Supporting Information for

Understanding the binding specificities of mRNA targets by the mammalian Quaking protein

Monika Sharma[1,*], Shakshi[1] and Apoorv Alawada[1]

[1] Department of Chemical Sciences, Indian Institute of Science Education and Research (IISER), Sector 81, Knowledge City, SAS Nagar, Punjab, India.

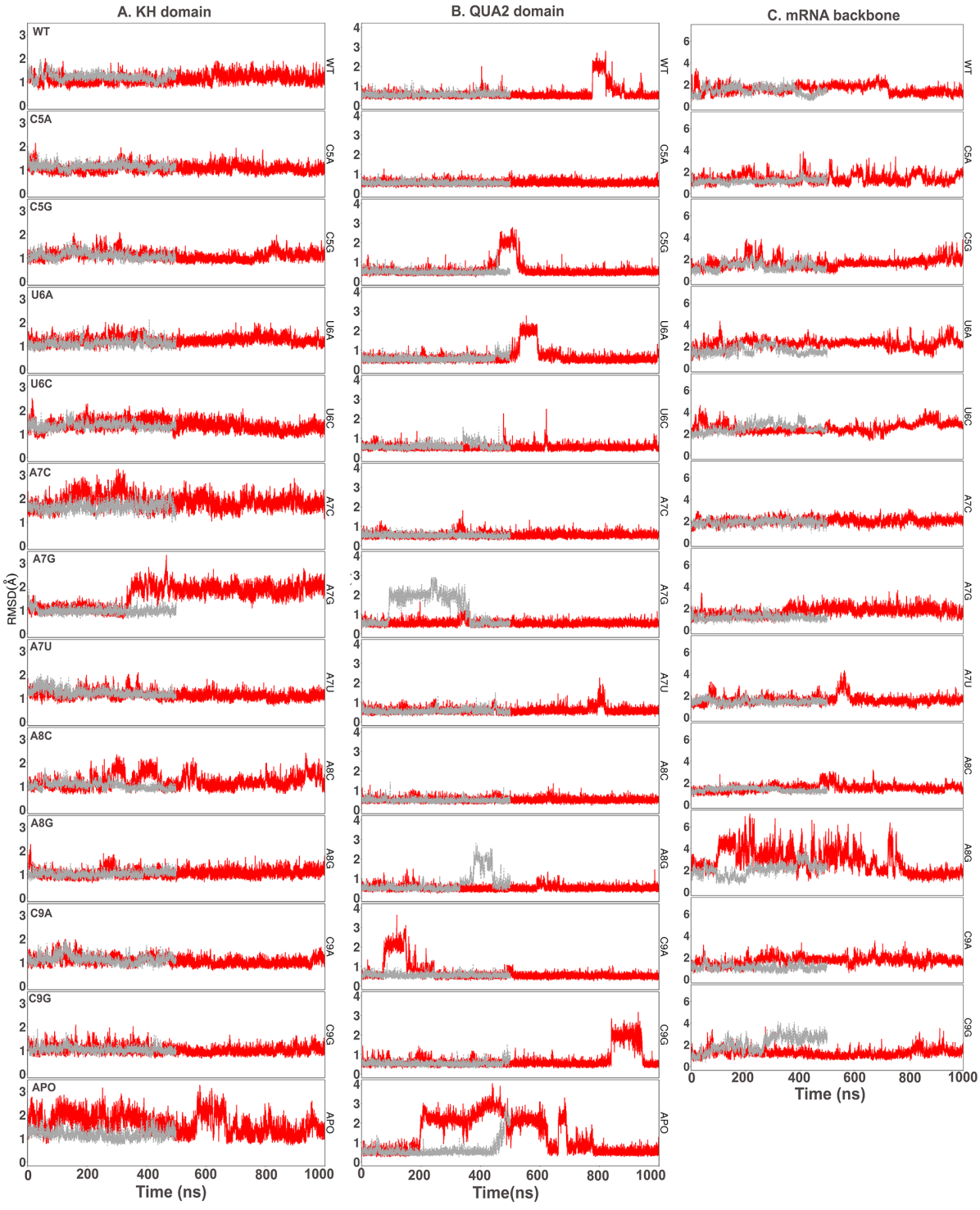


Figure SI1. Time evolution of backbone rmsd values (with respect to the crystal structure) of A. KH domain, B. C-terminal QUA2 domain and C. bound mRNA for conformations generated using AMBER forcefields. RMSD values during two independent runs with different timescales are shown in red and grey.

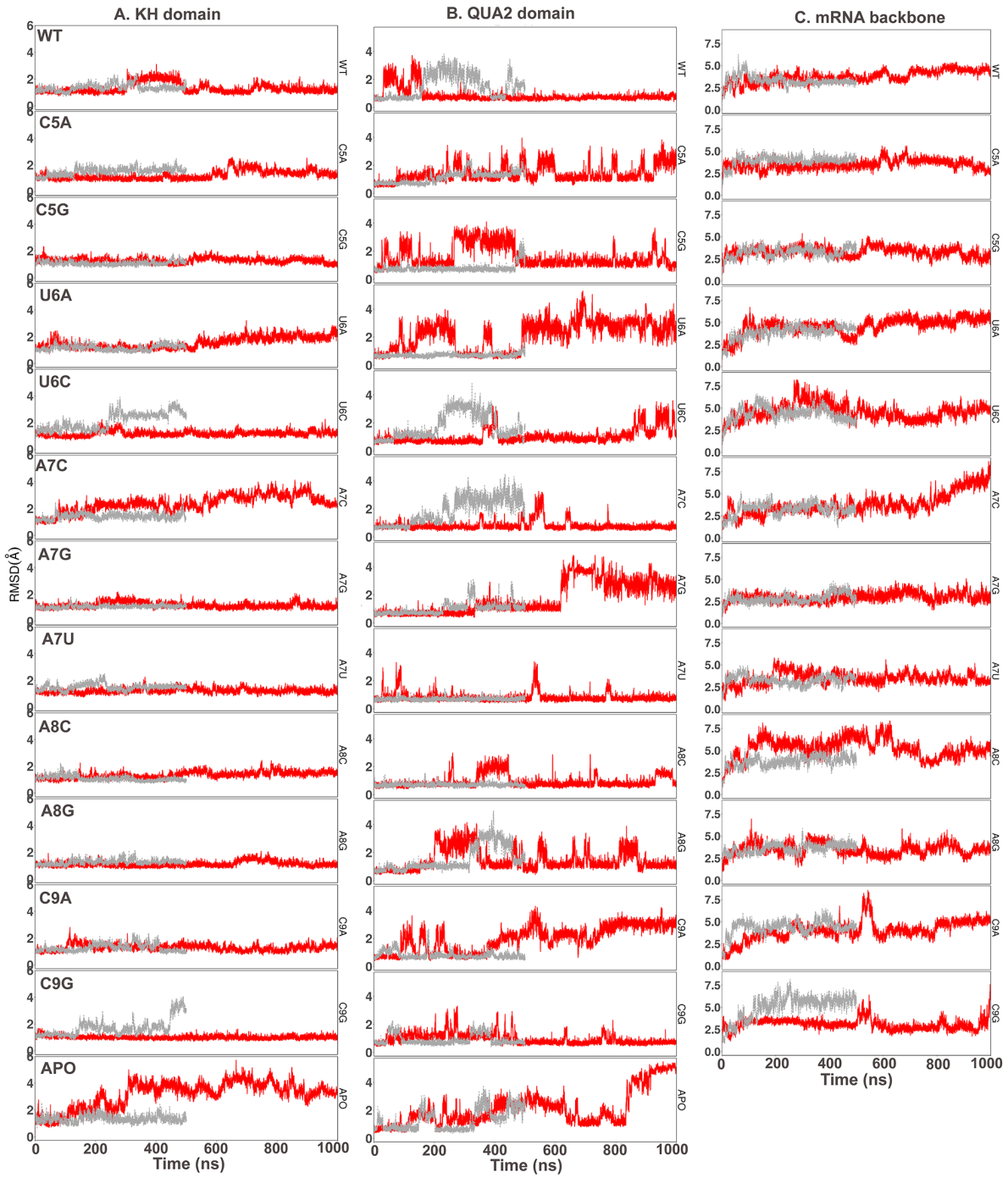


Figure SI2. Time evolution of backbone rmsd values (with respect to the crystal structure) of A. KH domain, B. C-terminal QUA2 domain and C. bound mRNA for conformations generated using CHARMM forcefields. RMSD values during two independent runs with different timescales are shown in red and grey.

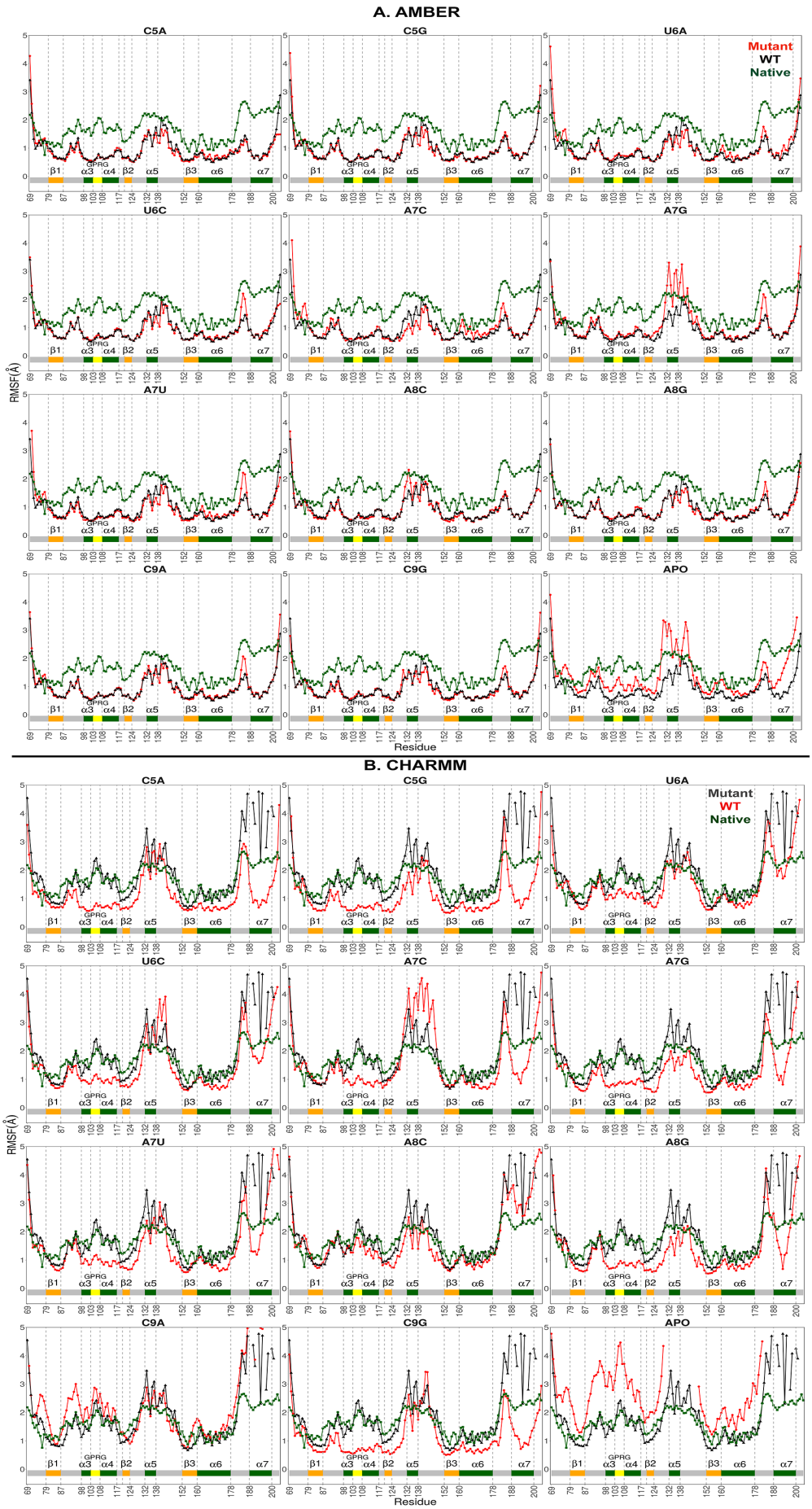

A. AMBER
C5A
C5G
U6A
U6C
A7C
A7G
A7U
A8C
A8G
C9A
C9G
APO
Mutant
WT
Native
RMSF(Å)
Residue
B. CHARMM
C5A
C5G
U6A
U6C
A7C
A7G
A7U
A8C
A8G
C9A
C9G
APO
Mutant
WT
Native
RMSF(Å)
Residue

Figure SI3. RMSF for Cα atoms of KH-QUA2 domain of QKI proteins averaged over all the simulation runs using (A) AMBER forcefields and (B) CHARMM forcefields. Data with black circles and black lines is for STAR domain bound with cognate mRNA, red colored triangles with red lines is for STAR domain bound with non-cognate mRNA, and green colored squares with green lines correspond to converted B-factors for crystal structure (4JVH). Secondary structure elements: alpha helices, beta sheets and loops are annotated with boxes of green color, orange color and grey color. 'GPRG' motif is annotated with yellow colored box.

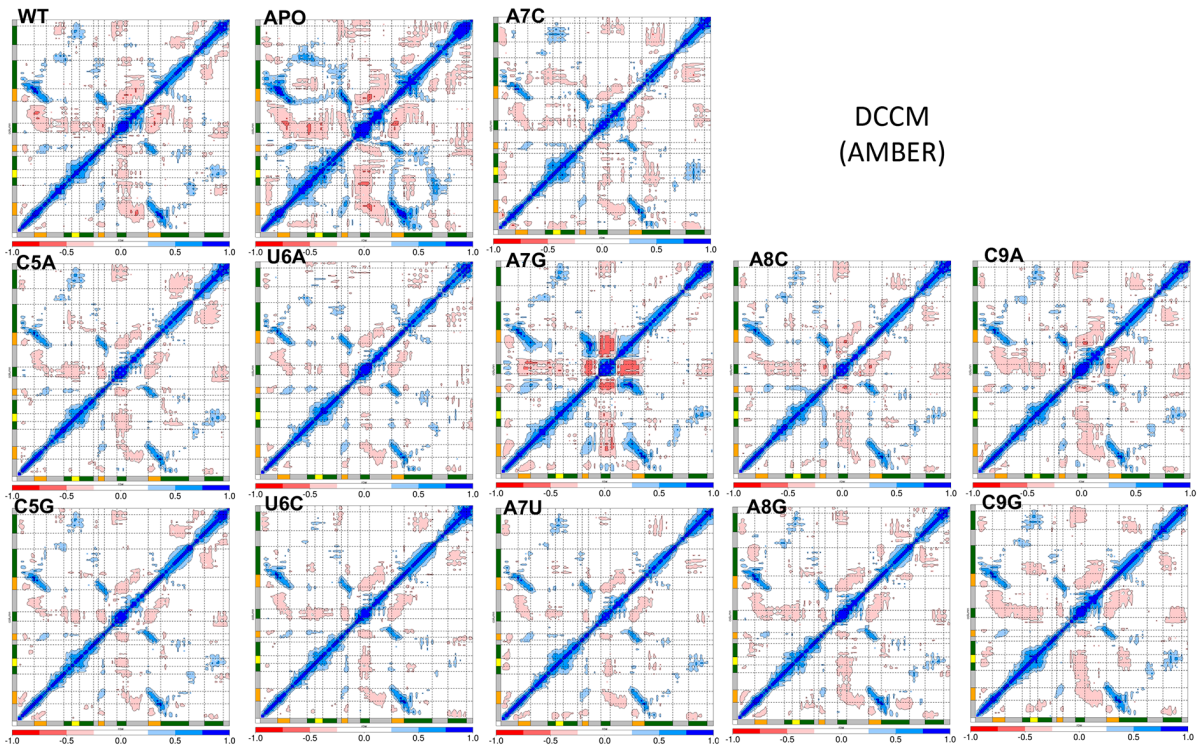


Figure SI4. Dynamic cross correlation maps calculated as time-averaged for Cα atoms of KH-QUA2 domain of QKI protein in presence of cognate (WT) and non-cognate mRNA sequences using AMBER forcefields. DCC map for domain in absence of mRNA sequence is also shown as APO. Secondary structure elements are shown as in Fig SI3. The whole range of correlation from −1 to + 1 is represented in three ranges: blue color corresponding to positive correlation values ranging from 0.25 to 1; red color corresponding to negative correlation values ranging from −0.25 to −1; and white color corresponding to weak or no-correlation values ranging from −0.25 to + 0.25. The extent of correlation or anti-correlation is indicated by variation in the intensity of respective blue or red color.

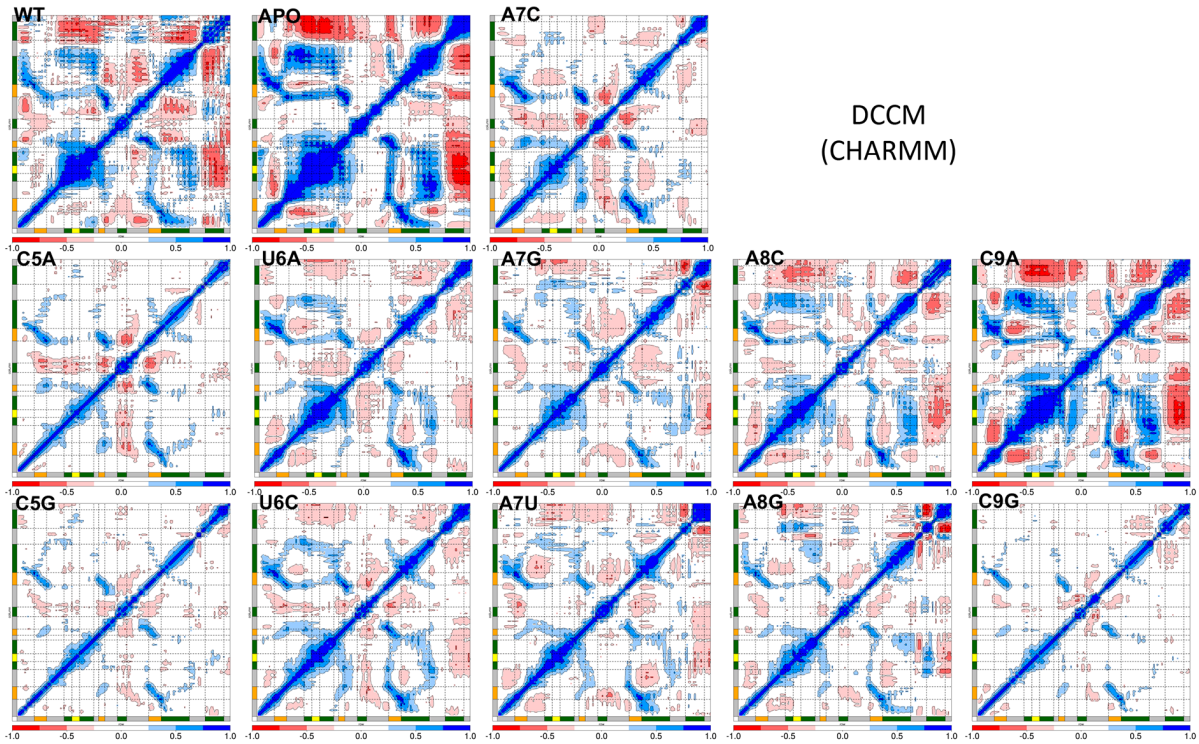


Figure SI5. Dynamic cross correlation maps calculated as time-averaged for Cα atoms of KH-QUA2 domain of QKI protein in presence of cognate (WT) and non-cognate mRNA sequences using CHARMM forcefields. DCC map for domain in absence of mRNA sequence is also shown as APO. Secondary structure elements are shown as in Fig SI3. The whole range of correlation from −1 to + 1 is represented in three ranges: blue color corresponding to positive correlation values ranging from 0.25 to 1; red color corresponding to negative correlation values ranging from −0.25 to −1; and white color corresponding to weak or no-correlation values ranging from −0.25 to + 0.25. The extent of correlation or anti-correlation is indicated by variation in the intensity of respective blue or red color.

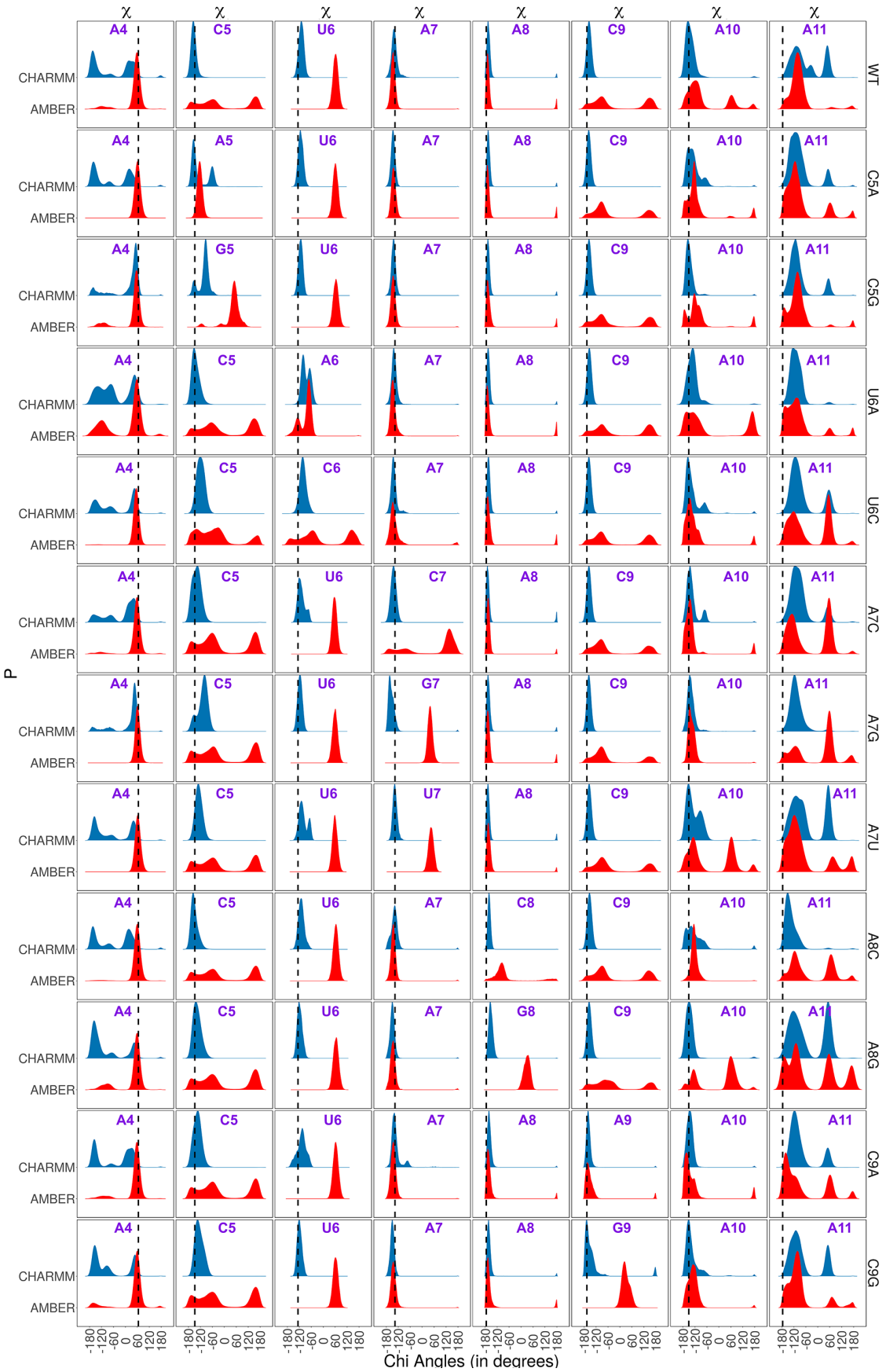


Figure SI6. Ridge plots showing probability distributions of glycosidic chi angles for mRNA sampled during the simulations. The vertical dashed lines within the faceted panels (CS) indicate χ values for mRNA bound to the crystal structure. (pdbid: 4JVH)

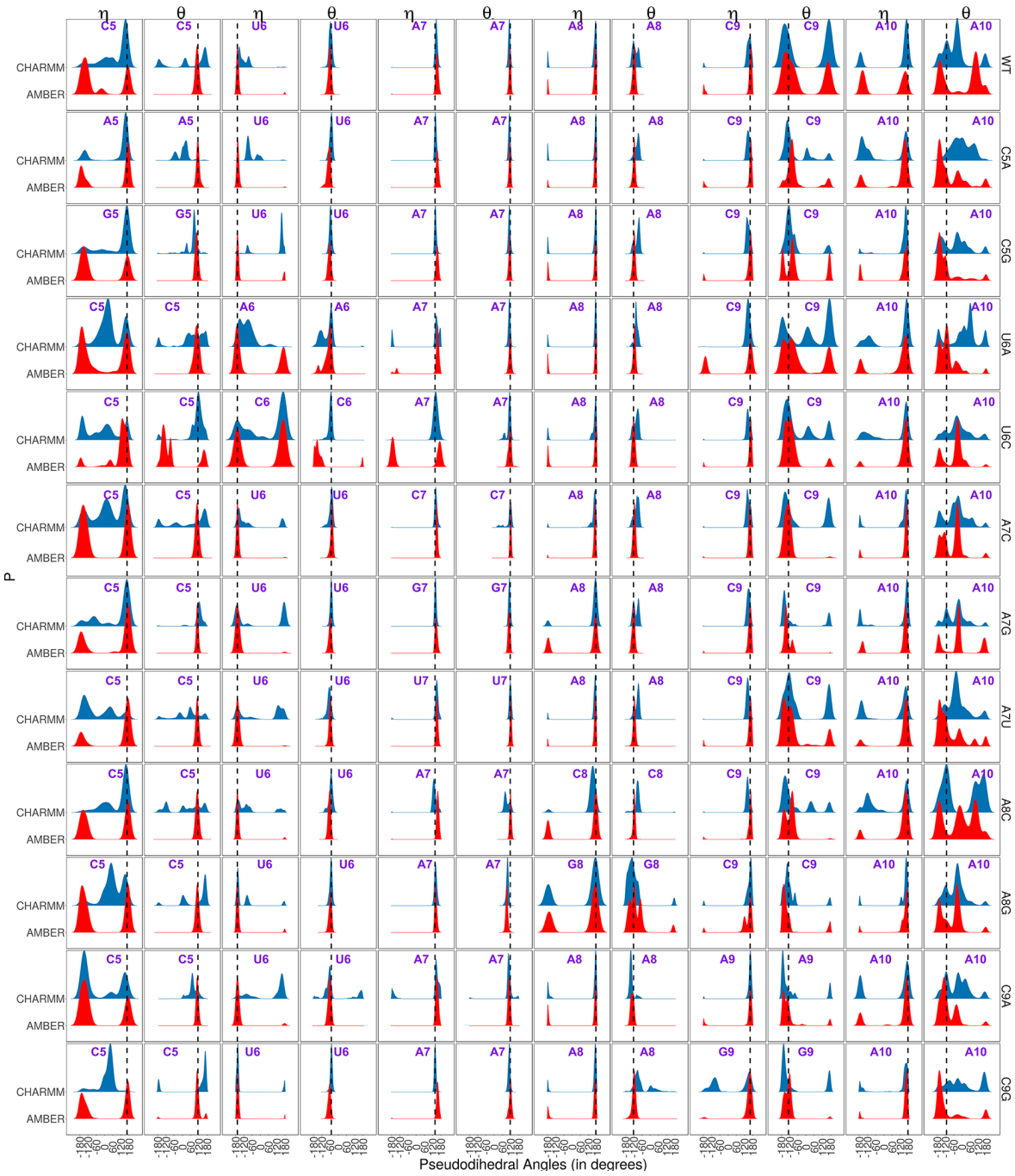


Figure SI7. Ridge plots showing probability distributions of pseudotorsion angles, η and θ for mRNA sampled during the simulations. The vertical dotted lines within the faceted panels (CS) indicates pseudo-torsion values for mRNA bound to the crystal structure. (pdbid: 4JVH)

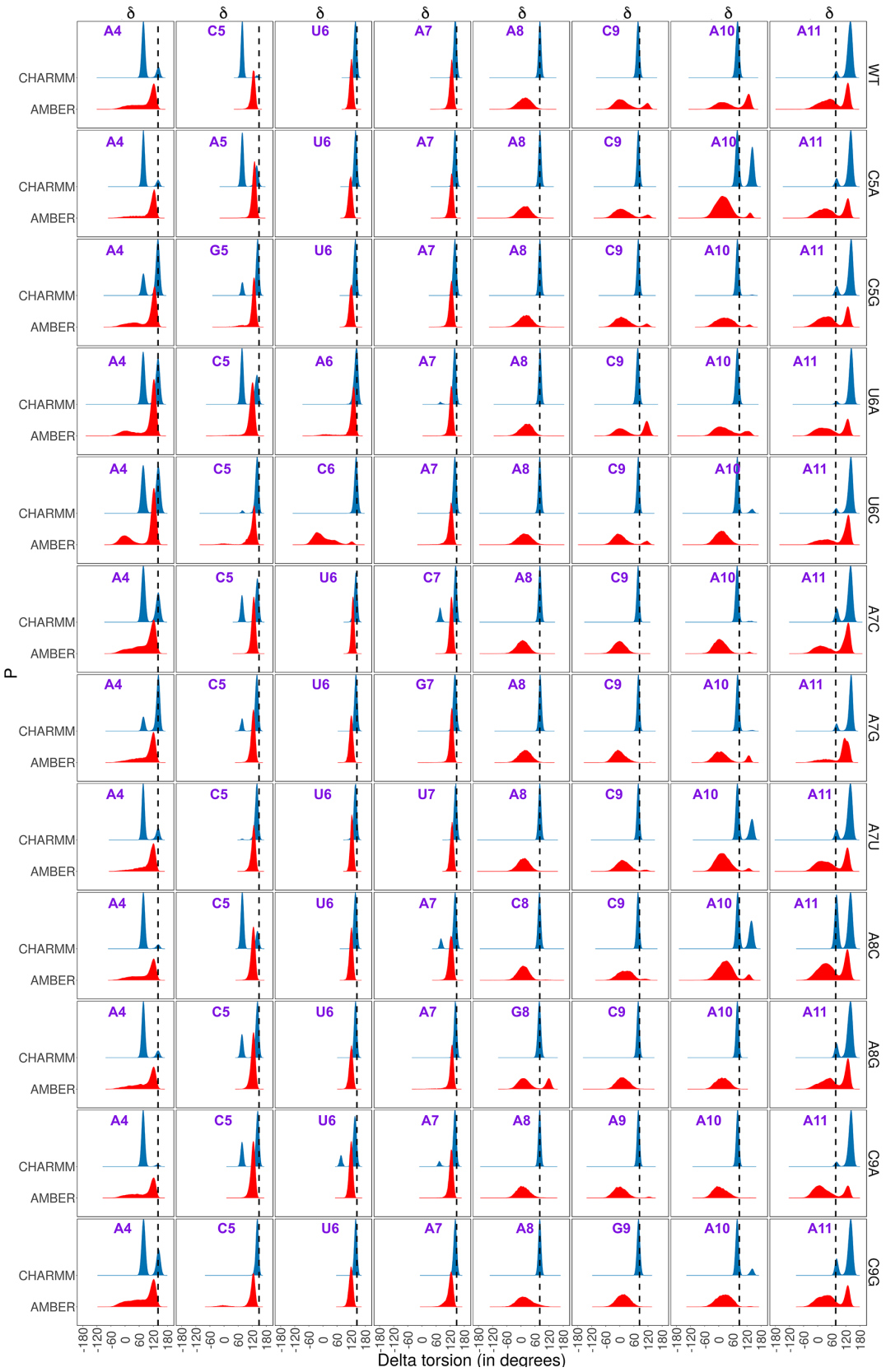


Figure SI8. Ridge plots showing probability distributions of delta (δ) torsions around C4′-C3′ bond for mRNA sampled during the simulations. The vertical dashed lines within the faceted panels (CS) indicate δ values for mRNA bound to the crystal structure. (pdbid: 4JVH)

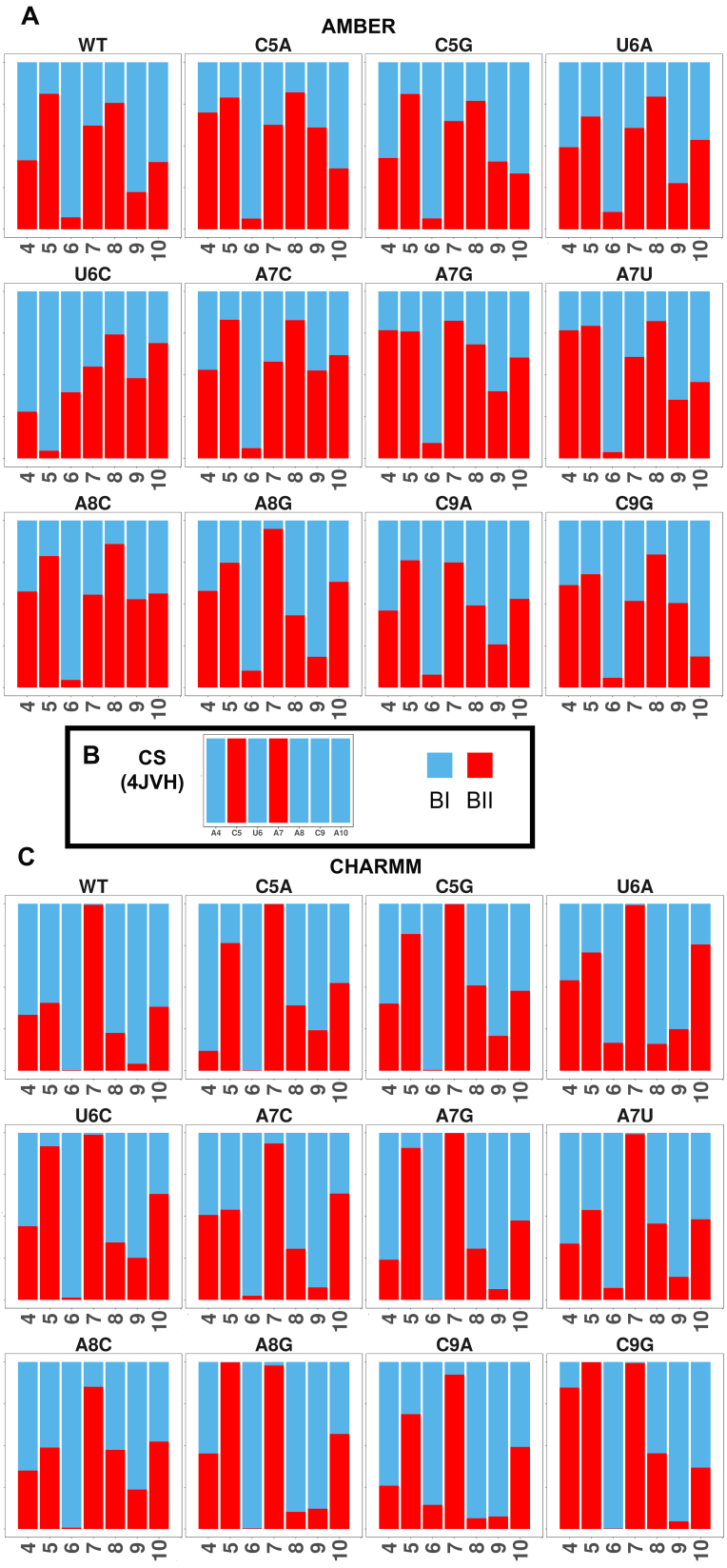


Figure SI9. Stacked histogram plots showing BI/BII transitions, as observed by variations in ε-ζ values for mRNA conformations sampled during simulations using A. AMBER and C.

CHARMM forcefields. B. The BI/BII transition values are shown for mRNA bound to crystal structure (pdbid: 4JVH).

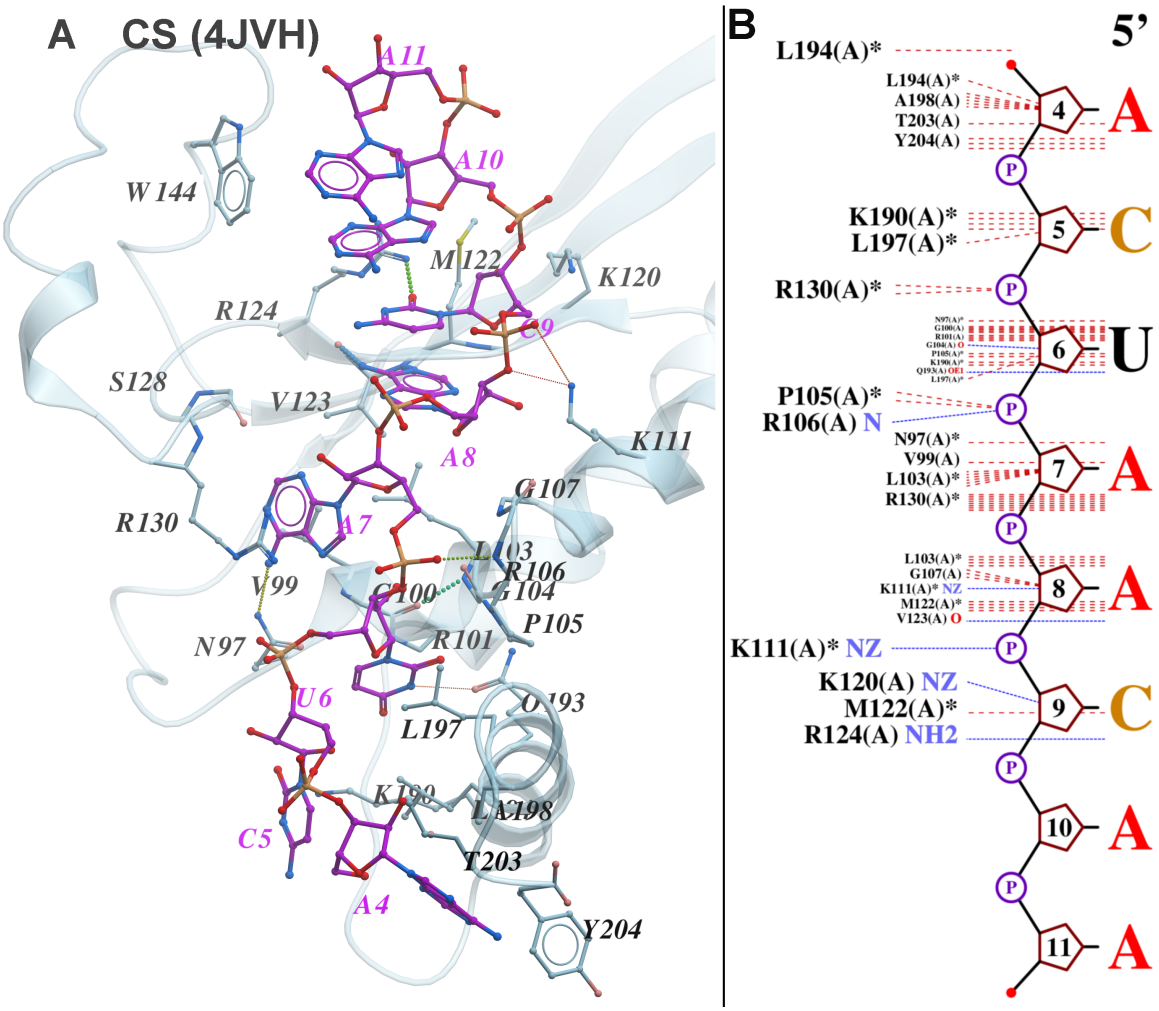


Figure SI10. A. Detailed interactions between mRNA and KH-QUA2 domain observed in experimental crystal structure (pdbid: 4JVH). B. Schematic representation as shown by NUCPLOT software (Luscombe NM, Laskowski RA, Thornton JM. NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. Nucleic Acids Res. 1997;25(24):4940–4945). Red lines correspond to nonbonded contacts and blue lines correspond to hydrogen bond interactions. A4 interacts with amino acid residues L194, A198, T203, and Y204. C5 interacts with amino acids R130, K190, and L197. U6 interacts with amino acids K190, Q193, and L197. Besides these, U6 also interacts with N97, G100, R101, and GPRG motif of KH domain. The interaction of nucleotide U6 with 'G^{104}PRG^{107}' motif is via hydrogen-bonding interactions with G104 and R106, and nonbonded contacts with P105. The nucleobase of A7 is observed sandwiched between the side chains of residues V99 and R130. A7 also interacts with N97 and L103. A8 forms nonbonded contacts with L103, M122, and V123, and backbone sugar-phosphate interactions with G107 and K111. C9 forms hydrogen bonding interactions with side chains of K120 and R124, and nonbonded interactions with M122. Nucleotides A10 and A11 do not form substantial interactions, except weak nonbonded interactions with N143 and W144.

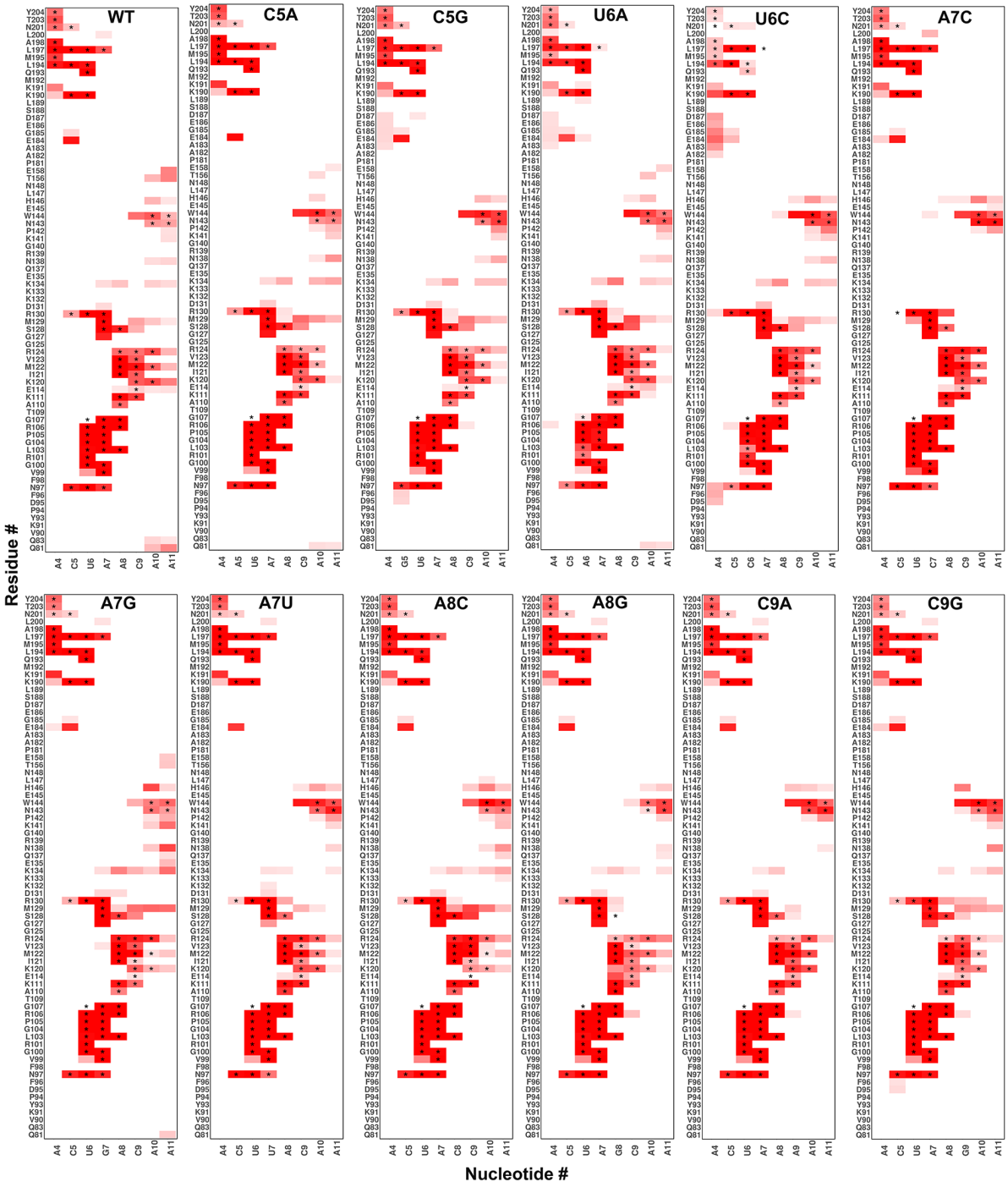


Figure SI11. Intermolecular contacts between mRNA and STAR domain of QKI protein computed for simulations using AMBER forcefields. The values are computed excluding the first 100 ns of simulation trajectories. The intensity of the color red is proportional to the relative occurrence of interaction present during the simulations. Interactions highlighted as '*' correspond to the interactions observed in the crystal structure.

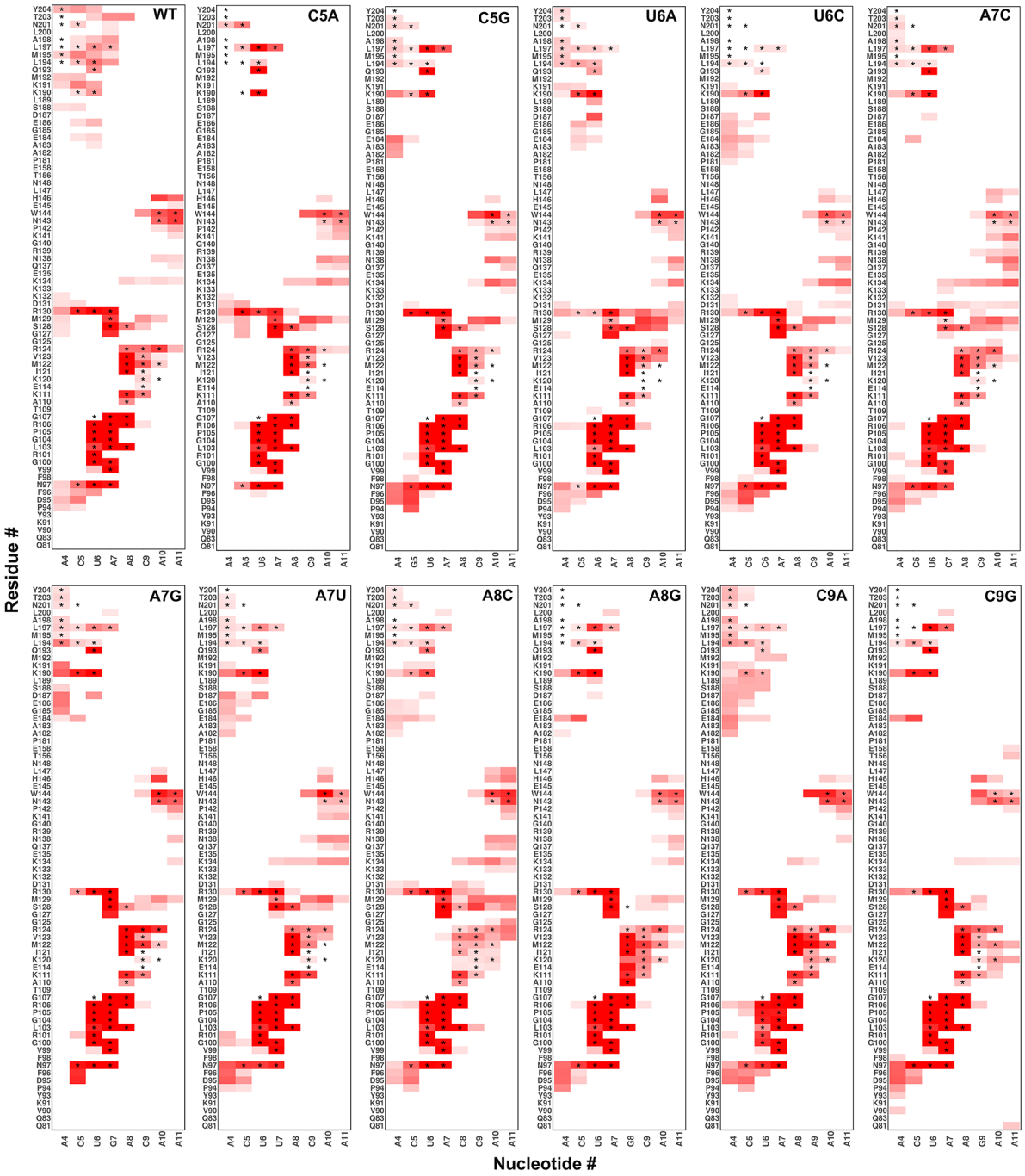
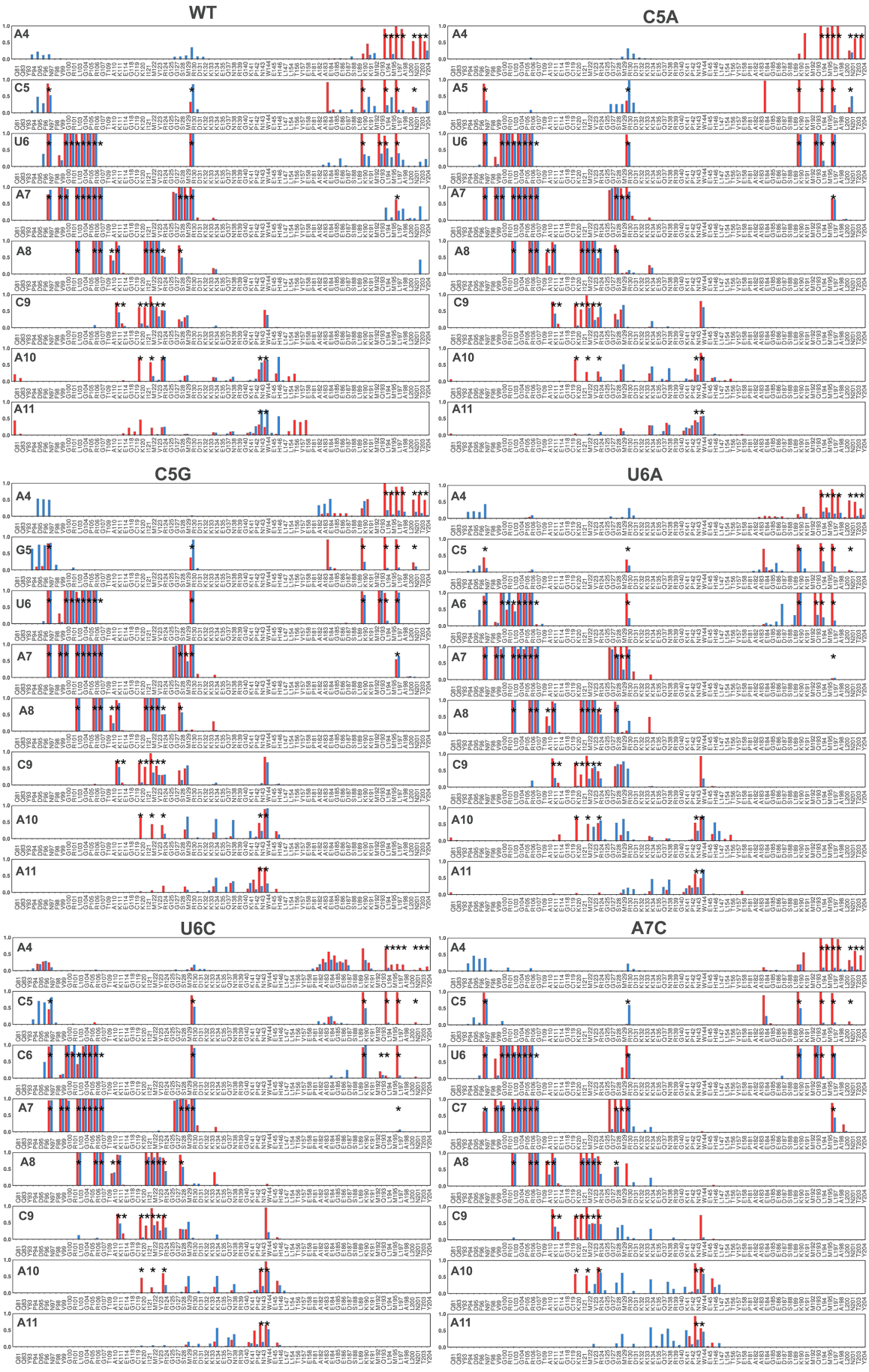


Figure SI12. Intermolecular contacts between mRNA and STAR domain of QKI protein computed for simulations using CHARMM forcefields. The values are computed excluding the first 100 ns of simulation trajectories. The intensity of the color red is proportional to the relative occurrence of interaction present during the simulations. Interactions highlighted as '*' correspond to the interactions observed in the crystal structure.


WT
C5A
C5G
U6A
U6C
A7C

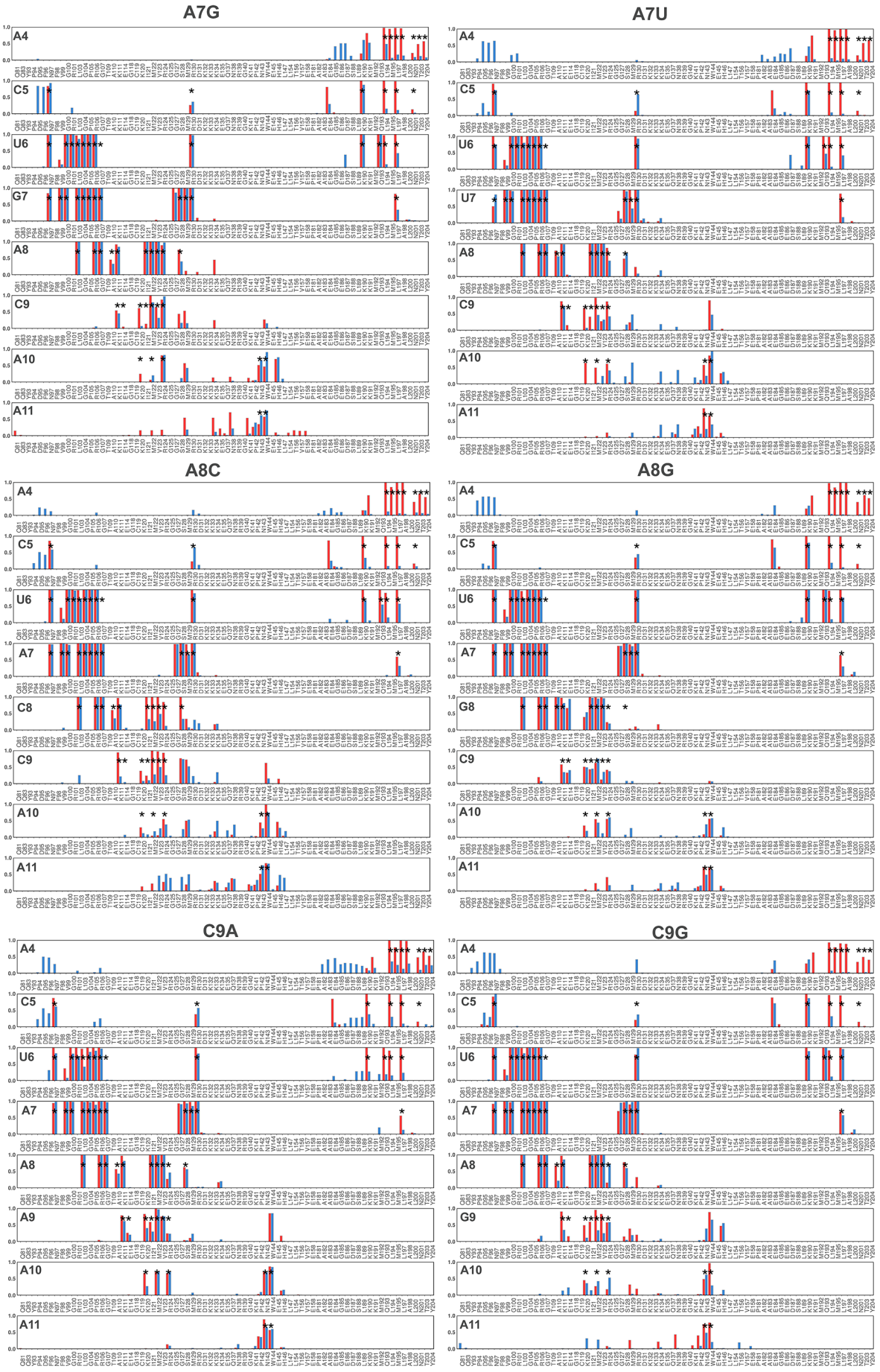
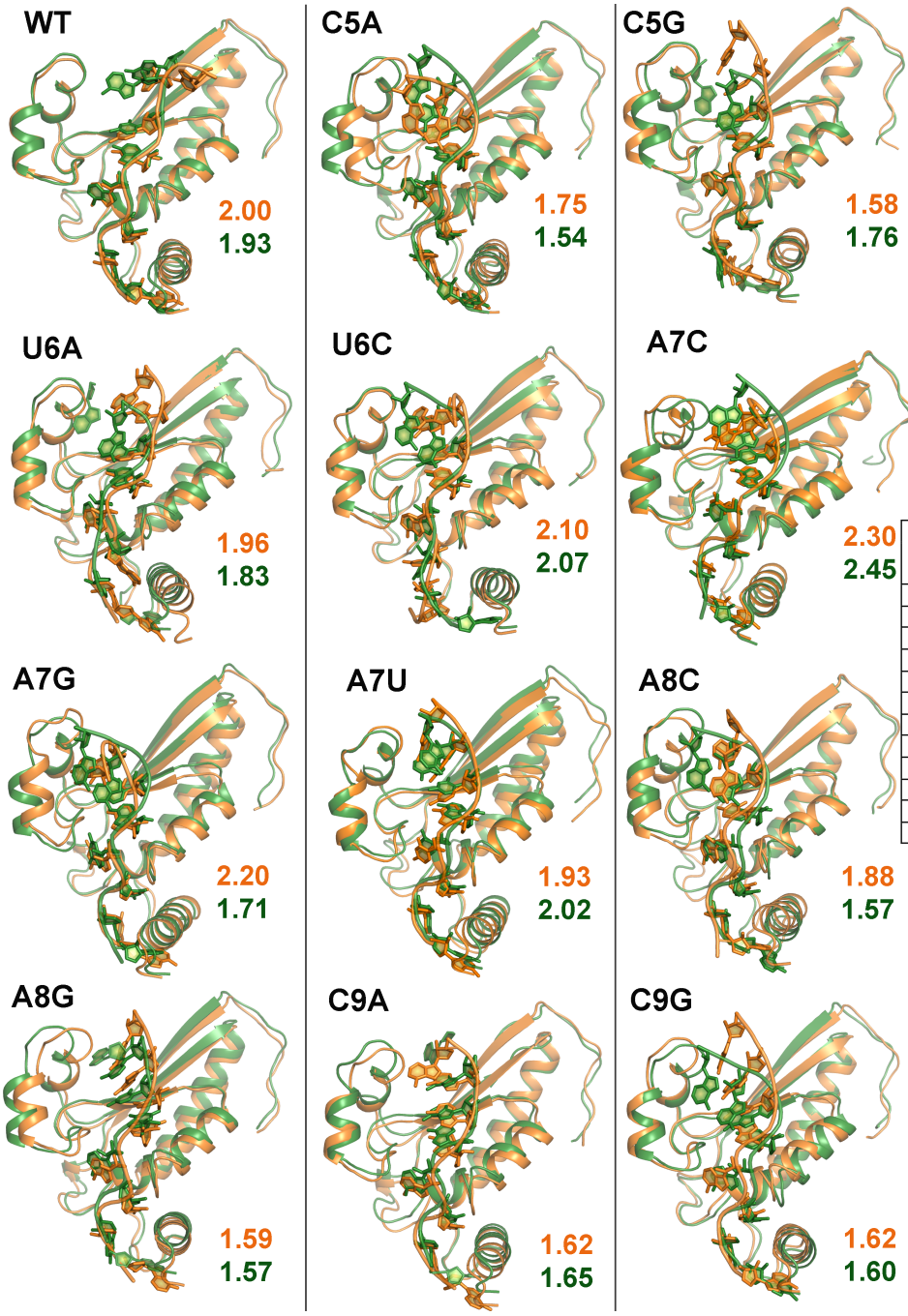
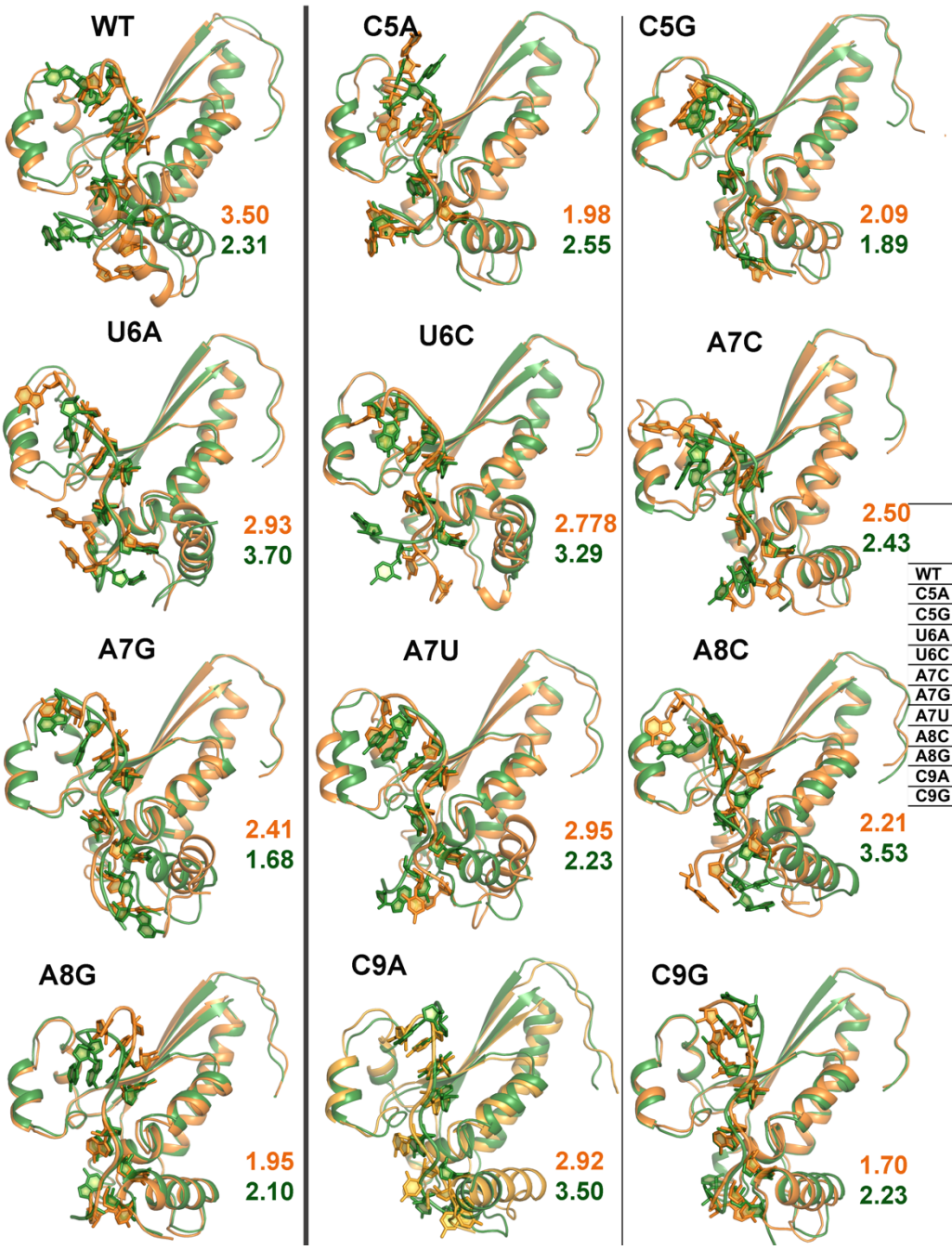

A7G
A4
C5
U6
G7
A8
C9
A10
A11
A7U
A4
C5
U6
U7
A8
C9
A10
A11
A8C
A4
C5
U6
A7
C8
C9
A10
A11
A8G
A4
C5
U6
A7
G8
C9
A10
A11
C9A
A4
C5
U6
A7
A8
A9
A10
A11
C9G
A4
C5
U6
A7
A8
G9
A10
A11

Fig SI13. Relative occurrence of residue-wise intermolecular contacts observed between protein residues (on x-axis) and mRNA nucleotides during simulations. Red bars represent values computed for simulations using AMBER forcefields, and blue bars corresponds to values computed for simulations using CHARMM forcefields. The contacts observed for the experimental structure are shown as '*'. The protein residues are arranged from N-terminal to C-terminal order.



	AMBER	
	Rep#1 (Orange)	Rep#2 (Green)
WT	28.4	23.4
C5A	40.4	18.0
C5G	45.8	26.1
U6A	44.7	27.8
U6C	59.9	12.9
A7C	46.3	15.7
A7G	35.8	20.2
A7U	39.1	33.5
A8C	26.9	26.7
A8G	31.0	30.7
C9A	48.3	35.6
C9G	32.0	30.4

Figure SI14. Representative structures of the two largest cluster centers obtained from the kmeans cluster analysis of conformations sampled during the simulations carried out using AMBER forcefields. Orange color corresponds to the largest cluster center and green color corresponds to the second largest cluster center. The two values beside the structures are the rmsd values in Å of backbone atoms of protein and mRNA with respect to the crystal structure (pdbid: 4JVH). The table provided on right shows the fractions (in %) of the conformations represented by these structures.



	CHARMM	
	Rep#1 (Orange)	Rep#2 (Green)
WT	42.3	29.2
C5A	40.0	26.5
C5G	52.7	17.3
U6A	30.0	26.8
U6C	51.1	23.5
A7C	39.6	23.3
A7G	41.1	32.6
A7U	36.3	28.4
A8C	53.2	20.3
A8G	38.1	34.0
C9A	58.7	28.2
C9G	44.0	32.6

Figure SI15. Representative structures of the two largest cluster centers obtained from the kmeans cluster analysis of conformations sampled during the simulations carried out using CHARMM forcefields. Orange color corresponds to the largest cluster center and green color corresponds to the second largest cluster center. The two values beside the structures are the rmsd values in Å of backbone atoms of protein and mRNA with respect to the crystal structure (pdbid: 4JVH). The table provided on right shows the fractions (in %) of the conformations represented by these structures.

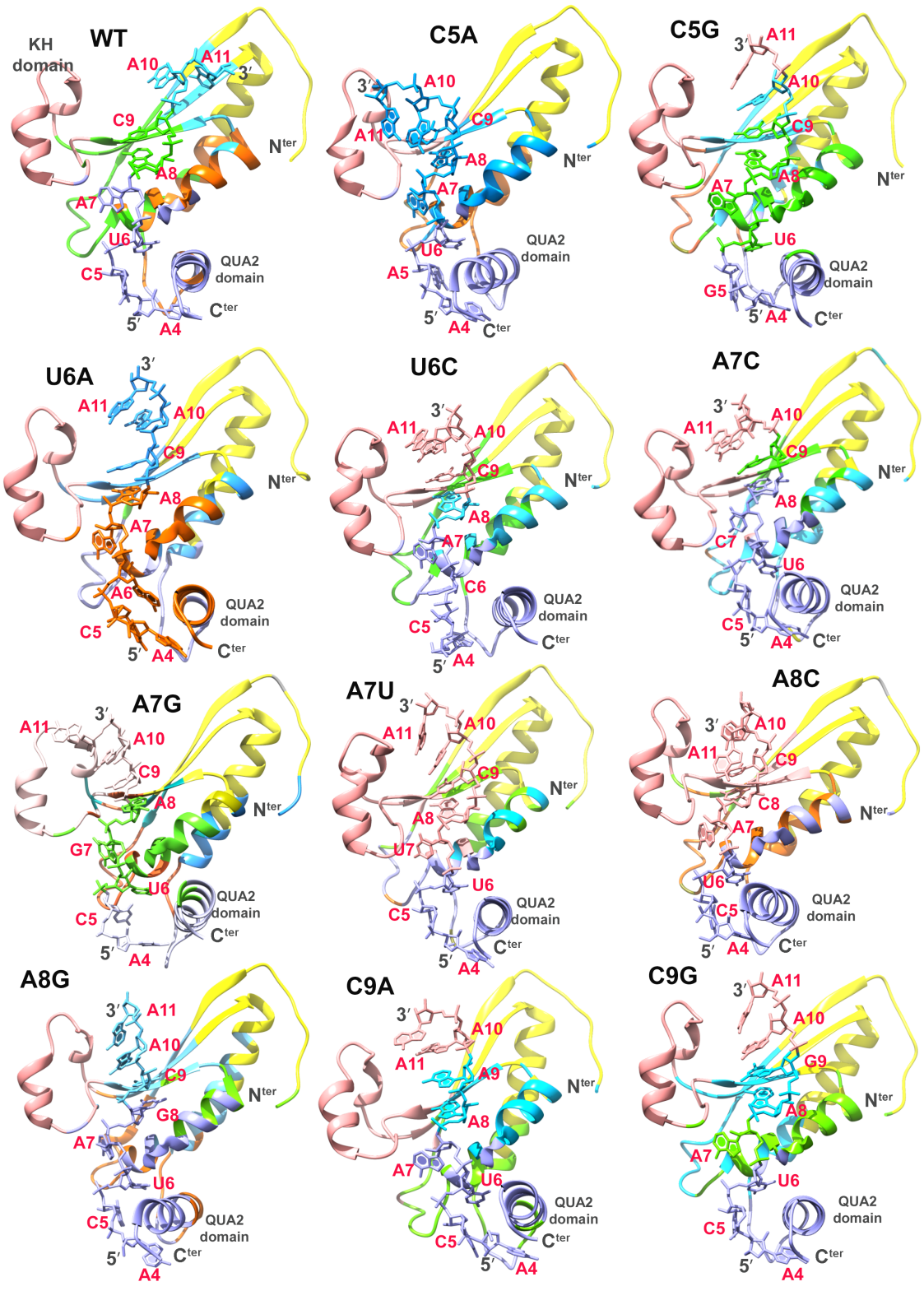


Figure SI16. Residue interaction networks for structures representing the largest cluster center as obtained from the kmeans clustering algorithm during the simulations carried out using AMBER forcefields. The residues are color-mapped according to the communities they belong. Full atom-representation of these networks are shown in Supplementary Figure SI17.

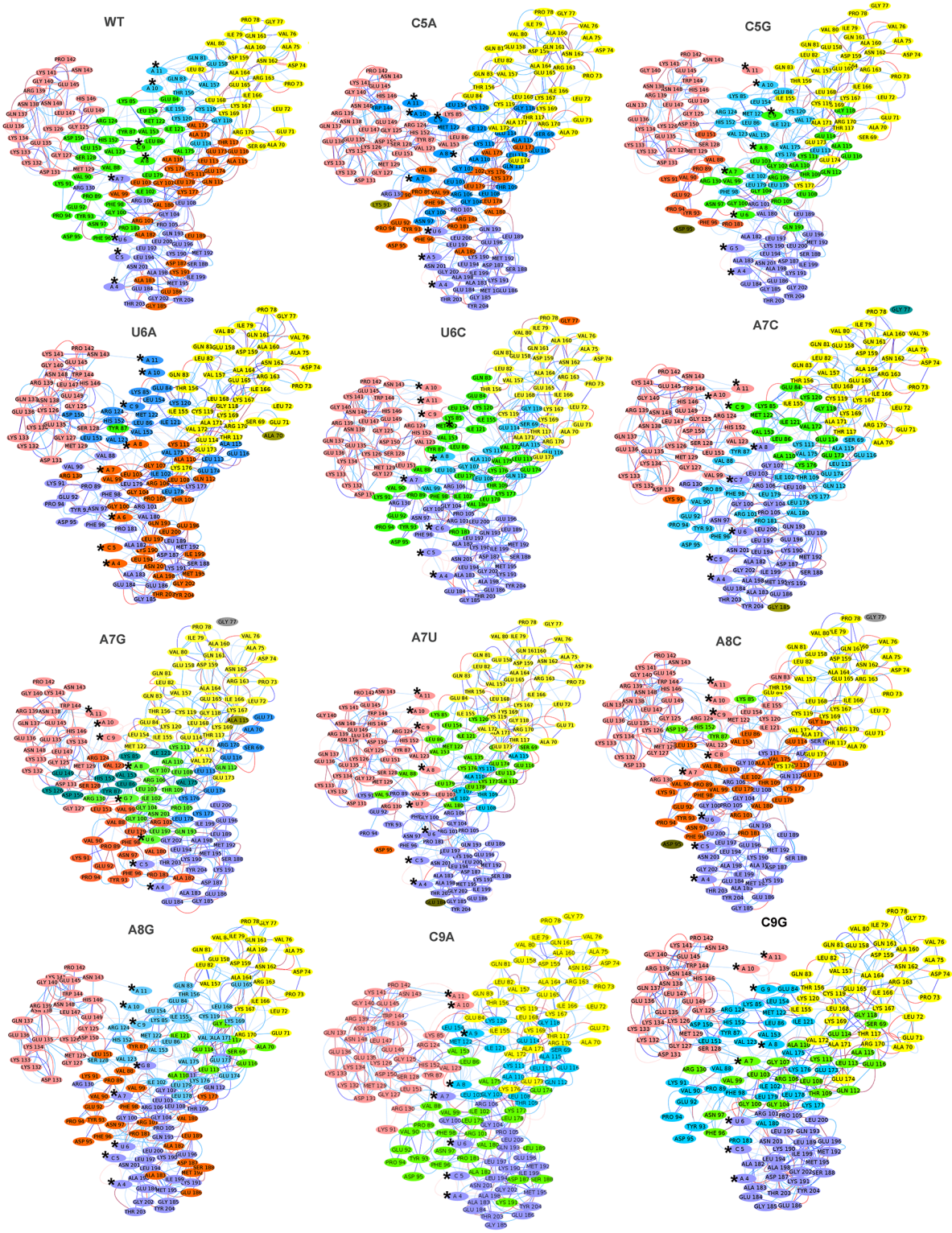


Figure SI17. Residue Interaction Networks observed for the mRNA and protein complexes simulated using AMBER forcefields. Structures representing the largest cluster center of conformations (obtained using kmeans clustering algorithm) are used. The residues are colored mapped to the community they belong to. The mRNA nucleotides are highlighted as '*'.

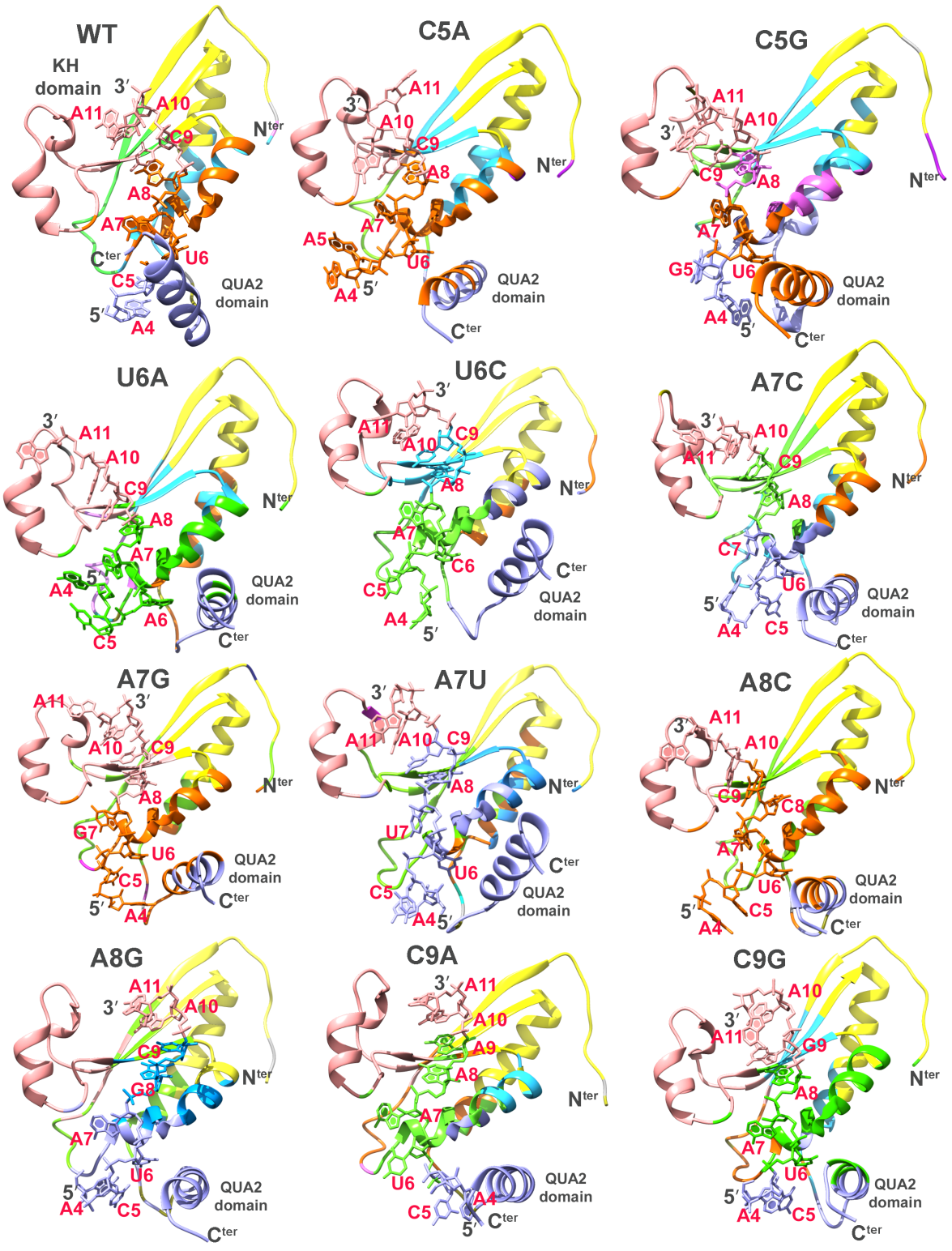


Figure SI18. Residue interaction networks for structures representing the largest cluster center as obtained from the kmeans clustering algorithm during the simulations carried out using CHARMM forcefields. The residues are color-mapped according to the communities they belong to. Full atom-representation of these networks are shown in Supplementary Figure SI19.

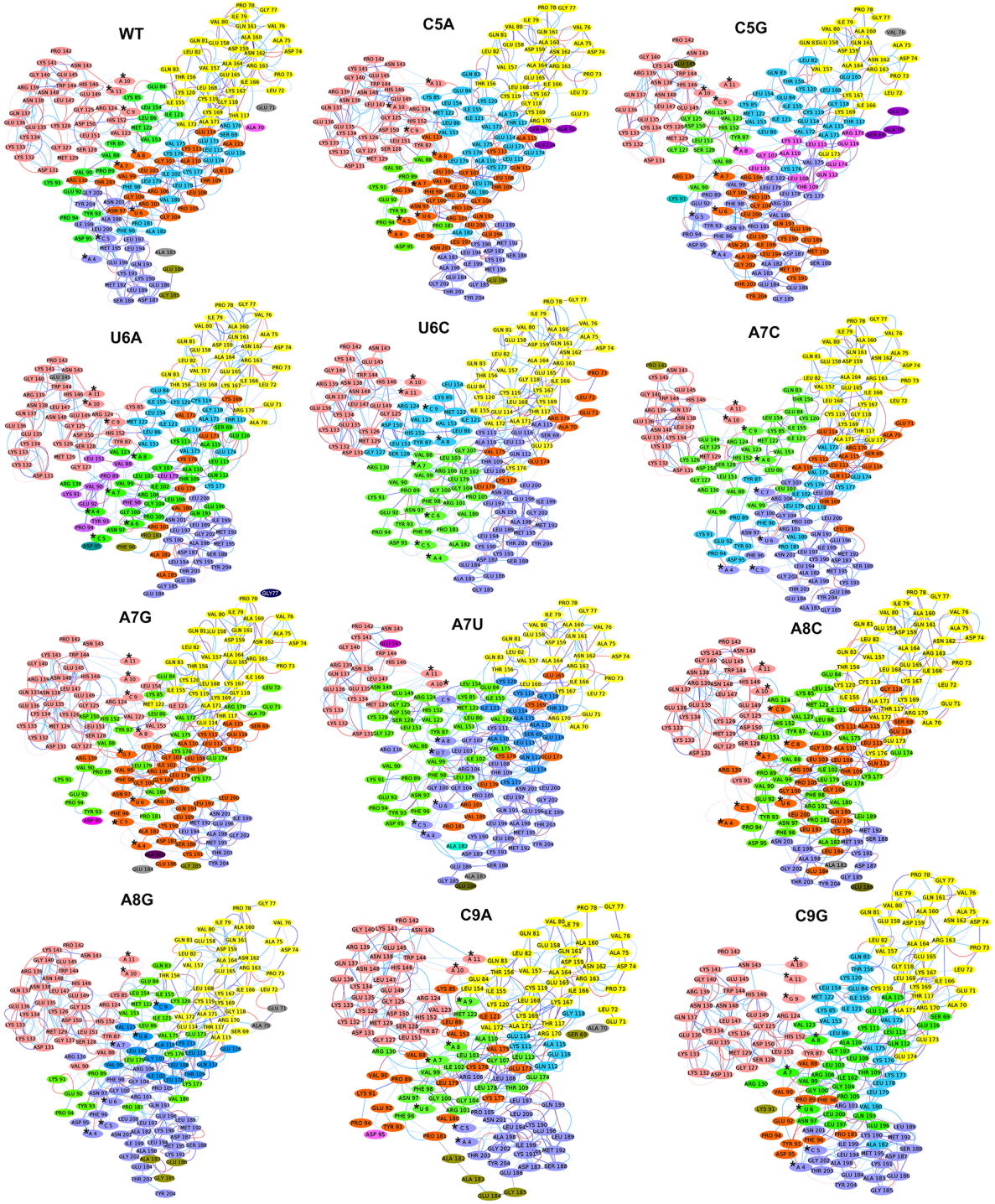


Figure SI19. Residue Interaction Networks observed for the mRNA and protein complexes simulated using CHARMM forcefields. Structures representing the largest cluster center of conformations (obtained using kmeans clustering algorithm) are used. The residues are colored mapped to the community they belong to.

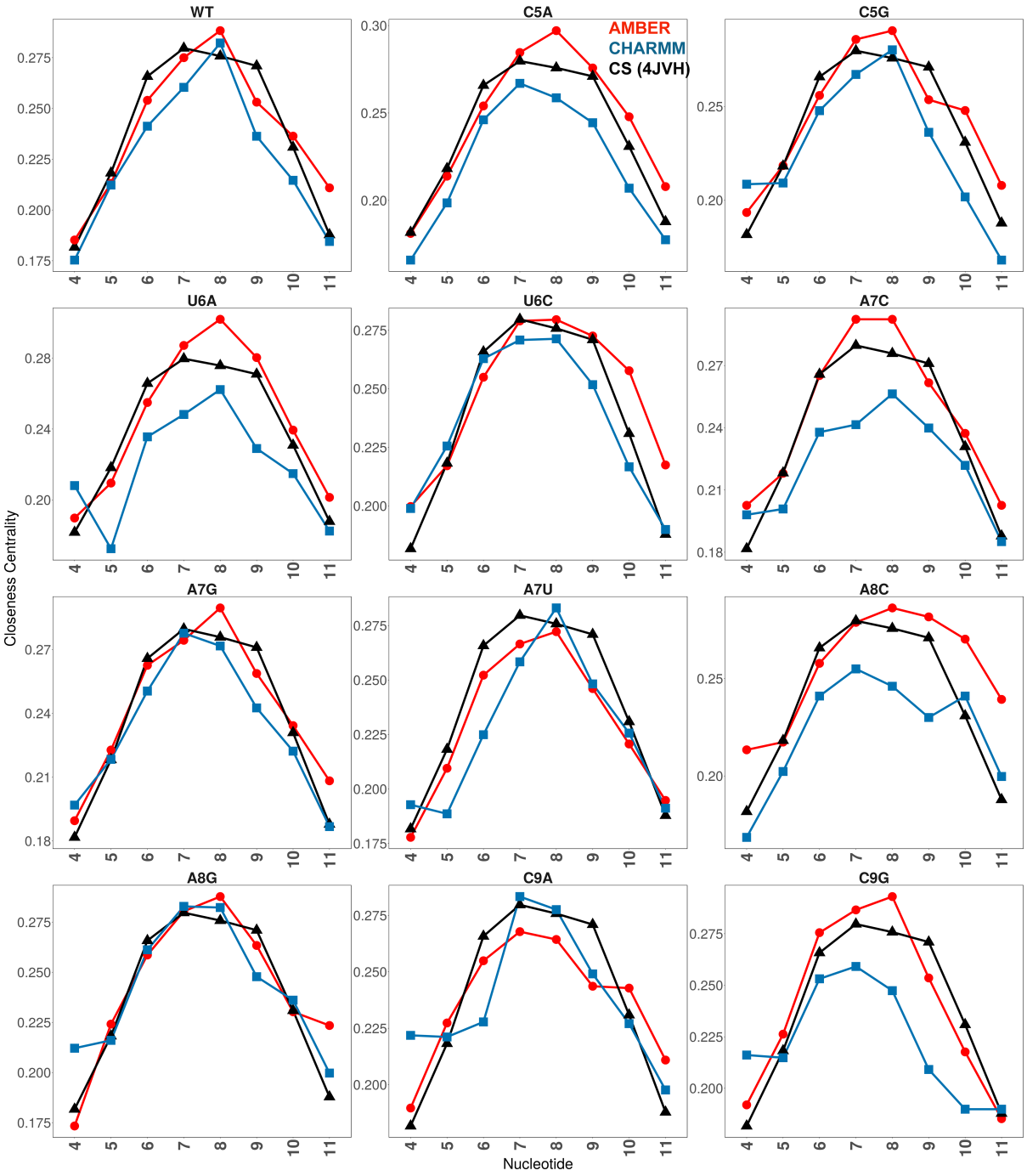


Figure SI20. Variation of closeness centrality values of the nucleotides as observed from simulations using CHARMM forcefields. Black colored triangles and lines correspond to closeness centrality values observed in crystal structure with bound cognate mRNA. Red colored circles and lines correspond to values observed in conformation representative of large cluster observed in simulations of non-cognate bound mRNA-QKI complexes.

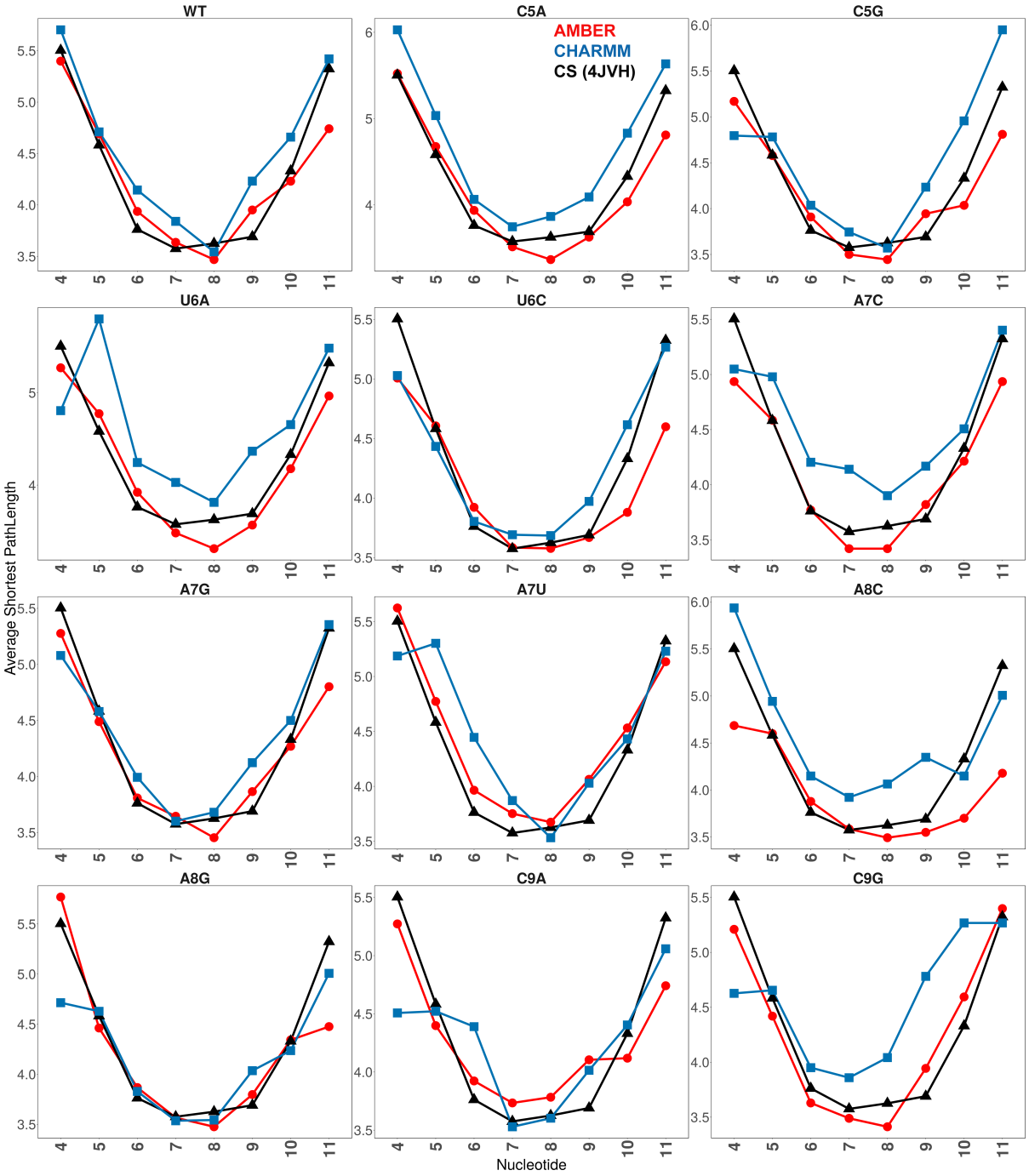


Figure SI21. Variation of average shortest path of the nucleotides as observed from simulations using CHARMM forcefields. Black colored triangles and lines correspond to closeness centrality values observed in crystal structure with bound cognate mRNA. Red colored circles and lines correspond to values observed in conformation representative of large cluster observed in simulations of non-cognate bound mRNA-QKI complexes.

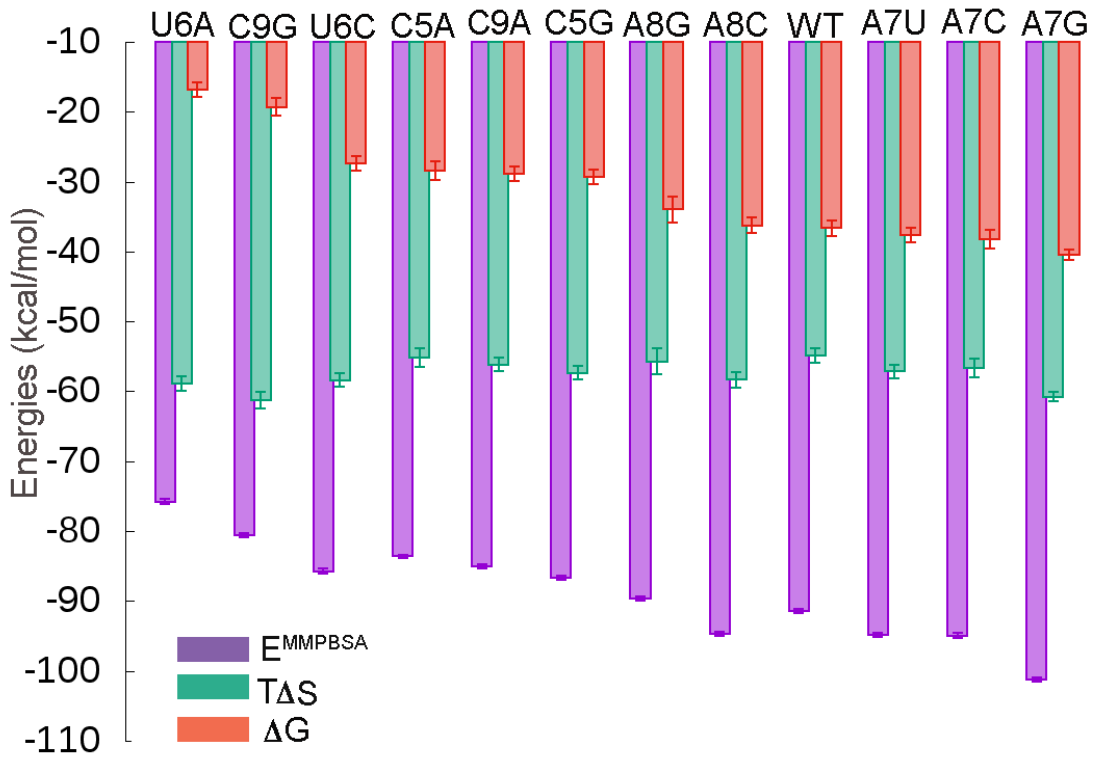


Figure SI22. Bar plots depicting binding energies calculated using MMPBSA approach for systems simulated using AMBER forcefields. The data is order according to increasing order of $\Delta G = E^{MMPBSA} - T\Delta S$, thus making U6A as least stable and A7G as most stable mutant. Purple bars corresponds to energies calculated using MMPBSA approach(E^{MMPBSA}), green bars corresponds to entropic contributions calculated from normal mode analysis ($T\Delta S$) and red bars corresponds to relative free energy of binding, ΔG. The error bars corresponds to standard error of mean values.

Table SI1: Table of relative IC_{50} values computed by experiments using quantitative gel mobility shift and fluorescence polarization assays. Taking the relation between relative IC_{50} values and relative binding energies as we computed

$\Delta\Delta G = -RT \ln\left(\frac{IC_{50}^{I}}{IC_{50}^{II}}\right) = \Delta G^{II} - \Delta G^{I}$.

The first relation is experimental data; and last one is from our studies, where $\Delta G = E^{MMPBSA} - T\Delta S$.

	Sequence used in experiments	Experimental study reference	Relative IC_{50}	$\Delta\Delta G^{exp}$ (kcal/mol)	Sequence used in our study	$\Delta\Delta G^{MMPBSA}$ (kcal/mol)
	FP/Mobility gel shift experiments *					
1	UACUCA	28	1	-	ACUCACAA (A7C)	-
2	UACUAA	28	41	-2.202	ACUAACAA (WT)	1.633
3.	UACUGA	28	<0.3 (~0.3)	> 0.714	ACUGACAA (A7G)	-2.223
4.	UACUUA	28	0.9	0.062	ACUUACAA (A7U)	0.617
	FP/EMSA experiments#					
5.	UACUAA	72	1	-	ACUAACAA (WT)	-
6.	UAAUAA	72	0.1	1.365	AAUAACAA (C5A)	8.191
7.	UAGUAA	72	< 0.01 (~0.01)	> 2.731	AGUAACAA (C5G)	7.283
8.	UACAAA	72	< 0.01 (~0.01)	> 2.731	ACAAACAA (U6A)	19.742
9.	UACCAA	72	< 0.01 (~0.01)	> 2.731	ACCAACAA (U6C)	9.264
10.	UACUCA	72	< 0.01 (~0.01)	> 2.731	ACUCACAA (A7C)	-1.633
11.	UACUGA	72	< 0.01 (~0.01)	> 2.731	ACUGACAA (A7G)	-3.856
12.	UACUUA	72	< 0.01 (~0.01)	> 2.731	ACUUACAA (A7U)	-1.015
13.	UACUAC	72	< 0.01 (~0.01)	> 2.731	ACUACCAA (A8C)	0.338
14.	UACUAG	72	< 0.01 (~0.01)	> 2.731	ACUAGAAA (A8G)	2.650
15.	UACUAA	72	< 0.01 (~0.01)	> 2.731	ACUAAAAA (C9A)	7.762
16.	UACUAA	72	< 0.01 (~0.01)	> 2.731	ACUAAGAA (C9G)	17.339

* Experiments (28) used 'UACUCA' as the base sequence. So mapping with our studies, we compared the data with ΔG^{I} for A7C and ΔG^{II} for other sequences: A7U, A7G, C5A and WT.
# Experiments (72) used 'UACUAA' as base sequence, thus we compared ΔG^{I} for WT and ΔG^{II} for other mutant sequences.

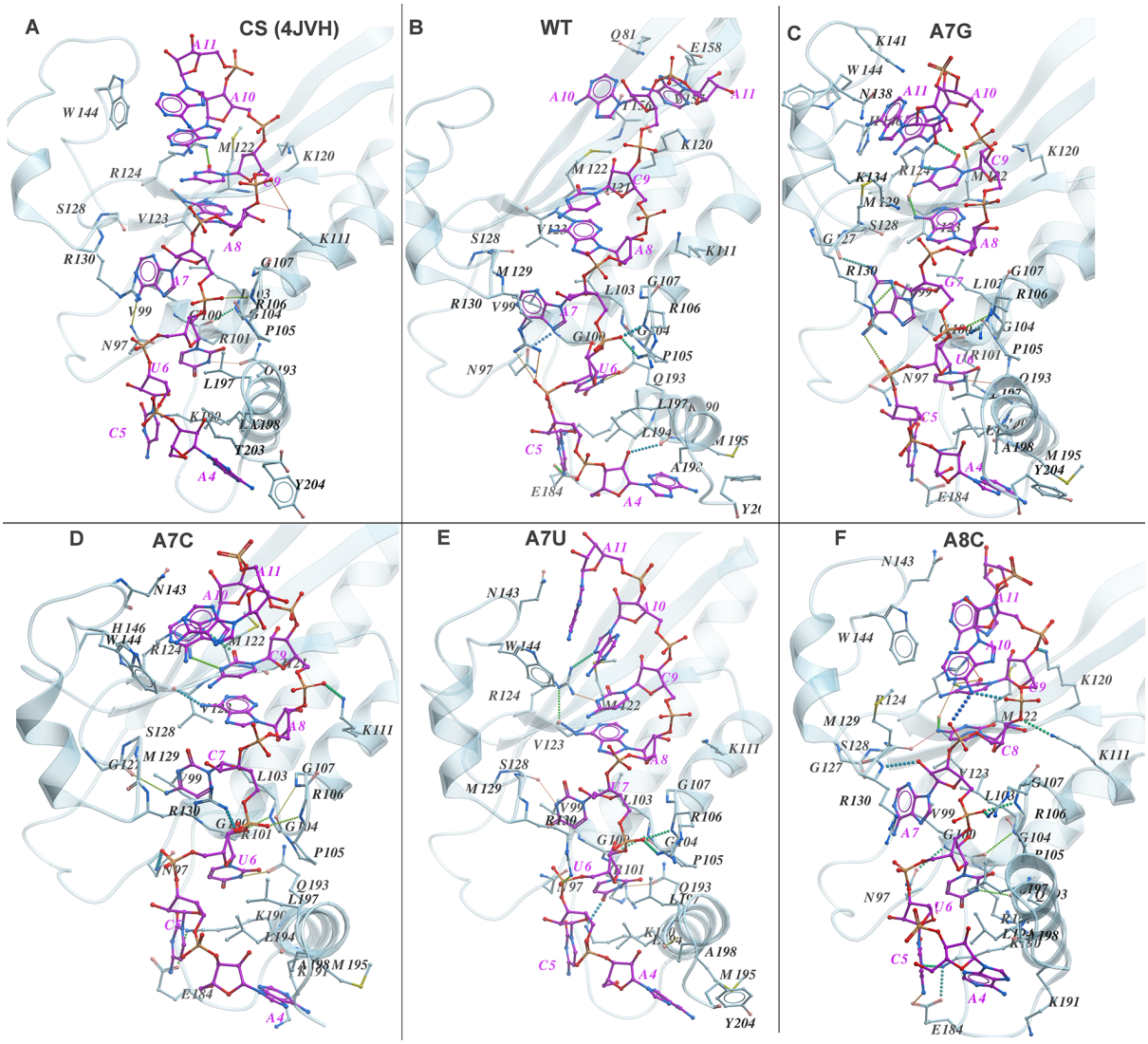


Figure SI23. Detailed interactions between mRNA and KH-QUA2 domain observed for mutants showing relative free binding energies as better than (A7G, A7C) or equivalent (A7U, A8C) to WT mRNA cognate sequence. CS corresponds to interactions in experimental crystal structure (pdbid: 4JVH). Interactions are shown for the structures representing the largest cluster center as obtained from the *kmeans* clustering algorithm during the simulations carried out using AMBER forcefields.

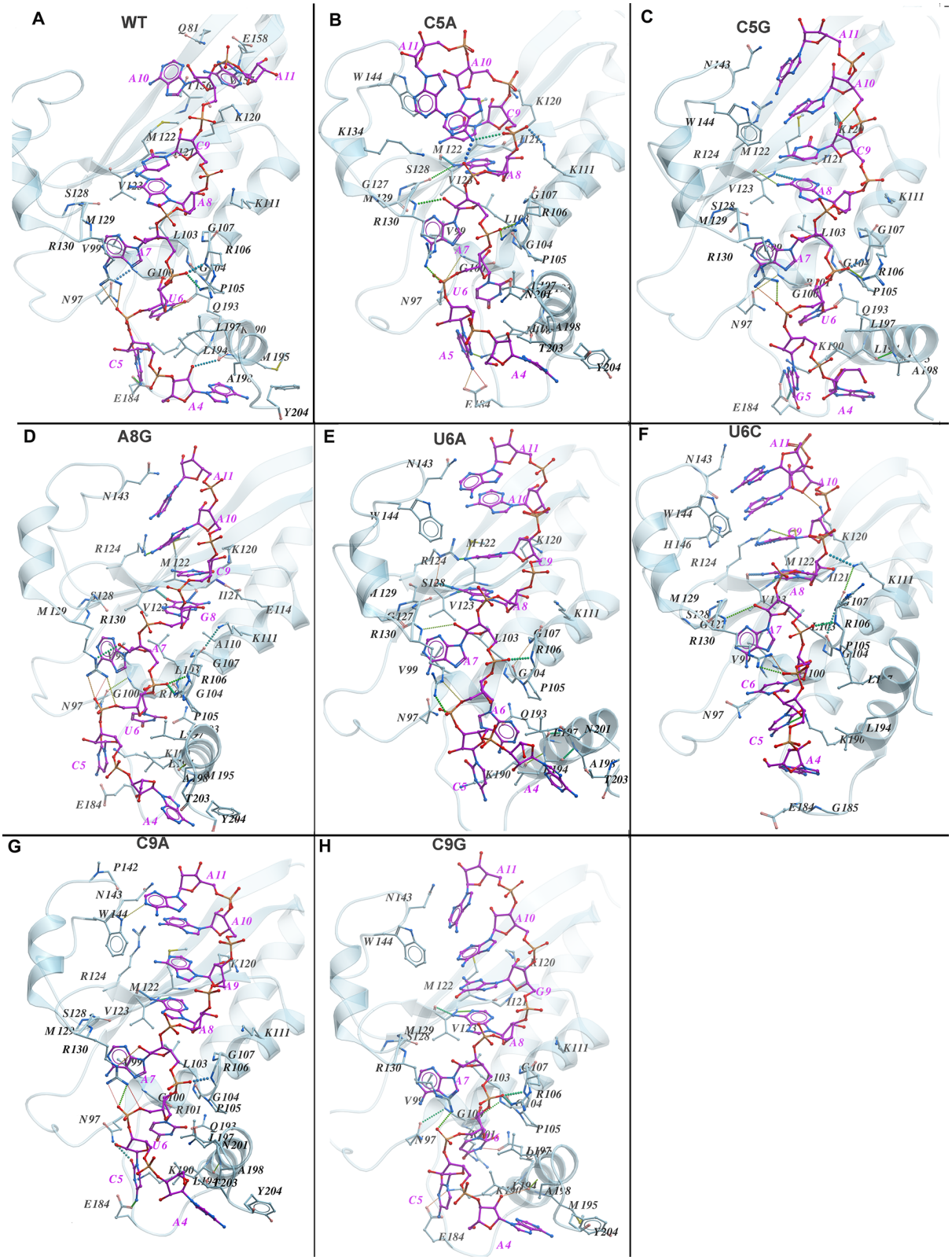


Figure SI24. Detailed interactions between mRNA and KH-QUA2 domain observed for mutants showing relative free binding energies lower than the WT mRNA cognate sequence. CS corresponds to interactions in experimental crystal structure. Interactions are shown for the structures representing the largest cluster center as obtained from the *kmeans* clustering algorithm during the simulations carried out using AMBER forcefields.