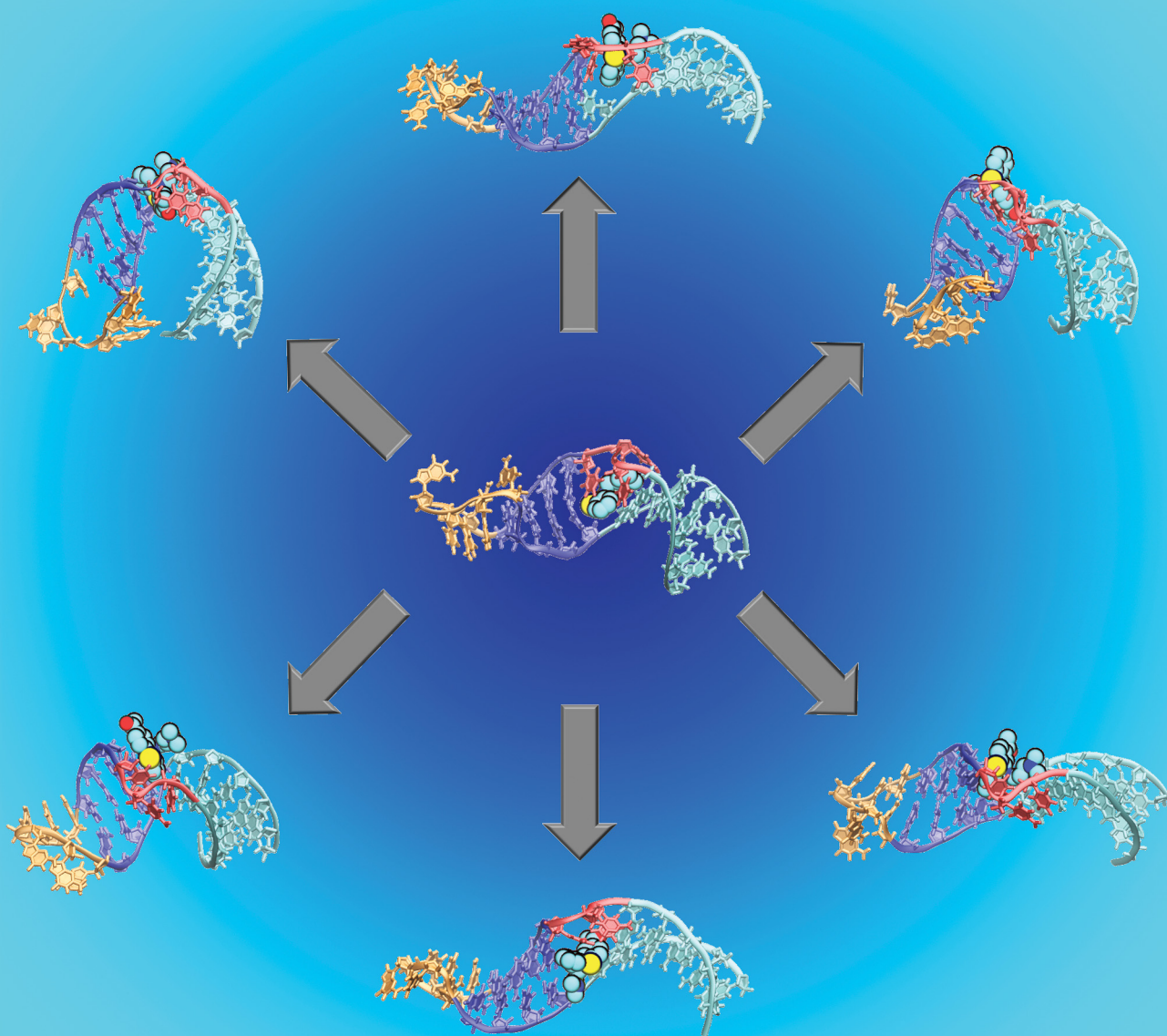


# PCCP

Physical Chemistry Chemical Physics

rsc.li/pccp



ISSN 1463-9076

**PAPER**

Lev Levintov and Harish Vashisth  
Role of conformational heterogeneity in ligand recognition  
by viral RNA molecules



Cite this: *Phys. Chem. Chem. Phys.*,  
2021, **23**, 11211

# Role of conformational heterogeneity in ligand recognition by viral RNA molecules†

Lev Levintov  and Harish Vashisth \*

Ribonucleic acid (RNA) molecules are known to undergo conformational changes in response to various environmental stimuli including temperature, pH, and ligands. In particular, viral RNA molecules are a key example of conformationally adapting molecules that have evolved to switch between many functional conformations. The transactivation response element (TAR) RNA from the type-1 human immunodeficiency virus (HIV-1) is a viral RNA molecule that is being increasingly explored as a potential therapeutic target due to its role in the viral replication process. In this work, we have studied the dynamics in TAR RNA in apo and liganded states by performing explicit-solvent molecular dynamics (MD) simulations initiated with 27 distinct structures. We determined that the TAR RNA structure is significantly stabilized on ligand binding with especially decreased fluctuations in its two helices. This rigidity is further coupled with the decreased flipping of bulge nucleotides, which were observed to flip more frequently in the absence of ligands. We found that initially-distinct structures of TAR RNA converged to similar conformations on removing ligands. We also report that conformational dynamics in unliganded TAR structures leads to the formation of binding pockets capable of accommodating ligands of various sizes.

Received 13th February 2021,  
Accepted 31st March 2021

DOI: 10.1039/d1cp00679g

rsc.li/pccp

## 1 Introduction

RNA molecules have long been considered primarily as passive carriers of genetic information but this conception has changed in recent years due to enhanced understanding of the roles of RNA in different cellular processes including translation and transcription,<sup>1</sup> regulation of gene expression,<sup>2</sup> and protein synthesis.<sup>3</sup> RNA is also implicated in various diseases, including cancers, neurological disorders, and viral infections.<sup>4–7</sup> This involvement of RNA presents an opportunity to target RNA by small-molecules for influencing the progression of various diseases.<sup>8</sup> Viral RNA molecules are a compelling target for small-molecule therapeutics since many viruses have RNA genomes, for example, human immunodeficiency virus (HIV), hepatitis C virus (HCV), influenza virus, and severe acute respiratory syndrome coronavirus (SARS CoV/CoV2).<sup>9–11</sup>

The variability in functions of RNA is rooted in its ability to undergo conformational changes that often lead to complex three-dimensional folds and an ensemble of structures which determine the function of RNA.<sup>12,13</sup> Conformational changes in RNA may be due to changes in physiological conditions or due

to binding of ligands (proteins, small ligands, and ions).<sup>14–16</sup> The conformational flexibility in RNA can also lead to formation of transient binding pockets that can be exploited for drug design.<sup>9,17,18</sup> While viral genomes encode for a limited number of protein targets, often considered “undruggable”,<sup>8</sup> conserved and structured RNA motifs in viral genomes are viable targets to discover binding pockets for small molecules.<sup>8,9</sup>

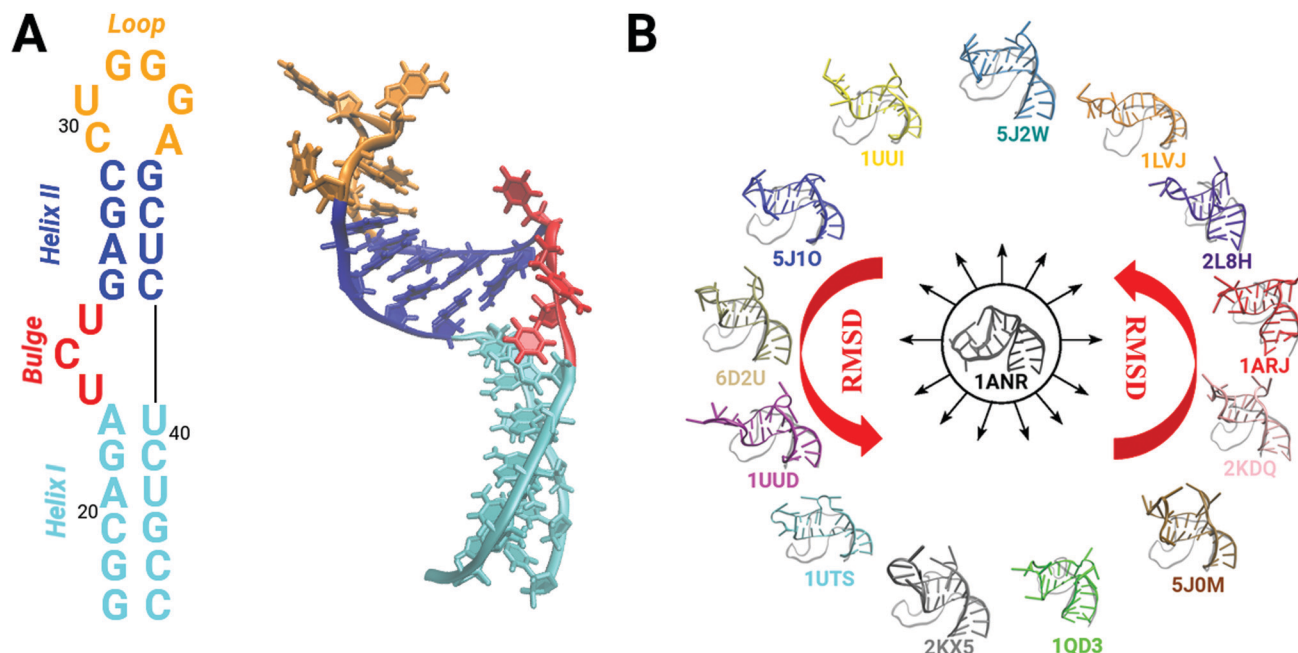
Although our understanding of RNA dynamics has increased *via* different experimental techniques,<sup>8,19–21</sup> especially those of RNA structure determination, the role of dynamics of all structural motifs in RNA and their coupling to ligand binding has not been fully explored to date.<sup>22</sup> Several experimental techniques including X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy can provide information on the dynamical properties of nucleic acids,<sup>23–25</sup> but even these methods are often limited in probing all possible parameters that can fully describe the dynamics in RNA molecules.<sup>16,26</sup> However, computational tools can further enhance our understanding of RNA dynamics by providing additional insights at the atomic level. Moreover, these tools can potentially assist in identifying binding pockets that can be explored in drug design.

TAR RNA (Fig. 1A) from HIV-1 is a key model system to study conformational transitions in RNA molecules due to its ability to adapt multiple states.<sup>12,25</sup> The TAR RNA is located at the 5' end of HIV-1 transcripts where it interacts with the viral transactivator

Department of Chemical Engineering, University of New Hampshire,  
Durham 03824, New Hampshire, USA. E-mail: harish.vashisth@unh.edu

† Electronic supplementary information (ESI) available: Additional analysis and snapshots from simulations are provided in Tables S1, S2 and Fig. S1–S18. See DOI: 10.1039/d1cp00679g





**Fig. 1** Sequence and structural details of HIV-1 TAR RNA. (A) Shown is the secondary structure and a snapshot of the three-dimensional structure of HIV-1 TAR RNA (PDB code 1ANR). Various structural motifs (Bulge, Helix I, Helix II, and Loop) are uniquely colored and labeled. (B) Shown are the snapshots of the initial states of TAR RNA (cartoon representation) in unliganded simulations. The *apo* conformation of TAR (PDB code 1ANR) is shown at the center (black cartoon) and superimposed onto other TAR RNA initial states. The initial conformations are placed in a circle such that the RMSD of the initial state relative to the *apo* conformation increases counterclockwise, with the 5J2W structure having the least RMSD and the 1LVJ structure having the highest RMSD.

(Tat) protein and the host cofactor cyclin T1 to promote efficient transcription of the downstream genome and is therefore considered to be an important drug target. Several previous studies have shown that TAR RNA can bind to peptide mimics,<sup>27–32</sup> to small molecules,<sup>33–36</sup> to proteins,<sup>37–40</sup> and to divalent cations<sup>41</sup> (examples are shown in Table S1, Fig. S1, ESI<sup>†</sup>). TAR RNA has been studied using NMR methods,<sup>25,27,42–51</sup> coarse-grained MD simulations,<sup>52</sup> electron paramagnetic resonance (EPR),<sup>53</sup> gel mobility,<sup>54</sup> combinations of NMR and MD methods,<sup>24,44</sup> and combinations of NMR and structure prediction software.<sup>55</sup> Collectively, these studies have shown that TAR RNA undergoes complex dynamics by sampling different interhelical conformations around the bulge junction, thus forming various conformational ensembles.<sup>12,43</sup> While detailed dynamics in TAR RNA have not been characterized in the presence of all known ligands, it has been suggested that ligands can potentially induce structural transitions in TAR by stabilizing pre-existing conformers or an ensemble of states in the *apo* TAR RNA structure.<sup>16,49,56,57</sup>

Several studies have revealed that a bulge motif in TAR RNA is especially critical for its recognition by the Tat protein.<sup>27,42,58</sup> Therefore, the bulge motif has been exploited in the design of inhibitors to disrupt the TAR/Tat interaction (Fig. S2A, ESI<sup>†</sup>).<sup>32,59</sup> As shown in Fig. S2A (ESI<sup>†</sup>), peptide ligands mostly interact with the apical loop (orange in Fig. 1A), helix II (blue in Fig. 1A), and the bulge motif (red in Fig. 1A), while small molecules are scattered between helices I and II (cyan and blue in Fig. 1A). The ligands differ in size and the charge value, which results in an increased buried surface area (BSA) as the

size of the ligand increases but the structural changes in TAR RNA are not correlated with the ligand size or BSA (Fig. S2B, ESI<sup>†</sup>).

In this work, we studied dynamics in TAR RNA by conducting long time-scale MD simulations that were initiated from 13 different initial conformations of TAR with ligands, and from 14 conformations without these ligands, 13 conformations after removing ligands and one conformation based on the experimental *apo* structure (Fig. 1B, Fig. S3 and Table S2, ESI<sup>†</sup>). By utilizing several different initial structures, we aimed to obtain a broader conformational mapping of TAR RNA which has not been carried out yet. Moreover, we reveal the effect of ligand binding on the dynamics in TAR RNA by comparing unliganded and liganded simulations.

## 2 Materials and methods

### 2.1 System preparation

We have studied the dynamics in HIV-1 TAR RNA using 27 different initial conformations with/without ligands (Fig. 1B and Fig. S3, ESI<sup>†</sup>). The initial coordinates for these conformations were obtained from the Protein Data Bank (PDB codes: 1ANR, 1ARJ, 1LVJ, 1QD3, 1UTS, 1UUD, 1UUI, 2KDQ, 2KX5, 2L8H, 5J0M, 5J10, 5J2W, 6D2U).<sup>27–36,42</sup> Several of these structures had either a different type of nucleotide or a different number of atoms in the deposited structure files. We selected the 1ANR conformation as our standard set of nucleotide



sequence and mutated or removed those nucleotides or atoms in the other 13 systems that were different from the *apo* structure (PDB code: 1ANR) which resulted in each TAR RNA system consistently having 29 nucleotides and 931 atoms (Table S2, ESI†). Each unliganded and liganded TAR structure was solvated in a periodic simulation domain of TIP3P water molecules where the overall number of atoms in various systems ranged between 19 443 and 25 647 (Table S2, ESI†). The overall charge in simulations of unliganded systems was neutralized with 29 Na<sup>+</sup> ions while the number of Na<sup>+</sup> ions in the liganded systems varied depending on the charge of the ligand.

## 2.2 Simulation details

All MD simulations were carried out and analyzed using software packages AMBER, CPPTRAJ and VMD<sup>60–62</sup> combined with the AMBER force-field for RNA (ff99OL3)<sup>63,64</sup> and for peptides (ff14sb).<sup>65</sup> For solvent, TIP3P water model<sup>66</sup> and for ions the Li/Merz parameters were used.<sup>67</sup> The Antechamber package was used to design force fields for small molecules by using the general AMBER force-field (GAFF) with the AM1-BCC charge method.<sup>68,69</sup> The temperature and pressure were maintained at 300 K and 1 atm using the Langevin thermostat and the Berendsen barostat. The steepest descent minimization was performed for 1000 steps followed by 100–500 steps of conjugate gradient minimization. The periodic boundary conditions were used with a cutoff of 9.0 Å for nonbonded interactions. Each of the 27 systems was subjected to a 2 μs long MD simulation in the *NPT* ensemble with a 2 fs timestep, which resulted in the

overall 54 μs dataset and the frames in each trajectory were saved every 20 ps.

## 2.3 Conformational metrics

**Torsional flexibility.** The overall torsional flexibility of each TAR RNA structure was investigated by computing (from MD simulation data) all backbone dihedral angles and the  $\chi$  dihedral angle (which describes the relative position of a nucleobase relative to the sugar group). The dihedral angles were computed for each nucleotide in each system across the entire trajectory (100 000 values per MD trajectory). The resulting values of dihedral angles were then used to compute their normalized distributions. These distributions were further compared against the typical ranges of values of dihedral angles known from experimentally determined structures of nucleic acids that were extracted from the Protein Data Bank and reported in the textbook by Tamar Schlick (Fig. 2A and Fig. S4, ESI†).<sup>70</sup>

**Buried surface area (BSA).** We calculated the BSA for each ligand in the liganded simulations. The BSA was computed using the following expression:

$$\text{BSA} = \text{SASA}_R + \text{SASA}_L - \text{SASA}_{RL}$$

where  $\text{SASA}_R$  represents the solvent accessible surface area (SASA) of the RNA,  $\text{SASA}_L$  represents the SASA of the ligand, and  $\text{SASA}_{RL}$  represents the combined SASA of the RNA/ligand complex. The BSA values indicate the area of contact between a ligand and the TAR RNA conformation.

**Root mean squared deviation (RMSD).** We calculated the all-atom RMSD for each system and for nucleotides in the bulge

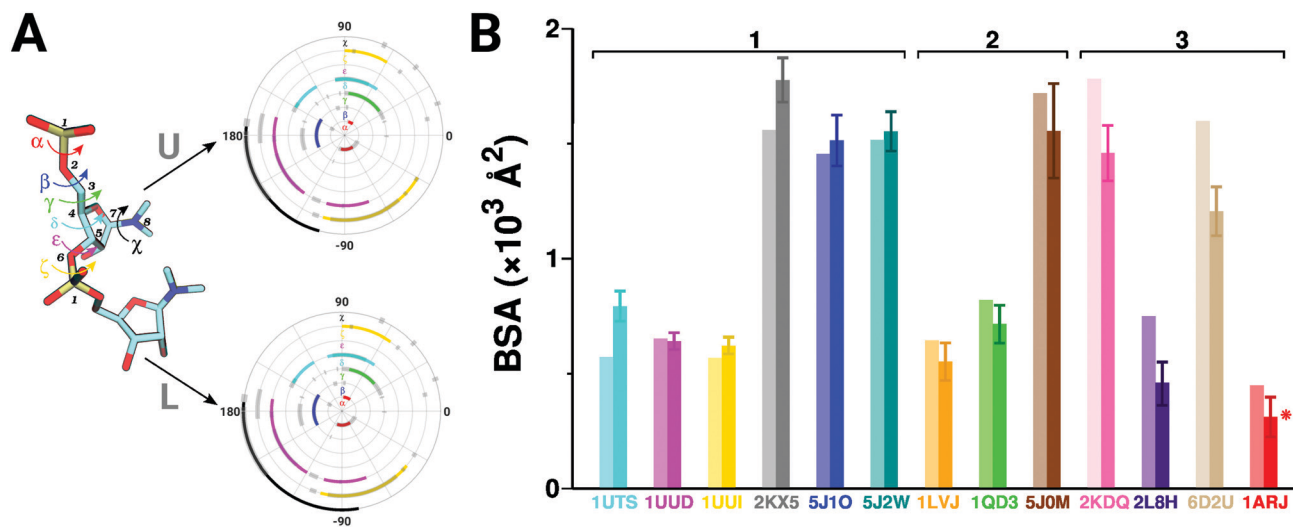


Fig. 2 Conformational metrics of torsional flexibility and BSA. (panel A; left) All dihedral angles are shown by an arrow and labeled on a snapshot of the polynucleotide chain. The atoms in the chain are labeled as follows: 1, P; 2, O5'; 3, C5'; 4, C4'; 5, C3'; 6, O3'; 7, C1'; and 8, N9/N1. (panel A; right, top and bottom) The normalized distributions of each RNA backbone dihedral angle ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$ ) and the glycosidic dihedral angle ( $\chi$ ) for unliganded (U) and liganded (L) simulations. The transparent and thicker gray lines represent expected ranges of dihedral angles based on experimentally known RNA structures. See also Fig. S4 (ESI†). (B) The histograms of mean BSA values based on liganded MD simulations (darker shades) and the initial liganded structures (lighter shades) are shown. The error bars (vertical lines marked on histograms) were computed based on each liganded simulation. The BSA histograms are organized into three groups (labeled 1, 2, and 3; marked by overbars). A red asterisk highlights a system (PDB code 1ARJ) which exhibited a partial dissociation of the ligand. See also Fig. S6 (ESI†).



motif (U23, C24, and U25) to understand the effect of ligand binding on the overall TAR RNA structure. The alignment of each structure was performed against all non-hydrogen atoms in the initial state. The RMSD values indicate changes in the TAR RNA structure relative to a reference state. We used the initial conformations as well as the average structures computed from each simulation as our reference states.

**Root mean squared fluctuation (RMSF).** We computed the backbone phosphorous (P) atom based RMSF per residue to further study the flexibility of each nucleotide and  $\Delta$ RMSF to compare the differences in dynamics between unliganded and liganded simulations. A negative value of  $\Delta$ RMSF signifies decreased fluctuations in the presence of ligands or increased fluctuations in the absence of ligands.

**Average structure.** We computed the average structures of TAR RNA from each unliganded and liganded simulation using the CPPTRAJ<sup>61</sup> program in the Amber software. The global rotational and translational motions were removed prior to computing the average structure. We then cross-compared all average structures using the RMSD as a comparison metric.

**Clustering analysis.** To determine the population of similar TAR RNA conformations in MD simulations, we performed a clustering analysis using the CPPTRAJ<sup>61</sup> tool with DBSCAN<sup>71</sup> clustering algorithm. We used the P-atom based RMSD in each TAR RNA conformation as a distance metric and a minimum of 25 conformations (RMSD within  $\sim 1.0$ – $1.5$  Å) were required to form a cluster. In addition to estimating clusters in individual trajectories, we performed combined cluster analysis (CCA) on the entire dataset of unliganded simulations and also on the entire dataset of liganded simulations by combining all trajectories. We used DBSCAN clustering algorithm with the minimum number of points set to 50 and the RMSD was set as a distance metric (RMSD within 1.9 Å). To conserve memory, we used every second frame of each trajectory resulting in 700 000 and 650 000 frames for combined clusters of unliganded and liganded simulations, respectively. The initial “sieve” value was set to 40 to form initial clusters, which means that every 40th frame was used to generate an initial cluster, resulting in 17 500 and 16 250 initial frames from unliganded and liganded simulations.

**Helical dynamics.** The TAR RNA structure consists of two helices (termed Helix I and II) that are linked by a flexible three nucleotide bulge motif (Fig. 1A). The dynamics in these helices were characterized using the  $\gamma_1$  and  $\gamma_2$  angles, which describe the twist of each helix around the helical axis, and by the  $\phi$  angle which describes the relative position of the helices. The  $\phi$  angle was defined between the centers of mass of the two helices. For the calculation of angles, Helix I was defined by the base pairs G17–C45, G18–C44, C19–G43, A20–U42, G21–C41, and A22–U40 (cyan in Fig. 1A), while Helix II was defined by the base pairs G26–C39, A27–U38, G28–C37, and C29–G36 (blue in Fig. 1A). The axes in helices were defined as passing through the centers of mass of the bottom and top base pairs in respective helices and the CPPTRAJ<sup>61</sup> program was used to compute the twist angles.

**Nucleotide flipping.** We characterized the flipping of nucleotides in the bulge motif (highlighted in red; Fig. 1A) using the

pseudo-dihedral angle computed between the centers of mass (COM) of four groups of atoms: the nitrogenous bases of U40 and A22, or C39 and C26, sugar moiety attached to A22 or C26, sugar moiety attached to U23, C24, or U25, and the nitrogenous base of U23, C24, or U25. The definition of the flipping angle was adopted from previous studies.<sup>72,73</sup> In all systems, the inward (flipped-in) state of a nucleotide corresponds to pseudo-dihedral angle values between  $-60^\circ$  and  $60^\circ$  but other values of the angle characterize an outward (flipped-out) state.

**Binding pocket analysis.** To characterize the binding pockets in TAR RNA conformations based on unliganded MD simulations, we used the MDpocket tool, which is an open-access pocket detection tool for MD trajectories.<sup>74</sup> Before analyzing each frame from MD simulations for pocket analysis, each trajectory was aligned to the initial structure based on the backbone P-atoms.

## 3 Results

### 3.1 Assessment of torsional flexibility and ligand stability

We first assessed the overall torsional flexibility of TAR by computing all backbone or glycosidic dihedral angles from unliganded and liganded simulations (Fig. 2A and Fig. S4, ESI<sup>†</sup>). Specifically, the distributions of these dihedral angles (Fig. 2A) were computed from the combined data of 14 unliganded simulations and similarly from the combined data of 13 liganded simulations. We found that the values spanned by these dihedral angles are consistent with the known ranges of dihedral angles in experimental structures of nucleic acids (marked by transparent gray lines on angle distributions in Fig. 2A), thereby highlighting that the TAR RNA conformations generated with the AMBER force-field are consistent with the expected dynamics in the backbone of RNA structures.

We also assessed whether the ligands remained associated with each TAR RNA structure during MD simulations. In Fig. 2B, we show the histograms of the mean values of BSA from liganded MD simulations (darker shades) along with the initial BSA values of ligands in various TAR structures (lighter shades). These BSA data are organized into three groups based on the distribution of BSA values in MD simulations in comparison to initial BSA values. We observed that most systems exhibited similar or higher ligand BSA values (*e.g.* group 1 systems with PDB codes 1UTS, 1UUD, 1UUI, 2KX5, 5J1O, and 5J2W) in comparison to the initial BSA due to conformational rearrangements of ligands in the binding pocket that led to a deeper burial of some ligands in the binding pocket (*e.g.*, see Fig. S5, ESI<sup>†</sup>). However, some systems exhibited fluctuations in nucleotides which allowed ligands to conformationally rearrange in the binding pocket and partially move out of the initial pocket, as indicated by the decreased BSA values (*e.g.* group 2 systems with PDB codes 1LVJ, 1QD3, and 5JOM; Fig. 2B).

A larger decrease in BSA of the ligand was observed in systems organized in group 3 (*e.g.* PDB codes 2KDQ, 2L8H, 6D2U, and 1ARJ) indicating that the ligands in these systems exhibited increased rearrangements in the binding pocket. For example,



the ligand arginine amide in one of the TAR RNA structures (PDB code 1ARJ; marked by a red asterisk in Fig. 2B) exhibited brief dissociation for about  $\sim 18$  ns before binding again in the initial binding pocket (Fig. S6, ESI<sup>†</sup>). This observation is consistent with the largest dissociation constant for this ligand<sup>27</sup> and the smallest size of this ligand among all ligands studied (Table S1, ESI<sup>†</sup>). However, we did not observe full dissociation of any ligand during MD simulations.

### 3.2 Ligands rigidify TAR RNA

To assess the differences in conformations of TAR RNA in unliganded and liganded states, we first computed the RMSD values with respect to the initial structure in each simulation (Fig. 3A; lighter and darker shades for unliganded and liganded simulations, respectively). We observed that the unliganded systems diverged from their respective initial states on average by  $5.28 \pm 1.12$  Å and the liganded systems by  $4.52 \pm 1.19$  Å. For several systems, we observed a significant decrease in mean

RMSD values and no overlap in error bars in liganded states compared to unliganded states (group 1; Fig. 3A). About half of liganded systems fluctuated more and thus had less distinct RMSD distributions in comparison to respective unliganded systems, as characterized by overlapping error bars (group 2; Fig. 3A). Despite having less distinct distributions, these liganded simulations still showed lower mean RMSD values in comparison to unliganded simulations (group 2; Fig. 3A). Two systems (PDB codes 1UUD and 2L8H) had almost no difference in their RMSD distributions in liganded and unliganded states (group 3; Fig. 3A). These data suggest that the TAR RNA structures became conformationally more rigid when ligands were present since liganded structures deviated to a smaller extent from their initial conformations in comparison to unliganded simulations.

However, one system (marked by an orange asterisk in group 3; Fig. 3A) showed increased fluctuations and a higher mean RMSD in the liganded state in comparison to the unliganded state. On further probing this structure (PDB code 1LVJ), we found that the structural deviation is likely a result of ligand rearrangements in the binding pocket that conformationally altered the TAR RNA structure from an initially bent conformation to a relaxed conformation with a higher RMSD value (Fig. S7, ESI<sup>†</sup>). The fact that the presence of a ligand caused higher perturbations in the TAR RNA structure that resulted in an unexpected conformational behavior needs to be considered in designing new inhibitors that target TAR RNA.

We further report the  $\Delta$ RMSF values per residue based upon unliganded and liganded simulations to understand the effect of ligand binding on the flexibility of a particular residue or a motif (Fig. S8, ESI<sup>†</sup>). In Fig. S8 (ESI<sup>†</sup>), we show the difference between the RMSF values ( $\Delta$ RMSF) of the unliganded and liganded simulations where a negative value corresponds to an increased flexibility on ligand removal. We observed that in eight systems, all of the residues became more flexible when ligands were removed (e.g. systems with PDB codes 1QD3, 1UUTS, 1UUI, 2KDQ, 2KX5, 5J0M, 5J1O, and 6D2U). In four systems, we observed larger flexibility in a portion of residues in the liganded simulations (e.g. systems with PDB codes 1ARJ, 1UUD, 2L8H, and 5J1O). Importantly, one of those systems (PDB code 5J1O) had only one residue (U25) with increased flexibility in the liganded state due to the flipping out of that nucleotide during the simulation. Finally, in one system (PDB code 1LVJ) all of the residues became more rigid after ligands were removed which was consistent with the aforementioned observation that the fluctuations and mean RMSD for this specific system was higher in the liganded conformation (Fig. 3A).

Due to the significance of the bulge motif in ligand recognition,<sup>27,58</sup> we also separately computed the RMSD values for the bulge motif (Fig. S9, ESI<sup>†</sup>). Overall, the bulge motifs in unliganded and liganded systems deviated by  $7.82 \pm 1.83$  Å and  $5.6 \pm 1.73$  Å, respectively. In particular, we observed that the majority of systems had higher mean RMSD values in the unliganded states with the exception of the system with the PDB code 1LVJ that had a higher mean RMSD value of the bulge

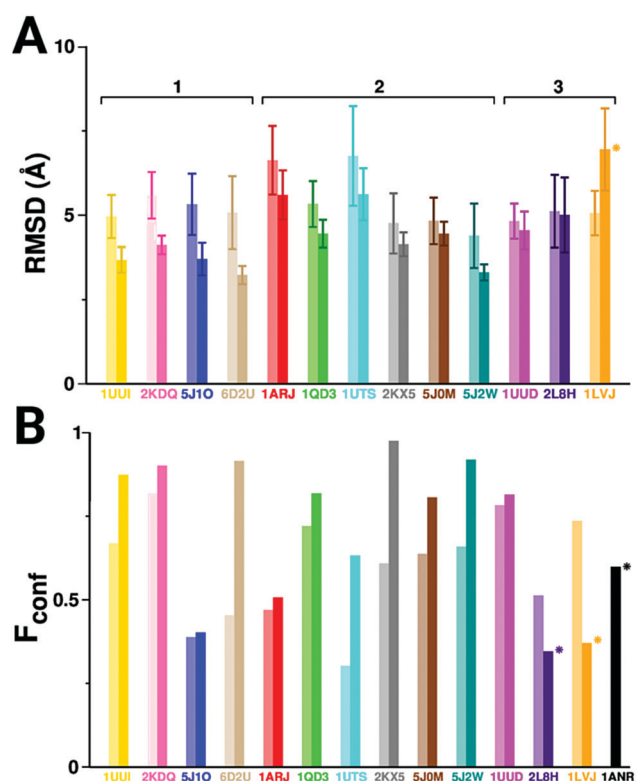


Fig. 3 RMSD and clustering analyses. (A) Shown are the histograms (with error bars) of mean values of RMSD for all unliganded (lighter shades) and liganded (darker shades) simulations. The RMSD histograms are organized into three groups (labeled 1, 2, and 3; marked by overbars). An orange asterisk marks a system (PDB code 1LVJ) which showed a different behavior in comparison to other systems. (B) The fraction of conformations ( $F_{conf}$ ) from a given simulation (100 000 conformations per simulation) that constitute the most populated cluster for each system in unliganded (lighter shades) and liganded (darker shades) simulations. Each bar corresponds to a unique system. The purple and orange asterisks indicate those systems in which  $F_{conf}$  was higher in unliganded simulations than in corresponding liganded simulations. A black asterisk marks the experimental apo TAR structure (PDB code 1ANR).



motif in the liganded state (Fig. S9, ESI†). This observation is also consistent with  $\Delta$ RMSF data that showed increased flexibility of the bulge nucleotides in the system with the PDB code 1LVJ (Fig. S8, ESI†). The behavior of the bulge motif in this system is coupled with the ligand movement in the binding pocket and the local rearrangements of bulge nucleotides that result in the overall change in the conformation of TAR RNA (Fig. S7, ESI†). These data suggest that the TAR RNA structures and the bulge nucleotides in liganded systems deviated to a smaller extent from their initial conformations compared to in the unliganded systems.

### 3.3 Comparison of average structures and formation of conformational clusters

We further investigated the conformational variability of our data by comparing the average structures from each simulation and by performing a cluster analysis. Using RMSD as a conformational metric, we cross-compared all initial TAR RNA structures before simulations were initiated (Fig. S10A, ESI†) as well as by obtaining the average structures of TAR RNA from each unliganded (Fig. S10B, ESI†) and liganded simulation (Fig. S10C, ESI†). This cross-comparison showed that the RMSD between a pair of structures decreased in unliganded simulations (Fig. S10B, ESI†), thereby indicating that the TAR RNA structures became on average more similar to each other in unliganded simulations. In liganded simulations, the RMSD between a pair of structures also decreased on average, but to a smaller extent in comparison to the unliganded simulations and the structures were still reasonably distinct (Fig. S10C, ESI†). For example, the initial RMSD between the structures with the PDB codes 1LVJ (orange) and 2L8H (purple) was 7.91 Å (dark purple bar in Fig. S10A, ESI†). After unliganded and liganded simulations, the RMSD between these systems for average structures was 3.8 Å and 4.8 Å, respectively (dark purple bars in Fig. S10B and C, ESI†). We also observed that the average structures of all systems obtained from unliganded MD simulations adopt conformations similar to the average structure obtained from an MD simulation of the experimental *apo* system (1ANR) (Fig. S11, ESI†).

To further assess the fluctuations and the flexibility in TAR RNA, we computed the RMSD values in the course of each simulation with respect to the average structure of the corresponding simulation (Fig. S12, ESI†). We observed that the unliganded systems diverged from their respective average structures by  $3.07 \pm 0.97$  Å and the liganded systems by  $2.31 \pm 0.57$  Å. The majority of the systems had a decrease in mean RMSD values and smaller magnitude of fluctuations in the liganded states compared to the unliganded states (group 1; Fig. S12, ESI†). Three systems exhibited very similar RMSD distributions in unliganded and liganded states (group 2; Fig. S12, ESI†), however, two of them still showed decreased mean RMSD values in the liganded state (PDB codes 1ARJ and 1UUD). Finally, one system (PDB code 1LVJ; Fig. S12, ESI†) showed increased fluctuations and a higher mean RMSD value in the liganded state in comparison to the unliganded state which is consistent with the observations described earlier.

Overall, this metric showed that the fluctuations in the TAR RNA structures decreased in the presence of ligands.

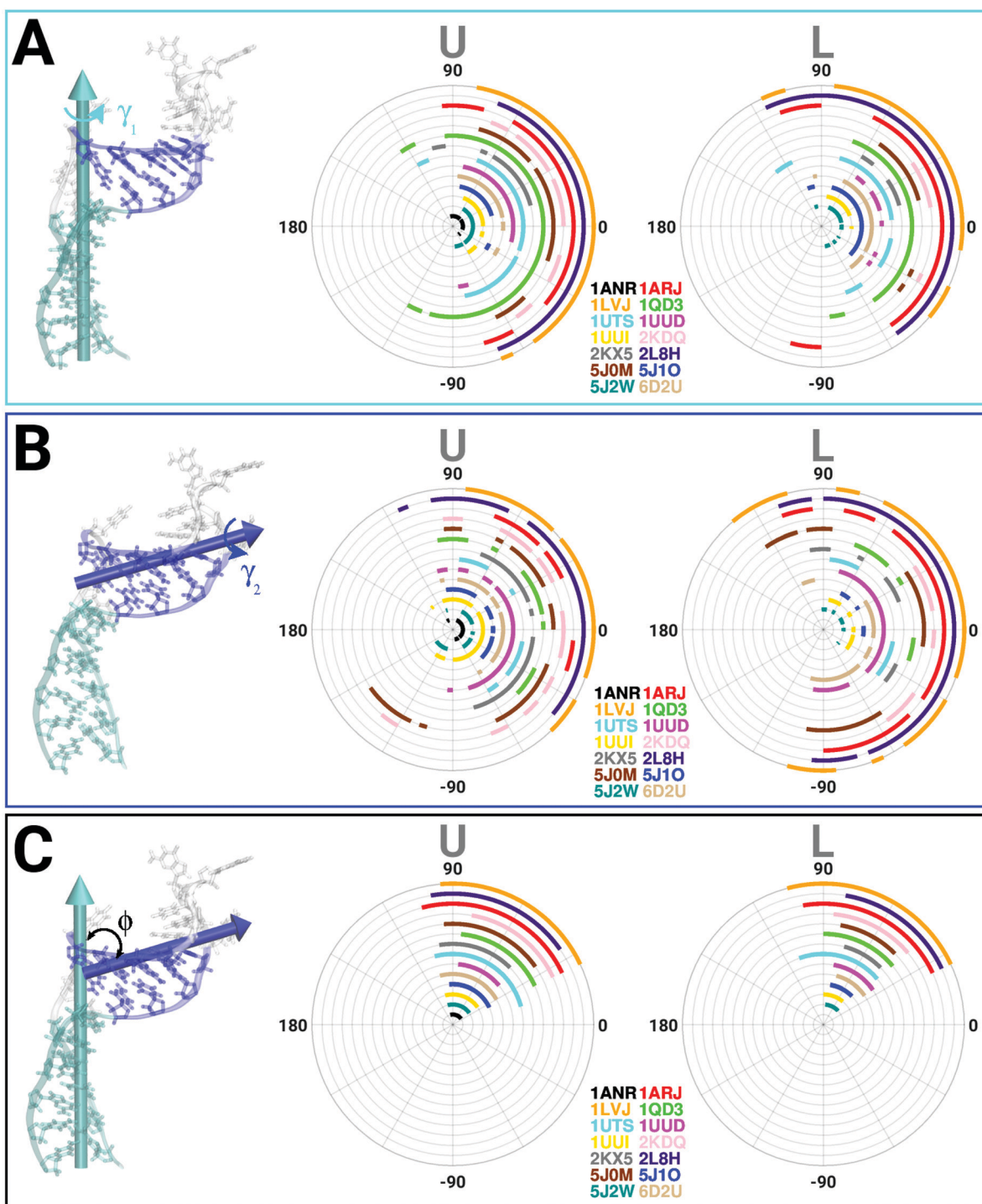
In addition to comparing the average structures from each simulation, we performed clustering analysis to detect similarities among structures within each simulation and to understand the effect of presence of ligands on conformational variability in TAR RNA. In Fig. 3B, we present the fraction of conformations in the most populated clusters derived from each unliganded and liganded simulation along with more comprehensive details on the distributions of clusters in Fig. S13 and S14 (ESI†). We observed a larger variation in conformational clusters in unliganded simulations in comparison to liganded simulations. For example, only two unliganded systems (PDB codes 1UUD and 2KDQ) had a cluster that contained at least 75% of structures while eight liganded simulations (PDB codes 1QD3, 1UUD, 1UUI, 2KDQ, 2KX5, 5J0M, 5J2W, and 6D2U) had a cluster of this type (Fig. 3B). The only exceptions were the unliganded systems with the initial structures based on PDBs 1LVJ (orange histograms in Fig. 3B) and 2L8H (purple histograms in Fig. 3B) that contained clusters with a higher fraction of conformations in the most populated cluster than in corresponding liganded simulations. Importantly, all liganded systems with peptide ligands (PDB codes 2KDQ, 2KX5, 5J0M, 5J1O, 5J2W, and 6D2U), except for the initial structure with the PDB code 5J1O, had the most populated cluster containing the majority of conformations (Fig. 3B). This analysis further supports the observation that ligands in general rigidify the TAR RNA structure by restricting its motion within an ensemble of structures that constitute the most populated cluster.

We also performed the combined cluster analysis (CCA) to further investigate conformational clusters in TAR simulations that were initiated with distinct initial structures. In Fig. S15 (ESI†), we show  $F_{\text{conf}}$  for the top clusters from the datasets of unliganded and liganded simulations. The CCA of unliganded simulations revealed three clusters that contain more than 5% of the total number of configurations each and are composed of multiple systems (Fig. S15A, ESI†). The CCA of liganded simulations, that was performed at the same value of the RMSD metric for constructing clusters as for the CCA of the unliganded simulations, revealed only one cluster which contains most of the systems (Fig. S15B, ESI†). These observations show that a number of our simulations, that were initiated from distinct initial structures, have conformations that are similar to each other.

### 3.4 Ligands alter helical dynamics in TAR RNA

The TAR RNA structure is comprised of two helices (termed Helix I and II in Fig. 1A) and the dynamics in these helices are described using the angles,  $\gamma_1$  and  $\gamma_2$  (Fig. 4A and B), which describe the twist of each helix around the helical axis and by the interhelical bending angle ( $\phi$ ) which describes the relative positioning of helices (Fig. 4C). We observed that the initial  $\gamma_1$  values for all systems were between  $29^\circ$  and  $34^\circ$  except for one system (PDB code 2L8H) where the angle was  $66.5^\circ$ . Similarly,





**Fig. 4** Intrahelical and interhelical dynamics in TAR RNA. (A; leftmost panel) A snapshot of the TAR RNA structure depicting the intrahelical angle  $\gamma_1$ , which describes the rotation of Helix I. The reference axis for the rotation of Helix I is marked by a cyan arrow. (A; middle and rightmost panels) The distributions of  $\gamma_1$  are shown for the unliganded (U) and liganded (L) simulations of each structure. (B and C) Data similar to panel A are shown for the intrahelical angle  $\gamma_2$ , which describes the rotation of Helix II (panel B), and  $\phi$ , the interhelical angle between Helix I and II (panel C). The reference axis for the rotation of Helix II is marked by a blue arrow. The color scheme in histograms is same as the PDB label.

the  $\gamma_2$  values were between  $27^\circ$  and  $38^\circ$  for all systems except one system (PDB code 1QD3) where it was  $16^\circ$ .

Based on the angle distributions from MD simulations, we observed that  $\gamma_1$  was mostly confined between  $-90^\circ$  and  $90^\circ$  for

the unliganded and liganded systems (Fig. 4A). Most of the liganded systems showed a decrease in the width of populated angles in comparison to the analogous systems in the unliganded form with the exception of the system with the PBD



code 2L8H (depicted in purple color in Fig. 4) which exhibited a higher twisting in the Helix I with  $\gamma_1$  angles between  $70^\circ$  and  $115^\circ$ . The system with the PDB code 1ARJ in the liganded form showed values similar to the unliganded conformation given conformational rearrangements in the ligand and a decreased BSA (Fig. 2 and Fig. S6, ESI<sup>†</sup>). The average  $\gamma_1$  values for the unliganded and liganded systems were estimated to be  $19 \pm 40^\circ$  and  $20 \pm 33^\circ$ , respectively. Overall, the presence of ligands decreased the standard deviation in  $\gamma_1$  by  $7^\circ$ , thereby leading to narrower  $\gamma_1$  distributions.

We also observed that Helix II showed more flexibility compared to Helix I in both liganded and unliganded simulations. In the unliganded simulations,  $\gamma_2$  was mostly distributed between  $-90^\circ$  and  $105^\circ$  with the exception of the structures with PDB codes 5J0M, 2KDQ, and 5J2W that spanned additional conformations between  $-105^\circ$  and  $-140^\circ$ , between  $-120^\circ$  and  $-130^\circ$ , and between  $-105^\circ$  and  $-145^\circ$ , respectively. The  $\gamma_2$  angle in the liganded systems was mostly confined between  $-105^\circ$  and  $120^\circ$  but even though the width of distributions were similar, the number of states that were populated decreased. The average  $\gamma_2$  values for the unliganded and liganded systems were estimated to be  $12 \pm 51^\circ$  and  $11 \pm 46^\circ$ . Overall, the presence of ligands decreased the standard deviation in  $\gamma_2$  by  $5^\circ$ .

We observed that  $\phi$  is mostly confined between  $25^\circ$  and  $105^\circ$  for the unliganded systems, with the exceptions of structures with the PDB codes 1QD3 and 2KDQ, which occupied states with angles between  $25^\circ$  and  $80^\circ$ , and 1UUD which occupied states between  $45^\circ$  and  $85^\circ$ . The distributions of  $\phi$  in the liganded systems became narrower (ranging between  $35^\circ$  and  $75^\circ$ ) with the exception of the structures with the PDB codes 1LVJ, 1ARJ, and 1UTS which spanned angles between  $25^\circ$  and  $105^\circ$ ,  $25^\circ$  and  $100^\circ$ , and  $45^\circ$  and  $110^\circ$ , respectively. Overall, the structural bending in TAR RNA decreased in the liganded systems, except for the systems with initial states based on the PDB codes 1UUD, 1UTS, 1LVJ and 1ARJ which have distributions similar to their unliganded systems. The average  $\phi$  for the unliganded and liganded systems was estimated to be  $70 \pm 14^\circ$  and  $60 \pm 11^\circ$ , respectively. While previous NMR analysis has suggested high amplitude bending and twisting motions in TAR RNA helices,<sup>24</sup> our data further suggest that the conformations of helices in TAR RNA are altered and stabilized on ligand binding.

### 3.5 Ligands stabilize nucleotide flipping in TAR RNA

Beyond global motions in TAR structures, we further probed local motions in key motifs such as the bulge region, which is considered important for the viral replication process because the rearrangements in nucleotides in this region (U23, C24, and U25) determine the orientation of helical motifs (Helix I and II) in TAR.<sup>27,42,58</sup> Specifically, the outward flipped conformations of nucleotides C24 and U25 facilitate coaxial stacking of Helices I and II, thereby rigidifying the TAR structure. In Fig. 5, we show the nucleotides used in defining the flipping angle ( $\theta$ ) and the time-traces of  $\theta$  for three bulge-nucleotides, as obtained from unliganded and liganded simulations. The inward flipping of a

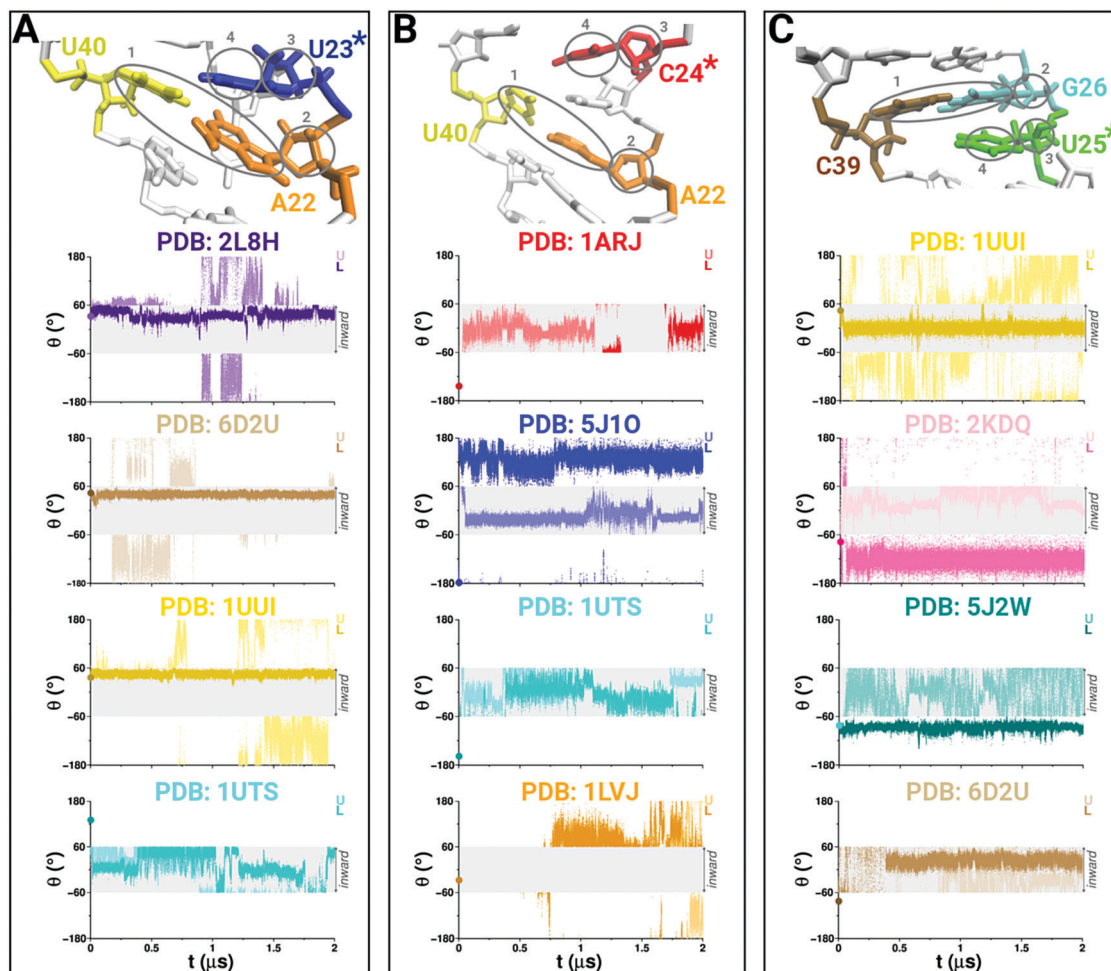
nucleotide is characterized by  $\theta$  values between  $-60^\circ$  and  $60^\circ$ , and the outward flipping for all other  $\theta$  values.

The first bulge-nucleotide U23 (Fig. 5A) was initially in a flipped-in state in most structures, except in two structures (PDB codes 1QD3 and 1UTS) where it was in a flipped-out state (as marked by a filled circle on the y-axis at  $t = 0$  in time-traces of  $\theta$  in Fig. 5A). We observed that U23 flipped out during five unliganded simulations (PDB codes 1ARJ, 1UUI, 2L8H, 5J2W, and 6D2U) in which U23 was initially in a flipped-in state (traces with lighter shades in Fig. 5A and Fig. S16, ESI<sup>†</sup>). In most of these systems U23 eventually returned to its initial position after briefly transitioning to a flipped-out conformation. For example, in one of the unliganded simulations (PDB code 2L8H; light purple time-traces in Fig. 5A) U23 flipped outward at  $t = \sim 0.9 \mu\text{s}$ , maintaining the flipped out state for  $\sim 0.5 \mu\text{s}$ , and then flipping inward, resuming its initial position.

We observed a similar conformational behavior in systems with PDB codes 6D2U (light brown time-traces in Fig. 5A) and 5J2W (light cyan time-traces in Fig. S16A, ESI<sup>†</sup>). In another unliganded system (PDB code 1UUI), U23 flipped out at  $t = \sim 1.25 \mu\text{s}$  and remained in a flipped-out state until the end of the simulation (light yellow time-traces in Fig. 5A). However, in one system (PDB code 1UTS) we observed conformational transitions in both unliganded and liganded simulations (cyan time-traces in Fig. 5A), where U23 flipped inward from an initially outward conformation with  $\theta = 130$ , retaining the inward position for almost the entirety of liganded simulation and flipping outward after  $\sim 1 \mu\text{s}$  in the unliganded simulation. In all liganded simulations, U23 transiently flipped outward only in one system (PDB code 1ARJ; Fig. S16A, ESI<sup>†</sup>). Overall, we observed that the presence of ligands significantly decreased conformational transitions in U23. This conformational behavior is consistent with observations from experiments, where a smaller pool of ligands (arginine amide and a linear as well as a cyclic peptide) were tested.<sup>75</sup> However, we consistently observed conformational stabilization of U23 for several ligands with different binding affinities.

The second bulge-nucleotide C24 (Fig. 5B) was initially in a flipped-out conformation in all systems, except in three systems (PDB codes 1ANR, 1LVJ, and 1QD3). We observed that C24 flipped inward in three unliganded simulations (PDB codes 1ARJ, 5J1O, and 1UTS; lighter shade time-traces shown in red, blue, and cyan in Fig. 5B). For example, C24 flipped inward in one of the unliganded systems (PDB code 1ARJ; Fig. 5B) and remained in the inward conformation until  $\sim 1.15 \mu\text{s}$ . However, in other unliganded systems (PDB codes 5J1O and 1UTS), C24 flipped inward at the beginning of simulations where it either remained flipped inward during the entire simulation (PDB code 5J1O; Fig. 5B) or flipped outward and then flipped back inward toward the end of the simulation (PDB code 1UTS; Fig. 5B). For two unliganded systems where C24 was initially in an inward flipped conformation (PDB codes 1LVJ and 1QD3), it flipped outward toward the end of simulations (Fig. 5B and Fig. S16B, ESI<sup>†</sup>). In liganded simulations, we observed that C24 flipped inward during two simulations (PDB codes 1ARJ and 1UTS; darker red and cyan time-traces in Fig. 5B) and flipped





**Fig. 5** Conformational transitions in bulge nucleotides. (A) A snapshot of TAR RNA nucleotides (stick representation) used in defining the flipping angle ( $\theta$ ) for U23 (blue sticks and marked by an asterisk) and the traces of  $\theta$  vs. simulation time ( $t$ ) are shown for four systems in which a conformational transition was observed either in the unliganded state (U; lighter shade) or in the liganded state (L; darker shade) or in both. The initial value of  $\theta$  for U23 in each system is marked on the y-axis by a filled circle in the same color as traces. The inward flipped state is characterized by  $\theta$  values between  $-60^\circ$  and  $+60^\circ$  (labeled and shown by a transparent gray rectangle). All other values of  $\theta$  indicate an outward flipped state. For those unliganded and liganded simulations where a transition occurred in both simulations, only those values of  $\theta$  are plotted where a transition was observed. In case the transition was observed only in an unliganded simulation (or *vice versa* in a liganded simulation), in addition to plotting  $\theta$  values in the unliganded simulation where the transition occurred, all values of  $\theta$  are shown for the corresponding liganded simulation (or *vice versa* corresponding unliganded simulation) where the transition was not observed. (B and C) Data similar to panel A are shown for the flipping of nucleotides C24 (red sticks and marked by an asterisk; panel B) and U25 (green sticks and marked by an asterisk; panel C). The flipping angle of a nucleotide is defined by the center of mass of each of the following four groups: the nitrogenous bases of base-paired nucleotides (labeled 1) neighboring the flipping base, sugar moiety (labeled 2) attached to the base that is stacked with the flipping base, sugar moiety (labeled 3) attached to the flipping base, and the nitrogenous base (labeled 4) of the flipping nucleotide. See also Fig. S16 (ESI $\dagger$ ).

outward during one simulation (PDB code 1LVJ; darker brown time-trace in Fig. 5B). Overall, we observed that C24 showed conformational transitions both in unliganded and liganded simulations, but less frequently in liganded simulations.

The third bulge-nucleotide U25 (Fig. 5C) was initially in a flipped-in conformation in most structures except in three structures (PDB codes 2KDQ, 5J2W, and 6D2U). We observed U25 to be significantly flexible in both unliganded and liganded simulations since it flipped inward or outward in most of the systems. For example, U25 flipped outward from an initially inward conformation in several unliganded simulations (PDB codes 1UUI, 1UTS, 2KX5, 2L8H, and 5J1O; Fig. 5C and Fig. S16C,

ESI $\dagger$ ). It also flipped inward from an initially outward conformation during unliganded simulations of several systems (PDB codes 2KDQ, 5J2W, and 6D2U) in which it remained in the inward conformation until the end of each simulation (Fig. 5C). In several liganded simulations (PDB codes 1ARJ, 1UTS, 2L8H, and 5J1O), U25 flipped outward from an initially inward conformation (Fig. S16C, ESI $\dagger$ ), while in other systems (*e.g.* PDB code 6D2U) it flipped inward from an initially outward conformation (Fig. 5C). Overall, we observed that all three bulge nucleotides (U23, C24, and U25) can conformationally transition between inward and outward states although ligands decrease the frequency of these transitions.



### 3.6 Conformational dynamics in TAR RNA reveal ligand binding pockets

Given that the knowledge of binding pockets is useful in developing novel inhibitors,<sup>8</sup> we probed all unliganded simulations for the presence of binding pockets that may form as a result of conformational dynamics. In Fig. 6A and Fig. S17 (ESI<sup>†</sup>), we show several binding pockets (depicted as cyan surfaces overlaid on the initial structure) that appear in various regions of each unliganded TAR RNA structure (labeled B, L, H1, and H2 for the bulge region, loop region, and helices I and II, respectively; see also Fig. 1A).

We also assessed whether the density of pockets observed in specific regions in a structure could accommodate ligands in conformations observed in liganded TAR RNA structures. We found that the observed pockets were sufficiently large in size to encapsulate the ligand known to bind to that specific TAR RNA

structure. For example, TAR RNA is known to bind to small-molecule ligands (acetylpromazine and RBT158) in the bulge region, where we observed binding pockets (labeled B for PDBs 1ANR and 5J1O in Fig. 6A) large enough to accommodate each ligand (Fig. 6B). Furthermore, TAR RNA is also known to bind other ligands (neomycin B and JB181) in the Helix I and the apical-loop/bulge regions, where we observed binding pockets (labeled H1 for PDB 6D2U and L/B for PDB 1UUD in Fig. 6A) large enough to accommodate respective ligands (Fig. 6C and D). We observed that unliganded simulations with different initial structures showed several similar binding pockets as well as previously unknown binding pockets (in the apical loop or Helices I and II) that accommodated ligands known to experimentally bind to other conformations (Fig. S18, ESI<sup>†</sup>).

## 4 Discussion

In this work, we have carried out long time-scale MD simulations (totaling 54  $\mu$ s) of the HIV-1 TAR RNA structure in unliganded and liganded states. Specifically, these simulations were conducted with initial coordinates derived from the experimentally resolved *apo* structure of TAR RNA (PDB code 1ANR) as well as 13 other structures of TAR RNA that were bound to a variety of ligands including small-molecules and peptides. To increase the pool of unliganded simulations, we also conducted simulations of 13 liganded TAR RNA structures by removing ligands and retaining the initial coordinates for RNA atoms. We aimed to probe conformational heterogeneity in ensembles of TAR RNA structures in unliganded and liganded states to understand the predisposition of the unliganded TAR conformations to ligand binding and the effect thereafter.

We initially assessed the overall torsional flexibility of TAR RNA by computing distributions of all backbone dihedral angles from unliganded and liganded simulations (Fig. 2A and Fig. S4, ESI<sup>†</sup>) and found these distributions to be consistent with the range of values from experimentally known structures of nucleic acids. This observation supports the ability of the interatomic potential in adequately capturing the dynamics in TAR RNA structures. By computing the buried surface area (BSA) of each ligand, we also assessed the stability of ligands during long time-scale MD simulations and found that in most TAR RNA structures ligands remained bound throughout simulations except in a few cases where ligands conformationally rearranged and/or partially dissociated.

We then probed the global dynamics in TAR RNA by comparing all unliganded and liganded conformations from MD simulations using global RMSD,  $\Delta$ RMSF, and clustering analyses (Fig. 3 and Fig. S8, S12, ESI<sup>†</sup>). The primary observation from the RMSD analysis was that the mean RMSD of unliganded conformations was higher than the mean RMSD of liganded conformations, thereby suggesting decreased conformational fluctuations in TAR RNA on ligand binding. The analysis of  $\Delta$ RMSF further supported this observation since the magnitude of fluctuations in nucleotides was smaller in the

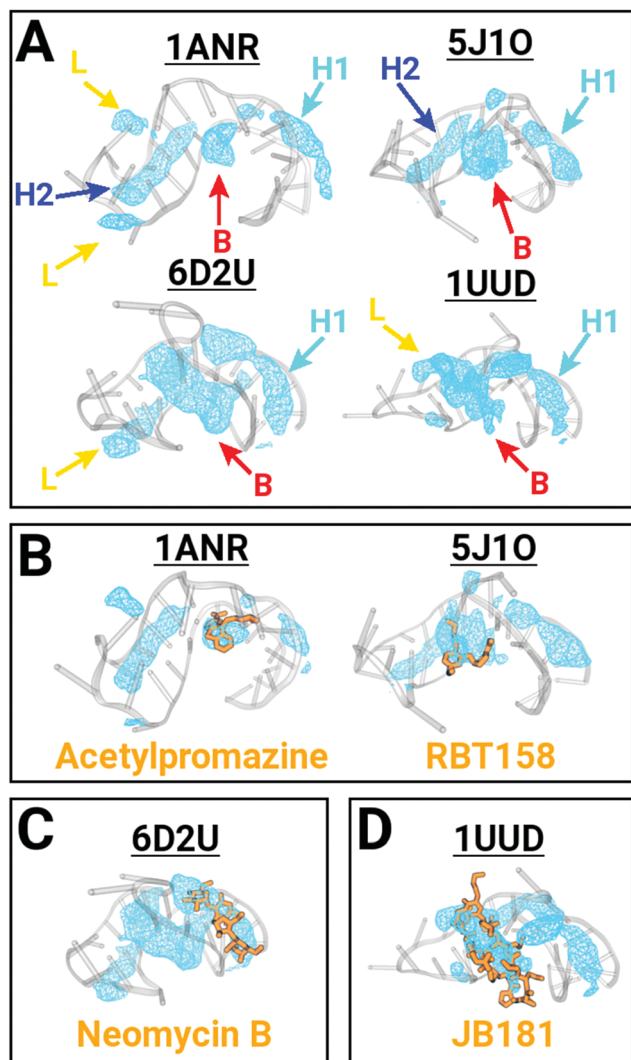


Fig. 6 Predicted binding pockets in unliganded TAR structures. (A) Predicted binding pockets (cyan surfaces) are shown overlaid on each TAR RNA structure (transparent white cartoon). (B–D) Snapshots of the overlay of each ligand (orange sticks) on the predicted binding pocket where the ligand is known to bind in each structure. See also Fig. S17 and S18 (ESI<sup>†</sup>).



liganded simulations than in the unliganded simulations. These observations are consistent with the notion that RNA molecules are stabilized by binding of ligands because liganded TAR RNA structures were conformationally more rigid compared to unliganded structures.

However, one of the small molecules, acetylpromazine, resulted in distinct perturbations in the overall structure of the liganded TAR RNA in comparison to other liganded systems and in comparison to the corresponding unliganded simulation. In the presence of acetylpromazine, the TAR RNA structure transitioned between two distinct (bent and stretched) conformations (Fig. S7, ESI<sup>†</sup>). We have previously also shown that the (un)binding process of acetylpromazine is associated with the flipping of nucleotides in the binding pocket.<sup>76</sup> The main structural difference between acetylpromazine and other small molecules is the presence of a sulfur moiety in one of the benzoic rings (Fig. S1, ESI<sup>†</sup>), which could be an important design feature for future development of inhibitory compounds.

The bulge motif which connects two helices in the TAR RNA structure (Fig. 1A) also became more rigid in the presence of ligands, which decreased the twisting and bending fluctuations in TAR RNA helices. The clustering analysis further supported these observations by showing a higher fraction of similar conformations in liganded structures in comparison to unliganded structures. In fact some liganded structures with peptide ligands exhibited a single cluster containing more than 90% of the RNA conformations, implying that these liganded simulations exhibited small conformational variability in the presence of peptides since the TAR RNA conformations were similar to each other (*e.g.* PDB code 2KX5; Fig. 3B and Fig. S14, ESI<sup>†</sup>). We also observed that various simulations have similar conformations that form combined clusters by performing combined cluster analysis (Fig. S15, ESI<sup>†</sup>) and by comparing average structures from each simulation (Fig. S10 and S11, ESI<sup>†</sup>). These observations further support the previously proposed hypothesis<sup>12,26,43</sup> that despite being a highly flexible molecule, TAR RNA potentially adopts a set of conformations forming an ensemble of structures that can recognize various ligands.

We further probed the local dynamics in bulge nucleotides (U23, C24, and U25), the conformational flipping motions in which facilitate ligand binding<sup>27,42</sup> as well as alter global dynamics in TAR RNA. As opposed to the notion that the binding of ligands may prevent conformational transitions in nucleotides, we observed that the bulge nucleotides can transition between inward and outward conformations in both unliganded and liganded states although the frequency of transitions significantly decreases in the presence of ligands. Overall, we found bulge nucleotides C24 and U25 to be more flexible than U23.

Importantly, as a result of conformational heterogeneity in TAR RNA structures and coupling between local and global dynamics, we observed the formation of ligand binding pockets near several structural motifs (bulge region and helices I/II). This observation is consistent with the suggestion that TAR RNA may adopt conformations with pre-existing binding pockets where ligands can fit.<sup>12</sup> We observed that these binding pockets

form consistently in all unliganded simulations with enough volume to accommodate different ligands (Fig. 6 and Fig. S17, S18, ESI<sup>†</sup>), including larger ligands (*e.g.* peptides). Moreover, we observed the formation of binding pockets in other structural motifs (*e.g.* the apical loop) in TAR RNA which are potentially useful to future inhibitor design. As an example, Patwardhan *et al.*<sup>59</sup> showed that the amiloride ligands can bind to nucleotides in the apical loop where we observed several binding pockets.

## 5 Conclusions

Although RNA molecules are known to undergo conformational changes during various cellular processes, the conformational dynamics in RNA molecules, with and without ligands, have been studied only to a limited extent. We used explicit-solvent MD simulations to study the dynamics in a model RNA system, the HIV-1 TAR RNA, which is known to recognize several types of ligands including small-molecules and peptides. We observed that the ligands rigidified TAR RNA structures by interacting with the bulge nucleotides and decreased the overall bending and twisting motions in helical motifs in TAR RNA. Therefore, we found that ligands overall decreased conformational heterogeneity in TAR structures. While RNA is considered a highly flexible molecule, we observed that TAR RNA structures on average became more similar to each other in the unliganded and liganded simulations compared to their initial conformations. We also observed that the conformational transitions leading to flipping of nucleotides in RNA molecules likely occur irrespective of the presence of ligands although the frequency of these transitions decreases on ligand binding. As a result of conformational heterogeneity, we also showed that unliganded RNA molecules possess ligand binding pockets that may be amenable to targeting by novel inhibitory molecules. For sharing with the scientific community, we have further made our simulation data available *via* the Zenodo platform (<https://doi.org/10.5281/zenodo.4521164>).

## Author contributions

L. L. and H. V. designed the research. L. L. performed the research and analyzed data. L. L. and H. V. wrote the article.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

We acknowledge the financial support from the National Science Foundation (CBET-1554558) and computational support through the following resources: Premise, a central shared HPC cluster at UNH supported by the Research Computing Center; BioMade, a heterogeneous CPU/GPU cluster supported by the NSF EPSCoR award (OIA-1757371); and the NSF supported



(ACI-1548562) Extreme Science and Engineering Discovery Environment (XSEDE)<sup>77</sup> Comet resource at San Diego Super-computer Center (SDSC) under grant TG-MCB160183 (HV).

## Notes and references

- P. Nissen, J. Hansen, N. Ban, P. B. Moore and T. A. Steitz, *Science*, 2000, **289**, 920–930.
- T. Cech and J. Steitz, *Cell*, 2014, **157**, 77–94.
- M. J. Fedor and J. R. Williamson, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 399–412.
- J. M. Watts, K. K. Dang, R. J. Gorelick, C. W. Leonard, J. W. Bess Jr, R. Swanstrom, C. L. Burch and K. M. Weeks, *Nature*, 2009, **460**, 711–716.
- A. Schmitt and H. Chang, *Cancer Cell*, 2016, **29**, 452–463.
- M. Esteller, *Nat. Rev. Genet.*, 2011, **12**, 861–874.
- J. Chen, D. VanEtten, M. Fountain, I. Yildirim and M. Disney, *Biochemistry*, 2017, **56**, 3463–3474.
- K. Warner, C. Hajdin and K. Weeks, *Nat. Rev. Drug Discovery*, 2018, **17**, 547–558.
- T. Hermann, *RNA Therapeutics. Topics in Medicinal Chemistry*, Springer, Cham, 2017, vol. 27.
- A. E. Gorbalenya, S. C. Baker, R. S. Baric, R. J. de Groot, C. Drosten, A. A. Gulyaeva, B. L. Haagmans, C. Lauber, A. M. Leontovich, B. W. Neuman, D. Penzar, S. Perlman, L. L. M. Poon, D. V. Samborskiy, I. A. Sidorov, I. Sola and J. Ziebuhr, *Nat. Microbiol.*, 2020, **5**, 536–544.
- N. M. Bouvier and P. Palese, *Vaccine*, 2008, **26**, D49–D53.
- L. Ganser, M. Kelly, D. Herschlag and H. M. Al-Hashimi, *Nat. Rev. Mol. Cell Biol.*, 2019, **20**, 474–489.
- E. A. Dethoff, J. Chugh, A. M. Mustoe and H. M. Al-Hashimi, *Nature*, 2012, **482**, 322–330.
- N. Leulliot and G. Varani, *Biochemistry*, 2001, **40**, 7947–7956.
- J. Williamson, *Nat. Struct. Biol.*, 2000, **7**, 834–837.
- L. R. Ganser, J. Lee, A. Rangadurai, D. K. Merriman, M. L. Kelly, A. D. Kansal, B. Sathyamoorthy and H. M. Al-Hashimi, *Nat. Struct. Mol. Biol.*, 2018, **25**, 425–434.
- J. A. Howe, H. Wang, T. O. Fischmann, C. J. Balibar, L. Xiao, A. M. Galgoci, J. C. Malinverni, T. Mayhood, A. Villafania, A. Nahvi, N. M. C. M. Barbieri, P. A. Mann, D. Carr, E. Xia, P. Zuck, D. Riley, R. E. Painter, S. S. Walker, B. Sherborne, R. de Jesus, W. Pan, M. A. Plotkin, J. Wu, D. Rindgen, J. Cummings, C. G. Garlisi, R. Zhang, P. R. Sheth, C. J. Gill, H. Tang and T. Roemer, *Nature*, 2015, **526**, 672–677.
- C. M. Connelly, M. H. Moon and J. S. Schneekloth Jr, *Cell Chem. Biol.*, 2016, **23**, 1077–1090.
- M. D. Disney, *J. Am. Chem. Soc.*, 2019, **141**, 6776–6790.
- M. D. Disney, I. Yildirim and J. L. Childs-Disney, *Org. Biomol. Chem.*, 2014, **12**, 1029–1039.
- A. U. Juru, N. Patwardhan and A. Hargrove, *ACS Chem. Biol.*, 2019, **14**, 824–838.
- T. Schlick and A. Pyle, *Biophys. J.*, 2017, **113**, 225–234.
- H. M. Al-Hashimi, *J. Magn. Reson.*, 2013, **237**, 191–204.
- L. Salmon, G. Bascom, I. Andricioaei and H. M. Al-Hashimi, *J. Am. Chem. Soc.*, 2013, **135**, 5457–5466.
- E. A. Dethoff, K. Petzold, J. Chugh, A. Casiano-Negroni and H. M. Al-Hashimi, *Nature*, 2012, **491**, 724–728.
- A. T. Frank, A. C. Stelzer, H. M. Al-Hashimi and I. Andricioaei, *Nucleic Acids Res.*, 2009, **37**, 3670–3679.
- F. Aboul-ela, J. Karn and G. Varani, *J. Mol. Biol.*, 1995, **253**, 313–332.
- A. Davidson, T. Leeper, Z. Athanassiou, K. Patora-Komisarka, J. Karn, J. Robinson and G. Varani, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 11931–11936.
- A. Davidson, K. Patora-Komisarka, J. Robinson and G. Varani, *Nucleic Acids Res.*, 2011, **39**, 248–256.
- A. Davidson, D. Begley, C. Lau and G. Varani, *J. Mol. Biol.*, 2011, **410**, 984–996.
- A. Borkar, M. Bardaro Jr, C. Camilloni, F. Aprile, G. Varani and M. Vendruscolo, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 7171–7176.
- M. Shortridge, P. Wille, A. Jones, A. Davidson, J. Bogdanovic, E. Arts, J. Karn, J. Robinson and G. Varani, *Nucleic Acids Res.*, 2019, **47**, 1523–1531.
- B. Davis, M. Afshar, G. Varani, A. Murchie, J. Karn, G. Lentzen, M. Drysdale, J. Bower, A. Potter, I. D. Starkey, T. Swarbrick and F. Aboul-ela, *J. Mol. Biol.*, 2004, **336**, 343–356.
- C. Faber, H. Sticht, K. Schweimer and P. Rösch, *J. Mol. Biol.*, 2000, **275**, 20660–20666.
- A. Murchie, B. Davis, C. Isel, M. Afshar, M. Drysdale, J. Bower, A. Potter, I. Starkey, T. Swarbrick, S. Mirza, C. Prescott, P. Vaglio, F. Aboul-ela and J. Karn, *J. Mol. Biol.*, 2004, **336**, 625–638.
- Z. Du, K. E. Lind and T. L. James, *Chem. Biol.*, 2002, **9**, 707–712.
- I. A. Belashov, D. W. Crawford, C. E. Cavender, P. Dai, P. C. Beardslee, D. H. Mathews, B. L. Pentelute, B. R. McNaughton and J. E. Wedekind, *Nucleic Acids Res.*, 2018, **46**, 6401–6415.
- V. V. Pham, C. Salguero, S. N. Khan, J. L. Meagher, W. C. Brown, N. Humbert, H. de Rocquigny, J. L. Smith and V. M. D'Souza, *Nat. Commun.*, 2018, **9**, 4266.
- U. Schulze-Gahmen and J. H. Hurley, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 12973–12978.
- S. S. Chavali, S. M. Mali, J. L. Jenkins, R. Fasan and J. E. Wedekind, *J. Biol. Chem.*, 2020, **49**, 16470–16486.
- J. A. Ippolito and T. A. Steitz, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 9819–9824.
- F. Aboul-ela, J. Karn and G. Varani, *Nucleic Acids Res.*, 1996, **24**, 3974–3981.
- J. Lee, E. A. Dethoff and H. M. Al-Hashimi, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 9485–9490.
- E. A. Dethoff, A. L. Hansen, C. Musselman, E. D. Watt, I. Andricioaei and H. M. Al-Hashimi, *Biophys. J.*, 2008, **95**, 3906–3915.
- R. Fonseca, D. V. Pachov, J. Bernauer and H. van den Bedem, *Nucleic Acids Res.*, 2014, **42**, 9562–9572.
- H. M. Al-Hashimi, S. W. Pitt, A. Majumdar, W. Xu and D. J. Patel, *J. Mol. Biol.*, 2003, **329**, 867–873.
- W. Huang, G. Varani and G. P. Drobny, *J. Am. Chem. Soc.*, 2010, **132**, 17643–17645.



- 48 B. Zhao and Q. Zhang, *Curr. Opin. Struct. Biol.*, 2015, **30**, 134–146.
- 49 Q. Zhang, A. C. Stelzer, C. K. Fisher and H. M. Al-Hashimi, *Nature*, 2007, **450**, 1263–1267.
- 50 N. I. Orlovsky, H. M. Al-Hashimi and T. G. Oas, *J. Am. Chem. Soc.*, 2020, **142**, 907–921.
- 51 L. R. Ganser, C.-C. Chu, H. P. Bogerd, M. L. Kelly, B. R. Cullen and H. M. Al-Hashimi, *Cell Rep.*, 2020, **30**, 2472–2480.
- 52 A. Mustoe, H. M. Al-Hashimi and C. L. Brooks, *J. Phys. Chem. B*, 2014, **118**, 2615–2627.
- 53 N. K. Kim, A. Murali and V. J. DeRose, *Chem. Biol.*, 2004, **11**, 939–948.
- 54 F. A. Riordan, A. Bhattacharyya, S. McAteer and D. M. J. Lilley, *J. Mol. Biol.*, 1992, **226**, 305–310.
- 55 H. Shi, A. Rangadurai, H. Abou Assi, R. Roy, D. A. Case, D. Herschlag, J. D. Yesselman and H. M. Al-Hashimi, *Nat. Commun.*, 2020, **11**, 5531.
- 56 H. M. Al-Hashimi, Y. Gosser, A. Gorin, W. Hu, A. Majumdar and D. J. Patel, *J. Mol. Biol.*, 2002, **315**, 95–102.
- 57 L. R. Ganser, M. L. Kelly, N. N. Patwardhan, A. E. Hargrove and H. M. Al-Hashimi, *J. Mol. Biol.*, 2020, **432**, 1297–1304.
- 58 D. Merriman, Y. Xue, S. Yang, I. Kimsey, A. Shakya, M. Clay and H. M. Al-Hashimi, *Biochemistry*, 2016, **55**, 4445–4456.
- 59 N. N. Patwardhan, L. R. Ganser, G. J. Kapral, C. S. Eubanks, J. Lee, B. Sathyamoorthy, H. M. Al-Hashimi and A. E. Hargrove, *Med. Chem. Commun.*, 2017, **8**, 1022–1036.
- 60 D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER 2018*, University of California, San Francisco, 2018.
- 61 D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
- 62 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.
- 63 A. Pérez, I. Marchán, D. Svozil, J. Šponer, T. E. Cheatham, III, C. A. Loughton and M. Orozco, *Biophys. J.*, 2007, **92**, 3817–3829.
- 64 M. Zgarbová, M. Otyepka, J. Šponer, A. Mládek, P. Banáš, T. E. Cheatham and P. Jurečka, *J. Chem. Theory Comput.*, 2011, **7**, 2886–2902.
- 65 J. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
- 66 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
- 67 P. Li, L. F. Song and K. M. Merz Jr, *J. Chem. Theory Comput.*, 2015, **11**, 1645–1657.
- 68 A. Jakalian, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2002, **23**, 1623–1641.
- 69 J. Wang, R. Wolf, J. Caldwell, P. Kollman and D. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1173.
- 70 T. Schlick, *Molecular modeling and simulation: an interdisciplinary guide*, Springer Science and Business Media, 2nd edn, 2010.
- 71 M. Ester, H. P. Kriegel, J. Sander and X. Xu, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 1996, pp. 226–231.
- 72 N. K. Banavali and A. D. MacKerell, Jr, *J. Mol. Biol.*, 2002, **319**, 141–160.
- 73 K. Hart, B. Nyström, N. Öhman and L. Nilsson, *RNA*, 2005, **11**, 609–618.
- 74 P. Schmidtke, A. Bidon-Chanal, J. Luque and X. Barril, *Bioinformatics*, 2011, **27**, 3276–3285.
- 75 M. Bardaro Jr, Z. Shajani, K. Patora-Komisarska, J. Robinson and G. Varani, *Nucleic Acids Res.*, 2009, **37**, 1529–1540.
- 76 L. Levintov and H. Vashisth, *J. Phys. Chem. Lett.*, 2020, **11**, 5426–5432.
- 77 J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott and N. Wilkins-Diehr, *Comput. Sci. Eng.*, 2014, **16**, 62–74.

