGCGGC**C**GA
ii2617-66	GV R45I	GV W45X	ATC	Ile	*GAT*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGT**ATC**CGTGAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-67	GV R46*	GV R46X	TAA	Stop	*TTA*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGG**TAA**GAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-68	GV R46H	GV R46X	CAC	His	*GTG*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGG**CAC**GAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-69	GV R46P	GV R46X	CCG	Pro	*CGG*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGG**CCG**GAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-70	GV R46P	GV R46X	CCC	Pro	*GGG*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGG**CCC**GAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-71	GV R46A	GV R46X	GCA	Ala	*TGC*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGG**GCA**GAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-72	GV R46H	GV R46X	CAC	His	*GTG*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGG**CAC**GAGCGTCAGCGTGCAGCTGCGGCGAA
ii2617-73	GV E47T	GV E47X	ACC	Thr	*GGT*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGT**ACC**CGTCAGCGTGCAGCTGCGGCGAA
ii2617-74	GV E47T	GV E47X	ACC	Thr	*GGT*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGT**ACC**CGTCAGCGTGCAGCTGCGGCGAA
ii2617-75	backbone	GV E47X				CCATGGGTCGTCCTCGTGGTACCCGCGGTAAAGGTCGCCGTATTCGCCGTGGTGGCGGGAATGCAGCAAA
ii2617-76	GV E47F	GV E47X	TTC	Phe	*GAA*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGT**TTC**CGTCAGCGTGCAGCTGCGGCGAA
ii2617-77	GV E47T	GV E47X	ACC	Thr	*GGA*	CCATGGCAACCGGACAGGGCCGTCGTGGCCGTAGACGTCGTTGGCGT**ACC**CGTCAGCGTGCAGCTGCGGCGAA
ii2617-78	mutated	GV E47X				CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTAGCCGTCAGCGTGCAGCTGCGGCGAA
ii2617-79	GV R48P	GV R48X	CCC	Pro	*GGG*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAG**CCC**CAGCGTGCAGCTGCGGCGAA
ii2617-80	GV R48S	GV R48X	TCA	Ser	*TGA*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAG**TCA**CAGCGTGCAGCTGCGGCGAA
ii2617-81	GV R48A	GV R48X	GCC	Ala	*GGC*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAG**GCC**CAGCGTGCAGCTGCGGCGAA
ii2617-82	backbone	GV R48X				CCATGGGTCGTCCTCGTGGTACCCGCGGTAAAGGTCGCCGTATTCGCCGTGGTGGCGGGAA
ii2617-83	GV R48A	GV R48X	GCA	Ala	*TGC*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAG**GCA**CAGCGTGCAGCTGCGGCGAA
ii2617-84	GV R48G	GV R48X	GGC	Gly	*GCC*	CCATGGCAACCGGACAGGCCCGTCGTGTCCGTAGACGTCGTTGGCGTGAG**GGC**CAGCGTGCAGCTGCGGCGAA

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

[a]The 15 WT and 15 GV plasmid libraries were transformed into cells without reporter. At least 6 colonies from each transformation were chosen randomly and used to prepare plasmid for sequencing to confirm the identities and diversity of targeted positions.
[b]The laboratory reference number of the sequenced samples.
[c]The expressed ARM identity of the sample as determined by sequencing using single-letter codes for amino acids and asterisks for stop codons. The absence of an entry means the sample did not yield a readable sequence; "mutated" refers to sequences with mutations that alter the expressed ARM at other than the targeted codon; "backbone" refers to a sequence of the cloning plasmid without a library insert; "unknown" refers to a sequence of uncertain origin matching common vectors and that may be located on the cloning vector used here.
[d]The library from which the clone was chosen.
[e]For each sequence representing a target, the codon of the randomized position is shown.
[f]For each sequence representing a target, the amino acid expressed at the randomized position is shown.
[g]For each sequence representing a target, the sequence of the randomized cloning oligonucleotide is shown: libraries 34-43 were constructed with the sense oligonucleotides being randomized, and libraries 44-48 were constructed with the antisense oligonucleotides being randomized. Of the 131 triplets comprising 393 nucleotides, 73 are A, 65 are C, 130 are G, and 125 are T.
[h]The sequence of the library insert is shown with any difference from the base sequence of respective original WT or GV indicated in bold, red font. Hyphens indicated deletions. In the WT and GV sequences, the codons of positions 35 and 40 are indicated with underlined, bold font.

For Peer Review

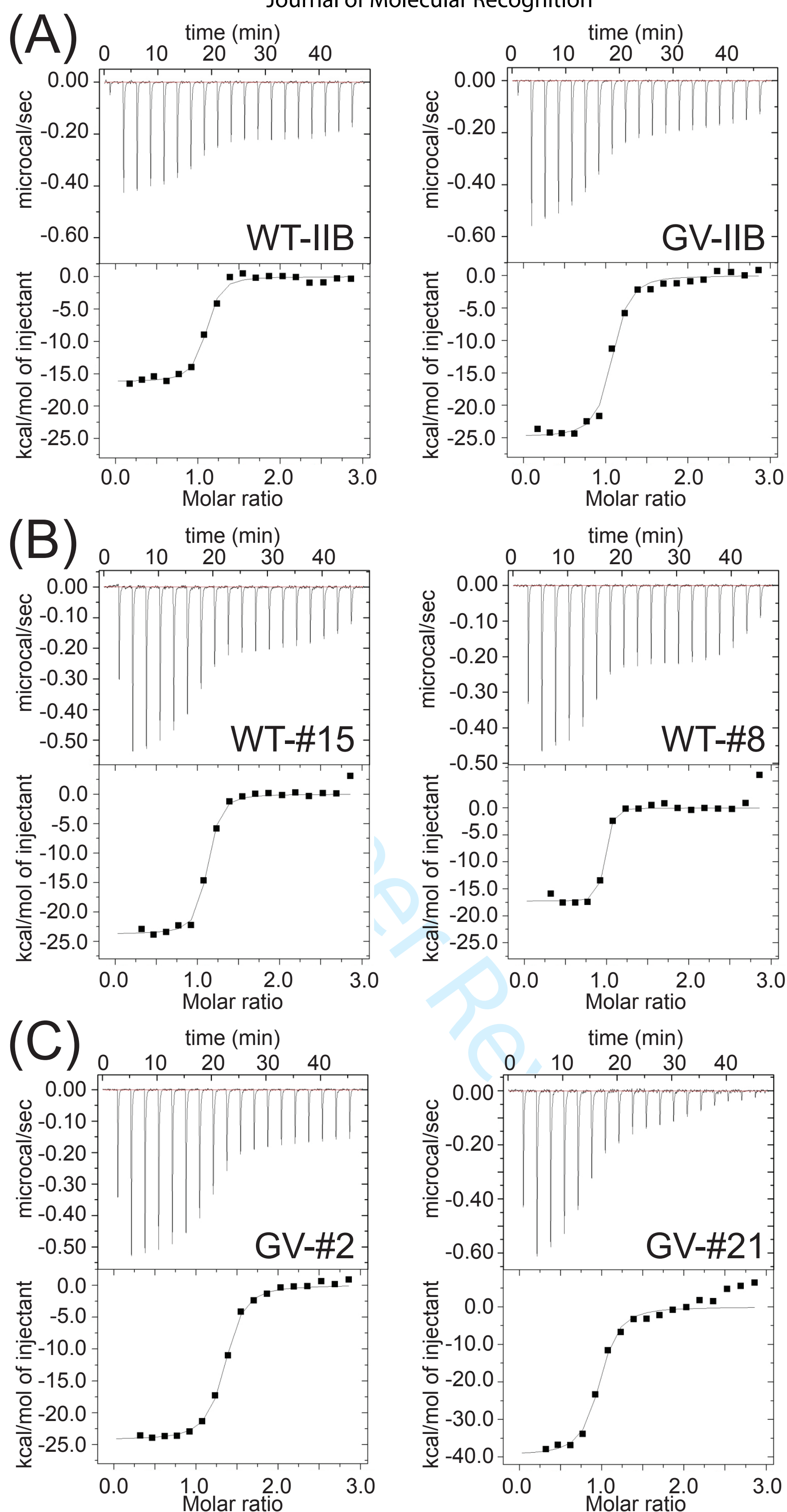


Figure S1. Isothermal titration calorimetry of WT and GV with RNAs. Top panels show raw calorimetry traces, and bottom panels show plots of integrated heat values after subtracting the estimated heat of non-specific binding. (A) Peptides with RRE IIB RNA, left: WT and right: GV. (B) WT peptides with mutant RNAs, left: #15 (G50A-C69A) and right: #8 (G48U). (C) GV peptides with mutant RNAs, left: #2 (U45G-A75U) and right: #21 (C51G-A52U-G53U-C65A-U66G-G67U). The heat of non-specific binding was estimated from the heat observed during the late, flat portion of the titration plot. Thermodynamic parameters in Table 3 were obtained by the curve fitting of the plots after subtracting the estimated heat of non-specific binding.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

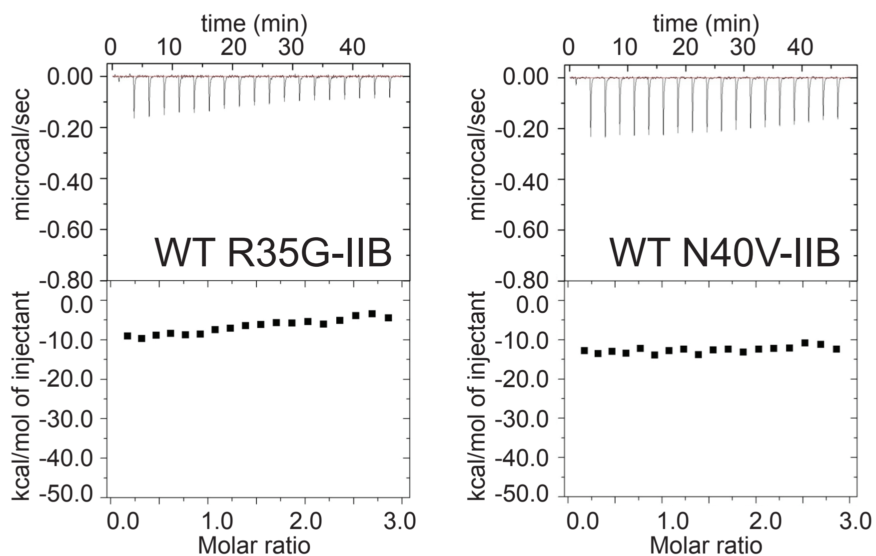
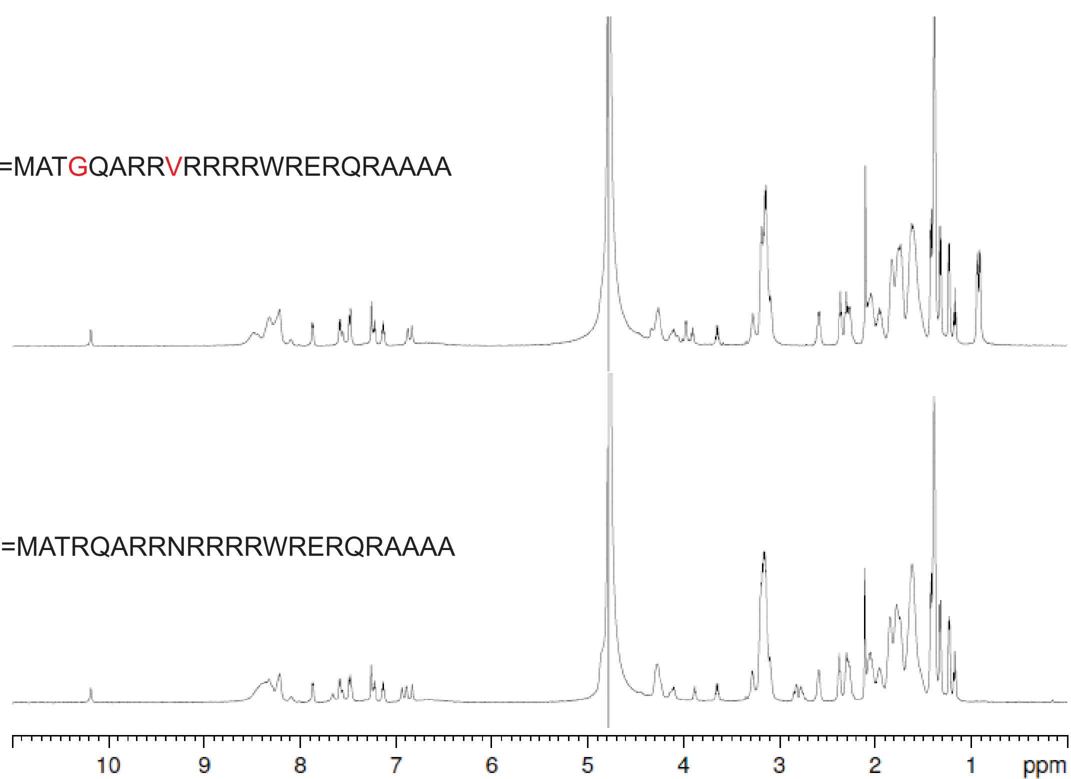


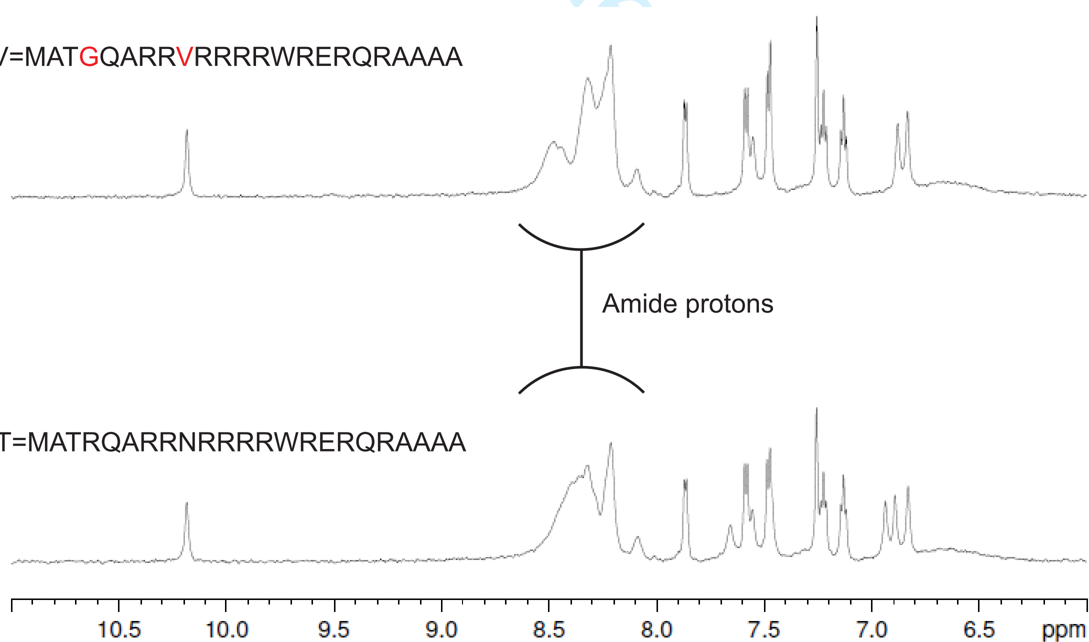
Figure S2. Isothermal titration calorimetry of WT R35G and WT N40V with RRE IIB RNA. Top panels show raw calorimetry traces, and bottom panels show plots of integrated heat values. Left panels show WT R35G and right panels show WT N40V.

For Peer Review

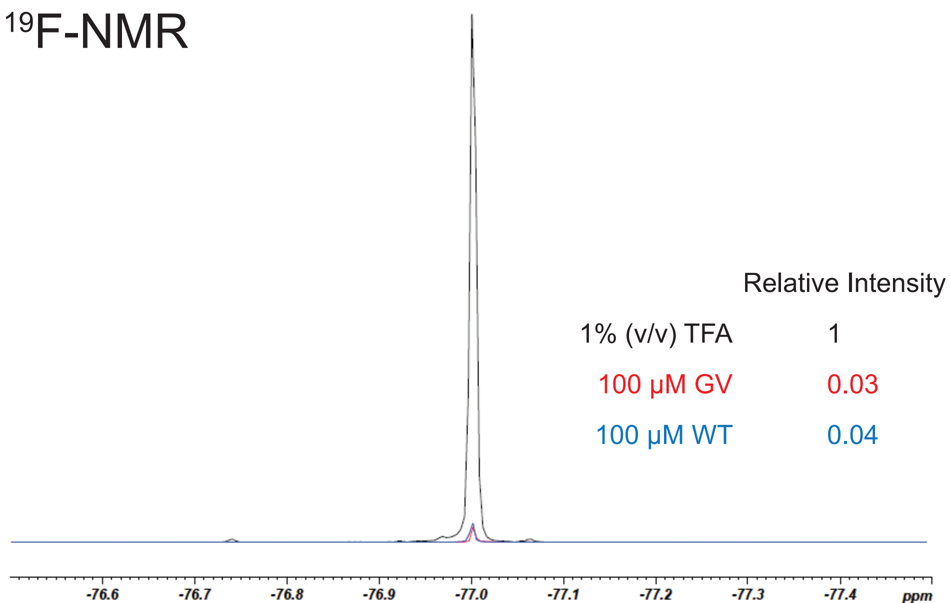
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

H-NMR





1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



For Peer Review