Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

ELECTRONIC SUPPLEMENTARY INFORMATION

Role of Conformational Heterogeneity in Ligand Recognition by Viral RNA Molecules

Lev Levintov[‡], and Harish Vashisth[‡*]

[‡]*Department of Chemical Engineering, University of New Hampshire, 33 Academic Way, Durham, NH 03824, United States*

[*] Corresponding author; Email: harish.vashisth@unh.edu

Table S1. Details on ligands studied in this work. For each ligand, shown is the PDB code, ligand name, chemical formula, number of atoms, molecular weight, and the dissociation constant (K_D). See also Fig. S1.

PDB ID	Name	Chemical Formula	Number of atoms	Molecular weight	K_D (μM)
1ARJ	Arginine amide	$C_6H_{15}N_4O_2$	27	175	1000
1LVJ	Acetylpromazine	$C_{19}H_{22}N_2OS$	45	326	0.1
1QD3	Neomycin B	$C_{23}H_{46}N_6O_{13}$	88	614	5.9
1UTS	RBT550	$C_{24}H_{33}N_5O$	63	407	0.039
1UUD	RBT203	$C_{16}H_{31}N_7O_2$	56	353	1.54
1UUI	RBT158	$C_{16}H_{29}N_5O_2$	52	323	>50
2L8H	MV2003	$C_{17}H_{25}N_5O_2$	49	331	NA
2KDQ 5J0M 5J1O 5J2W	L-22	$C_{76}H_{145}N_{33}O_{15}$	269	1759	0.03
2KX5	KP-Z-41	$C_{94}H_{179}N_{41}O_{19}S_2$	335	2249	0.001
6D2U	JB181	$C_{72}H_{140}N_{31}O_{15}$	258	1678	<0.00018

Table S2. Details on all simulation systems. For each TAR structure (see PDB codes), listed are RNA sequences in the original structure (column labeled *RNA Before*) and in the modeled structure (column labeled *RNA After*). The last column lists the total number of atoms in solvated and ionized unliganded as well as liganded (marked in bold) systems except for the PDB code 1ANR (first row) which is the experimental *apo* structure. See also materials and methods.

	PDB ID	RNA Before	RNA After	System Size
apo	1ANR	930 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	22959
liganded/unliganded	5J2W	929 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	20586 **20542**
	1UUI	930 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	23076 **23078**
	5J1O	929 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	21159 **21127**
	6D2U	930 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	21742 **21761**
	1UUD	929 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	24633 **24645**
	1UTS	930 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	24165 **24159**
	2KX5	930 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	23475 **25647**
	1QD3	927 G**C**CAGAU**U**UGAGCCUGGGAGCUCUCUG**G**C	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	19443 **19483**
	5J0M	929 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	22446 **22444**
	2KDQ	930 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	24471 **24529**
	1ARJ	930 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	22845 **22865**
	2L8H	930 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	20691 **20714**
	1LVJ	999 GGC**C**AGAUCUGAGCCUGGGAGCUCUCU**G**GCC	931 GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	22023 **22053**

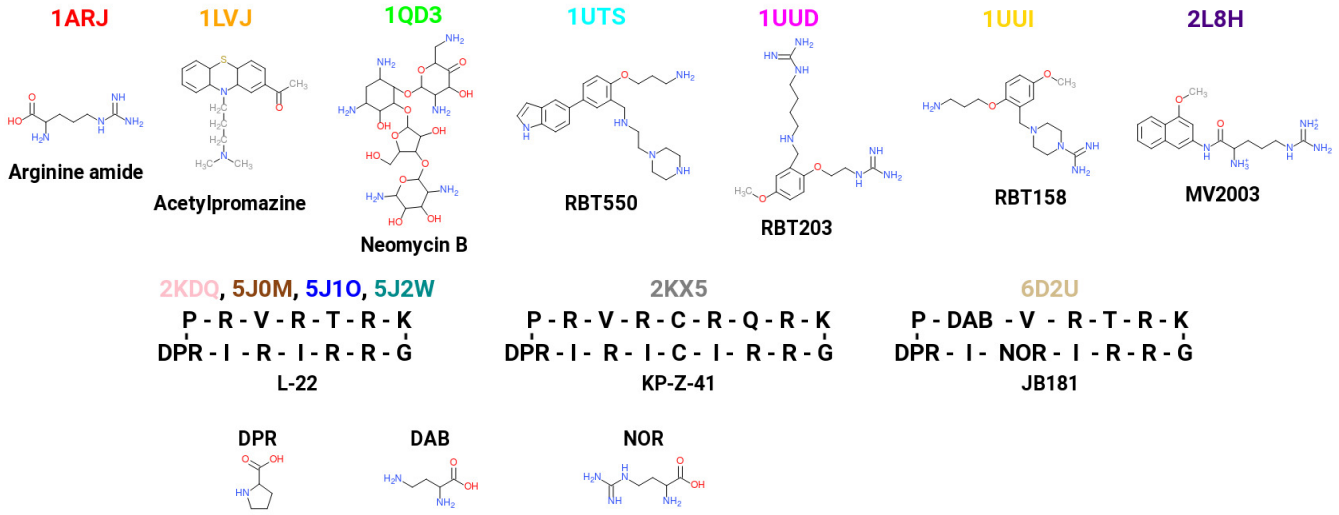

Figure S1. Chemical structures of ligands studied in this work: (*top row*) small molecules; (*bottom row*) peptides.

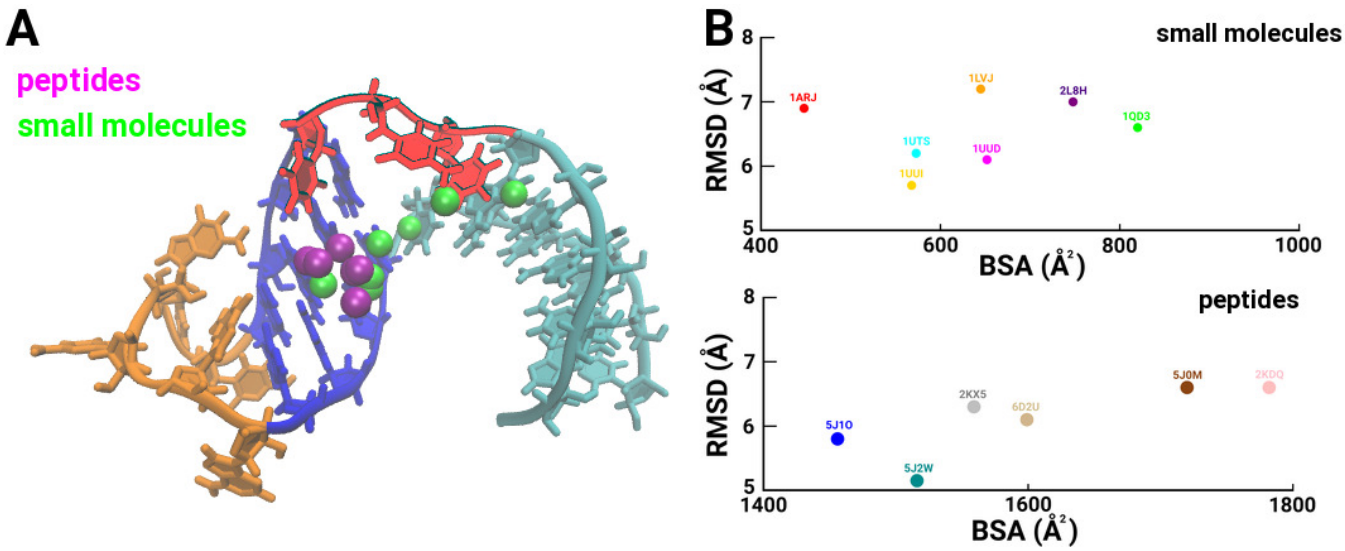

Figure S2. Conformational metrics of ligands. (A) The centers of mass of ligands are represented by spheres (colored and labeled) and overlaid on the *apo* structure of TAR. (B) The all-atom root-mean-squared-deviation (RMSD) of each liganded TAR structure relative to the apo structure (PDB code 1ANR) *vs.* buried surface area (BSA) of each ligand are shown for small-molecules (*top panel*) and peptides (*bottom panel*).

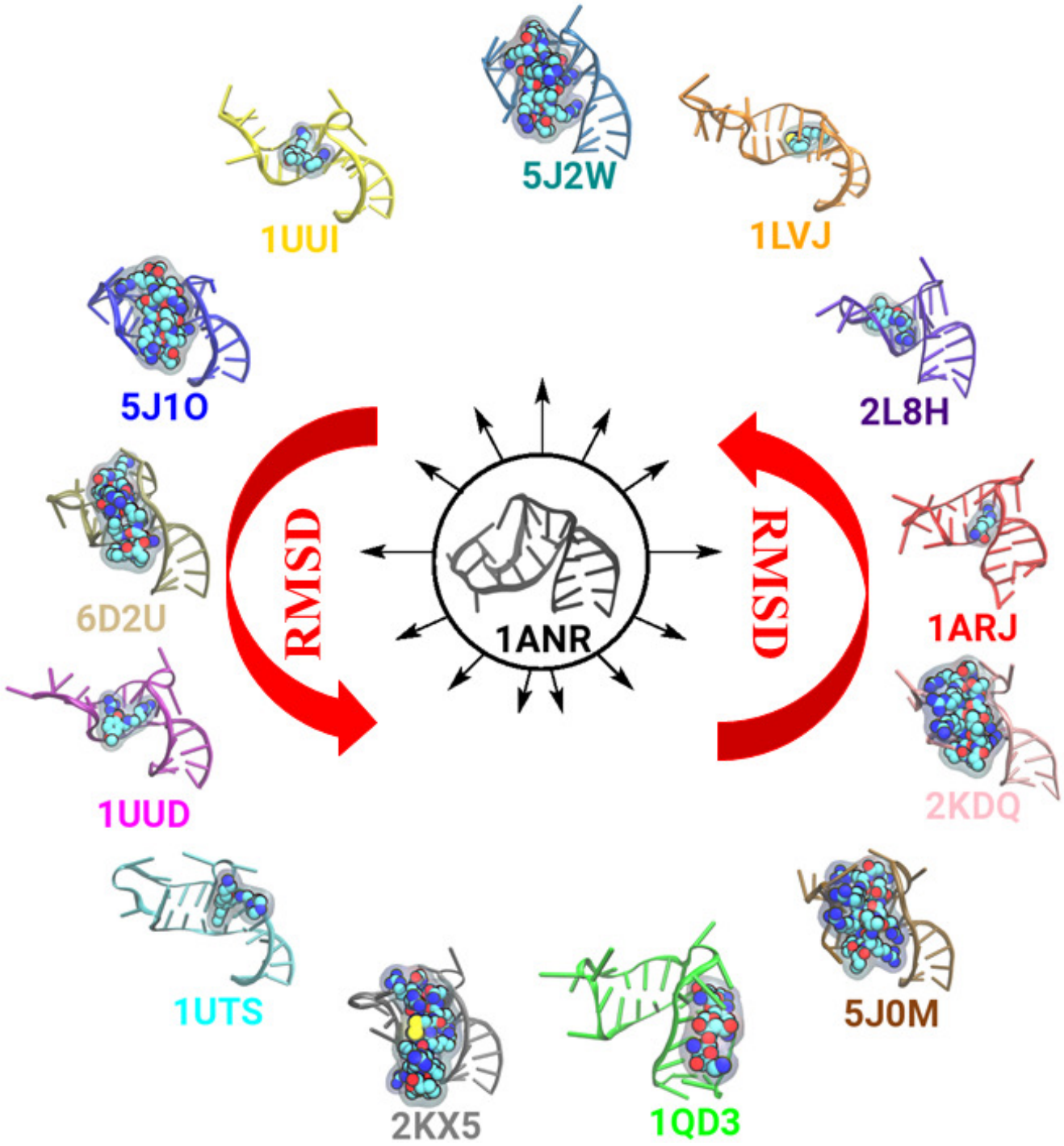

Figure S3. Snapshots of the initial systems in liganded simulations: RNA, cartoon representation; ligands, space-filling. Snapshot of the *apo* TAR structure (PDB code 1ANR) is located at the center (black cartoon). See also Fig. 1.

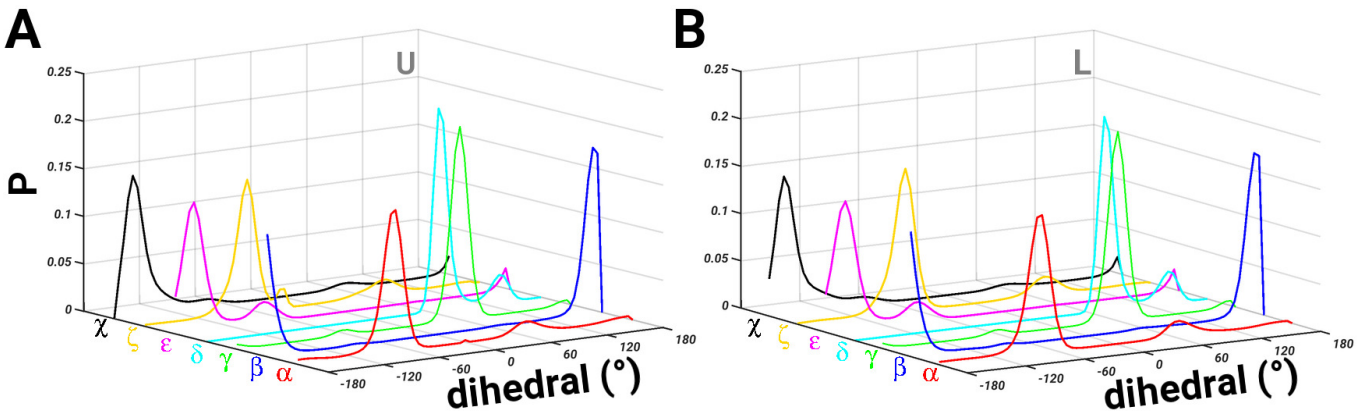

Figure S4. Torsional flexibility. Th normalized distributions of each RNA backbone dihedral angle (α, β, γ, δ, ϵ, ζ) and the glycosidic dihedral angle (χ) for unliganded (labeled U in panel A) and liganded (labeled L in panel B) simulations.

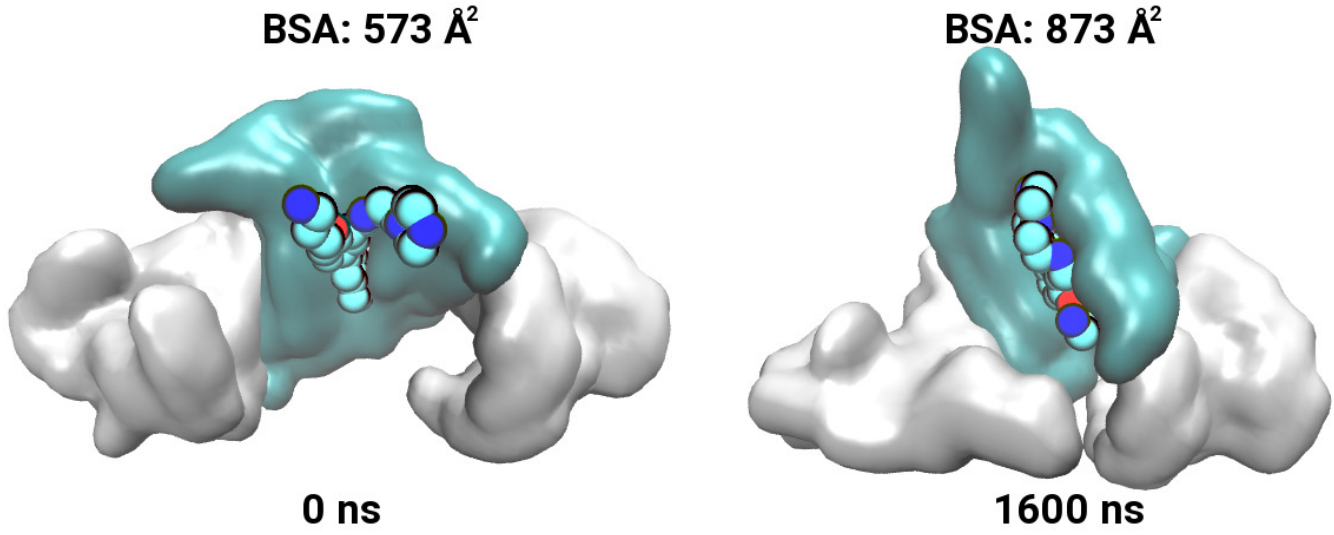

Figure S5. Snapshots highlighting an increase in BSA. Shown are the snapshots of the TAR RNA conformations (surface map) and the ligand (RBT550; space-filling) from a liganded simulation (PDB code 1UTS) highlighting an increase in BSA in comparison to the initial BSA due to conformational rearrangements of the ligand in the binding pocket. A cyan surface indicates the nucleotides of the binding pocket in close contact with the ligand and a white surface represents the rest of the RNA structure.

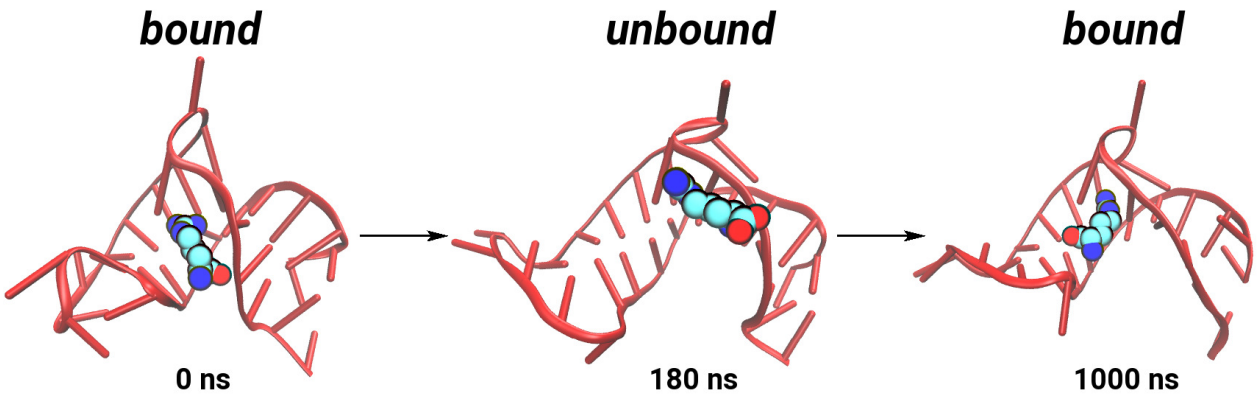

Figure S6. Snapshots highlighting partial ligand dissociation. Shown are the snapshots of the TAR RNA conformations (red cartoon) and the ligand (arginine amide; space-filling) from a liganded simulation (PDB code 1ARJ) highlighting the partial dissociation of the ligand at t = 180 ns and then rebinding again in the original binding pocket (t = 1000 ns).

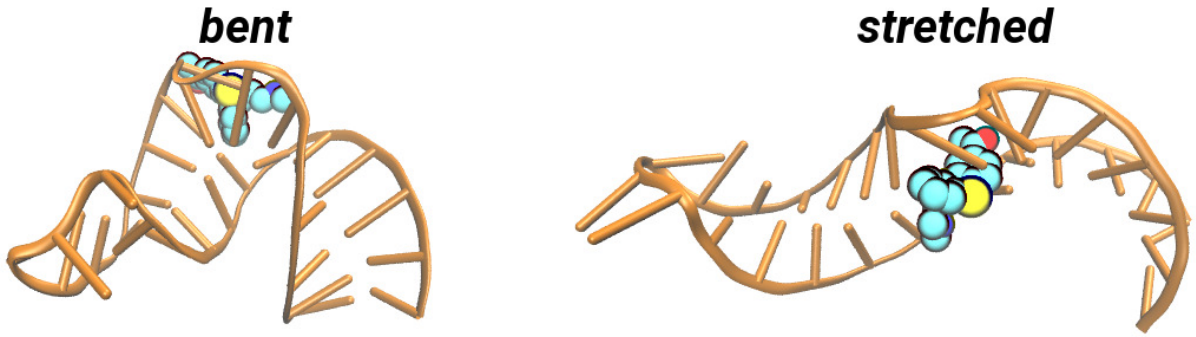

Figure S7. Conformational change in TAR RNA in a liganded simulation: Shown are the *bent* and *stretched* conformations of TAR (orange cartoon) with the ligand (space-filling) from a liganded simulation (PDB code 1LVJ).

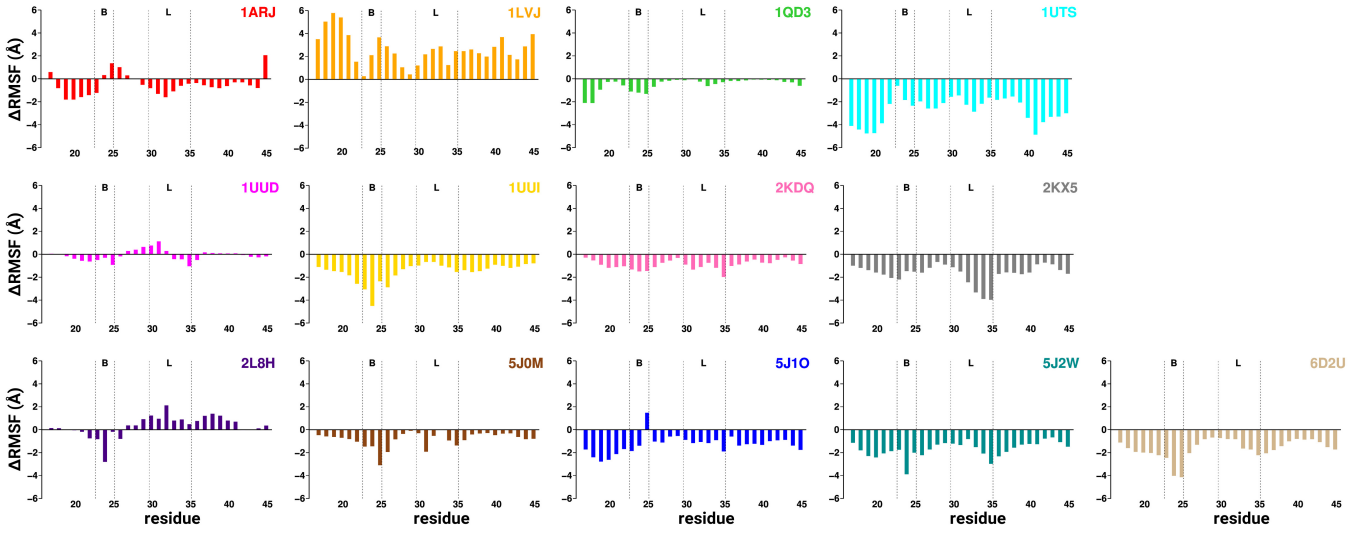

Figure S8. The ΔRMSF per residue data are presented highlighting the differences between the unliganded and liganded simulations. Each system is uniquely colored. The bulge (B) and the loop (L) motifs are marked with the dashed lines.

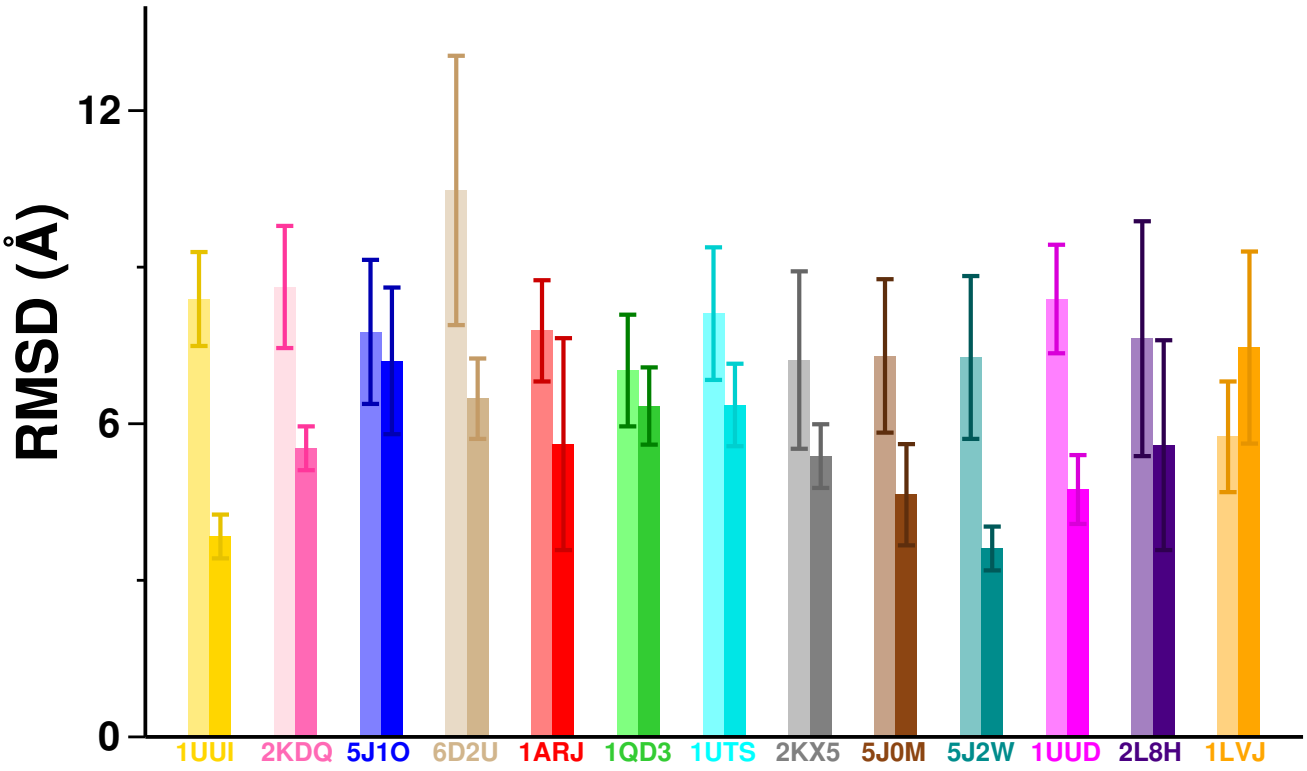

Figure S9. Flexibility of the bulge motif in unliganded and liganded states. RMSD data with error bars, similar to Figure 3A, are shown for the bulge motif nucleotides (*lighter shades*, unliganded; *darker shades*, liganded).

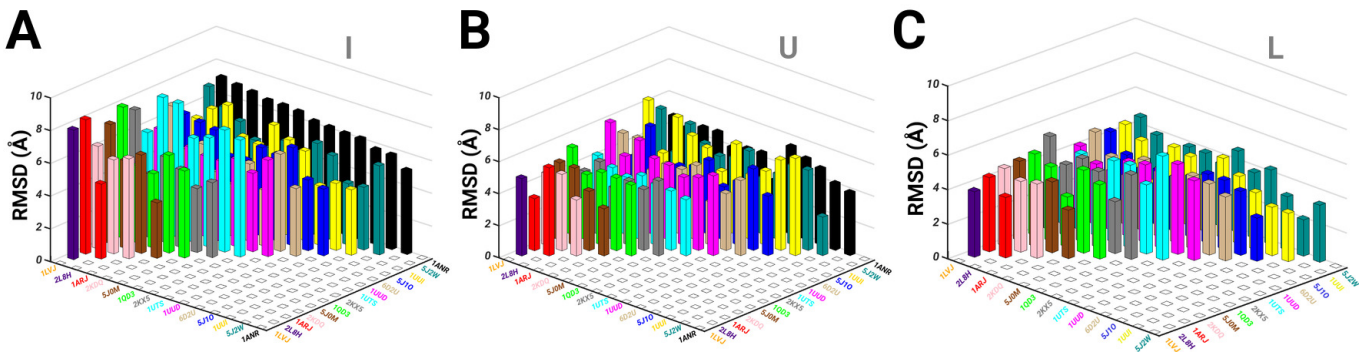

Figure S10. Cross-comparison of initial, unliganded, and liganded TAR structures. A cross-comparison of TAR RNA conformations via RMSD is highlighted for all structures in the initial states (panel A; labeled I) and based upon average structures derived from unliganded (panel B; labeled U) and liganded (panel C; labeled L) MD simulations.

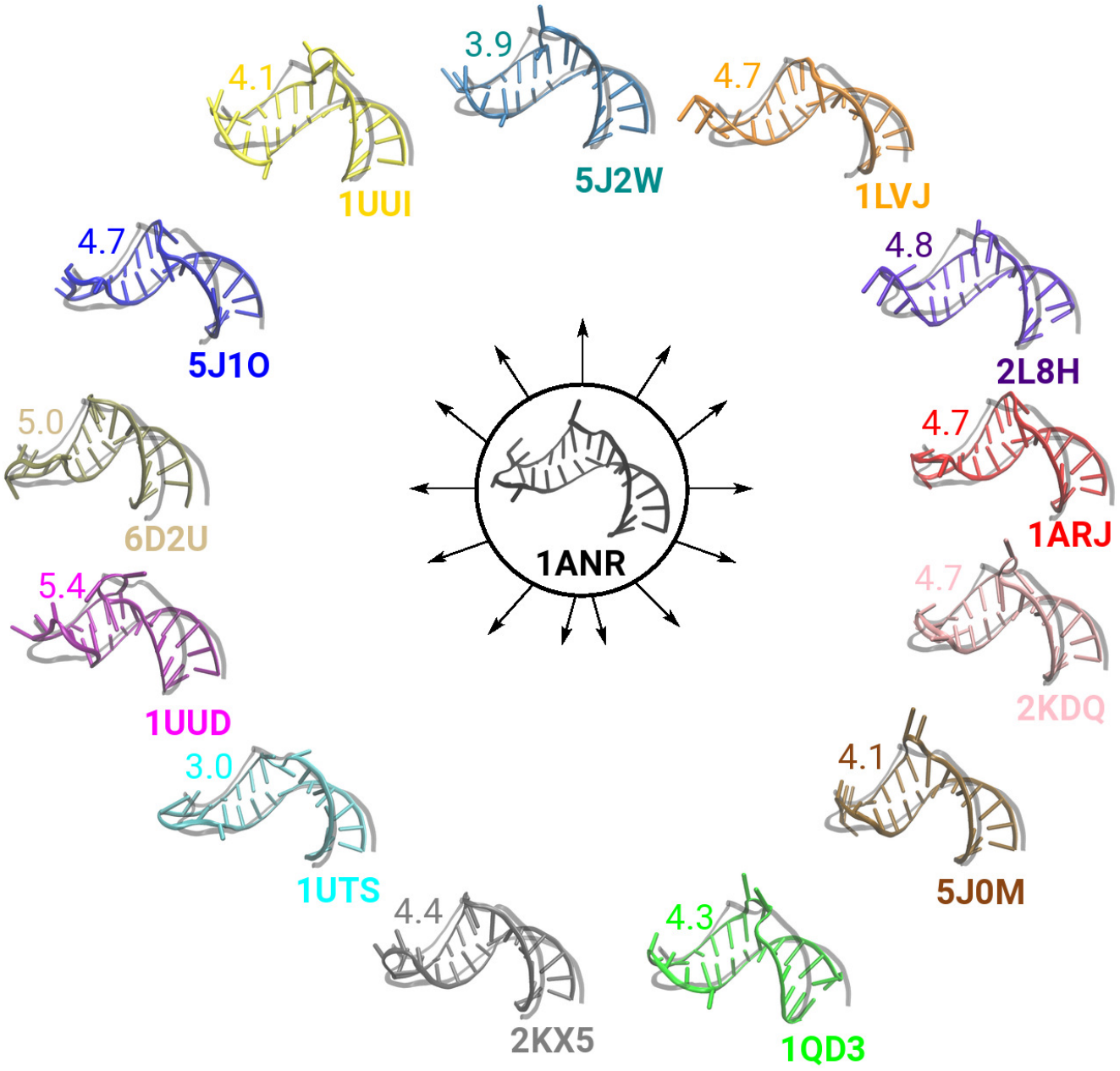

Figure S11. Comparison of average structures of TAR RNA. The average structures of TAR RNA from unliganded simulations are overlaid on the average structures from the simulation of the *apo* TAR RNA structure (PDB code 1ANR). The RMSD values between the average structures are also labeled in color along with the PDB codes.

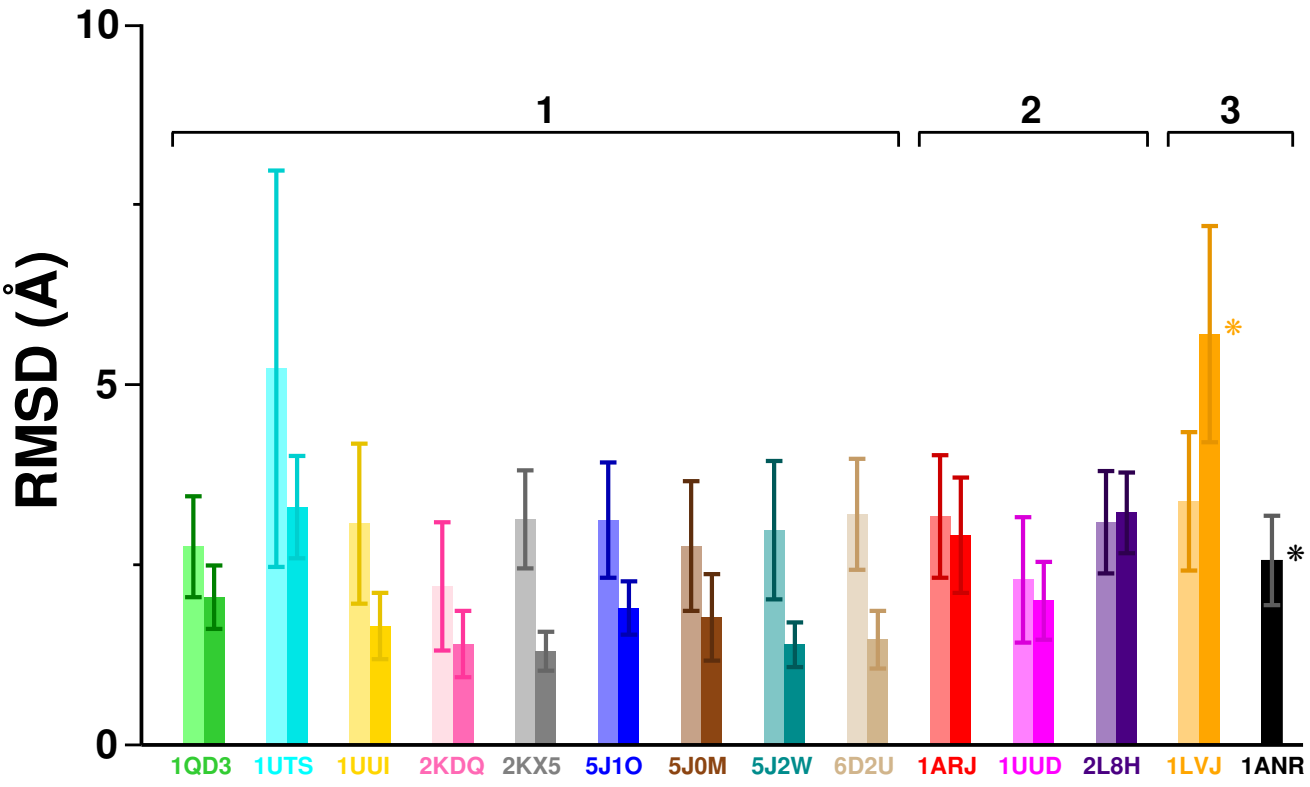

Figure S12. Shown are the histograms (with error bars) of mean values of RMSD computed with respect to the average structure for all unliganded (*lighter shades*) and liganded (*darker shades*) simulations. The RMSD histograms are organized into three groups (labeled 1, 2, and 3; marked by overbars). An orange asterisk marks a system (PDB code 1LVJ) which showed a different behavior in comparison to other systems. A black asterisk marks the experimental *apo* TAR structure (PDB code 1ANR).

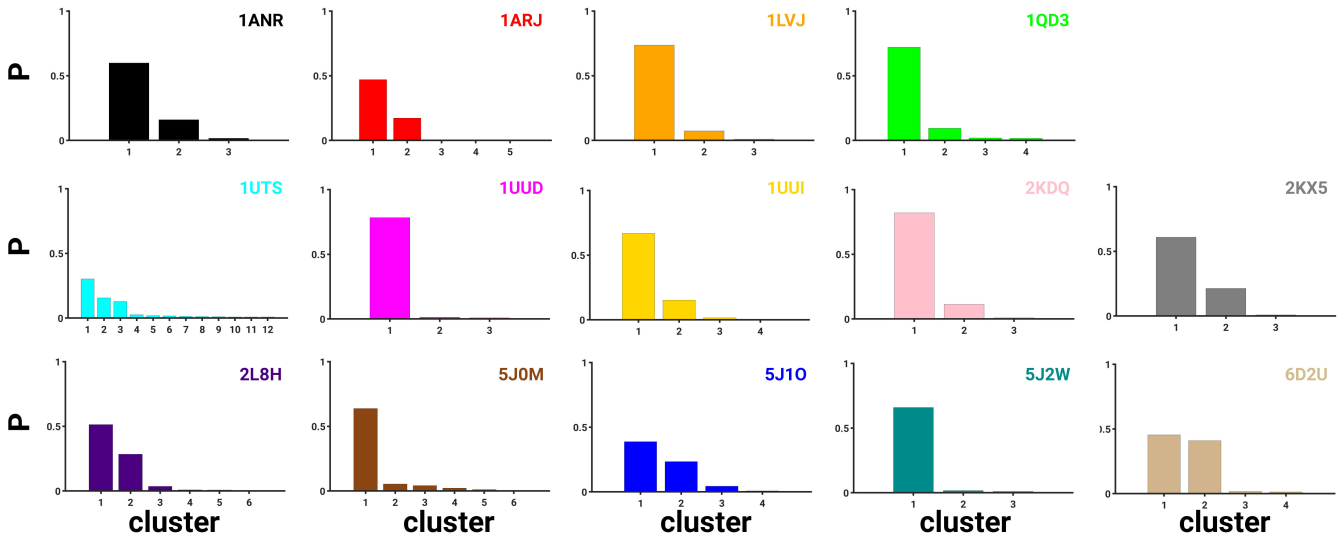

Figure S13. Cluster analysis of unliganded simulations. The distributions of all clusters computed from conformations sampled via MD simulations are shown for the unliganded state of each system. Histograms are shown in the same color as the labeled PDB code. See also Fig. 3B.

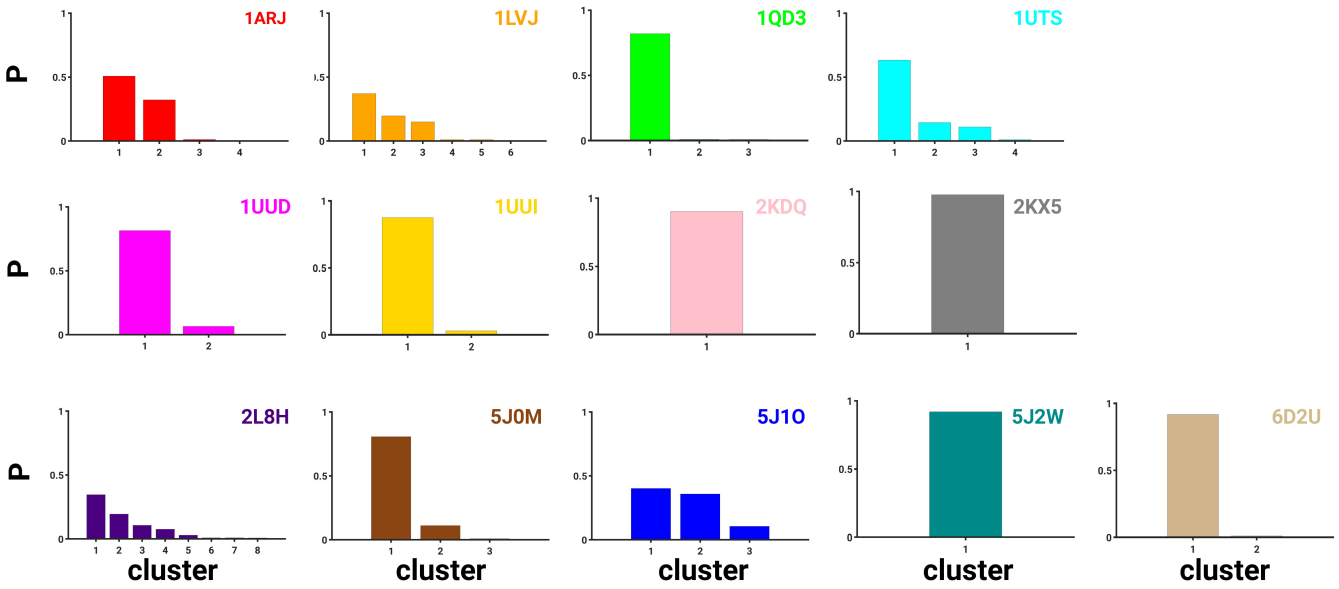

Figure S14. Cluster analysis of liganded simulations. Data similar to Fig. S13 are shown for the liganded state of each system. See also Fig. 3B.

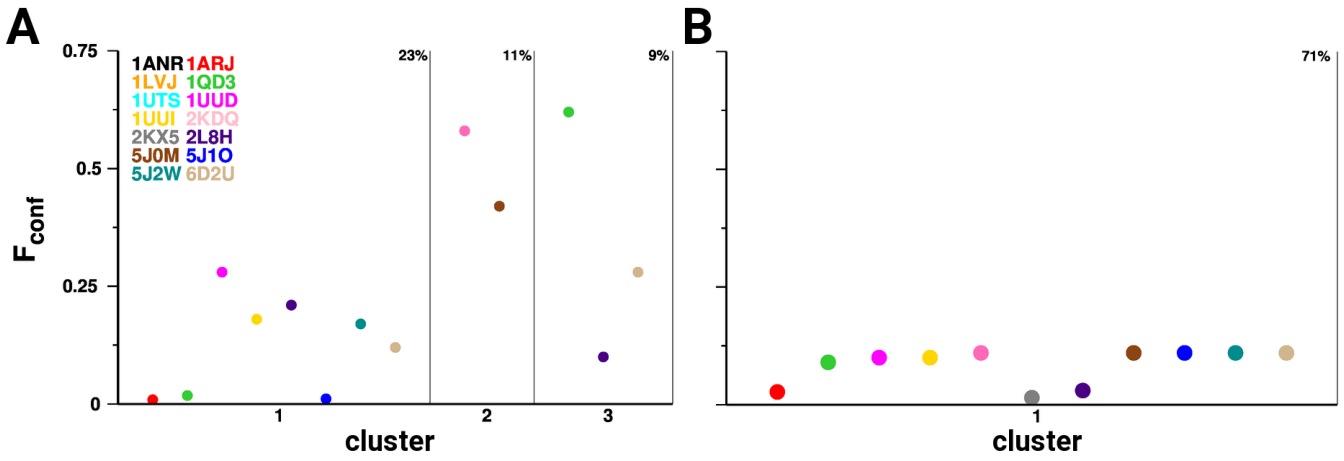

Figure S15. Combined cluster analysis. The fraction of conformations (F_{conf}) from each system that populate each cluster from a set of (A) unliganded and (B) liganded simulations. Each system is uniquely colored. The numbers at the top of each cluster signify the percentage of the total number of frames that constitute that specific cluster.

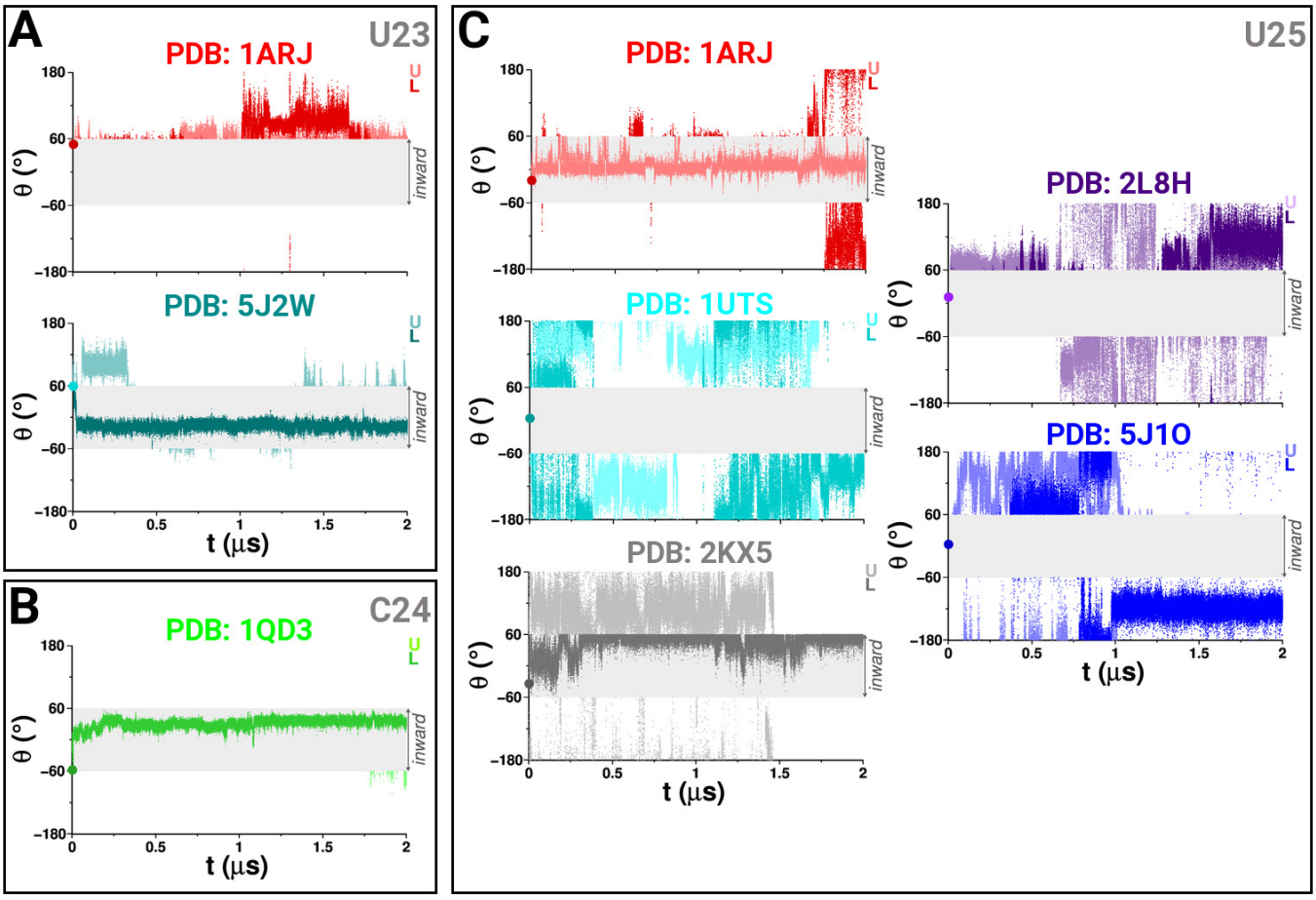

Figure S16. Conformational transitions in bulge nucleotides (U23, C24, and U25) in unliganded (U) and liganded (L) simulations. Data similar to Fig. 5 are shown for additional systems.

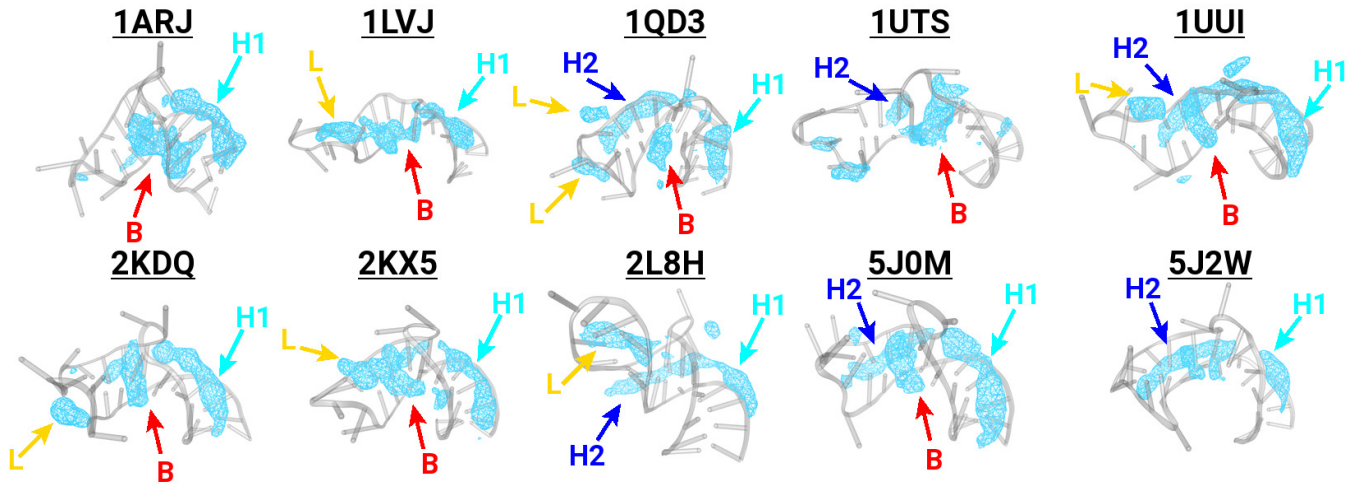

Figure S17. Predicted binding pockets in unliganded TAR structures. Data similar to Fig. 6A are shown for additional systems.

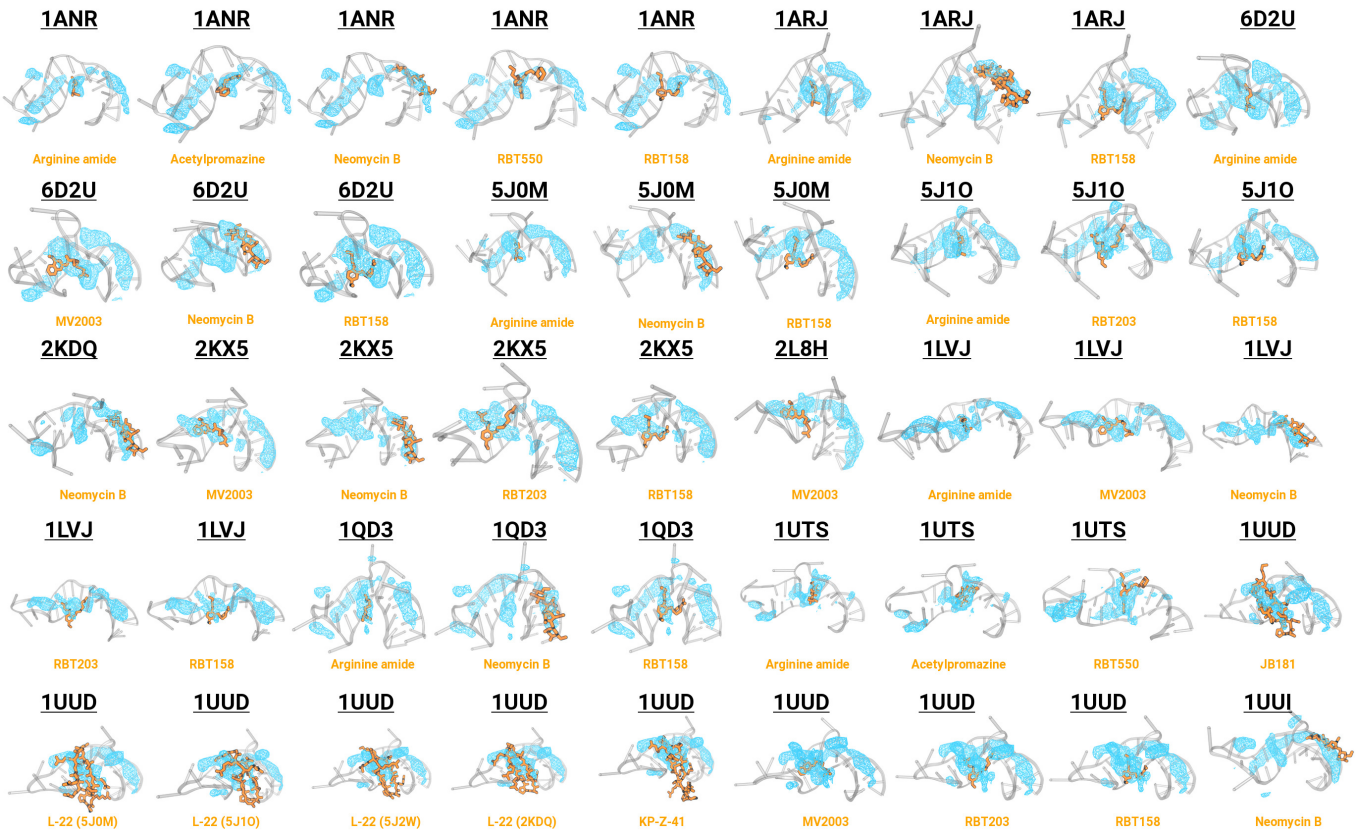

Figure S18. Overlays of ligands in predicted binding pockets. Shown are the snapshots of predicted binding pockets (cyan surfaces) with an overlay of each ligand (orange sticks) on various TAR structures (transparent gray cartoons). See also Fig. 6.