Role of Mutations in Differential Recognition of Viral RNA Molecules by Peptides

Amit Kumar and Harish Vashisth*

Department of Chemical Engineering, University of New Hampshire, Durham 03824, New Hampshire, USA

E-mail: harish.vashisth@unh.edu

Phone: +1-603-862-2483

Abstract

The conserved non-coding RNA elements in viral genomes interact with proteins to regulate various events during viral replication. We report studies on the recognition mechanisms of two helical peptides, a native (Rev) peptide and a lab-evolved (RSG1.2) peptide, by a highly conserved viral RNA element from the HIV-1 genome. Specifically, we investigated the physical interactions between the viral RNA molecule and helical peptides by computing free energy changes on mutating key amino-acid residues involved in recognition of an internal loop in the viral RNA molecule.

Introduction

Ribonucleic acid (RNA) is an essential biomolecule for encoding and transmitting genetic information in RNA viruses, including Ebola,[1] Influenza,[2] HIV,[3] and SARS-CoV-2.[4] As a common feature, most of the RNA functions rely on interactions with proteins. The highly conserved non-coding RNA (ncRNA) elements in viral genomes fold into complex three-dimensional structures to regulate various aspects of infection, including virus replication, persistence, and pathogenesis through their molecular interactions with RNA binding proteins.[5–9] As a result, viral RNA is associated with various proteins throughout its life cycle. Therefore, understanding the RNA-protein interaction mechanisms is essential to deciphering various aspects of the viral infection as well as for designing novel anti-viral therapeutics.

Many RNA binding proteins recognize RNA molecules through arginine-rich motifs (ARMs) as they are relatively shorter in length, have little sequence similarity (aside from containing many arginine residues), and adopt diverse conformations (including an α-helix and a β-hairpin) upon association with the viral RNA.[10–16] The experimental work using x-ray crystallography and nuclear magnetic resonance (NMR) methods on the peptide-RNA complexes has demonstrated that the RNA backbone interacts with peptides more commonly than the nucleotide bases, suggesting that the majority of peptide-RNA interactions are non-specific.[17,18] Overall, these studies showed that the peptide-RNA interactions occur through dynamic rearrangements in both molecules, which often entail backbone shifts and the flipping of nucleobases and amino-acid residues.[19,20] Moreover, recent studies have demonstrated that the β-hairpin peptides could be synthesized with ultra-high affinity toward viral RNA molecules.[21–24] However, *de novo* design of α-helical peptides that potently and selectively recognize viral RNA molecules is limited due to a poor understanding of the RNA recognition mechanism.

In this regard, structured RNA elements from the HIV genome[3,25,26] serve as good model systems[16,27,28] to study the mechanisms of recognition of viral RNA by peptides. The HIV-1 Rev-responsive RNA element (RRE RNA), a 350 nucleotide long RNA found in the *env*

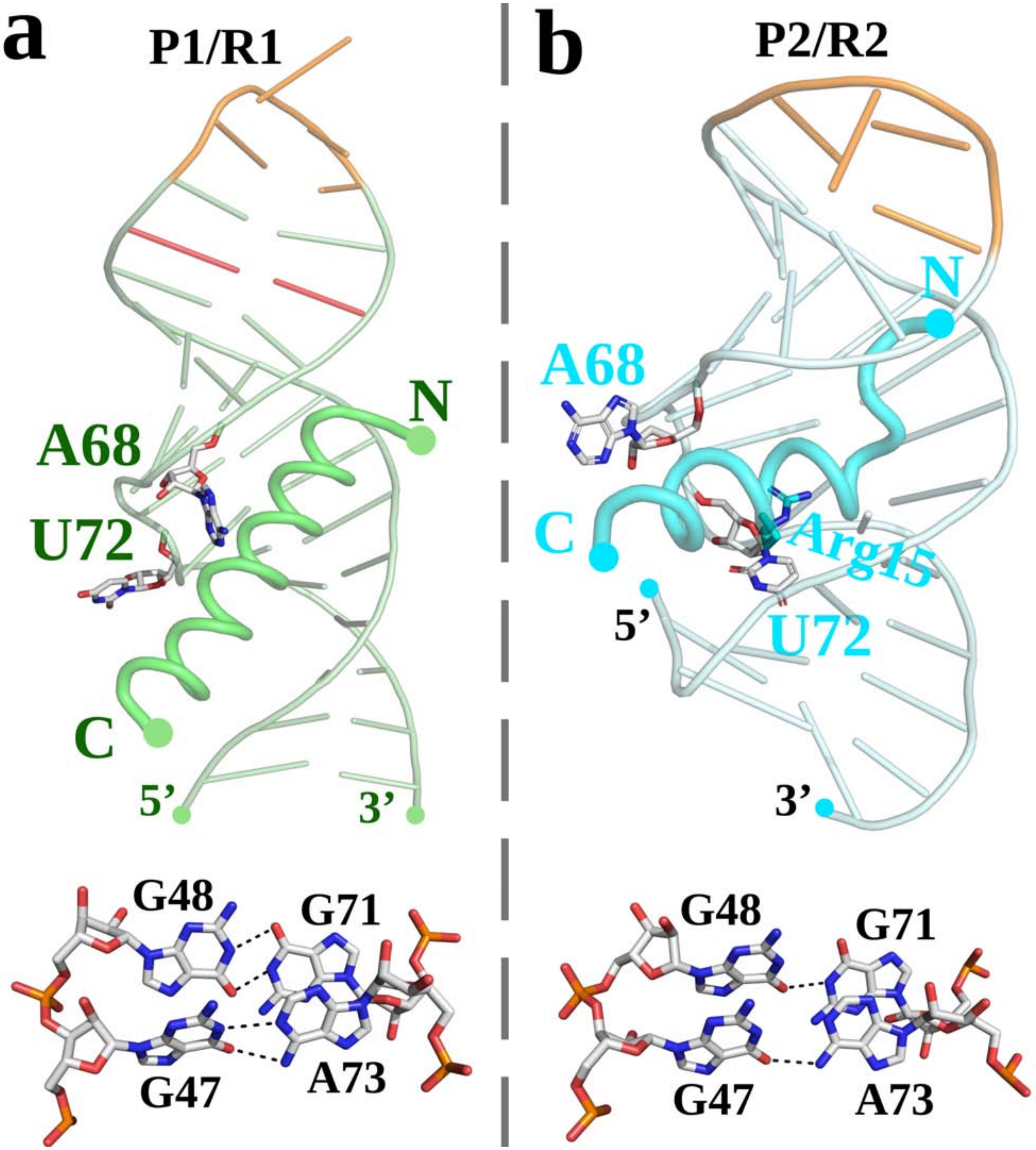

Figure 1: **Structural details of the peptide-RNA complexes.** Cartoon representations of the tertiary structures of peptides P1 (panel a) and P2 (panel b) bound to RRE RNA. The bulge nucleotides (ADE68 and URA72), purine-purine base-pairs, and key amino-acids are shown in stick representations. In the P1/R1 complex (panel a), the unpaired nucleobase URA72 is flipped away from the major groove of RNA, while in the P2/R2 complex (panel b) it is stablilized in the RNA major groove via stacking interactions with the side-chain of Arg15.

gene, interacts with the arginine-rich motif of the Rev protein to regulate nuclear export of the unspliced mRNA during the HIV life cycle and is therefore considered to be an important drug target.[29–34]

The NMR structures of oligonucleotide constructs (**Figure 1** and **Figure S1**), RRE-TR-0 (termed R1) and RRE-TR-4 (termed R2), representing the stem IIB Rev-binding site of HIV-1 RRE in complex with the native Rev peptide (termed P1)[11] and the lab-evolved RSG1.2 peptide (termed P2)[35] have been reported. The P2 peptide was evolved via selection of the P1 peptide against the RRE target and subsequent mutations.[36] The P2 peptide binds RRE RNA with a 7-fold higher affinity and 15-fold higher specificity than the P1 peptide. The NMR structures of the P1/R1 and P2/R2 complexes suggest that these peptides bind in an α-helical conformation to the RNA major groove formed by the non-canonical purine-purine base pairs, GUA47-ADE73 and GUA48-GUA71 (**Figure 1, bottom panels**). Specifically, these NMR structures reveal that the P1 and P2 peptides employ a combination of base-specific hydrogen bonding, purine-purine base-pair shape recognition, and some van der Waals interactions for recognizing the same RNA target.[11,35,36] Furthermore, the NMR structures show that the binding of the P2 peptide results in the flipping of the nucleotide URA72 into the major groove, thereby facilitating the stacking of URA72 with the Arg15 residue (**Figure 1b**), whereas in the P1/R1 complex URA72 is flipped-out and away from the major groove (**Figure 1a**).[11,35,37,38] Biochemical studies have suggested that a network of both electrostatic and hydrophobic interactions is involved in the recognition of RNA major groove by the P2 peptide.[35,36] Further, these studies suggested that replacing alanine residues with glycine residues in the P2 peptide results in a loss of binding affinity for RRE RNA.[35,36] However, the basis of a higher affinity and specificity of the P2 peptide in comparison to the P1 peptide for RRE RNA is still unknown. We have probed these questions using all-atom explicit-solvent molecular dynamics (MD) simulations (**Figure S2**).

Specifically, we report free energy changes on mutations in each peptide to provide a detailed explanation for a higher affinity of the P2 peptide over the P1 peptide for RRE

RNA. To compute the free energy changes, we used the free energy perturbation (FEP) method[24,39,40] with an appropriate thermodynamic cycle (**Figure S3**) to quantify the relative changes in the binding affinities of peptides for RRE RNA. We have also studied the dynamics of unliganded RNA molecules (R1 and R2) and the peptide-RNA complexes (P1/R1 and P2/R2) by all-atom MD simulations to further probe conformational diffferences (**Table S1** and **Figure S2**). With this aim, we have performed an aggregate of ∼50 μs of free energy calculations and all-atom unbiased MD simulations. Based on these data, we provide the energetic and conformational basis for an improved understanding of RRE RNA recognition by each of the two peptides.

Materials and Methods

Molecular Dynamics Setup

We conducted all-atom MD simulations of two unbound RRE RNA molecules (RRE-TR-0, termed R1; and RRE-TR-4 termed R2) and two peptides (Rev, termed P1; and RSG1.2 termed P2) bound RRE complexes (P1/R1 and P2/R2). We derived the initial coordinates for the P1/R1 complex from the NMR structure (PDB code: 1ETF) by Battiste et al.,[11] and for the P2/R2 complex from the NMR structure (PDB code: 1G70) by Gosser et al.[35] We obtained the coordinates of RRE RNA molecules (R1 and R2) by removing the peptides from the peptide-RRE complexes. We then solvated all systems with explicit TIP3P water molecules in periodic simulation domains (**Figure S2**). To neutralize each system, we placed Mg^{2+} and Cl^- ions into the minima of the electrostatic potential computed using the *meadionize* plugin in VMD.[41] We further maintained ionic concentration by adding K^+ and Cl^- ions (150 mM KCl) to bulk water. The structures of P1/R1[11] and P2/R2[35] complexes were determined at a pH of 5.5 and 6.0, respectively. The low pH could result in the protonation of charged residues (such as Asp and Glu). Therefore, we used PropKa server for calculating the pKa values of charged residues in the peptide-RNA complexes

(P1/R1 and P2/R2).[42] The proPka server predicted the pKa values for Asp and Glu lower than 4.0, which are significantly lower than the pH used for determining the structures. This suggests unprotonated states for Asp and Glu residues in the P1/R1 and P2/R2 complexes. Therefore, we considered unprotonated states of negatively charged residues (i.e., Asp and Glu) in our simulation setup. The final system sizes are given in **Table S1**, and the overall simulation setup is shown in **Figure S2**.

Before conducting all-atom MD simulations, we energy minimized each system via the conjugate gradient minimization algorithm for 2000 steps. During the initial phase (30 ns) of MD equilibration, we restrained ($k = 10$ kcal mol^{-} Å^{2}) the C$_{\alpha}$ atoms for peptides and the P-atoms in RNA. We did not use any restraints during production MD simulations. We carried out all conventional MD simulations using the CHARMM36[43] force field with a 2-fs time-step in the NPT ensemble. We maintained the temperature and pressure at 310 K and 1 bar using the Langevin thermostat and Nose-Hoover barostat. We used periodic boundary conditions in all simulations and computed the long-range electrostatic interactions using the Particle Mesh Ewald method.[44] For the van der Waals interactions, we used a cut-off of 16 Å with smooth switching taking effect at 15 Å.

We subjected each simulation model (two unbound RNA constructs and two RNA-peptide complexes) to a 2 μs long MD trajectory and saved configurations every 20 ps. We generated four trajectories for each model with an aggregate time of 32 μs. We used the VMD software[41] for generating input files. We performed visualization, analysis, and post-processing of the simulation trajectories using VMD,[41] CPPTRAJ,[45] and Pymol.[46] We used NAMDv2.13[47] for conducting conventional MD simulations and performed free energy calculations using a new GPU implementation of the free energy perturbation (FEP) method in NAMDv3.0.[48]

Alchemical Free Energy Calculations

Using the alchemical free energy simulation method,[39] we determined the energetic contribution of each amino-acid residue in both peptides for their binding to RRE RNA. Specifically, we designed a thermodynamic cycle (**Figure S3**) where the vertical arms of the cycle correspond to binding of each peptide to RNA, and the horizontal arms correspond to the alchemical transformation of a wild-type amino-acid into a mutated amino-acid in a given peptide. Since the physical binding of a peptide (vertical arms in **Figure S3**) is not the focus of this study, we computed the free energy changes along the unphysical pathways in complex (upper horizontal arm in **Figure S3**, ΔG_{comp}) and free peptide in an aqueous environment (lower horizontal arm in **Figure S3**, ΔG_{free}) as $\Delta\Delta G = \Delta G_{comp} - \Delta G_{free} = \Delta G_{bind}$ (wild-type) - ΔG_{bind} (mutant). These alchemical simulations enable us in computing the estimates of thermodynamic quantities, namely, the relative binding free energy ($\Delta\Delta G$) due to a mutation in an amino-acid residue.

We used a hybrid energy function (U) to represent a mixture of two endpoint states of a particular horizontal arm of the thermodynamic cycle **(Figure S3)**. The molecular topology follows the dual topology strategy. A coupling parameter λ connects the initial (I) and the final (F) states by a series of equispaced intermediate states. The coupling parameter values $\lambda = 0$ and 1, corresponds to the physical end states, whereas an intermediate value corresponds to a mixed unphysical state. Using the previously described[40] FEP method, we obtained the total free energy change along the horizontal paths by summing over the intermediate states in the following way: $\Delta G\ (I \rightarrow F) = G_F - G_I = -\beta^{-1} \sum_{m=1}^{n-1} ln\langle exp[-\beta\ (U_{m+1} - U_m)]\ \rangle_m$, where $U_m = (1-\lambda_m)\ U_I + \lambda_m\ U_F$ with the coupling parameter λ_m varying from 0 to 1, and the total number of intermediate points m=1,...,(n-1). Here, β is $1/k_BT$, with k_B as Boltzmann's constant and T as the temperature.

To scale the non-bonded interactions along the alchemical coordinate, we set "alchDecouple" to "OFF" in our free energy calculation protocol. This protocol allowed us to scale the non-bonded interactions of the mutated residue with their environment and within the mutated residue, which contribute to the cumulative free energy. Further, in our free energy calculations, we decoupled the charges from $\lambda = 0$ to $\lambda = 0.5$, while the van der Waals parameters were decoupled from $\lambda = 0$ to $\lambda = 1$ using 25 equally spaced λ_{m} windows distributed over 30 ns simulations. We simulated each λ_{m} window for 1.2 ns, and used the last 1 ns for the free energy estimation. We averaged the total free energy change (ΔG_{free}, ΔG_{comp}) over forward and backward simulations and repeated in triplicate with different initial velocities, yielding a minimum of 180 ns of simulation data per transformation **(Tables S2-S6)**. We report free energy calculations with an aggregate time of $\sim$18 μs. We have reported the $\Delta\Delta G$ as the difference of the averaged ΔG^{comp} and ΔG_{free} values **(Figure 2)**.

We estimated the free energy differences using the bidirectional approach by incorporating samples from both forward and backward transformations and estimated the associated statistical error using the Bennet Acceptance Ratio (BAR) estimator implemented in Parse-FEP toolkit[49] in VMD. To ensure the convergence, we compared the graphical representation of the underlying probability distributions characterizing the forward and backward transformations. We have reported the uncertainities in the averaged ΔG_{comp} and ΔG_{free} as the standard error of the mean (from 3 replicas) and have calculated the error in the final $\Delta\Delta G$ by computing the standard error of the mean associated with the averaged ΔG values. We obtained a good convergence and a reasonable statistical uncertainty (<1 kcal/mol) of the computed energetics ($\Delta\Delta G$).[24,40]

Results

Structure-Based Energetics of Peptide Mutations

Experimental studies[35,36] suggest that the P2 peptide binds RRE RNA with a higher affinity and specificity than the P1 peptide, but the precise free energy changes involving mutations in each peptide in the peptide/RNA complexes remain unknown. Further, biochemical studies have also suggested that a network of electrostatic and hydrophobic interactions are involved in the recognition of RRE RNA by the P2 peptide and replacing alanine residues with glycine residues results in a loss of binding affinity.[35,36]

To test these hypotheses, we computed the changes in the binding affinity upon mutations in two ways: (i) by mutating the non-alanine amino acids into alanine amino acids to quantify the binding contribution of electrostatic interactions, and (ii) by mutating alanine amino acids into glycine amino acids to quantify the binding contribution of hydrophobic interactions in each peptide-RRE complex. We then used an appropriate thermodynamic cycle (**Figure S3**) to calculate the relative changes in the binding affinity upon mutations in each of the two peptides. The computed free energy change (ΔG) resulting from the FEP calculations are listed in **Tables S2-S5**.

In **Figure 2a**, we show the calculated relative changes in the binding affinities upon mutations in the P1 peptide. These calculations reveal the following key features:

- The mutations R35A (ΔΔG = 5.12 ± 0.51 kcal/mol), R38A (ΔΔG = 4.70 ± 0.50 kcal/mol), and R42A (ΔΔG = 4.58 ± 0.32 kcal/mol) impose the highest energetic penalties (∼5 kcal/mol).

- The mutations T34A (ΔΔG = 2.39 ± 0.50 kcal/mol), R39A (ΔΔG = 3.97 ± 0.99 kcal/mol), N40A (ΔΔG = 3.73 ± 0.64 kcal/mol), R43A (ΔΔG = 3.58 ± 0.41 kcal/mol), R44A (ΔΔG = 4.24 ± 0.48 kcal/mol), R46A (ΔΔG = 2.81 ± 0.38 kcal/mol), and R48A (ΔΔG = 2.46 ± 0.43 kcal/mol) reduce the RRE binding affinity by ∼2-4 kcal/mol relative to the WT P1 peptide.

- The mutations Q36A (ΔΔG = 0.52 ± 0.37 kcal/mol), Q49A (ΔΔG = 0.01 ± 0.23 kcal/mol), R50A (ΔΔG = 1.28 ± 0.48 kcal/mol), and R55A (ΔΔG = 0.66 ± 0.12 kcal/mol) impose lower energetic penalties with respect to the WT P1 peptide.

- The mutations D33A (ΔΔG = -0.20 ± 0.0.50 kcal/mol), W45A (ΔΔG = -0.55 ± 0.21 kcal/mol), and E47A (ΔΔG = -0.32 ± 0.54 kcal/mol) favor the binding of the mutated peptide to RRE RNA relative to the WT P1 peptide.

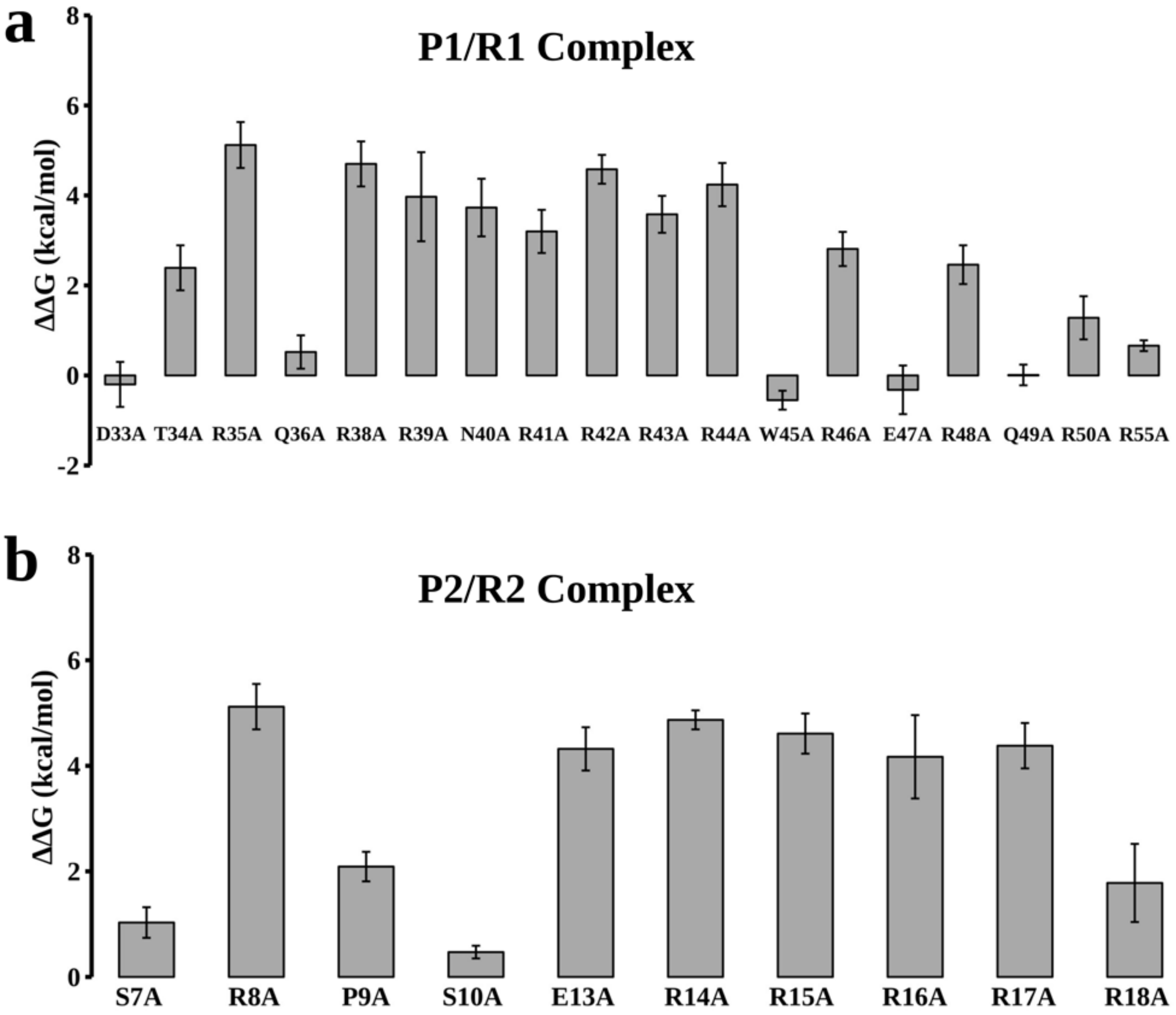

Figure 2: **Energetics of peptide mutations in each peptide/RNA complex.** The relative binding free energy (ΔΔG) of mutations in peptides P1 (panel a) and P2 (panel b). The error bars represent the standard error of the mean, and the binding free energies are in kcal/mol.

In **Figure 2b**, we show the changes in binding affinities upon mutations in the P2 peptide. We make the following key observations:

- All arginine mutations, R8A (ΔΔG = 5.12 ± 0.0.43 kcal/mol), R14A (ΔΔG = 4.87 ± 0.18 kcal/mol), R15A (ΔΔG = 4.61 ± 0.38 kcal/mol), R16A (ΔΔG = 4.17 ± 0.79 kcal/mol), and R17A (ΔΔG = 4.38 ± 0.43 kcal/mol) impose significant energetic penalties of ∼4-5 kcal/mol except for R18A (ΔΔG = 1.78 ± 0.74 kcal/mol) mutation which reduces the RRE RNA binding affinity by ∼2 kcal/mol relative to the WT P2 peptide.

- The mutation E13A ($\Delta\Delta$G = 4.32 ± 0.41 kcal/mol) imposes an energetic penalty of ∼4 kcal/mol similar to the arginine mutations R16A and R17A.
- The alicyclic Proline mutation (P9A, $\Delta\Delta$G = 2.09 ± 0.28 kcal/mol) reduces the binding affinity by ∼2 kcal/mol.
- The mutations S7A ($\Delta\Delta$G = 1.03 ± 0.29 kcal/mol) and S10A ($\Delta\Delta$G = 0.47 ± 0.12 kcal/mol) impose significantly lower energetic penalties.

We also mutated the alanine residues into glycine residues in both P1 and P2 peptides to quantify the contributions of hydrophobic contacts formed by Ala residues in the overall binding affinity of these peptides for RRE RNA (**Tables S4 and S5**). In **Figure S4**, we show the computed energetics of mutations (alanine residues into glycine residues) in both peptides. These data reveal that only R12A ($\Delta\Delta$G = 1.42 ± 0.26 kcal/mol) mutation in the P2 peptide reduces the RRE binding affinity by ∼1 kcal/mol whereas other Ala mutations (P1 peptide: A37G, A51G, A52G, A53G, and A54G; P2 peptide: A12G, A19G, A20G, and A21G) do not reduce the RRE binding affinity of mutated peptides. In **Figure S5**, we show the comparative analysis of changes in binding affinities after non-alanine residue mutations in the P2 peptide calculated using the standard CHARMM36 force field[43] and CHARMM36 force field with improved non-bonded parameters (CHARMM36+NBFix).[50–52] These data suggest a consistent pattern of relative change in binding affinities after mutations in the P2 peptide using both CHARMM36 and CHARMM36+NBFix force fields (**Table S6**). Additionally, we have probed the effect of using different cut-off values for computing interatomic-interactions in free-energy calculations as well as the role of protonation states of some residues (Asp33 and Glu47). The free-energy values computed using different cut-off values show only a marginal difference (**Table S7**). Moreover, the free-energy values for mutating residues Asp33 and Glu47 into alanine residues by considering different protonation states of Asp/Glu residues are similar (**Table S8**). This result is consistent with the observation that Asp33 and Glu47 do not directly participate in peptide/RNA interactions.

Structural and Conformational Insights into the Mutated Peptide-RRE Complexes

In this study, we performed free energy calculations using the alchemical FEP[39] method and repeated these calculations in triplicate with different initial velocities. To check the structural changes after each mutation in both peptides P1 and P2, we analyzed the end point structures from each of the three independent simulations. In Figure 3, we show the common interactions present in at least two out of three endpoint structures obtained from free energy simulations of the peptide-RNA complexes. In **Figure 3a**, we show the structures from the mutated P1/R1 complexes, and in panels (i) through (viii) we show the interaction pattern of key peptide residues with nucleotides in the major groove of RNA after mutations. In **Figure 3a, panel i**, we show that the side-chain of Asp33 is oriented toward the side-chain of Arg38 to form salt-bridging interactions after the T34A mutation, and we observed a similar interaction pattern for the R35A mutation. Further, we note that after the R38A mutation, the N-terminus of the P1 peptide is oriented such that the side-chain of Arg41 forms salt-bridging interactions with the non-bridging phosphate oxygen of GUA46 and the side-chain of Asp33, and the side-chain of Arg35 recognizes the Hoogsteen edge of GUA67 and forms a salt-bridge with its phosphate oxygen (**Figure 3a, panel ii**). The R39A mutation results in the formation of a salt-bridge interaction network involving the side-chains of residues Asp33, Arg38, and Arg41 (**Figure 3a, panel iii**). From the structures of the N40A mutated peptide, it is observed that the side-chain of Arg41 is oriented toward the phosphate oxygen of GUA47 to form salt-bridging interactions (**Figure 3a, panel iv**).

In the R41A mutant, the side-chain of Asn40 establishes a hydrogen bonding interaction with the oxygen atom of the ribose sugar (**Figure 3a, panel v**). The R43A mutation leads to hydrogen bonding between the side-chain of Arg44 and the Hoogsteen edge of GUA46

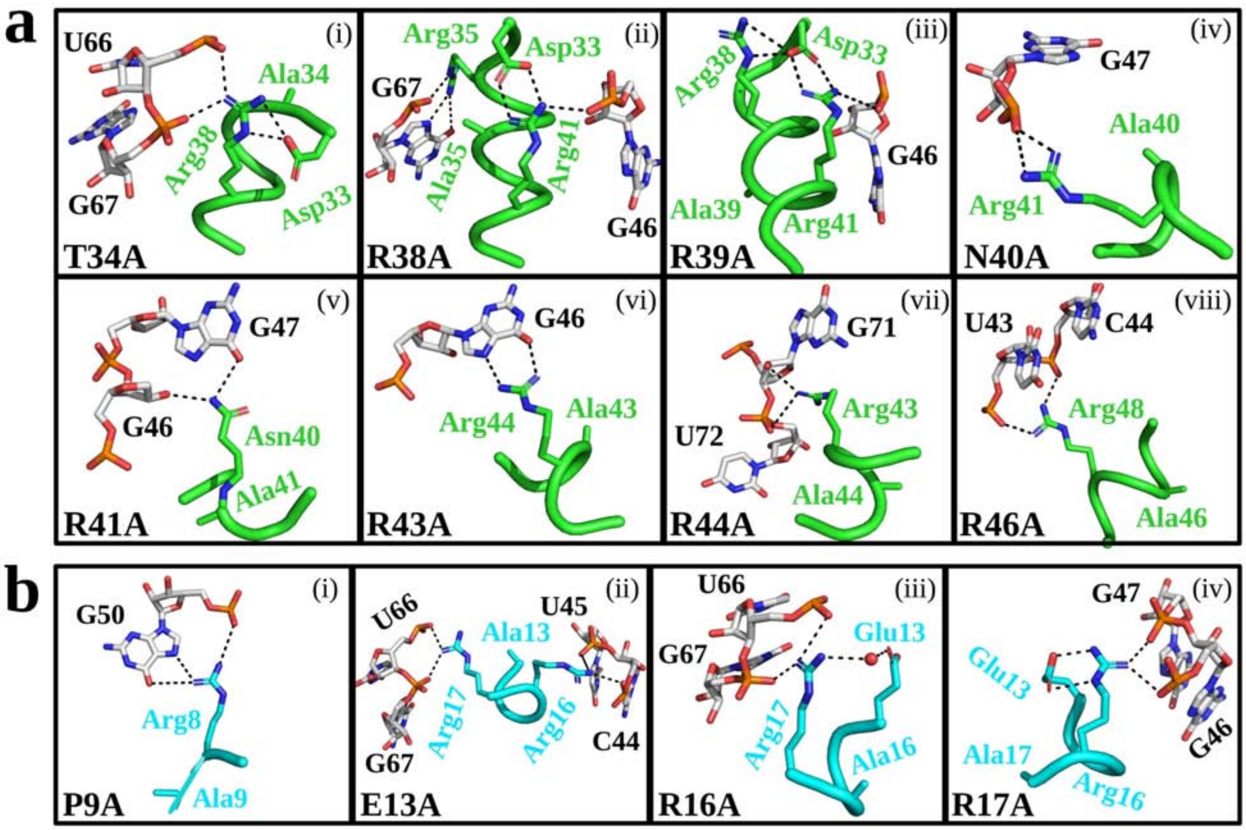

Figure 3: **Structural changes in the peptide-RNA complexes on mutations.** (a) Structural insights from the mutated P1/R1 complexes. In panels (i) through (viii), we show the interaction pattern of key peptide residues with nucleotides in the major groove of RNA after the following mutations: T34A, R38A, R39A, N40A, R41A, R43A, R44A, and R46A, respectively. (b) Structural insights from the mutated P2/R2 complexes. Shown in panels (i) through (iv) are the interaction patterns of key residues after the following mutations: P9A, E13A, R16A, and R17A. The interactions shown are those that are consistently observed in at least two end point structures from three independent free energy simulations of the peptide-RNA complexes. The key interactions are highlighted by black dotted lines and water molecules are depicted by a red sphere.

(**Figure 3a, panel vi**), while the R44A mutation disrupts base-specific interactions between the side-chain of Arg43 and RRE RNA and results in the formation of new salt-bridging interactions with the phosphate oxygen of the nucleotides GUA71 and URA72 (**Figure 3a, panel vii**). Furthermore, the R46A mutation places the side-chain of Arg48 toward the phosphate oxygen of the nucleotides URA43 and CYT44 to enable salt-bridging interactions (**Figure 3a, panel viii**). These structural analyses of the P1/R1 complex suggest that the mutations in amino-acid residues in the peptide P1 disrupt interactions between the side-chains of mutated residues and RRE RNA. The mutations T34A, R38A, R39A, N40A, R41A, R43A, R44A, and R46A also alter the interaction patterns of neighboring non-mutated amino-acid residues (**Figure 3a, panels i-viii**).

In **Figure 3b**, we show the structures from the mutated P2/R2 complex, and in panels (i) to (iv) we show the interaction patterns of key peptide residues with RNA after mutations. We observed that after the P9A mutation, the side-chain of Arg8 established interactions with the Hoogsteen edge and the phosphate oxygen atom of the nucleotide GUA50 (**Figure 3b, panel i**). Further, we observed that after the E13A mutation, the side-chain of Arg16 forms salt-bridging interactions with the phosphate oxygen atoms of nucleotides CYT44 and URA45, and the side-chain of residue Arg17 salt-bridges to the phosphate oxygen atoms of nucleotides URA66 and GUA67 (**Figure 3b, panel ii**). We have observed similar interactions of Arg17 with RNA after the R16A mutation, where the side-chain of Arg17 also forms a water-mediated interaction with the side-chain of Glu13 (**Figure 3b, panel iii**). Moreover, in the R17A mutant, the side-chain of Arg16 establishes salt-bridging interactions with the side-chain of Glu13 and with the phosphate oxygen atoms of the nucleotides GUA46 and GUA47 (**Figure 3b, panel iv**). These structural analyses reveal that the mutations in amino-acid residues in the P2 peptide disrupt interactions between the side-chain of mutated residues and the RRE RNA molecule. In contrast, the interaction patterns of non mutated residues in the complex remain intact except for the P9A, E13A, R16A, and R17A mutations (**Figure 3b, panels i-iv**).

Dynamics of Unbound RNAs and Peptide-Bound RNA Complexes

Biochemical and NMR studies have suggested that RRE RNA undergoes conformational changes upon binding to each peptide, thereby leading to the flipping of ADE68 and URA72 nucleobases and *anti* conformations in the purine-purine base pairs (GUA48-GUA71 and GUA47-ADE73).[11,35] Therefore, to investigate the conformational changes in RNA induced by the binding of both peptides, we first simulated unliganded RRE RNA and then peptide-bound RNA complexes in explicit solvent (see Supporting Methods, **Figure S2**, and **Table S1**). The RMSD data showed that the MD structures of unbound RNA and peptide-bound RNA complexes are stable during structural relaxation with marginal conformational deviations (<5 Å) from their initial structures (**Figure S6** and **Figure S7**). Further, RMSF data reveal larger fluctuations for residues located in the tetraloop (R1: GUA60, CYT61, ADE62, and ADE63; R2: URA60, URA61, CYT62, and GUA63) and the unpaired nucleotides (ADE68 and URA72) in RRE RNAs (**Figure S8**).

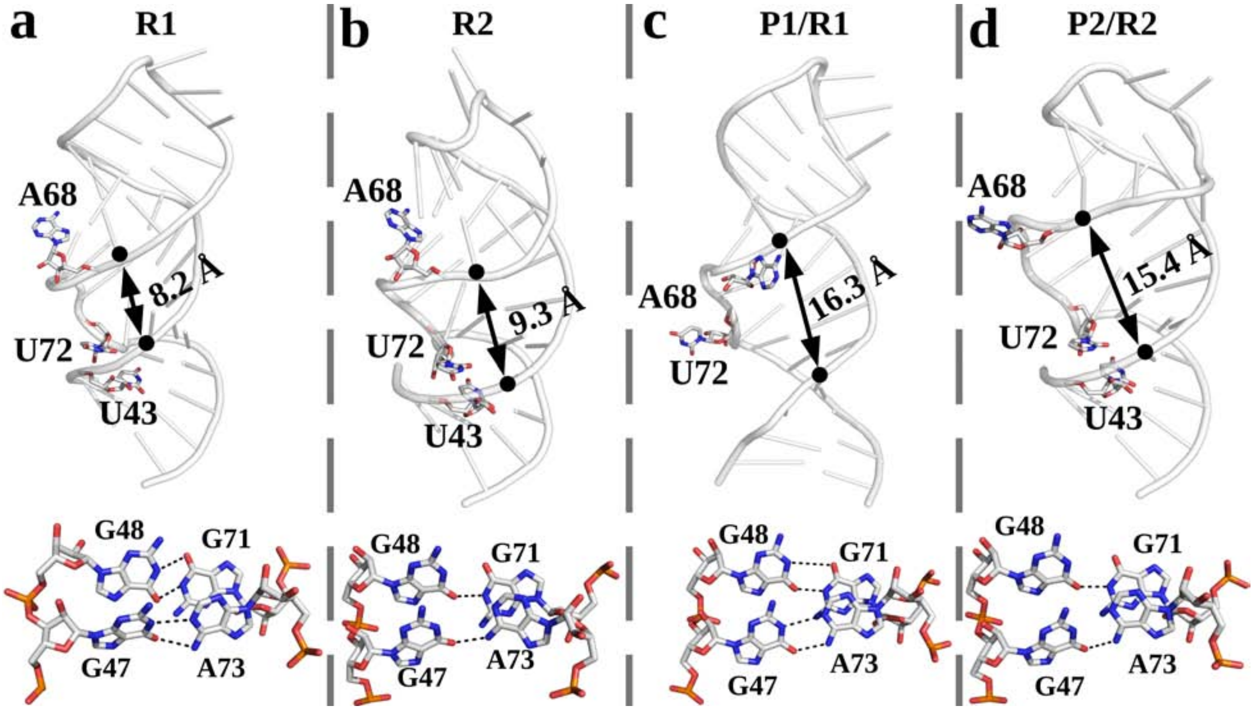

Figure 4: **The structures of unbound and peptide-bound RRE RNA.** The averaged structures from MD simulations of RRE RNA constructs (R1 and R2) when unbound (panels a and b) and bound to the P1 and P2 peptides (panels c and d). The mean distances between the P-P atoms (depicted by filled black circles) of the major groove nucleotides GUA46 and GUA67 are labeled and shown by arrows. The key nucleotides are represented as sticks. The interaction patterns in purine-purine base pairs (GUA48-GUA71 and GUA47-ADE73) when unbound (a and b bottom panels) and bound to the peptides (c and d bottom panels) are shown. The hydrogen atoms are not shown for clarity. The purine-purine interactions are highlighted by black dotted lines and their fractional occupancies are shown in Figure S9.

In **Figure 4** and **Figure 5**, we show the key interactions between peptide and RNA on the basis of their fractional occupancies. The fractional occupancy is the fraction of simulation frames, sampled at 1 ns, in which a specific interaction is occupied. We determined the hydrogen bond occupancy using a cutoff of 3.5 Å for heavy atoms (oxygen and nitrogen) and averaged over four MD trajectories.

In **Figure 4**, we show the structures of RRE RNA molecules derived from MD simulations of unbound (R1 and R2) and peptide-bound RNA conformations. The P-P interatomic distances between the major groove nucleotides GUA46 and GUA67 in unbound and peptide-bound forms are 8.2 ± 0.32 Å (R1), 9.3 ± 0.38 Å (R2), 16.2 ± 0.41 Å (P1/R1), and 15.4 ± 0.49 Å (P2/R2) (**Figure 4**). These data suggest a wider major groove of RNA in peptide-bound forms compared to the unbound RNA forms. Further, in **Figure 4, bottom panels**, we show the interaction patterns in purine-purine base pairs in unbound and peptide-bound RRE RNA conformations which suggest that these base-pairs utilize the Watson-Crick edges of both nucleotides to form two hydrogen bonds in R1 RNA and one hydrogen bond in R2 RNA in unbound and bound forms. In **Figure S9**, we show the data on fractional occupancies of the purine-purine base pairs (GUA48-GUA71 and GUA47-ADE73) interactions. These data suggest that all interactions in both base-pairs are stable. Further, in **Figure 4**, we show that URA72 stacks on URA43 in the unbound and peptide-bound RRE RNA molecules except in RRE RNA bound to the P1 peptide (panel c), where URA72 is oriented away from the major groove of RNA.

These structures of unbound and peptide-bound RRE RNA suggest that URA72 changes its conformation on binding of the P1 peptide to RRE. In a previous study of the P2 peptide unbinding from the RRE RNA, we showed that URA72 stacks on URA43 in the unbound and peptide-bound RRE RNA (**Figure S10**),[16] thereby confirming these mechanistic observations. The structural analyses also revealed that the nucleotide ADE68 is oriented away from the major groove in RRE RNA in unbound states (**Figure 4a, b**), while it is oriented away toward the major groove of RRE RNA in the peptide-bound states (**Figure 4c, d**). These data suggest that the nucleotide ADE68 changes its conformation on binding of peptides to RRE RNA for establishing interactions with the peptide (**Figure S11**).

Further, the average dihedral angle values of purine-purine base-pairs (GUA48-GUA71 and GUA47-ADE73) in R1 RNA and in RNA in the P1/R1 complex suggest an *anti* conformation of each nucleotide (**Figure S12**). The average dihedral angle values of these purine-purine base-pairs in R2 RNA suggest *syn* conformations of nucleotides GUA47 and GUA71 and *anti* conformations of nucleotides GUA48 and ADE73; whereas in the P2/R2 complex, the average dihedral angle values suggest *syn* conformations of nucleotides GUA71 and ADE73 and *anti* conformations of nucleotides GUA47 and GUA48 (**Figure S13**). Overall, our structural analyses of RRE RNA in unbound and peptide-bound forms suggest that RRE RNA undergoes conformational changes, primarily widening the major groove on binding of peptides.

In **Figure 5**, we show the MD averaged structures to highlight the key interactions between peptides (P1 and P2) and RRE RNA. These key interactions are shown on the basis of fractional occupancy data (**Figure S14**). We have also shown the representative structures from the dominant cluster of each peptide-RNA complex obtained by clustering the trajectories using cpptraj.[45,53] In Figure S15, the snapshots from the dominant cluster represent interactions similar to Figure 5.

In **Figure 5a**, we show the MD averaged structures derived from MD simulations of the P1/R1 complex with panels (i) through (xi) highlighting the interaction patterns of key peptide residues with nucleotides in the major groove of RNA. In **Figure 5a**, we show that the amino acid residues ranging between Thr34 and Arg48 anchor the P1 peptide within the internal loop of RRE RNA. The residues Thr34, Arg35, Arg38, Arg41, Arg42, Arg46, and Arg48 (**Figure 5a** panels **i, ii, iii, vi, vii, x, and xi**, respectively) adopt conformations to form salt-bridging interactions with the non-bridging phosphate oxygen atoms of the RNA backbone. The side-chain of Arg39 stacks on the CYT69 nucleobase and makes hydrogen bonds with the Hoogsteen edge of GUA70 (**Figure 5a, panel iv**). The side-chain of Asn40 sits near the nucleotides GUA46 and GUA47 in an orientation to form hydrogen bonds between the amine group of Asn40 and the carbonyl oxygen atom (O6) of each base (see

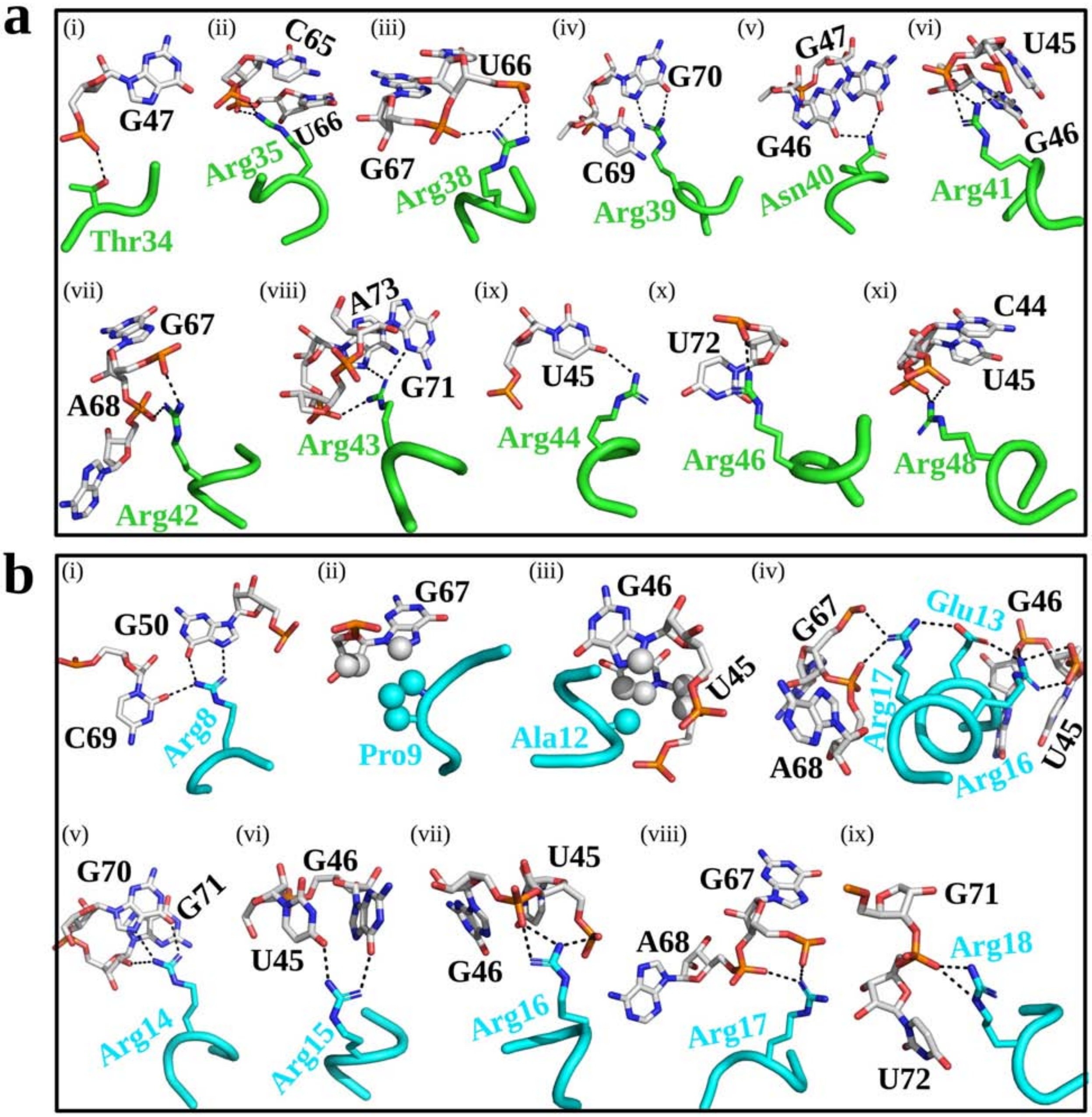

Figure 5: **Structural insights from MD simulations of the Peptide-RRE complexes.** The structures from MD simulations of the P1/R1 and the P2/R2 complexes are shown in panels (a) and (b), respectively. Key RRE nucleotides and peptide residues are represented as sticks. The interaction patterns of key peptide residues with the major groove in RRE RNA is shown with black dotted lines. The atoms in RRE RNA and peptides involved in hydrophobic interactions are represented as spheres. The fractional occupancies of interactions in panels a and b are shown in Figure S14.

Figure 5a, panel v). A second hydrogen-bonding motif in the P1/R1 complex involves the side-chains of Arg43 and Arg44, where Arg43 makes hydrogen bonds with the Hoogsteen edge of GUA71 and ADE73 nucleobases, and the side-chain of Arg43 also forms a salt-bridge with the non-bridging phosphate oxygen of GUA71 (**Figure 5a, panel viii**); as well as the side-chain of Arg44 hydrogen bonds with the carbonyl oxygen atom in URA45 (**Figure 5a, panel ix**). We also observed that the side-chain of Arg46 stacks on the URA72 nucleobase (**Figure 5a, panel x**). These data show that electrostatic interactions dominate in the binding of the P1 peptide to RRE RNA.

In **Figure 5b**, we show the MD averaged structures derived from MD simulations of the P2/R2 complex with panels (i) through (ix) highlighting the interaction patterns of key peptide residues with nucleotides in the major groove of RNA. We observed that the amino acid residues between Arg8 and Arg18 anchor the P2 peptide within the internal loop of RRE RNA through specific intermolecular interactions (**Figure 5b**). The side-chain of residue Arg8 makes hydrogen bonds with the Hoogsteen edge of GUA50 and the carbonyl oxygen of CYT69 (**Figure 5b, panel i**). The side-chain carbon atoms of alicyclic Pro9 rest on a hydrophobic surface-patch formed by the atoms C8, C2′, and C3′ of GUA67 (**Figure 5b, panel ii**). The side-chain of Ala12 (methyl group) rests on a hydrophobic region of RRE RNA comprising the C5/C6 edge of the URA45 nucleobase along with the C2′/C3′ edge of the URA45 ribose and the C8 atom of the GUA46 nucleobase (**Figure 5b, panel iii**).

In **Figure 5b, panel iv**, we show that the side-chain of Glu13 forms salt-bridging interactions with the side-chains of residues Arg16 and Arg17. The side-chain of Arg14 forms hydrogen bonds with the Hoogsteen edge of GUA70 and ribose sugar of GUA71 (**Figure 5b, panel v**). The side-chain of residue Arg15 makes hydrogen bonds with the carbonyl oxygen of URA45 and GUA46 nucleobases (see **Figure 5b, panel vi**). Further, the side-chains of three other Arg residues (Arg16, Arg17, and Arg18 as shown in **Figure 5b** panels **vii, viii, and ix**, respectively) adopt conformations to form salt-bridging interactions with non-bridging phosphate oxygen atoms of the RNA backbone. We also observed that the side-chain of Arg18 stacks upon the URA72 nucleobase (**Figure 5b, ix**). These data show that although polar groups dominate in RRE RNA, we observed specific hydrophobic interactions in the P2/R2 complex. These data also reveal that the P2/R2 complex contains intrapeptide interactions beyond the hydrogen bonds, salt-bridges, and hydrophobic interactions.

In **Figure S14**, we show the data on fractional occupancies of the peptide-RNA interactions. These data suggest that all interactions in both P1/R1 and P2/R2 complexes are stable except the interactions between the side-chain of Arg18 and the phosphate oxygen atoms of GUA71, which are transient with an average occupancy less than 0.5. The pattern of hydrophobic interactions observed for the residue Ala12 in this study is similar to the interaction pattern observed for the same residue in a previous NMR study.[35] We observed that the network of salt-bridging interactions involving residues Glu13, Arg16, and Arg17 of the P2 peptide and nucleotides URA45 and GUA67 of RRE RNA are stable with an average occupancy above 0.85 (**Figure S14**); however, that same interaction network was not observed in the NMR structures[35] and was observed transiently in MD simulations of a previous study.[54] Our simulations suggest that the binding of both peptides to the major groove of RRE RNA in different orientations leads to unique and distinct interaction patterns.

Discussion

Biochemical and structural studies suggest that the major contribution to a higher binding affinity of the P2 peptide (compared to the P1 peptide) for RRE RNA comes from the stacking interactions between the URA72 nucleobase and the side-chain of Arg15 residue of the P2 peptide.[11,35,55] However, the mutational studies on the role of arginines in the P1 and P2 peptides are lacking. Therefore, in this work, we have reported free energy calculations for quantitatively deciphering the energetics of mutations in each peptide and have performed long time-scale MD simulations of the HIV RRE RNA in unbound and peptide-bound forms to provide insights into their dynamics, thereby linking energetics with the structural details.

We calculated free energy changes on mutations in each peptide using alchemical free energy simulations.[39] Our calculations suggest that there are 5 Arg residues (Arg35, Arg38, Arg39, Arg42, and Arg44) in the P1 peptide as well as in the P2 peptide (Arg8, Arg14, Arg15, Arg16, and Arg17), the mutations of which decrease the binding affinities of these peptides toward RRE by more than 4 kcal/mol (**Figure 2**). Therefore, these arginine residues dominate the RNA recognition process in both peptides. Our MD simulations suggest that the disruption of key interactions (**Figure 5**) between these Arg residues and RRE RNA is responsible for the larger penalties upon arginine mutations (**Figure 2**). Our calculations also revealed that the mutation of Glu13 into Ala13 residue in the P2 peptide decreases the RRE binding affinity by ∼4 kcal/mol (**Figure 2b**). Our MD simulations of the P2/R2 complex suggest that the residue Glu13 forms salt-bridges with the residues Arg16 and Arg17 and keeps them oriented toward the backbone of the major groove of RNA (**Figure 5b, panel iv**). In the P2 peptide, the salt-bridge network established by Glu13 effectively holds the peptide in place in the major groove of RNA to make stable peptide-RNA interactions, thereby resulting in an increased energetic penalty on mutation in Glu13.

Furthermore, our free energy calculations suggest that the residues Pro9 and Ala12 in the P2 peptide are involved in hydrophobic contacts with RRE RNA and decrease the overall binding affinity of the P2 peptide on their mutations. However, we have not observed a similar hydrophobic interaction network in simulations of the P1/R1 complex. Further, the sequence alignment of these peptides (**Figure S1**) suggests that the P1 peptide would place the stretch of polar residues at the N-terminus of the P2 peptide, abolishing all specific hydrophobic interactions with RRE RNA and leading to steric clashes involving the bulky side-chains of Arg residues within the peptide-binding pocket of RNA. The specific hydrophobic contacts and intrapeptide interactions in the P2 peptide explain its distinct binding mode compared to the P1 peptide. Experimental studies[36] have also suggested that the replacement of alanine residues with glycine residues in the P2 peptide results in an overall decrease in the binding affinity but the magnitudes of changes in the free energy on mutations are not known. Hence, the calculated relative binding free energies of mutations cannot be directly compared with available experimental data.[36] However, the signs of relative binding free energies are correct and magnitudes we report are biochemically plausible.

MD simulations of RRE RNA in both unbound and the peptide-bound forms suggest that the purine-purine base pairs (GUA47-ADE73 and GUA48-GUA71) in the internal loop of RRE RNA are unaffected on peptide binding and their Watson-Crick edges are involved in hydrogen-bonding interactions (**Figure 4, bottom panels**). Also, the glycosidic dihedral angle data suggest that the purine-purine base-pairs in RRE RNA do not change their conformations upon binding of the peptides except in the GUA47-ADE73 base-pair, which changes its conformation from GUA47(*syn*)-ADE73(*anti*) to GUA47(*anti*)-ADE73(*syn*) upon binding of the P2 peptide (**Figure S12** and **Figure S13**). Further, we discovered that binding of the P1 peptide to RRE RNA opens the major groove by ∼8.1 Å whereas the binding of the P2 peptide to RRE RNA opens the major groove by ∼6.1 Å, thereby suggesting a wider RNA major groove opening upon binding of the P1 peptide compared to the P2 peptide (**Figure 4**). MD simulations also revealed that the binding of the P1 peptide disrupts the URA72 and URA43 stacking in RRE RNA, which is intact for the P2 peptide (**Figure 4** and **Figure S16**). These data suggest that disruption of the URA72-URA43 stacking interaction offsets the favorable interactions between the P1 peptide and RRE RNA, thereby decreasing its overall binding affinity.

Further, we probed the P1/R1 and P2/R2 complexes to decipher hydrogen bonding patterns for binding of each peptide to RNA. MD simulations revealed that both peptides bind to the major groove of RRE RNA in different orientations, thereby accounting for different hydrogen bonding patterns, although similar regions in RRE RNA are involved in recognition of each peptide. The P1/R1 complex contains 22 hydrogen bonds between the peptide and RNA, while the P2/R2 complex contains 18 hydrogen bonds (**Figure 5**). In the P1/R1 complex, 15 contacts occur between the peptide P1 and the phosphate groups or sugar moieties of RNA, while only 7 are base-specific contacts, thereby resulting in a roughly 2:1 preference for the backbone contacts over bases (**Figure 5a**). However, in the P2/R2 complex, the ratio of the backbone to base contacts of the peptide is 1:1, where 10 contacts are base-specific, and only 8 hydrogen bonds are between the peptide P2 and the RNA backbone (**Figure 5b**). The structural study[11] of the P1/R1 complex also suggested a similar distribution of hydrogen bonds, where 6 contacts are base-specific and 11 hydrogen bonds are between the P1 peptide and the RNA backbone. Thereby, resulting in a roughly 2:1 ratio of the backbone to base contacts. These experimental[11] data suggest that the ratio of the backbone to base contacts is preserved in MD simulations of the P1/R1 complex. Further, the difference in the ratio of the backbone to base contacts for P1 (2:1) and P2 (1:1) peptides explain the earlier experimental[35,36] observations that the P2 peptide binds RRE with a higher specificity than the P1 peptide.

Conclusions

In this study, our results reveal that the native Rev peptide interacts with the major groove of RRE RNA through some base-specific contacts and the majority of interactions occur through contacts of amino acid side-chains with the sugar-phosphate backbone, suggesting that the native peptide recognizes the shape of the phosphate backbone. This hydrogen-bonding pattern in the native peptide-bound RRE complex suggests that the native peptide employs in part an indirect RNA readout recognition mechanism,[56,57] which is usually associated with the binding affinity. Our results further show that the distinct binding orientation of the native peptide in the major groove of RRE RNA disrupts the URA72-URA43 base-stacking interaction, which offsets the favorable interactions between the native peptide and RRE RNA, thereby decreasing the overall binding affinity.

In contrast, our results on the lab-evolved RSG1.2 peptide suggest that it employs a complex set of strategies to recognize RRE RNA that include base-specific hydrogen bonding, cation-pi interactions, specific hydrophobic interactions, as well as shape-selective RRE phosphate backbone recognition and intrapeptide interactions. The distinct binding mode of the lab-evolved peptide in the major groove of the RNA leads to specific hydrophobic and intrapeptide interactions. These intrapeptide interactions effectively hold the lab-evolved peptide in the major groove of RRE RNA in such an orientation that the Arg residues extending from either side of the peptide contact the phosphate groups and nucleotide bases on both sides of the major groove of RNA, thereby contributing to a higher binding affinity and specificity of the lab-evolved peptide over the Rev peptide. Therefore, our work provides a basis for the observation that targeting the nucleotides in RRE RNA as interaction sites not used by the native peptides might contribute to the additional binding specificity and affinity of lab-evolved peptides. We suggest that the results from this study are potentially useful in designing synthetic peptides that target viral RNA molecules with higher affinities and specificities.

Data and Software Availability

All methodological details for all data generated in this study are available in the Method section including PDB codes of input files, simulation conditions, error analysis. The simulation and visualization softwares[41,46–48] and the force-field[43,50–52] used to carry out calculations are openly available.

Acknowledgement

AK thanks Lev Levintov for providing simulation data on the RSG1.2 peptide unbinding and helpful discussions. We acknowledge the financial support provided by the National Science Foundation through grant CBET-1554558 and the National Institutes of Health (NIH) through grant R35GM138217. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We are grateful for computational support through the following resources: Premise, a central shared HPC cluster at UNH supported by the Research Computing Center; and BioMade, a heterogeneous CPU/GPU cluster supported by the NSF EPSCoR award (OIA-1757371).

Supporting Information

Supporting Results, Tables S1-S8, and Figures S1-S16.

References

(1) Sanchez, A.; Kiley, M. P. Identification and Analysis of Ebola virus Messenger RNA. *Virology* **1987**, *157*, 414–420.

(2) Gultyaev, A. P.; Fouchier, R. A.; Olsthoorn, R. C. Influenza Virus RNA Structure: Unique and Common Features. *Int Rev Immunol* **2010**, *29*, 533–556.

(3) Watts, J. M.; Dang, K. K.; Gorelick, R. J.; Leonard, C. W.; Bess Jr, J. W.; Swanstrom, R.; Burch, C. L.; Weeks, K. M. Architecture and Secondary Structure of an Entire HIV-1 RNA Genome. *Nature* **2009**, *460*, 711–716.

(4) Robertson, M. P.; Igel, H.; Baertsch, R.; Haussler, D.; Ares Jr, M.; Scott, W. G. The Structure of a Rigorously Conserved RNA Element Within the SARS Virus Genome. *PLoS Biol* **2005**, *3*, e5.

(5) Lindenbach, B. D.; Rice, C. M. Unravelling Hepatitis C Virus Replication from Genome to Function. *Nature* **2005**, *436*, 933–938.

(6) Enjuanes, L.; Almazán, F.; Sola, I.; Zuñiga, S. Biochemical Aspects of Coronavirus Replication and Virus-Host Interaction. *Annu Rev Microbiol* **2006**, *60*, 211–230.

(7) Patters, B. J.; Kumar, S. The Role of Exosomal Transport of Viral Agents in Persistent HIV Pathogenesis. *Retrovirology* **2018**, *15*, 1–13.

(8) Piasecka, J.; Lenartowicz, E.; Soszynska-Jozwiak, M.; Szutkowska, B.; Kierzek, R.; Kierzek, E. RNA Secondary Structure Motifs of the Influenza A Virus as Targets for SiRNA-Mediated RNA Interference. *Mol Ther Nucleic Acids* **2020**, *19*, 627–642.

(9) Rangan, R.; Zheludev, I. N.; Hagey, R. J.; Pham, E. A.; Wayment-Steele, H. K.; Glenn, J. S.; Das, R. RNA Genome Conservation and Secondary Structure in SARS-CoV-2 and SARS-Related Viruses: A First Look. *RNA* **2020**, *26*, 937–959.

(10) Sibanda, B.; Thornton, J. β-Hairpin Families in Globular Proteins. *Nature* **1985**, *316*, 170–174.

(11) Battiste, J. L.; Mao, H.; Rao, N. S.; Tan, R.; Muhandiram, D.; Kay, L. E.; Frankel, A. D.; Williamson, J. R. α Helix-RNA major Groove Recognition in an HIV-1 Rev Peptide-RRE RNA Complex. *Science* **1996**, *273*, 1547–1551.

(12) Cléry, A.; Blatter, M.; Allain, F. H. RNA Recognition Motifs: Boring? Not Quite. *Curr Opin Struct Biol* **2008**, *18*, 290–298.

(13) Frankel, A. D. Fitting Peptides into the RNA World. *Curr Opin Struct Biol* **2000**, *10*, 332–340.

(14) Frenkel-Pinter, M.; Haynes, J. W.; Mohyeldin, A. M.; Martin, C.; Sargon, A. B.; Petrov, A. S.; Krishnamurthy, R.; Hud, N. V.; Williams, L. D.; Leman, L. J. Mutually Stabilizing Interactions Between Proto-Peptides and RNA. *Nat Commun* **2020**, *11*, 1–14.

(15) Walker, M. J.; Varani, G. Design of RNA-Targeting Macrocyclic Peptides. *Meth Enzymol* **2019**, *623*, 339–372.

(16) Levintov, L.; Vashisth, H. Role of Salt-Bridging Interactions in Recognition of Viral RNA by Arginine-Rich Peptides. *Biophys J* **2021**, *120*, 5060–5073.

(17) Corley, M.; Burns, M. C.; Yeo, G. W. How RNA-Binding Proteins Interact With RNA: Molecules and Mechanisms. *Mol Cell* **2020**, *78*, 9–29.

(18) Jones, S.; Daley, D. T.; Luscombe, N. M.; Berman, H. M.; Thornton, J. M. Protein-RNA Interactions: A Structural Analysis. *Nucleic Acids Res* **2001**, *29*, 943–954.

(19) Leulliot, N.; Varani, G. Current Topics in RNA- Protein Recognition: Control of Specificity and Biological Function Through Induced Fit and Conformational Capture. *Biochem* **2001**, *40*, 7947–7956.

(20) Matthews, M. M.; Thomas, J. M.; Zheng, Y.; Tran, K.; Phelps, K. J.; Scott, A. I.; Havel, J.; Fisher, A. J.; Beal, P. A. Structures of Human ADAR2 Bound to dsRNA Reveal Base-Flipping Mechanism and Basis for Site Selectivity. *Nat Struct Mol Biol* **2016**, *23*, 426–433.

(21) Belashov, I. A.; Crawford, D. W.; Cavender, C. E.; Dai, P.; Beardslee, P. C.; Mathews, D. H.; Pentelute, B. L.; McNaughton, B. R.; Wedekind, J. E. Structure of HIV TAR in Complex With a Lab-Evolved RRM Provides Insight into Duplex RNA Recognition and Synthesis of a Constrained Peptide that Impairs Transcription. *Nucleic Acids Res* **2018**, *46*, 6401–6415.

(22) Shortridge, M. D.; Wille, P. T.; Jones, A. N.; Davidson, A.; Bogdanovic, J.; Arts, E.; Karn, J.; Robinson, J. A.; Varani, G. An Ultra-High Affinity Ligand of HIV-1 TAR Reveals the RNA Structure Recognized by P-TEFb. *Nucleic Acids Res* **2019**, *47*, 1523–1531.

(23) Chavali, S. S.; Mali, S. M.; Jenkins, J. L.; Fasan, R.; Wedekind, J. E. Co-crystal Structures of HIV TAR RNA Bound to Lab-Evolved Proteins Show Key Roles for Arginine Relevant to the Design of Cyclic Peptide TAR Inhibitors. *J Biol Chem* **2020**, *295*, 16470–16486.

(24) Kumar, A.; Vashisth, H. Conformational Dynamics and Energetics of Viral RNA Recognition by Lab-Evolved Proteins. *Phys Chem Chem Phys* **2021**, *23*, 24773–24779.

(25) Knoepfel, S. A.; Berkhout, B. On the Role of Four Small Hairpins in the HIV-1 RNA Genome. *RNA Biol* **2013**, *10*, 540–552.

(26) Liu, Y.; Chen, J.; Nikolaitchik, O. A.; Desimmie, B. A.; Busan, S.; Pathak, V. K.; Weeks, K. M.; Hu, W.-S. The Roles of Five Conserved Lentiviral RNA Structures in HIV-1 Replication. *Virology* **2018**, *514*, 1–8.

(27) Jain, C.; Belasco, J. G. A structural Model for the HIV-1 Rev–RRE Complex Deduced from Altered-Specificity Rev Variants Isolated by a Rapid Genetic Strategy. *Cell* **1996**, *87*, 115–125.

(28) Nifosì, R.; Reyes, C. M.; Kollman, P. A. Molecular Dynamics Studies of the HIV-1 TAR and its Complex with Argininamide. *Nucleic Acids Res* **2000**, *28*, 4944–4955.

(29) Olsen, H. S.; Nelbock, P.; Cochrane, A. W.; Rosen, C. A. Secondary Structure is the Major Determinant for Interaction of HIV Rev Protein with RNA. *Science* **1990**, *247*, 845–848.

(30) Kjems, J.; Brown, M.; Chang, D. D.; Sharp, P. A. Structural Analysis of the Interaction Between the Human Immunodeficiency Virus Rev Protein and the Rev Response Element. *Proc Natl Acad Sci USA* **1991**, *88*, 683–687.

(31) Mann, D. A.; Mikaélian, I.; Zemmel, R. W.; Green, S. M.; Lowe, A. D.; Kimura, T.; Singh, M.; Jonathan, P.; Butler, G.; Gait, M. J., et al. A Molecular Rheostat: Cooperative Rev Binding to Stem I of the Rev-Response Element Modulates Human Immunodeficiency Virus Type-1 Late Gene Expression. *J Mol Biol* **1994**, *241*, 193–207.

(32) Williamson, J.; Battiste, J.; Mao, H.; Frankel, A. Interaction of HIV Rev Peptides with the Rev Response Element RNA. Nucleic acids symposium series. 1995; pp 46–48.

(33) Fischer, U.; Pollard, V. W.; Lührmann, R.; Teufel, M.; Michael, M. W.; Dreyfuss, G.; Malim, M. H. Rev-Mediated Nuclear Export of RNA is Dominant Over Nuclear Retention and is Coupled to the Ran-GTPase Cycle. *Nucleic Acids Res* **1999**, *27*, 4128–4134.

(34) Malim, M. H.; Tiley, L. S.; McCarn, D. F.; Rusche, J. R.; Hauber, J.; Cullen, B. R. HIV-1 Structural Gene Expression Requires Binding of the Rev Trans-Activator to its RNA Target Sequence. *Cell* **1990**, *60*, 675–683.

(35) Gosser, Y.; Hermann, T.; Majumdar, A.; Hu, W.; Frederick, R.; Jiang, F.; Xu, W.; Patel, D. J. Peptide-Triggered Conformational Switch in HIV-1 RRE RNA Complexes. *Nat Struct Biol* **2001**, *8*, 146–150.

(36) Harada, K.; Martin, S. S.; Tan, R.; Frankel, A. D. Molding a Peptide into an RNA Site by in Vivo Peptide Evolution. *Proc Natl Acad Sci USA* **1997**, *94*, 11887–11892.

(37) Battiste, J. L.; Tan, R.; Frankel, A. D.; Williamson, J. R. Binding of an HIV Rev peptide to Rev Responsive Element RNA Induces Formation of Purine-Purine Base Pairs. *Biochemistry* **1994**, *33*, 2741–2747.

(38) Peterson, R. D.; Feigon, J. Structural Change in Rev Responsive Element RNA of HIV-1 on Binding Rev Peptide. *J Mol Biol* **1996**, *264*, 863–877.

(39) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. *J Chem Phys* **1954**, *22*, 1420–1426.

(40) Kumar, A.; Mukherjee, D.; Satpati, P. Mutations in Parkinson's Disease Associated Protein DJ-1 Alter the Energetics of DJ-1 Dimerization. *J Chem Inf Model* **2019**, *59*, 1497–1507.

(41) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14*, 33–38.

(42) Li, H.; Robertson, A. D.; Jensen, J. H. Very Fast Empirical Prediction and Rationalization of Protein pKa Values. *Proteins: Structure, Function, and Bioinformatics* **2005**, *61*, 704–721.

(43) Huang, J.; MacKerell Jr, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J Comput Chem* **2013**, *34*, 2135–2145.

(44) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J Chem Phys* **1993**, *98*, 10089–10092.

(45) Roe, D. R.; Cheatham III, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **2013**, *9*, 3084–3095.

(46) Schrödinger, L. The PyMOL Molecular Graphics System, Version 1.8. 2015.

(47) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J Comput Chem* **2005**, *26*, 1781–1802.

(48) Phillips, J. C.; Hardy, D. J.; Maia, J. D.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W., et al. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J Chem Phys* **2020**, *153*, 044130.

(49) Liu, P.; Dehez, F.; Cai, W.; Chipot, C. A Toolkit for the Analysis of Free-Energy Perturbation Calculations. *J Chem Theory Comput* **2012**, *8*, 2606–2616.

(50) Yoo, J.; Aksimentiev, A. Improved Parametrization of Li+, Na+, K+, and Mg2+ Ions for All-Atom Molecular Dynamics Simulations of Nucleic Acid Systems. *J Phys Chem Lett* **2012**, *3*, 45–50.

(51) Yoo, J.; Aksimentiev, A. Improved Parameterization of Amine-Carboxylate and Amine-Phosphate Interactions for Molecular Dynamics Simulations Using the CHARMM and AMBER Force Fields. *J Chem Theory Comput* **2016**, *12*, 430–443.

(52) Yoo, J.; Aksimentiev, A. New Tricks for Old Dogs: Improving the Accuracy of Biomolecular Force Fields by Pair-Specific Corrections to Non-Bonded Interactions. *Phys Chem Chem Phys* **2018**, *20*, 8432–8449.

(53) Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334.

(54) Michael, L. A.; Chenault, J. A.; Miller III, B. R.; Knolhoff, A. M.; Nagan, M. C. Water, Shape Recognition, Salt bridges, and Cation-Pi Interactions Differentiate Peptide Recognition of the HIV Rev-Responsive Element. *J Mol Biol* **2009**, *392*, 774–786.

(55) Kumar, S.; Bose, D.; Suryawanshi, H.; Sabharwal, H.; Mapa, K.; Maiti, S. Specificity of RSG-1.2 Peptide Binding to RRE-IIB RNA Element of HIV-1 Over Rev Peptide is Mainly Enthalpic in Origin. *PLoS One* **2011**, *6*, e23300.

(56) Coulocheri, S. A.; Pigis, D. G.; Papavassiliou, K. A.; Papavassiliou, A. G. Hydrogen Bonds in Protein-DNA complexes: Where Geometry Meets Plasticity. *Biochimie* **2007**, *89*, 1291–1303.

(57) Sarai, A.; Kono, H. Protein-DNA Recognition Patterns and Predictions. *Annu Rev Biophys Biomol Struct* **2005**, *34*, 379–398.

TOC Graphic