

3D Structure of RNA Monomeric G-Quadruplex Containing ALS and FTD Related G4C2 Repeat and It's Binding with TMPyP4 Probed by Homology Modelling based on Experimental Constraints and Molecular Dynamics Simulations

Kelly Mulholland[†], Holli-Joi Sullivan[†], Joseph Garner[†], Jun Cai, Brian Chen, and Chun Wu*

College of Science and Mathematics, Rowan University, Glassboro, NJ, 08028 USA

* To whom correspondence should be addressed. Tel: (856) 256-5462 Email: wuc@rowan.edu

† These authors contributed equally to this work

Present Address: Kelly Mulholland, Bioinformatics, University of Delaware, Newark, Delaware, 19716

ABSTRACT

The G-Quadruplex-forming hexanucleotide repeat expansion (HRE), d(G4C2)_n, within the human *C9orf72* gene is the root cause for familial amyotrophic lateral sclerosis-frontotemporal dementia (ALS-FTD). A recent study has shown that TMPyP4 has the good potential to work as a RNA G-quadruplex binder in treating ALS and FTD. Although the high-resolution structure of the monomeric DNA antiparallel G-quadruplex form of the monomeric hexanucleotide repeat was recently solved, the RNA parallel G-quadruplex structure and its complex with TMPyP4 are not available yet. In this study, we first constructed the homology model for the parallel monomeric RNA G-quadruplex of r(G4C2)₃G4 based on experimental constraints and the parallel monomeric G-quadruplex DNA crystal structure. Although the G-tetra core of the homology model was stable observed in 15 μ s molecular dynamics (MD) simulations, we observed the loops adopt additional conformations besides the initial crystal conformation, where TMPyP4 binding was found to reduce the loop fluctuation of the RNA Monomeric G-quadruplex. Next, we probed the elusive binding behaviour of TMPyP4 to the RNA Monomeric G-quadruplex. Encouragingly, the binding modes observed are similar to the modes observed in two experimental complexes of a parallel DNA G-quadruplex with TMPyP4. We also constructed a Markov state model to provide insights into the binding pathways. Together, the findings from our study may assist future development of G-quadruplex-specific ligands in the treatment of neurodegenerative diseases like ALS and FTD.

Keywords: Neurodegeneration, RNA G-quadruplex, Homology Modeling, Molecular Dynamics Simulation, MM/GBSA

INTRODUCTION

Guanine-rich DNA and RNA sequences can readily fold into stable four-stranded secondary structures known as G-quadruplexes, which are therapeutically involved in the development of anti-HIV(1), anti-cancer (2-4) and, in regard to the purpose of this study, anti-neurodegenerative (5) therapies. These nucleic acid secondary structures consist of a core of stacked guanine-quartets (G-tetrad) linked by loops of varying length and sequences (6). Both *in vitro* and *in vivo* data strongly support the physiological relevance of this nucleic acid secondary structure at the telomere and the promoter region of oncogenes, a signature guanine-rich region of the genome(7). Promotor and telomeric G-quadruplexes have been identified as potential therapeutic targets for human cancers and other diseases (8-13).

Two neurodegenerative pathologies involving the G-quadruplex motif include familial amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Both disorders cause progressive muscular tissue weakness, uncontrolled atrophy and change in personality as well as cognitive impairment through the degeneration of neurons (14,15). The prevailing genetic cause of the two diseases has been identified as a hexanucleotide repeat expansion (HRE), d(G4C2)_n, within the non-coding first intron of the *C9orf72* gene (5). Typically, individuals have HRE repeat lengths of about 2–23 hexanucleic acid units (16), however patients affected by this disease have an estimated 700-1600 units (17). The disease cascade is under extensive investigation and key contributors to the pathogenesis of ALS and FTD have recently been identified. These contributors include DNA G-quadruplexes and transcription leading to a loss of full-length products and accumulation of abortive RNA transcripts(18); loss of function of the *C9orf27* gene (17,19,20); the sense r(G4C2)_n or antisense r(C4G2)_n RNA transcripts of d(G4C2)_n repeats, that form nuclear foci and aggregate to create toxicity by sequestering different RNA-binding proteins(17,21,22); and finally, the “c9RAN” protein, that originates from repeat-associated non-ATG (RAN) translation, which form neuronal and cytoplasmic dipeptide inclusions of the central nervous system(23,24).

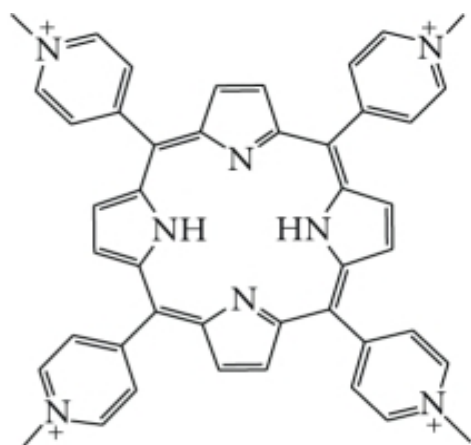


Figure 1. Structure of TMPyP4.

Given the crucial role that HRE plays in the development of familial ALS and FTD, one major research effort is focused on the development of therapeutic small-molecules that can neutralize the toxic RNA G-quadruplex and inhibit RNA foci formation and RAN translation(5), with some studies showing promising therapeutic effects (25,26). In particular, Zamiri *et al.* have shown that TMPyP4 (**Figure 1**), a well-known DNA G-quadruplex-binding ligand (27-30), binds to the RNA G-quadruplex structures formed by the $r(G_4C_2)_n$ repeat (5). This binding distorts the secondary structure of the repeat, inhibiting interactions with mutant repeat-binding proteins that may have resulted in the formation of neuronal and cytoplasmic dipeptide inclusions in the brain. Although the authors suggested that TMPyP4 has good potential to work as a RNA G-quadruplex binder in treating ALS and FTD, a detailed understanding of ligand-RNA G-quadruplex interactions at the molecular level remains uninvestigated. Encouragingly, in 2014 Haeusler *et al.* showed that, while $d(G_4C_2)_n$ HRE form an antiparallel monomeric DNA G-quadruplex, $r(C_4G_2)_n$ HRE form a parallel monomeric RNA G-quadruplexes(18). In 2015, two high-resolution structures of DNA G-quadruplexes of the sequence $d(G_4C_2)_3G_4$ were solved (PDB ID: 2N2D and 5OPH) by NMR(31) and both structures indeed adopt an antiparallel scaffold. However, the high-resolution structure of the RNA monomeric G-quadruplex of $r(G_4C_2)_3G_4$ still remains to be elusive. Homology modelling could be a solution. In two previous studies, Collie *et al.* have successfully generated homology models of RNA G-quadruplex using the solved DNA quadruplex structures (32,33).

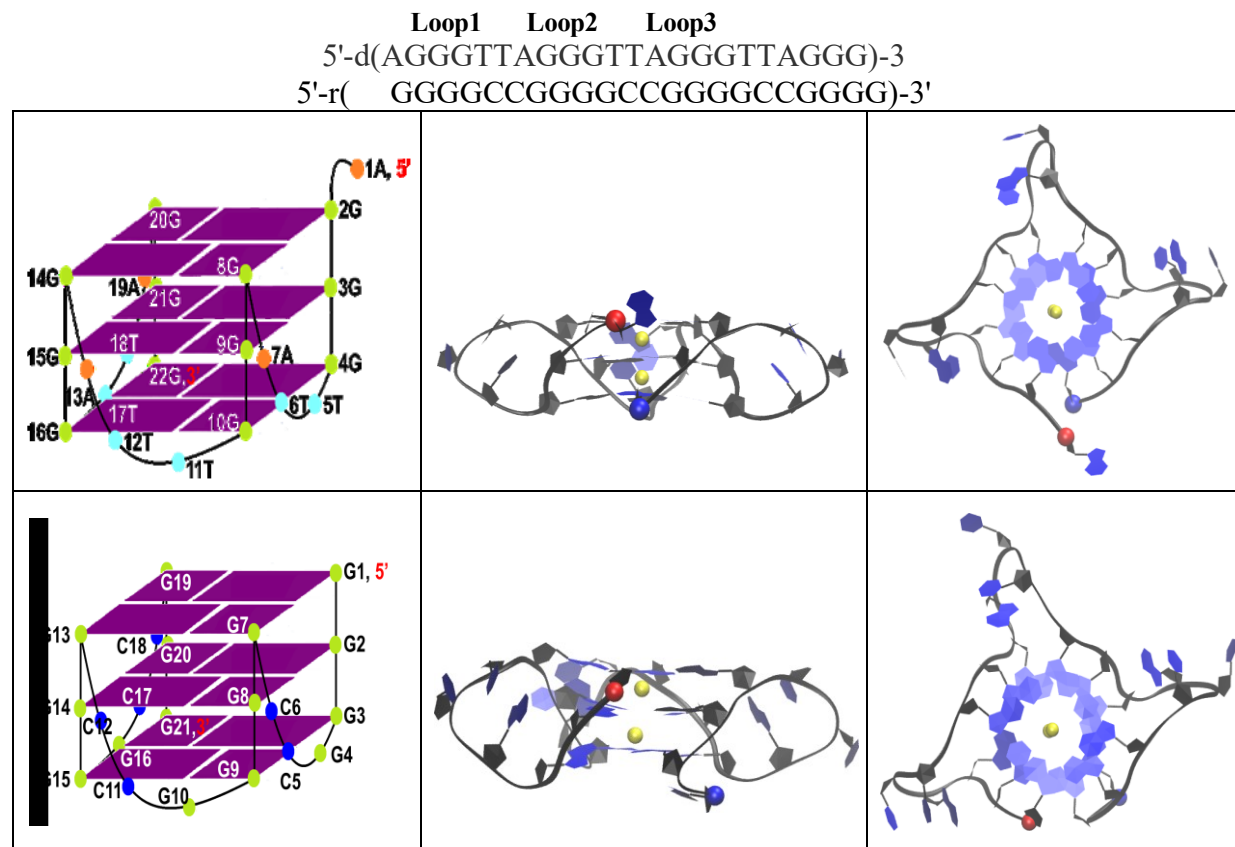


Figure 2 Sequence Alignment between $r(G_4C_2)_3G_3$ and Human Telomeric Repeat Sequence
 Structure of $r(G_4C_2)_3G_3$ and Human Telomeric Repeat Sequence

Molecular dynamics (MD) simulations are powerful computational tools that provide insights at a molecular level. Several recent studies have used this approach to determine the 3D structures of DNA G-quadruplexes (34), identify the importance of ion binding in the stability and folding of quadruplexes (35), resolve ligand-binding specificity using absolute free binding energy calculations for both c-MYC (36) and human telomeric G-quadruplex DNA (37), and to examine binding interactions between G-quadruplexes and therapeutic agents (1,38). More specifically, Hou et al. used this approach to probe the stability of six ligand-G-quadruplex DNA complexes, including TMPyP4, that have been structurally determined by experimental approaches. Ali *et al.* investigated the binding of TMPyP4 to three different scaffolds of human telomeric DNA G-quadruplex using MD simulations and determined that the binding free energy analysis suggested the parallel scaffold is preferred by TMPyP4 over antiparallel and hybrid scaffold (39). Using MD simulations in our previous studies, we have comprehensively examined the major binding poses of various G-quadruplex ligands to the human telomeric DNA G-quadruplexes starting from an unbound state (40-44). In this present study, we carefully generate and validate a homology model for the parallel RNA G-quadruplex of r(G₄C₂)₃G₄ (**Figure 2**) based on experimental constraints (18,45,46) and the available crystal structure of human telomeric parallel G-quadruplex DNA (PDB ID: 1KF1). Using Molecular dynamics (MD) simulations (**Table 1**), we provide a thorough evaluation of the detailed structural properties of the RNA G-quadruplex and the conformation changes of the RNA G-quadruplex when binding to TMPyP4. Our proposed model of the parallel RNA G-quadruplex of r(G₄C₂)₃G₄, as well as its binding interactions with TMPyP4, provide knowledge that will assist future studies aimed at developing new treatments of neurodegenerative diseases such as ALS-FTD.

Table 1. Molecular dynamics simulations.

System ID	Sequence	No. of ligand	No. of run	Drug Initial State	NPT equ. (ns)	NVT (ns)	Total time (ns)
1	d(AG ₃ T ₂) ₃ G ₃ *	0	4	-	1	2000	8000
2	r(G ₄ C ₂) ₃ G ₄	0	15	-	1	1000	15000
3	r(G ₄ C ₂) ₃ G ₄	1	3/37	Unbound	1	2000/200	13400
4	N/A	1	5		1	500	2500

* Human Telomeric parallel G-quadruplex (PDB id: 1KF1)

RESULTS

A homology model of parallel G-quadruplex RNA structure of the r(G₄C₂)₃G₄ repeat was built based on experimental evidences and a high-resolution parallel G-quadruplex DNA template.

Multiple CD spectra studies, a RNase T1 protection assay and a gel mobility shift assay suggest that the RNA repeat shows characteristics of a parallel G-quadruplex structure (18,45,46). The RNA digestion pattern from the endonuclease RNase T1 assay indicates that the parallel topology contains three G-tetrads, compared to four G-tetrads in the DNA anti-parallel topology (18). Since the only

high resolution structures of RNA parallel G-quadruplexes available are formed from multiple RNA strands (PDB IDs: 3MIJ, 2KBP, 3IBK), we chose a single stranded human telomeric parallel DNA G-quadruplex structure (PDB ID: 1KF1) for this study, as it most closely matched the proposed structure. Although a number of structures are available for single stranded parallel DNA G-quadruplexes that contain the same or similar sequences (**Table S1**), they include additional biasing factors such as being in complex form, formed from multiple strands or formed under crowded solution, which are not factors to consider for the human telomeric parallel DNA G-quadruplex structure chosen in this paper (PDB ID: 1KF1). Therefore this parallel DNA quadruplex structure was utilized as a template for homology modeling of the G-quadruplex structure of the RNA repeat in this study (**Figure 2**). The major differences are as follows: 1) the 5' terminus of the RNA repeat does not have A; 2) the loop sequence of RNA repeat is GCC instead of TTA; 3) the 3' terminus of the RNA has additional G. Positioning of G4, G10, G16 and G22 at the loops is consistent with the RNase T1 protection assay results(18).

The G-quadruplex RNA homology model showed good stability throughout multiple simulations. To probe the stability of our homology model, the RNA G-quadruplex only system was built and subjected to 15 runs of 1.0 μ s MD simulations (**Table 1**) as described in the method section. The last snap shots of all 15 simulation runs are available in the supporting document (**Figure S2**), as well as two addition simulation runs including the order plot (**Figure S3**), in the second part of this figure, we present one of the three simulation runs that we observed a potassium ion moving out of the ion pore. The representative trajectories and three order parameters were closely analyzed (**Figure 3** and **Figure S3**). These parameters consisted of calculations of H-bonds for three G-tetrads, RMSD of the whole G-quadruplex backbone, and the center-to-center distance of the two potassium ions within the G-quadruplex. 12 of 15 runs show a stable structure throughout the simulation with no large fluctuations in any of the three order parameters over the total 1000 ns (See a representative trajectory in **Figure 3**).

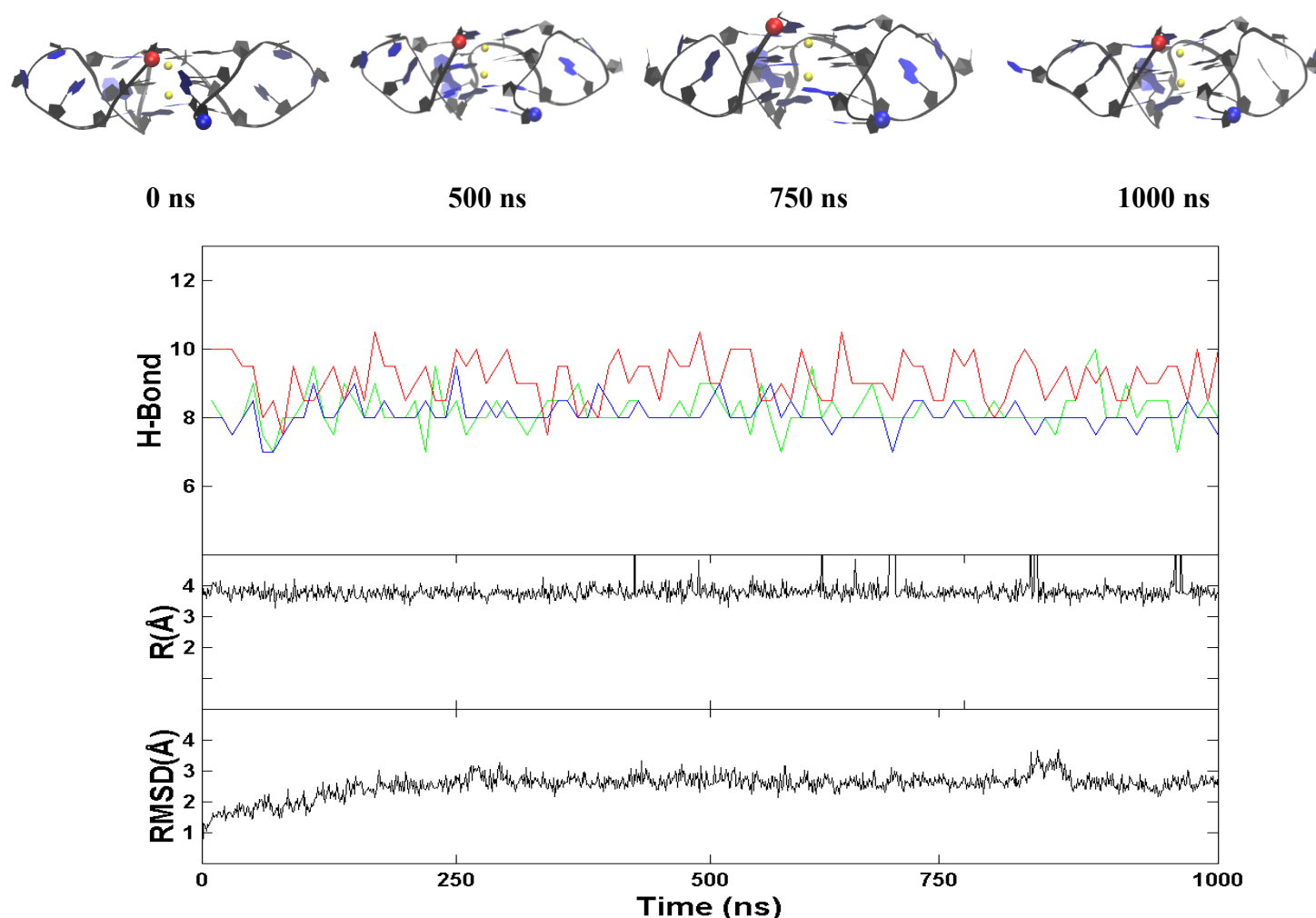


Figure 3. Stability simulation of RNA quadruplex built from the NMR DNA quadruplex. **A)** RNA quadruplex, **B)** Hydrogen bonds present in first (green), second (red), and third (blue), layers of G-tetrads, **C)** K⁺-K⁺ distance and **D)** Backbone RMSD. 5' and 3' of the RNA chain are indicated by a red and blue ball, respectively.

Geometric characterization of the RNA G-quartets showed a right handed helical twist; ligand binding can cause subtle changes. Clear from the top view of the RNA G-quadruplex core (**Figure 4A**), the G-quartets have a right handed helical rotation. In this work, we calculated the rise, H-rise, and H-twist of the apo form DNA G-quadruplex, the apo form RNA G-quadruplex as well as for the RNA G4/TMPyP4 top and bottom binding modes (**Table 2**), as detailed in the methods section. For the RNA-only simulated system, the rise, H-rise and H-twist averaged over the two layer steps (i.e. bottom to middle and middle to top) are ~3.53 Å, ~1.82 Å and ~13.7°, respectively, which are close to the values of the DNA only system (~3.49 Å, ~1.88 Å, and ~12.32°). Clearly, H-rise and H-twist in both the RNA and DNA G-quadruplex are smaller than those in standard B-DNA (3.32 Å and 34.3° for H-rise and H-twist). The binding of TMPyP4 to the RNA G-quadruplex did not change these parameters much to both the top (~3.55 Å, ~2.35 Å and ~16.60°) and the bottom binding modes (~3.49 Å, ~2.63 Å and ~21.17°). H-twist was the parameter most changed upon ligand binding, with ~7Å difference from the DNA apo form. Thus, subtle changes may occur upon ligand binding.

Table 2. DNA G-quadruplex G4 layer geometry parameters. 1-3: top layer, middle and G4 bottom, respectively.

Layers	Parameter ¹	Apo 1KF1	Apo $r(G_4C_2)_3G_4$	Top Binding	Bottom Binding
3:2	Rise	3.56±0.01	3.63±0.08	3.70±0.11	3.52±0.10
3:2	H-Rise	1.91±0.41	1.88±0.17	2.35±0.21	2.61±0.21
3:2	H-Twist	12.76±2.55	12.11±0.36	14.14±2.16	19.15±3.15
2:1	Rise	3.42±0.01	3.44±0.01	3.40±0.07	3.46±0.09
2:1	H-Rise	1.86±0.32	1.77±0.12	2.35±0.13	2.65±0.25
2:1	H-Twist	11.86±1.63	15.13±1.00	19.06±1.58	23.20±3.10

¹ Value for Rise and H-rise reported in Å; Values for H-twist reported in degree.

The interaction between K⁺ and the G-quadruplex did not change with ligand binding. We calculated the distance between each potassium ion and the eight neighboring oxygen atoms (**Figure 4C**) and the distance between each G-quartet oxygen relative to one-another (**Figure 4D**) for the DNA G-quadruplex, as well as the three RNA G-quadruplex systems (**Table S2-S3**). The apo form of the DNA G-quadruplex had an average oxygen-potassium distance of 2.69 Å, 2.83 Å, and 2.74 Å, for the top, middle, and bottom G-quartets, respectively. For the apo form of the RNA G-quadruplex, the average oxygen-potassium distance for the top, middle, and bottom layers were 2.70 Å, 2.47 Å, and 2.70 Å, respectively. Compared to the apo forms, both complex forms with TMPyP4 showed no significant change in potassium-oxygen distance. For the top binding mode, the average distance was 2.78 Å, where there was an average distance of 2.70 Å, 2.85 Å, and 2.72 Å for the top, middle, and bottom G-quartets, respectively. For the bottom binding mode, there was an average potassium-oxygen distance of 2.79 Å, where the top, middle, and bottom G-quartets were 2.71 Å, 2.86 Å, and 2.71 Å away from the local potassium ion, respectively.

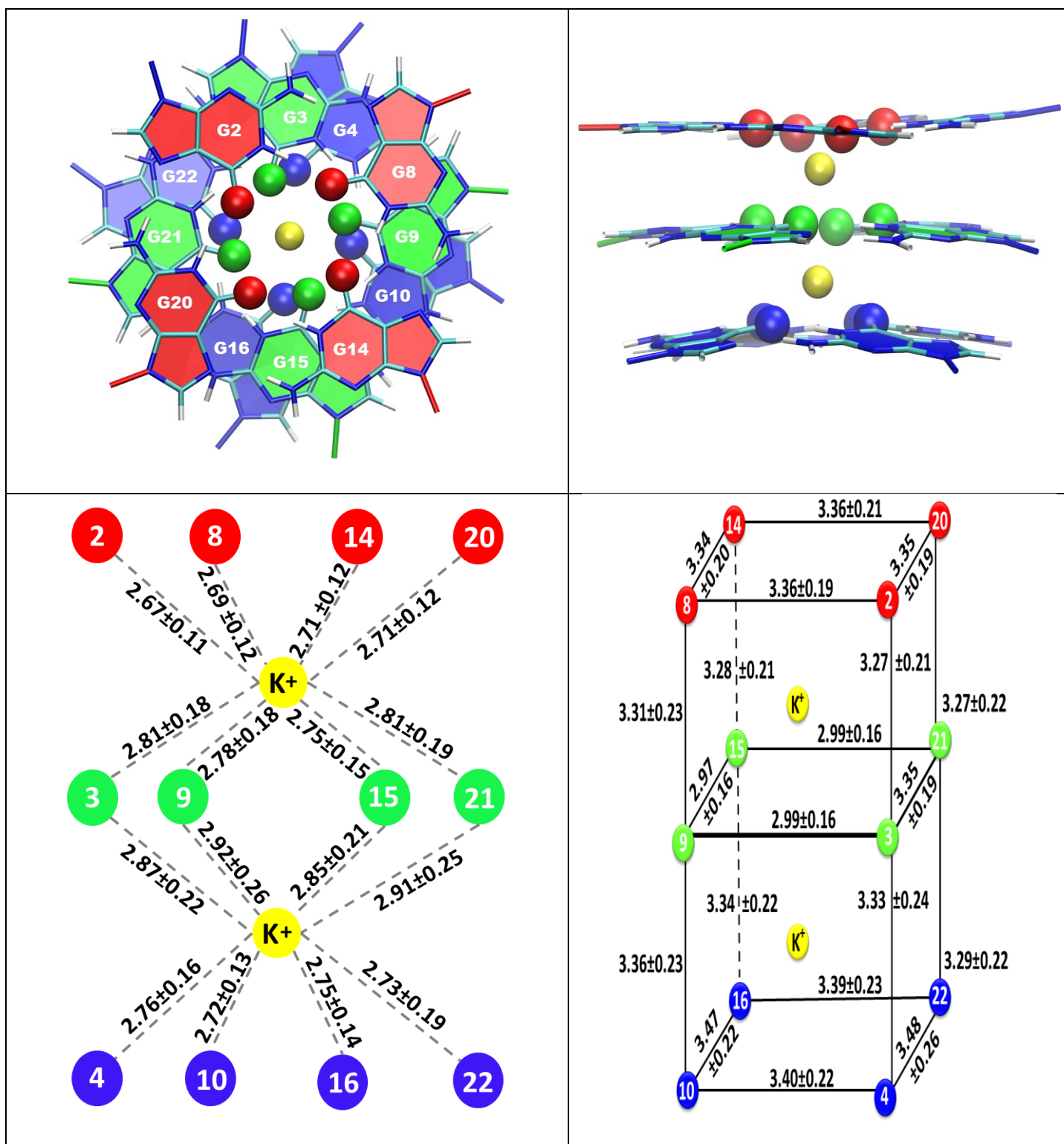


Figure 4. Oxygen-Potassium Distance Parameters in the 3-layers of the RNA G-quadruplex: **(A)** Top view of the three G4 layers. Oxygen is represented by a like colored ball, K^+ cations are represented by a yellow ball; **(B)** Side view of the three G4 layers; **(C)** Distance (Å) of the oxygen from each residue to the nearest K^+ cations; **(D)** Distance (Å) of each oxygen relative to the nearest side.

By calculating the distance between each oxygen residue lining the ion pore of the G4 core, we observed no significant difference in oxygen-oxygen distance upon ligand binding. For the DNA and RNA apo forms, the average distance between each residue of the same G-tetrad was 3.29 Å and 3.31 Å, and the average distance between each residue to the residue of the neighboring G-tetrad was 3.31 Å and 3.32 Å, respectively. For the top and bottom binding modes, the average distance between each residue of the same G-tetrad was 3.29 Å and 3.30 Å, and the average distance between each residue to the residue of the neighboring G-quartet was 3.34 Å and 3.33 Å, respectively. These calculations clearly show that ligand binding did not significantly affect the interaction between the potassium cations and the G-quadruplex when the potassium ion remained bound in the ion pore.

Besides tribase stacking, various loop conformations were observed in the stability simulations of the monomeric G-quadruplex of the $r(G_4C_2)_3G_4$ repeat. Because our homology model of the RNA repeat was built using the crystal structure of the human telomeric DNA G-quadruplex monomer (**Figure 5a**), its GCC loop initially adopts a tribase stacking conformation (**Figure 5b**) that resembles the TAT stacking conformation of the TTA loop in the DNA G-quadruplex. We note that because the crystal structure is a dimer structure, the TAT stack conformation might be partially stabilized by a π - π stacking between the loop A and the 5' A of the other G-quadruplex monomer. If the dimer is not the major form of the G-quadruplex under physiology conditions, the loop conformations may become more dynamic in the absence of these additional stabilizing interactions. Indeed, we observed the GCC loop adopted additional conformations in our stability simulations of our homology model. We present an example of the additional loop conformations sampled in our simulations in **Figure 5c-5e**, using the third loop C as a representative for each case. The statistics on the three loops from the last snapshot of the 15 stability simulations are summarized in **Table S4** of the supporting document. For roughly 60% of the loop population maintained three base stacking, as reported in the crystal structure (**Figure 5b**). In ~22% of the loop population, two bases are stacking and one base flips outward. An example of this is presented in Figure 5c where residue C17 flips outward and residues G16 and C18 maintain their initial conformations. As depicted in **Figure 5d**, ~14% of the loop population showed two bases flip outward. In this example, C17 and C18 flip away from the G-Quadruplex, as G16 stacks onto the G-quadruplex core. Least common, was the case of all three bases flipping outward, away from the G-quadruplex core which occurred in ~4% of the loop population and only occurred in loop C. This data shows that although a majority of the loop conformations maintained a tri-base stacking as in the crystal structure, alternative conformations are possible and our MD simulation provides a good sampling of various loop conformations which are important in studying the interactions between G-quadruplex and ligand.

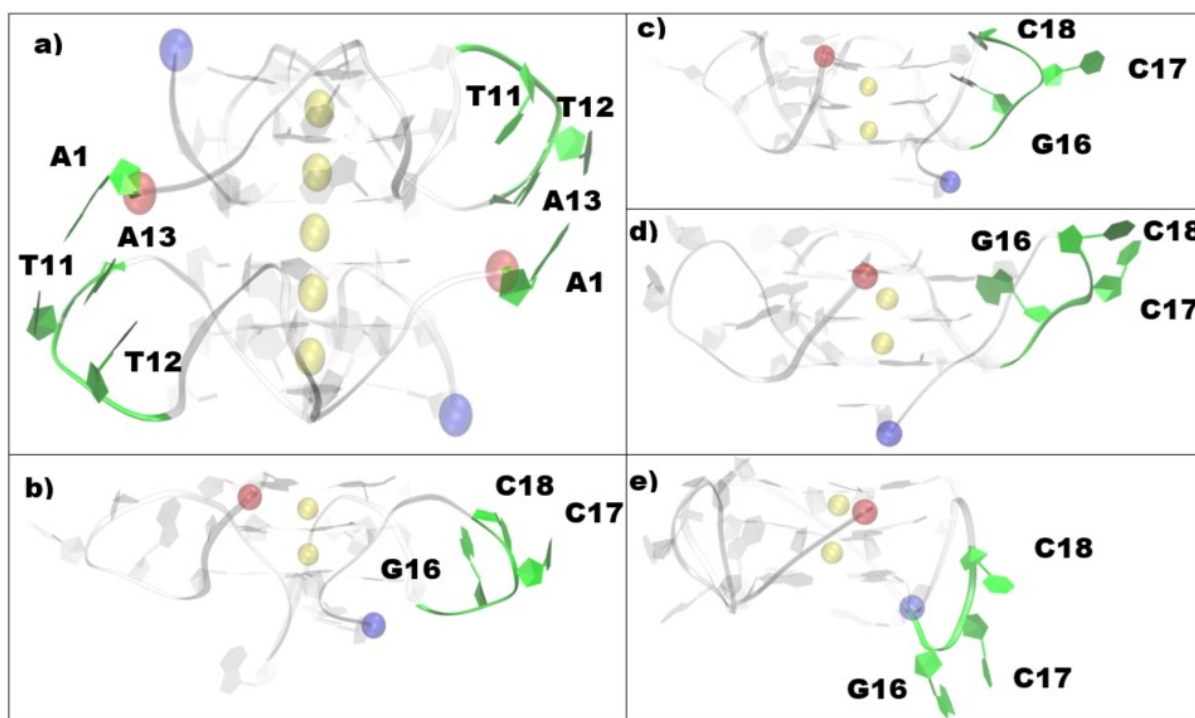


Figure 5. Various Loop Conformations (green). **a)** In the crystal structure of human telomeric quadruplex (pdb id: 1KF1), the TTA loops adopts a TAT stack conformation, which might be also stabilized by partial stacking between the loop A and the 5' A. The GCC loops of the $r(G_4C_2)_3G_4$ repeat adopt various conformations observed in the simulations: **b)** GCC stacking **c)** 1 C flipping out **d)** 2 C flipping out **e)** GCC all flipping out. 5' and 3' of the RNA chain are indicated by a red and blue ball, respectively.

Three TMPyP4 binding modes were observed in free ligand binding simulations for G-quadruplex RNA. Beginning from an unbound state (see the methods section), forty free ligand binding simulation runs (200 ns of each) for each G-quadruplex-ligand system were carried out. The convergence of the binding simulations was confirmed (see the methods section). At 200 ns, TMPyP4 was bound to the G-quadruplex: the ligand bound to the two ends of the RNA G-quadruplex in thirty-six runs and bound to the side of the G-quadruplex RNA in the remaining four runs (**Figure S6**). A plot was generated to show the position of TMPyP4, represented by one atom, in the 40 simulation runs which showed a good sampling of the G-quadruplex (**Figure S7**). Interestingly, due to the positive charge on each 1-Methylpiperidinium side chain of TMPyP4, we observed ligand binding within the first 100 ns of the simulation run, which is likely due to electrostatic steering. A figure was generated illustrating the electrostatically accelerated binding process of the three binding modes in the first 5 ns of the simulation (**Figure S8**). The stable complexes, extracted from the forty trajectories, were categorized into structural families based on a clustering analysis, as described in the methods section. By setting a threshold of 5% population, five structural families of complexes were identified in the RNA G-quadruplex (**Figure S9**). These five structural families for RNA were further merged into three binding modes: top stacking, bottom stacking and side binding (**Figure 6**). Top stacking accounted for 33% of the total population, bottom stacking accounted for 50% of the total population and side binding accounted for 14% of the total population. Two dimensional diagrams were

generated to show which residues play a role in each of the three binding poses (**Figures S16-18**). For example, G1, G13 and G19 are responsible for π - π stacking observed in the top stacking mode, while G15, G22 and G9 are responsible for the π - π stacking observed in the bottom stacking mode.

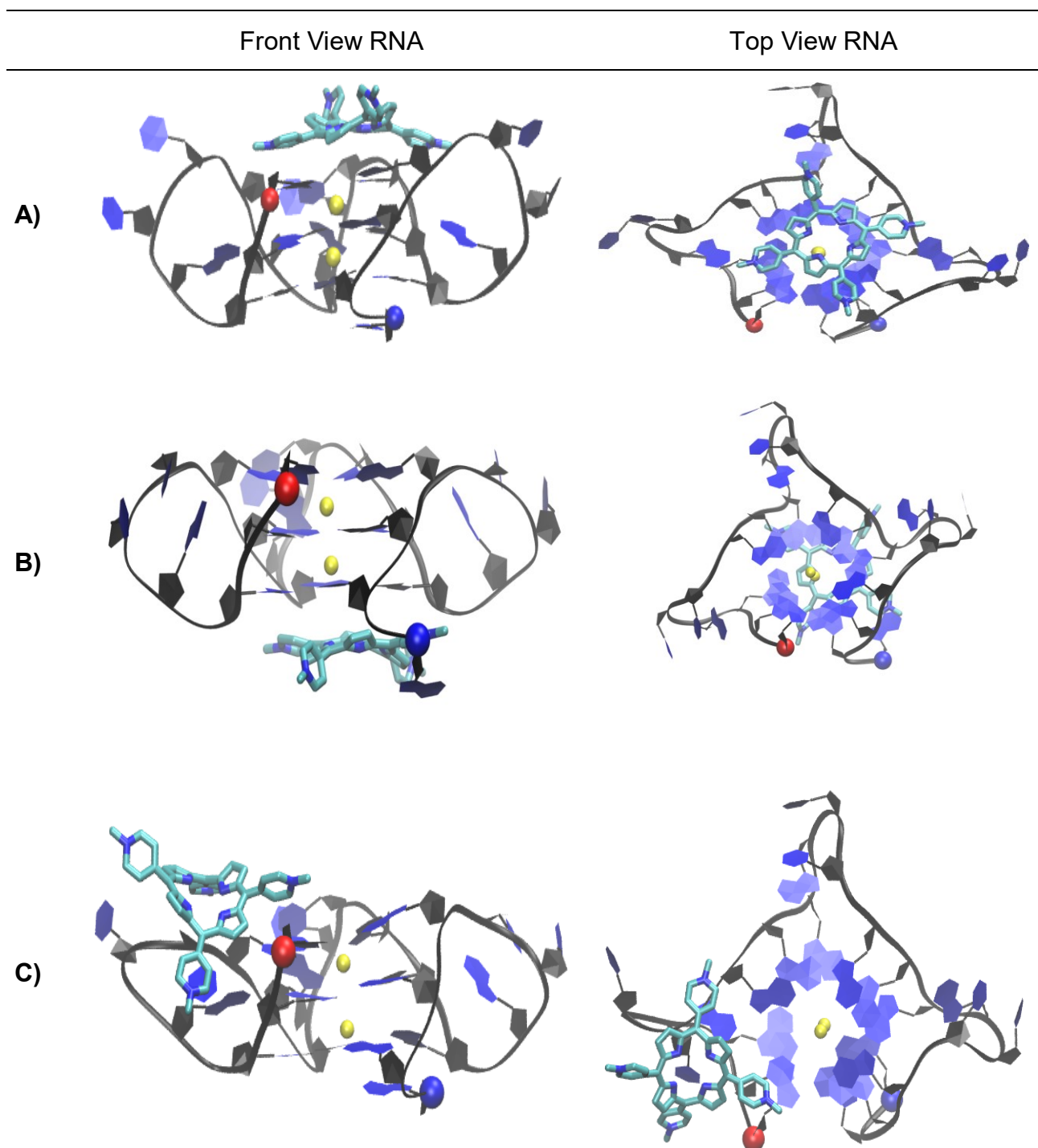


Figure 6. Top (A), Bottom (B), and Side (C) binding modes of TMPyP4 to RNA G-quadruplex. 5' and 3' of the nucleic acid chain are indicated by a red and blue ball, respectively.

MM-PBSA binding energy calculations determined that the bottom binding mode was most energetically favorable. To examine the relative stability for the major binding modes observed, MM-PBSA binding energy calculations were conducted for the three ligand binding modes (**Table 3**). These calculations show that the most favorable binding pose is bottom stacking (-28.4 ± 2.4 kcal/mol) which is followed by top stacking (-20.4 ± 3.1 kcal/mol) and side binding (-16.5 ± 0.5 kcal/mol). It is clear that the hydrophobic interactions play a role in the ability of TMPyP4 to stack to the ends of the quadruplex as seen when comparing bottom stacking (-34.1 ± 1.9) and top stacking (-23.7 ± 3.9 kcal/mol) modes to side binding (-12.2 ± 0.5 kcal/mol). The change in binding energy ($\Delta\Delta E_{TOT}$) between bottom stacking and top stacking (-8.0 kcal/mol) is less than between bottom stacking and side binding (-11.9 kcal/mol) which could be due to TMPyP4's stacking ability. As a result, hydrophobic interaction energy makes up the majority of the total MM-PBSA binding energy for all three binding poses.

Table 3. MMPBSA binding energy of TMPyP4 to G-quadruplex RNA

Unit (kcal/mol)	ΔE_{VDW}	ΔE_{SUR}	ΔE_{GBELE}	ΔE_{CONF}	ΔE_{TOT}	$\Delta\Delta E_{TOT}$
RNA Top	0.9 ± 8.9	-23.7 ± 3.9	2.8 ± 1.0	-0.3 ± 1.3	-20.4 ± 3.1	8.0
RNA Side	0.7 ± 0.9	-12.2 ± 0.5	-1.0 ± 0.5	-4.0 ± 1.5	-16.5 ± 0.5	11.9
RNA Bottom	2.2 ± 3.8	-34.1 ± 1.9	5.2 ± 3.1	-1.7 ± 4.8	-28.4 ± 2.4	0

Δ_{VDW} = Change of VDW energy in gas phase upon complex formation (Units: kcal/mol)

Δ_{SUR} = Change of energy due to surface area change upon complex formation (Units: kcal/mol)

Δ_{GBELE} = Change of GB reaction field energy + gas phase Elec. energy upon complex formation (Units: kcal/mol)

Δ_{CONF} = Conformation change (Units: kcal/mol)

$\Delta_{TOT} = \Delta_{VDW} + \Delta_{SUR} + \Delta_{GBELE} + \Delta_{CONF}$ Change of potential energy in water upon complex formation (Units: kcal/mol)

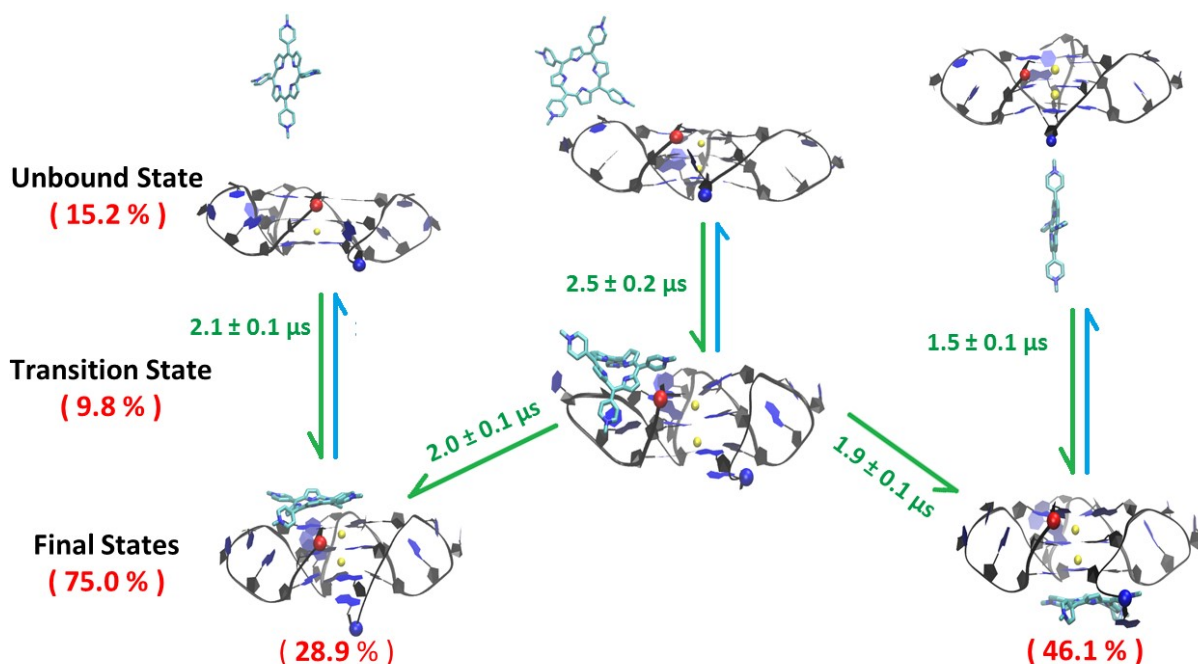


Figure 7. The mean first passage times between the four states (unbound, side transition, top, and bottom) of the TMPyP4-RNA G-quadruplex complex system.

Parallel binding pathways toward stable top and bottom binding states were observed for the TMPyP4-RNA G-quadruplex complex system. To decipher the kinetics pathways, Markov state model (MSM) was constructed from 40 binding trajectories using transition path theory as described in the method section. The implied time scales of each cluster for all lag times (**Figure S10**), Chapman-Kolmogorov test (**Figure S11**), and the network model (**Figure S12**) is available in the supporting document. Consistent with the thermodynamics analysis, there were three observed kinetic binding states: top, bottom, and side (groove). We can observe that the transition from unbound directly to the bottom/top binding state is slightly faster than the parallel pathway that involves the side transition state. The mean first passage times between the four states (unbound, side transition, top, and bottom) are shown in **Figure 7**. Forward transitions are indicated by green arrows while the reverse pathways are indicated by blue arrows. Arrows are shown for any edge that had at least 5 transitions in that direction. The transition time of unbound to bottom is the fastest of any pathway leading to a final binding state, while the transition time of unbound to top is only slightly slower. Transitioning from unbound to the transition state and then to a final binding state is approximately two-fold the time that it takes for a direct transition from unbound to a final binding state. The bottom and top binding poses are the final binding states which collectively make up approximately 75% of the simulation. The approximate interstate flux for unbound to bottom binding was 2:1, unbound to side binding was 1:1, unbound to top binding was 3:1, side binding to top binding or bottom binding were both unidirectional.

A representative trajectory was chosen for each binding pose to analyze the binding dynamics using order parameters. These order parameter calculations include: hydrogen bond analysis, center-to-center distance (D), drug-base dihedral angle, receptor and ligand RMSD and MM-PBSA binding energy (ΔE). The bottom stacking pose is the most favorable overall according to the MM-PBSA binding energy data. In the trajectory chosen for the top stacking mode (**Figure 8**) TMPyP4 makes initial contact with G-quadruplex RNA as early as 3 ns with minor repositions around 90 ns and again at 807 ns where it stayed bound for the remainder of the trajectory. The representative trajectory for the bottom pose (**Figure 9**) shows TMPyP4 making initial contact with the bottom of the G4 around 4 ns and maintaining this stable binding pose for the length of the trajectory. We also present a comparison for the backbone torsion angles of residue G22 from this trajectory (**Figure 10**). Additional trajectories for the top and bottom binding modes are provided in the supporting document (**Figure S19-S20**). In several trajectories we observed TMPyP4 binding to the side initially and ending in either a top (**Figure S21A-21B**) or bottom (**Figure S21C**) stacking mode. Several side binding trajectories were extended from 200 ns to 2000 ns to determine whether this binding pose is an intermediate step in the top or bottom stacking modes. The representative trajectory for the side binding pose (**Figure 11**) shows initial contact within 3 ns. With TMPyP4 repositioning to the top of the G4 around 1200 ns where it stays bound for the remainder of the trajectory. For each trajectory, once the final pose is achieved, very few changes in the position of the ligand are seen. These limited fluctuations are supported in the order calculations. Our findings suggest that side binding mode is an intermediate step leading to the more stable top or bottom stacking modes of TMPyP4 to the RNA quadruplex.

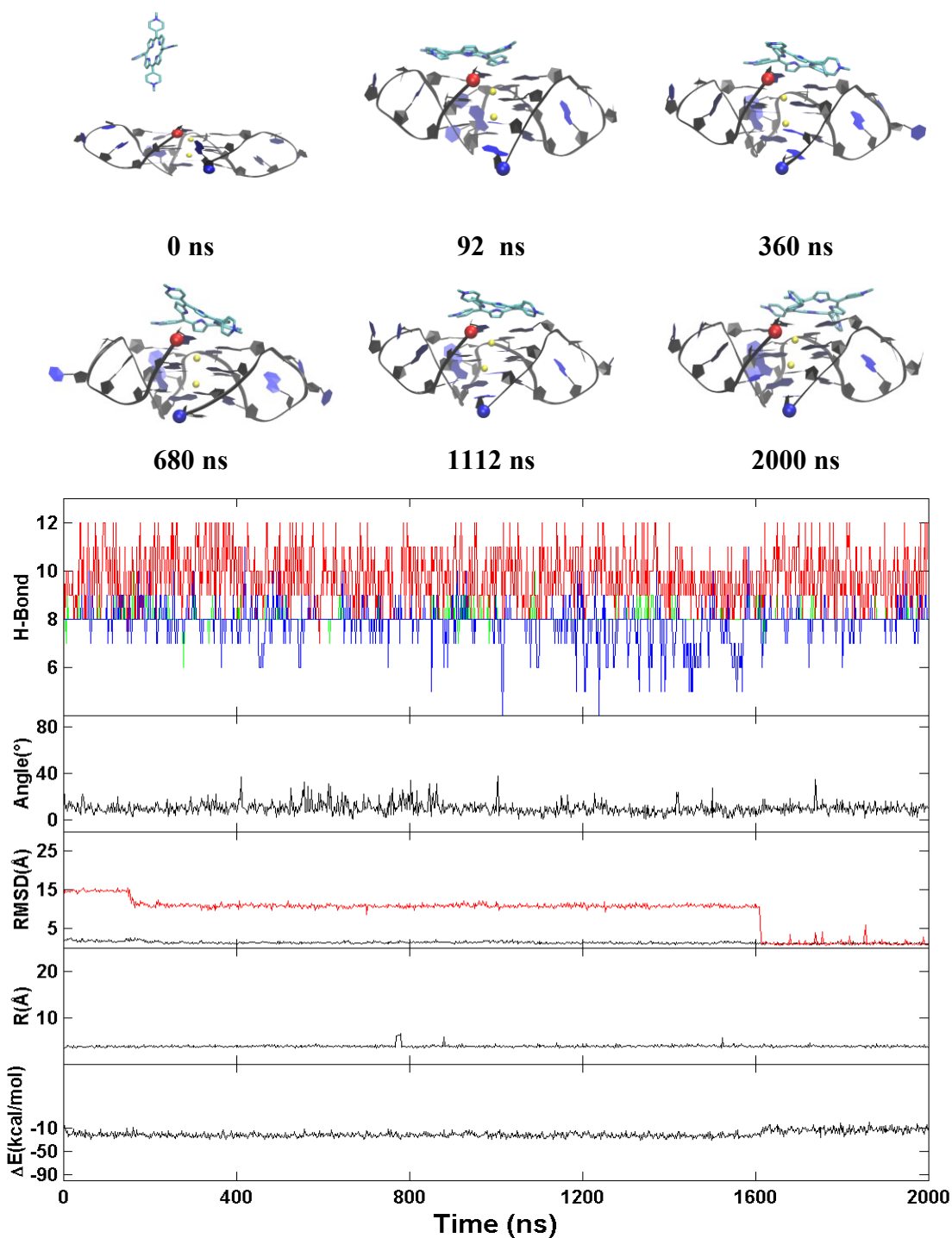


Figure 8. A representative trajectory of the top stacking mode of TMPyP4 to the RNA quadruplex (A): B) hydrogen bonds present in first (green), second (red), and third (blue), of G-tetrad layer of quadruplex (H-Bond), C) the drug-base dihedral angle, D) RMSD of ligand (red) and receptor (black), E) center-to-center distance (R), and F) MM-PBSA binding energy (ΔG). 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively.

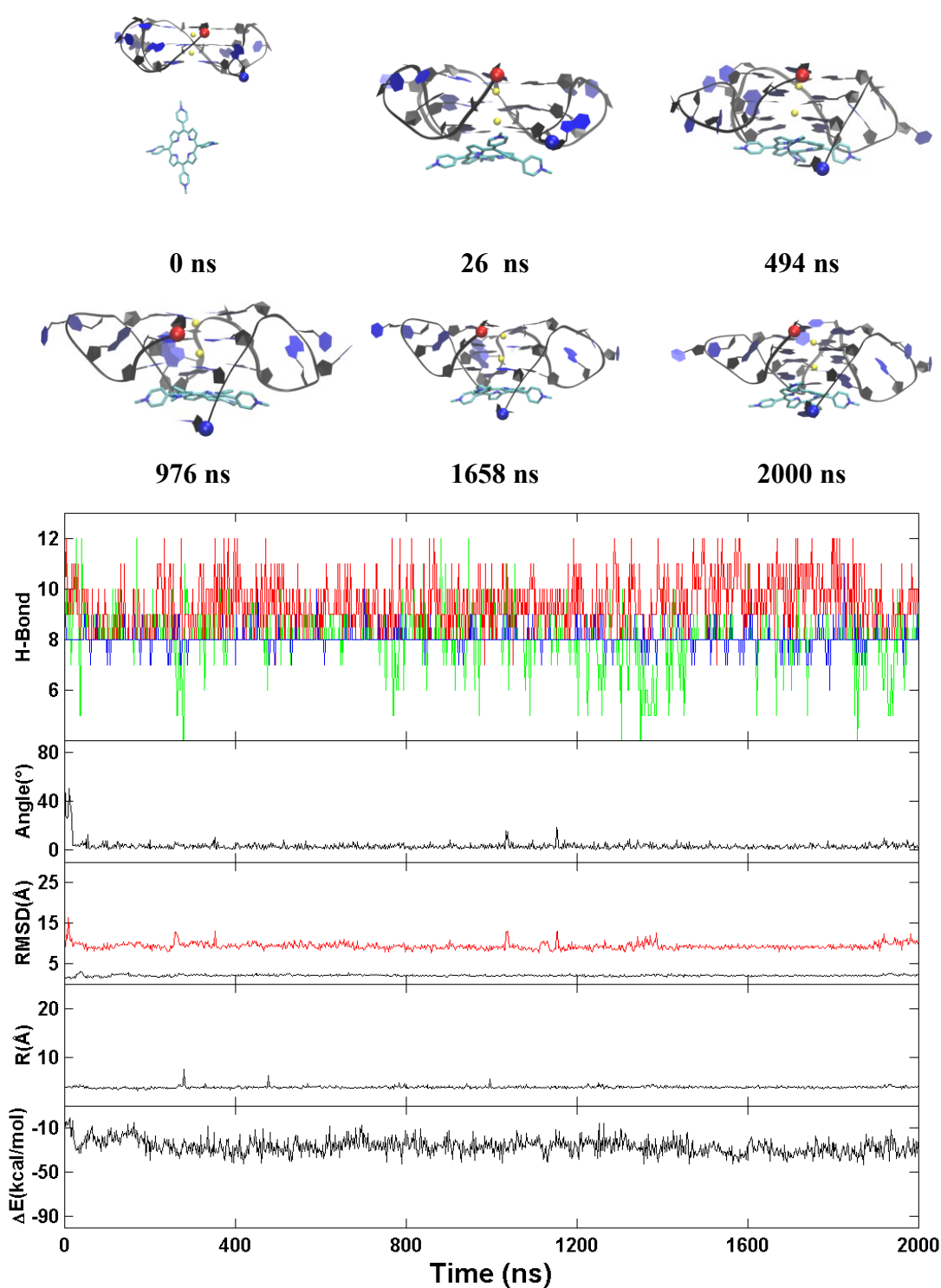


Figure 9. **A)** A representative trajectory of the bottom stacking mode of TMPyP4 to the RNA quadruplex **B)** hydrogen bonds present in first (green), second (red), and third (blue), of G-tetrad layer of quadruplex (H-Bond), **C)** the drug-base dihedral angle, **D)** RMSD of ligand (red) and receptor (black), **E)** center-to-center distance (R), and **F)** MM-PBSA binding energy (ΔG). 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively.

GGGGCCGGGGCCGGGGCCGGGG

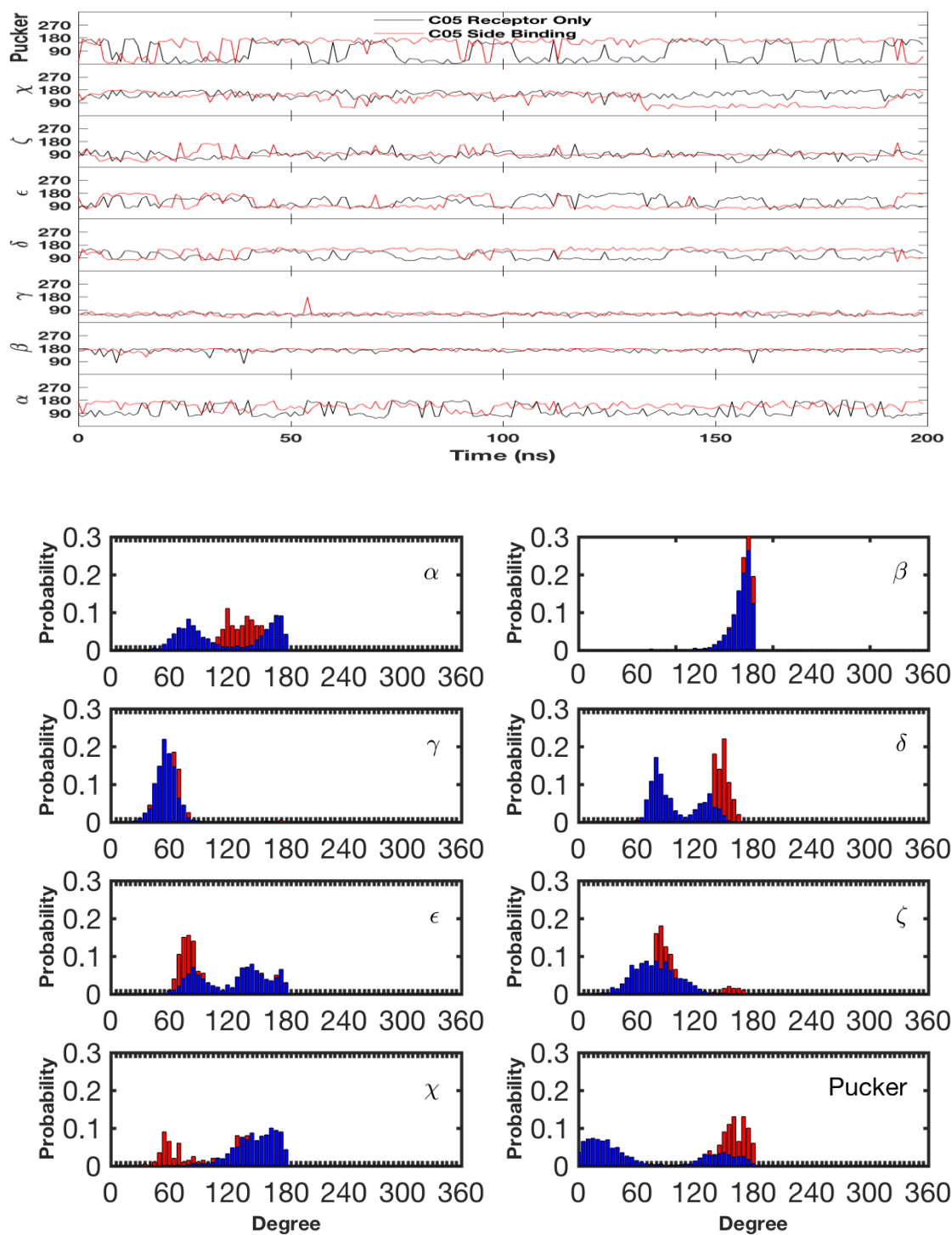


Figure 10. Comparison for the backbone torsion angles of residue G22 from the first loop between the free ligand binding simulation (red/red) of the side binding mode and the stability simulation of the crystal pose (black/blue) of the apo form.

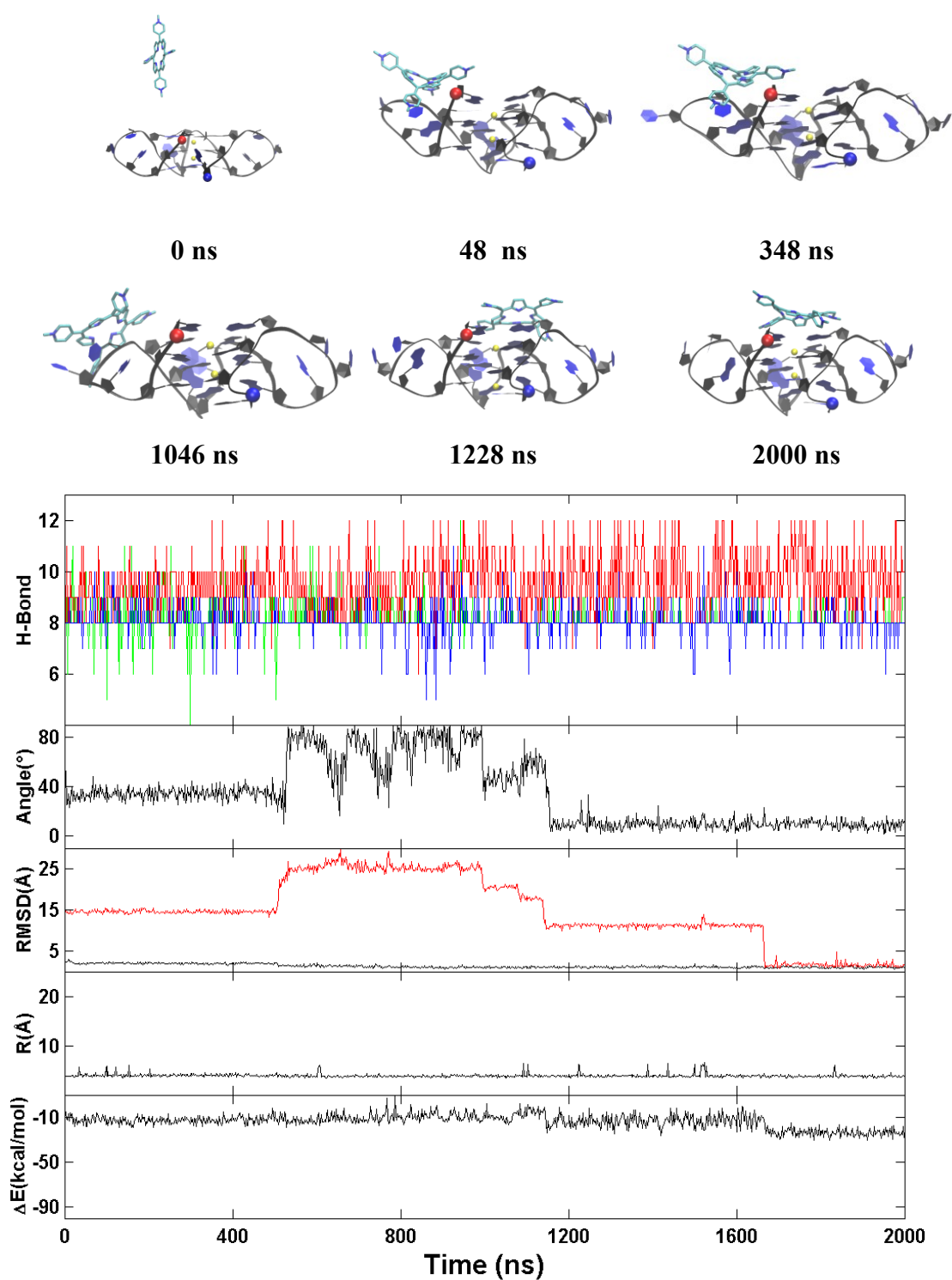


Figure 11. A representative trajectory of the side binding mode of TMPyP4 to the RNA quadruplex(A): **B)**hydrogen bonds present in first (green), second (red), and third (blue), of G-tetrad layer of quadruplex (H-Bond), **C)** the drug-base dihedral angle, **D)** RMSD of ligand (red) and receptor (black), **E)** center-to-center distance (R), and **F)** MM-PBSA binding energy (ΔG). 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively.

Backbone Dihedral Angles Provide Insight Into the Complexes. Backbone dihedral angles were not included in the previous order parameters, but are essential in elucidating the stability and conformational changes that occur in a complex. To further characterize the backbone of each complex, the standard backbone dihedral angles (α , β , γ , δ , ϵ , χ and pucker) (**Figure S24**) around the covalent bonds of the deoxyribose were calculated. Since we observed one potassium ion move out in three RNA apo form simulation runs we wanted to determine whether the potassium ion affects the stability of the quadruplex. To do this the conformational change induced by the potassium ion moving out of the quadruplex was compared to the potassium ion remaining in the quadruplex (**Figures S25-S45**). The top layer appears to show the most conformational change when the potassium ion moves out of the quadruplex (**Figures S25-S28**), however the loops and the middle and bottom layer are unaffected. To determine whether ligand binding has an effect on quadruplex stability, the conformational change induced by the ligand was calculated for the bottom binding mode. Histograms were generated to display the differences between the complex where TMPyP4 is in the bottom binding pose and the apo form. Conformational changes were only observed in residue G19 of the top layer, which is the only residue in this layer that makes contact with TmPyP4 at some point in the trajectory (**Figure S49**). Major conformational changes were observed in the middle layer due to TMPyP4 making initial contact with the side (**Figures S51-S53**), specifically at residues G8, G14 and G20. TMPyP4 only makes direct contact with residue G20 during its binding (**Figure S21C**), so it is interesting that both direct and allosteric effects are induced by the bottom stacking mode. No major changes were observed in the bottom layer.

Root Mean Square Fluctuation (RMSF) showed the RNA G-quadruplex had less overall fluctuations than the DNA G-quadruplex; binding of TMPyP4 further reduced the loop fluctuation. The RMSF plots (**Figure 12 A-B**) revealed the two apo forms both showed major fluctuations in the loop regions (G4-C6, G10-C12, G16-C18) (**Figure 12A**). However, the magnitude ($\text{\AA}$) of the first and third peaks were significantly lower in the RNA. Specifically, the DNA showed $\sim 4.5 \text{ \AA}$, $\sim 3.5 \text{ \AA}$, $\sim 4.5 \text{ \AA}$ peaks for the first, second, and third loop regions. Whereas the RNA showed $\sim 2.5 \text{ \AA}$, $\sim 3.5 \text{ \AA}$, $\sim 3.0 \text{ \AA}$ peaks for the first, second, and third loop regions. The overall smaller fluctuations suggest that the RNA structure is more stable, which may be attributed to the ability of the RNA to form additional hydrogen bonds (**Figure S15**). The RMSF plot comparing the bottom and top binding modes to the apo form shows that in both the top and bottom binding modes the major fluctuations were reduced by the binding of TMPyP4. This indicates the binding of TMPyP4 has a stabilizing effect on the loops of the RNA G-quadruplex (**Figure 15B**).

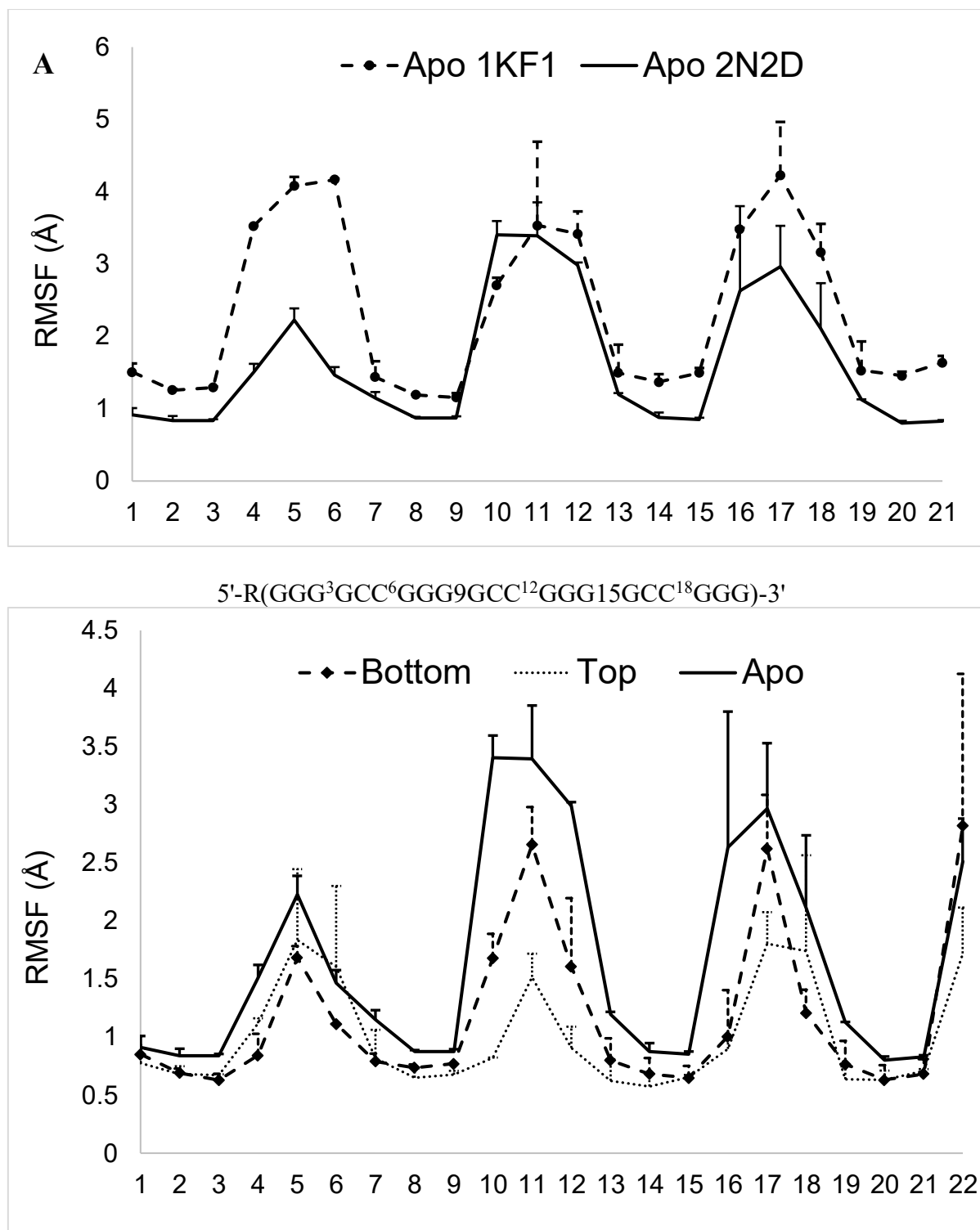


Figure 12. Root Mean Square Fluctuation. A) RMSF of the Apo form simulation runs. Data is aligned in reference to the RNA. B). RMSF of the top and bottom binding poses in comparison to the RNA apo form.

DISCUSSION

G-Quadruplex-forming repeat expansions HRE in the *C9orf72* gene is the root cause of two neurological diseases: familial ALS and FTD. Small-molecule targeting of the d(G4C2)_n or r(G4C2)_n

is a promising therapeutic strategy for treating these neurological diseases. TMPyP4, a well-known G-quadruplex ligand, is able to bind the $r(G_4C_2)_n$ repeat, diminishing its interactions with hnRNPA1 and SF2 proteins (5). Although previous studies stress the importance of using TMPyP4 to inhibit both DNA and RNA G-quadruplexes to treat ALS and FTD, a high-resolution structure of the RNA G-quadruplex does not yet exist.

G-quadruplex topologies have been accurately proposed by the identification of scaffold specific peaks in circular dichroism (CD) absorptivity experiments (18,45,47,48). Experimental evidence suggests that the DNA and RNA hexanucleotide repeats adopt different G-quadruplex scaffolds, antiparallel and parallel, respectively. In 2014, Haeusler *et al.* proposed the scaffold of the DNA G-quadruplex of $d(G_4C_2)_n$ using a CD absorptivity spectra which showed maximum and minimum absorbance characteristic of an anti-parallel scaffold (18). Their polyacrylamide gel electrophoresis (PAGE) shift assay further supported a conformational change into G-quadruplex form, showing that the formation of the DNA G-quadruplex was dependent on K^+ . In addition, their dimethyl sulphate (DMS) foot printing provided evidence of a four-stack anti-parallel G-quadruplex. Within the same study(18), Haeusler *et al.* also characterized the RNA using CD absorption spectroscopy which showed maximum and minimum absorbance characteristic of a parallel scaffold (18,45-47). Similar to DNA G-quadruplex, their PAGE assay highlighted the dependency of K^+ for RNA G-quadruplex formation. They also performed an RNase protection assay which showed every fourth guanine was susceptible to single stranded cleavage, supporting a three-stacked parallel stranded RNA G-quadruplex that contains a guanine and two cysteine's in the single stranded loop region. Two NMR solution structures of the DNA G-quadruplex were published (PDB ID: 2N2D and 5OPH) based the original proposal of the DNA G-quadruplex of $d(G_4C_2)_n$ (49,50). Each of the high-resolution structures have anti-parallel scaffolds with four G-quartets and three edgewise loops as proposed in the low-resolution model. The closely matching topologies of the two recently solved high-resolution structures of the anti-parallel $d(G_4C_2)_3G_4$ G-quadruplex with the proposal made by Haeusler *et al.* provide a motivation to build high resolution structure model of their proposed parallel $r(G_4C_2)_3G_4$ G-quadruplex based on the low resolution experimental data, unlike the antiparallel scaffold of the DNA, the high resolution structure of the parallel scaffold of the RNA repeat is not yet available.

In this study, a homology model pertaining to RNA G-quadruplex of $r(G_4C_2)_3G_4$ was developed from the crystal structure of a human telomeric parallel DNA G-quadruplex, including all necessary potassium ions. Adhering to the scaffold of the RNA repeat proposed by experimental studies (18), our homology model adopts a parallel G-quadruplex consisting of three tetrads with one guanine and two cytosine's in the loop region. Stability simulations for analysis of our homology model were generated for a total of 15 μs . An *in silico* analysis of the binding interactions between RNA G-quadruplex and TMPyP4 starting from an unbound state was conducted using 13.4 μs MD simulations and the MM-PBSA binding energy calculations and was compared to the experimentally determined, high-resolution structure.

A representative trajectory for each of the binding poses was chosen and the binding dynamics, as well as the six order parameters, were analyzed. To further characterize the backbone of each complex, the standard backbone dihedral angles (α , β , γ , δ , ϵ , χ ζ and pucker) were defined to characterize the conformational changes and stability (**Figures S25-S66**). A complex with the two potassium ions in the quadruplex (**Figures S25-S45**) was compared to a complex with one of the two potassium ions moved out in order to observe the role of this ion plays in the stability of the quadruplex. Our dihedral data shows that the move out of one potassium ion slightly changed the dihedral angle preferences for some bases, supporting the important role of potassium ions in maintaining a stable G-quadruplex. Additionally, the bottom binding mode was compared to the apo form of the quadruplex (**Figure S46-S66**). In this comparison, the drug binding was found to slightly change the dihedral distribution.

There are many factors that contribute to the overall stability of the RNA and DNA G-quadruplex structure. One integral feature is the coordination of O6 carbonyls by cations from within the ion pore. It has been reported that after the bipyramidal antiprism is formed as a result of cation stabilization, the O6 carbonyls maintain an average interquartet distance of 3.3 Å (51). We have defined some parameters to understand the position of the potassium cation relative to the surrounding G4 DNA (**Figure 4; Table S2-S3**). From our oxygen-potassium and oxygen-oxygen distance calculations we observed that each residue maintained 3.3 Å distance during the simulations. The similar mean and low standard deviations provide support that our simulation parameters were appropriately set despite one potassium cation moving out of the ion pore in 3 of the 15 apo form trajectories and 3 out 20 of binding trajectories.

Literature suggests that the overall helical structure was most accurately defined by the H-twist and H-rise parameters. In this study, we compared these parameters in the three RNA G-quadruplex systems to the DNA G-quadruplex system. For the DNA G-quadruplex we identified an average value of 3.49 Å, 1.88 Å, and 12.31 Å for the rise, H-rise and H-twist parameters respectively. For the RNA G-quadruplex systems we identified average values of 3.52 Å, Å, 2.26 and 17.13 Å for the rise, H-rise, and H-twist parameters, respectively. The closely comparable averages in each system provide qualitative support that the helical structure of the RNA G-quadruplex was maintained throughout the simulations.

The current state of MD simulations show that modeling the experimentally solved parallel DNA/RNA G-quadruplex propeller loops is most accurately done using the latest AMBER parmbsc0 force field (52) on the μ s timescale (53). Although tri-base stacking is a commonly observed propeller loop conformation for both DNA and RNA G-quadruplexes (32) there has been a number of loop conformations presented using NMR, X-ray crystallography and MD simulation which leaves reason for understanding a variety of possible loop conformations. The original crystal structure of a double stranded parallel DNA G-quadruplex published by Parkinson et al. (54) shows that in complex with TMPyP4, two of three propeller loop bases flip outward. Martadinata and Phan (55) published structures of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA

sequences in K⁺ solution where all three of the bases are flipped outward. Havrila et. al (56) studied the structural dynamics of RNA G-quadruplex propeller loops of and presented clustering results which highlighted the highly dynamic conformations of the propeller loops of the TERRA RNA G-quadruplex, whose structure has been used to rationalize ligand binding (33).

In the X-ray diffraction study of TMPyP4 bound to a parallel DNA G-quadruplex of human telomeric sequence (PDB ID: 2HRI), both top stacking and side binding were observed (**Figure S22 A-E**). In another NMR study of TMPyP4 bound to a parallel DNA G-quadruplex of human MYC promoter (PDB ID: 2A5R), only top stacking was observed (**Figure S22 F-G**). Using molecular dynamics binding simulations with a free ligand, we determined that there are three binding modes that TMPyP4 can form with the RNA G-quadruplex structure of d(G4C2)₃G₄. These poses include: top stacking (33% of the total population), bottom stacking (50% of the total population) and side binding (14% of the total population). Our top stacking mode is very similar to the mode determined experimentally (PDB ID: 2A5R and 2HRI), with a binding affinity of -20.4 kcal/mol. However, two major differences were observed when comparing the TMPyP4-RNA G-quadruplex complex system to the TMPyP-DNA G-quadruplex complex system (PDB ID: 2HRI and 2A5R) which are illustrated in the supporting document (**Figure S23**). First, neither published TMPyP4-DNA G-quadruplex complex structures show a bottom binding pose, which had the most favorable binding energy of the three poses (-28.4 kcal/mol) and was the most abundant conformation reported in our study. Second, the most abundant side binding mode shows clear differences in binding where it is able to stack on top of the G4 G5 C6 loop, whereas in the DNA system, one of the four 1-Methylpiperidinium side chains of TMPyP4 interacts with the T6 T7 A8 loop, as the rest of the structure remains perpendicular to the G4 core. Interestingly, the stability of the core G-tetrad containing three layers (G1-G7-G13-G19; G2-G8-G14-G20; and G3-G9-G15-G21) was maintained for each binding pose (**Figure S14**), consistent with the RNase T1 protection assay results that these guanine residues were protected and involved in the formation of the G-quadruplex core (18).

Literature suggests there is a divide as to whether TMPyP4 stabilized or destabilizes RNA G-quadruplexes. The work done by Zamiri et al. has shown that TMPyP4 can bind and distort the G-quadruplex formed by r(GGGGCC)_{n=2,5,8}, reducing the melting temperature from 95 °C to 67 °C. However, they noticed the melting was irreversible, likely caused by induction of multimeric G-quadruplexes (i.e. G-quadruplex formed by four nucleic acid molecules) rather than monomeric G-quadruplex. Their early work has also shown that r(GGGGCC)_{n=4,6,8} forms length-dependent unimolecular, multimolecular RNA G-quadruplexes and stable hairpins(46). Therefore, the exact molecular nature of the destabilization is not yet clear. Another work done by Morris et al. has also shown the destabilizing behavior of TMPyP4 on a RNA quadruplex(57). On the other hand, a number of other experimental studies, suggest a stabilization upon binding of TMPyP4 to RNA G-quadruplexes (58-61). Our work suggests a stabilization of Monomeric G-Quadruplex at an infinite low ligand concentration. Further study is required to clarify this issue.

In addition to identifying the binding modes of TMPyP4 to the RNA G-quadruplex, the detailed pathway analysis of each binding mode offers valuable information about the mechanisms for which the binding occurs. Clear from our results, the side binding mode showed a significantly less favorable binding energy when compared to the top and bottom stacking modes. To investigate this pose further, several side binding trajectories were extended from 200 ns to 2000 ns. From this we observed TMPyP4 binding to the side initially and ultimately ending at one of the two end stacking poses (**Figure 11** and **Figure S21**). Combining our MSM with the MM-PBSA binding energy analysis suggests that side binding is in fact a meta-stable intermediate step that proceeds the top or bottom stacking modes.

CONCLUSION

G-Quadruplex-forming repeat expansions in the *C9orf72* gene have been experimentally determined to directly contribute to the pathogenesis of familial ALS and FTD. Through the inhibitory mechanisms of select G-quadruplex-specific small molecules, such as TMPyP4, these toxic RNA G-quadruplex structures of the hexanucleotide repeat can be modulated to inhibit RNA foci formation and RAN translation toward potential treatment for these neurodegenerative diseases. Two high resolution parallel DNA G-quadruplex structures of human MYC promoter and human telomeric sequence in complex with TMPyP4 already exist. However, neither the RNA G-quadruplex structure of the hexanucleotide repeat nor its complex with TMPyP4 exist. This study utilizes homology modelling to carefully construct the first RNA G-quadruplex structure of the hexanucleotide repeat from the experimentally derived topology and a crystal structure of the parallel DNA G-quadruplex (PDB ID: 1KF1). Molecular dynamics stability simulations were carried out to test the stability of the homology model and the structure was characterized thoroughly. A free ligand binding simulation was conducted to examine the elusive binding behaviour of TMPyP4 to this RNA G-quadruplex. Our results indicate that TMPyP4 binds to the RNA G-quadruplex similar to the two experimental TMPyP4-Quadruplex DNA complex in three binding modes (top, bottom and side binding). Interestingly, our kinetics analysis suggests that the side mode appeared to be a meta-stable intermediate step that proceeds the top or bottom stacking modes. This information may assist future development of G-quadruplex-specific ligands in the treatment of neurodegenerative diseases such as ALS and FTD.

MATERIAL AND METHODS

A full description on the methods used in this study is provided in the Supporting Information. In brief, we constructed four systems (**Table 1**): A DNA only system (PDB id: 1KF1), an RNA only system (r(G4C2)₃G₄) built from a homology model using the X-ray solved DNA G-quadruplex structure (PDB id: 1KF1) as a template, a ligand only system, and an unbound RNA-ligand system using the RNA G-quadruplex structure and a TMPyP4 molecule that was 10 Å away from the RNA (**Figure S1**). Each

system was solvated in a water box of truncated octahedron with 10 Å water buffer plus K⁺ as counter ions to neutralize the system and 0.1 M KCl. A refined OL15 version of the AMBER nucleic acid force field was used represent the DNA and RNA fragments (i.e. For DNA parm99bsc0(63) + χ_{OL4} (64)+ ϵ/ζ_{OL1} (65)+ β_{OL1} (66) updates; For RNA with parm99bsc0(63) + χ_{OL3} (67,68)), TIP3P model(69) was used represent water, and the K⁺ model by Cheatham group was used for the K⁺ ions (70). The partial charges for a TMPyP4 molecule were obtained using standard AMBER protocol: the electrostatic potential of the TMPyP4 molecule was obtained at the HF/6-31G* level after geometry optimization at the same level; the electrostatic potential using the RESP (Restrained ElectroStatic Potential) method determined the partial charges(71); and other force field parameters were taken from the AMBER GAFF2 force field(72). The TMPyP4 force field in Mol2 format can be found in the supporting document (**Figure S67**). These AMBER force fields are commonly used in nucleic acid simulations (73-81). The simulation runs (**Table 1**) were conducted using the AMBER 14 simulation package(82). The simulation parameters followed our early studies (43,44,83) which are briefly described here. Each unbound RNA-ligand system underwent an additional 1000 ps pre-run at 500K to ensure the position and orientation of the free ligand was randomized before a production run at 300K; during this pre-run, the receptors position remained fixed, followed by multiple independent runs with random initial velocities. A run at 300 K, included a short 1.0 ns molecular dynamics in the NPT ensemble mode (constant pressure and temperature) to equilibrate the system density and production dynamics in the equivalent NVT ensemble mode (constant volume and temperature). SHAKE was applied to constrain all bonds connecting hydrogen atoms, enabling a 2.0 fs time step in the simulations(84). The particle-mesh Ewald method was used to treat long-range electrostatic interactions under periodic boundary conditions (charge grid spacing of ~1.0 Å, the fourth order of the B-spline charge interpolation; and direct sum tolerance of 10⁻⁵)(85). The cutoff distance for short-range non-bonded interactions was 10 Å, with the long-range van der Waals interactions based on a uniform density approximation. To reduce the computation, non-bonded forces were calculated using a two-stage RESPA approach where the short range forces were updated every step and the long range forces were updated every two steps(86). Temperature was controlled using the Langevin thermostat with a coupling constant of 2.0 ps. The trajectories were saved at 50.0 ps intervals for analysis.

Featurization and Clustering. All 40 trajectories (201 ns each) were combined into one trajectory. Using VMD, all frames in which there were less than 20 interactions, at a distance less than 3Å, between the G-quadruplex and the ligand were separated as the unbound state (87). The trajectory was then superimposed based on the nucleic backbone using MDtraj and calculations for RMSD as well as center of mass of the ligand heavy atoms were performed (88). K-means clustering, performed using scikit-learn, was then used to classify the remaining frames into various states (89). Clustering was performed for K between 2 and 30 inclusively, using the silhouette index as the metric

for similarity of clusters (89,90). Analysis of the silhouette indices indicated that K between 2 to 6 inclusively were all likely to be distinctly clustered and further validation of the clustering determined that a K of 6 would give the best results.

The most representative frame for each cluster was determined by calculating the mean RMSD for each cluster and finding the frame with the least difference from the mean. Further validation of the clustering was performed by creating a trajectory for each of the clusters containing all of the frames in each cluster and visually confirming the similarity within each cluster. Through visual analysis of the cluster representative frames, clusters that were determined to be highly similar were combined. The unbound frames were then reintroduced as a single cluster.

Transition Path Theory: Count matrices were then created for lagtimes (τ) of 1, 10, 20, 30, ... 200 ns by counting the number of observed transitions between discrete states such that the count of transitions from state i to state j (c_{ij}) is the sum of the number of times each of the trajectories were observed in state i at time t and in state j at time $t + \tau$, for all $t \leq t_{max} - \tau$

(91). The count matrices were symmetrized (sym_{ij}) such that $sym_{ij} = sym_{ji} = \frac{c_{ij} + c_{ji}}{2}$

$sym_{ij} = sym_{ji} = \frac{c_{ij} + c_{ji}}{2}$ and then row-normalized ($norm_{ij}$) such that $norm_{ij} = \frac{sym_{ij}}{\sum_{j=1}^n sym_{ij}}$

$norm_{ij} = \frac{sym_{ij}}{\sum_{j=1}^n sym_{ij}}$. For the purpose of determining the lag time at which the model has

converged, the implied timescale of each cluster was calculated for all lag times and plotted (**Figure S10**). The implied timescale of the first cluster is not included in the plot as the eigenvalue is always 1 and thus contributes no information (92). Further validation that the model had been converged was performed through the Chapman-Kolmogorov test (**Figure S11**)(91). Network models (**Figure S12**) were then generated based on the count matrices at an optimal lag time of 100ns with the cutoff for a directed edge in the network being set at 5 transitions (93). Edges in the network diagram are drawn if there were at least 5 transitions (in any single direction) between two nodes. Outgoing transitions are labelled on the edges as the same color as the originating node. The system nodes are labelled 1 for unbound, 2 for side binding, 3 for bottom binding, and 4 for top binding. Thereafter, the mean first passage times (F_{if}) at the optimal lag time and the standard deviations within 20 ns of the optimal lag time were calculated according to the formula $F_{if} = \tau + \sum_{j \neq f} P_{ij} F_{jf} F_{if} = \tau + \sum_{j \neq f} P_{ij} F_{jf}$, with

the boundary condition $F_{ff} = 0F_{ff} = 0$, where τ is the lag time used to construct the transition matrix $P(\tau)$. A summary plot was generated showing the mean first passage times between the four states (unbound, side transition, top, and bottom) of the system (**Figure 7**).

The expected convergence time of the implied timescales should be significantly greater than that of a model with a greater number of clusters since we choose to cluster into a handful of “macrostates” directly and skip over the experimentally unverifiable thousand “microstates”. This provides a coarse grained model that traded finer detail for greater experimental testability and easier human understanding (62). Directly clustering into “macrostates” is thought to still maintain the integrity of the MSM as verification through the Chapman-Kolmogorov test (**Figure S11**) which indicates that the model closely resembles the observed simulation data.

AUTHOR INFORMATION

Conceptualization, C.W.; Formal Analysis, K.M., H.S. and J.G.; Writing – Original Draft Preparation, K.M., H.S., J.G., J.C., and B.C.; Writing – Review & Editing, C.W.

FUNDING SOURCES

This work was supported by Rowan CSM SEED grant and the National Science Foundation under Grants NSF ACI-1429467/RUI-1904797 and XSEDE MCB 170088.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

SUPPLEMENTARY DATA

The Supporting Information is available free of charge on the ACS Publications website.

The detailed structural and dynamic information of the simulated systems.

The AMBER GAFF2 force field of TMPyP4 in Mol2 format

The movies files of the parallel RNA G-quadruplex r(G4C2)₃G₄ and its complex with TMPyP4.

REFERENCES

1. Li, M.H., Zhou, Y.H., Luo, Q. and Li, Z.S. (2010) The 3D structures of G-Quadruplexes of HIV-1 integrase inhibitors: molecular dynamics simulations in aqueous solution and in the gas phase. *Journal of Molecular Modeling*, **16**, 645-657.
2. Tauchi, T., Shin-ya, K., Sashida, G., Sumi, M., Nakajima, A., Shimamoto, T., Ohyashiki, J.H. and Ohyashiki, K. (2003) Activity of a novel G-quadruplex-interactive telomerase inhibitor, telomestatin (SOT-095), against human leukemia cells: involvement of ATM-dependent DNA damage response pathways. *Oncogene*, **22**, 5338-5347.
3. Elmore, L.W., Rehder, C.W., Di, X., McChesney, P.A., Jackson-Cook, C.K., Gewirtz, D.A. and Holt, S.E. (2002) Adriamycin-induced senescence in breast tumor cells involves functional p53 and telomere dysfunction. *Journal of Biological Chemistry*, **277**, 35509-35515.
4. Read, M.A., Wood, A.A., Harrison, J.R., Gowan, S.M., Kelland, L.R., Dosanjh, H.S. and Neidle, S. (1999) Molecular Modeling Studies on G-Quadruplex Complexes of Telomerase Inhibitors: Structure-Activity Relationships. *Journal of Medicinal Chemistry*, **42**, 4538-4546.
5. Zamiri, B., Reddy, K., Macgregor, R.B., Jr. and Pearson, C.E. (2014) TMPyP4 Porphyrin Distorts RNA G-Quadruplex Structures of the Disease-Associated r(GGGGCC)_n Repeat of the C9orf72 Gene and Blocks Interaction of RNA-Binding Proteins. *J. Biolol. Chem.*, **289**, 4653-4659.
6. Husby, J., Todd, A.K., Platts, J.A. and Neidle, S. (2013) Small-molecule G-quadruplex interactions: Systematic exploration of conformational space using multiple molecular dynamics. *Biopolymers*, **99**, 989-1005.
7. Shay, J.W. and Wright, W.E. (2006) Telomerase therapeutics for cancer: challenges and new directions. *Nature reviews. Drug discovery*, **5**, 577-584.
8. Qin, Y., Fortin, J.S., Tye, D., Gleason-Guzman, M., Brooks, T.A. and Hurley, L.H. (2010) Molecular cloning of the human PDGFR- β promoter and drug targeting of the G-quadruplex-forming region to repress PDGFR- β expression. *Biochemistry*, **49**, 4208-4219.
9. Incles, C.M., Schultes, C.M., Kempiski, H., Koehler, H., Kelland, L.R. and Neidle, S. (2004) A G-quadruplex telomere targeting agent produces p16-associated senescence and chromosomal fusions in human prostate cancer cells. *Molecular Cancer Therapeutics*, **3**, 1201.
10. Riou, J.F., Guittat, L., Mailliet, P., Laoui, A., Renou, E., Petitgenet, O., Mégnin-Chanet, F., Hélène, C. and Mergny, J.L. (2002) Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 2672-2677.
11. Taka, T., Huang, L., Wongnoppavich, A., Tam-Chang, S.-W., Lee, T.R. and Tuntiwechapikul, W. (2013) Telomere shortening and cell senescence induced by perylene derivatives in A549 human lung cancer cells. *Bioorganic & Medicinal Chemistry*, **21**, 883-890.
12. Burger, A.M., Dai, F., Schultes, C.M., Reszka, A.P., Moore, M.J., Double, J.A. and Neidle, S. (2005) The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Research*, **65**, 1489.
13. Chen, Z.-F., Qin, Q.-P., Qin, J.-L., Zhou, J., Li, Y.-L., Li, N., Liu, Y.-C. and Liang, H. (2015) Water-Soluble Ruthenium(II) Complexes with Chiral 4-(2,3-Dihydroxypropyl)-formamide Oxoaporphine (FOA): In Vitro and in Vivo Anticancer Activity by Stabilization of G-Quadruplex DNA, Inhibition of Telomerase Activity, and Induction of Tumor Cell Apoptosis. *Journal of Medicinal Chemistry*, **58**, 4771-4789.
14. Simone, R., Fratta, P., Neidle, S., Parkinson, G.N. and Isaacs, A.M. (2015) G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. *Febs Letters*, **589**, 1653-1668.
15. Gendron, T.F., Belzil, V.V., Zhang, Y.J. and Petrucelli, L. (2014) Mechanisms of Toxicity in C9FTLD/ALS. *Acta neuropathologica*, **127**, 359-376.

16. Rutherford, N.J., Heckman, M.G., DeJesus-Hernandez, M., Baker, M.C., Soto-Ortolaza, A.I., Rayaprolu, S., Stewart, H., Finger, E., Volkening, K., Seeley, W.W. *et al.* (2012) Length of normal alleles of C9ORF72 GGGGCC repeat do not influence disease phenotype. *Neurobiology of Aging*, **33**.
17. DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J. *et al.* (2011) Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron*, **72**, 245-256.
18. Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A.J., Shaw, P.G., Kim, M.S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R. *et al.* (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, **507**, 195-+.
19. Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L. *et al.* (2011) A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron*, **72**, 257-268.
20. Gijssels, I., Van Langenhove, T., van der Zee, J., Sleegers, K., Philtjens, S., Kleinberger, G., Janssens, J., Bettens, K., Van Cauwenberghe, C., Pereson, S. *et al.* (2012) A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. *Lancet Neurology*, **11**, 54-65.
21. Renoux, A.J. and Todd, P.K. (2012) Neurodegeneration the RNA way. *Progress in Neurobiology*, **97**, 173-189.
22. Xu, Z.H., Poidevin, M., Li, X.K., Li, Y.J., Shu, L.Q., Nelson, D.L., Li, H., Hales, C.M., Gearing, M., Wingo, T.S. *et al.* (2013) Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 7778-7783.
23. Mori, K., Weng, S.M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H.A., Cruts, M., Van Broeckhoven, C. *et al.* (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science (New York, N.Y.)*, **339**, 1335-1338.
24. Ash, P.E., Bieniek, K.F., Gendron, T.F., Caulfield, T., Lin, W.L., DeJesus-Hernandez, M., van Blitterswijk, M.M., Jansen-West, K., Paul, J.W., 3rd, Rademakers, R. *et al.* (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron*, **77**, 639-646.
25. Su, Z., Zhang, Y., Gendron, T.F., Bauer, P.O., Chew, J., Yang, W.Y., Fostvedt, E., Jansen-West, K., Belzil, V.V., Desaro, P. *et al.* (2014) Discovery of a Biomarker and Lead Small Molecules to Target r(GGGGCC)-Associated Defects in c9FTD/ALS. *Neuron*, **83**, 1043-1050.
26. Simone, R., Balendra, R., Moens, T.G., Preza, E., Wilson, K.M., Heslegrave, A., Woodling, N.S., Niccoli, T., Gilbert-Jaramillo, J., Abdelkarim, S. *et al.* (2018) G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. *EMBO molecular medicine*, **10**, 22-31.
27. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A*, **99**, 11593-11598.
28. Kim, M.Y., Gleason-Guzman, M., Izbicka, E., Nishioka, D. and Hurley, L.H. (2003) The different biological effects of telomestatin and TMPyP4 can be attributed to their selectivity for interaction with intramolecular or intermolecular G-quadruplex structures. *Cancer Research*, **63**, 3247-3256.
29. Mikami-Terao, Y., Akiyama, M., Yuza, Y., Yanagisawa, T., Yamada, O., Kawano, T., Agawa, M., Ida, H. and Yamada, H. (2009) Antitumor activity of TMPyP4 interacting G-quadruplex in retinoblastoma cell lines. *Experimental Eye Research*, **89**, 200-208.
30. Mikami-Terao, Y., Akiyama, M., Yuza, Y., Yanagisawa, T., Yamada, O. and Yamada, H. (2008) Antitumor activity of G-quadruplex-interactive agent TMPyP4 in K562 leukemic cells. *Cancer Letters*, **261**, 226-234.

31. Brcic, J. and Plavec, J. (2015) Solution structure of a DNA quadruplex containing ALS and FTD related GGGGCC repeat stabilized by 8-bromodeoxyguanosine substitution. *Nucleic Acids Res.*, **43**, 8590-8600.
32. Collie, G.W., Haider, S.M., Neidle, S. and Parkinson, G.N. (2010) A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res*, **38**, 5569-5580.
33. Collie, G., Reszka, A.P., Haider, S.M., Gabelica, V., Parkinson, G.N. and Neidle, S. (2009) Selectivity in small molecule binding to human telomeric RNA and DNA quadruplexes. *Chemical Communications*, 7482-7484.
34. Hou, J.Q., Chen, S.B., Tan, J.H., Luo, H.B., Li, D., Gu, L.Q. and Huang, Z.S. (2012) New insights from molecular dynamic simulation studies of the multiple binding modes of a ligand with G-quadruplex DNA. *Journal of Computer-Aided Molecular Design*, **26**, 1355-1368.
35. Islam, B., Stadlbauer, P., Krepl, M., Koca, J., Neidle, S., Haider, S. and Sponer, J. (2015) Extended molecular dynamics of a c-kit promoter quadruplex. *Nucleic acids research*, **43**, 8673-8693.
36. Deng, N., Wickstrom, L., Cieplak, P., Lin, C. and Yang, D. (2017) Resolving the Ligand-Binding Specificity in c-MYC G-Quadruplex DNA: Absolute Binding Free Energy Calculations and SPR Experiment. *The Journal of Physical Chemistry B*, **121**, 10484-10497.
37. Deng, N., Xia, J., Wickstrom, L., Lin, C., Wang, K., He, P., Yin, Y. and Yang, D. (2019) Ligand Selectivity in the Recognition of Protoberberine Alkaloids by Hybrid-2 Human Telomeric G-Quadruplex: Binding Free Energy Calculation, Fluorescence Binding, and NMR Experiments. *Molecules*, **24**.
38. Luo, D. and Mu, Y. (2015) All-Atomic Simulations on Human Telomeric G-Quadruplex DNA Binding with Thioflavin T. *The Journal of Physical Chemistry B*, **119**, 4955-4967.
39. Ali, A., Bansal, M. and Bhattacharya, S. (2015) Ligand 5,10,15,20-Tetra(N-methyl-4-pyridyl)porphine (TMPyP4) Prefers the Parallel Propeller-Type Human Telomeric G-Quadruplex DNA over Its Other Polymorphs. *The Journal of Physical Chemistry B*, **119**, 5-14.
40. Sullivan, H.-J., Readmond, C., Radicella, C., Persad, V., Fasano, T.J. and Wu, C. (2018) Binding of Telomestatin, TMPyP4, BSU6037, and BRACO19 to a Telomeric G-Quadruplex-Duplex Hybrid Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent. *ACS Omega*, **3**, 14788-14806.
41. Machireddy, B., Kalra, G., Jonnalagadda, S., Ramanujachary, K. and Wu, C. (2017) Probing the Binding Pathway of BRACO19 to a Parallel-Stranded Human Telomeric G-Quadruplex Using Molecular Dynamics Binding Simulation with AMBER DNA OL15 and Ligand GAFF2 Force Fields. *Journal of Chemical Information and Modeling*, **57**, 2846-2864.
42. Mulholland, K. and Wu, C. (2016) Binding of Telomestatin to a Telomeric G-Quadruplex DNA Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent. *Journal of Chemical Information and Modeling*, **56**, 2093-2102.
43. Mulholland, K., Siddiquei, F. and Wu, C. (2017) Binding modes and pathway of RHPS4 to human telomeric G-quadruplex and duplex DNA probed by all-atom molecular dynamics simulations with explicit solvent. *Physical Chemistry Chemical Physics*, **19**, 18685-18694.
44. Mulholland, K. and Wu, C. (2016) Binding of Telomestatin to a Telomeric G-Quadruplex DNA Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent. *J. Chem. Inf. Model*, **56**, 2093-2102.
45. Fratta, P., Mizielińska, S., Nicoll, A.J., Zloh, M., Fisher, E.M.C., Parkinson, G. and Isaacs, A.M. (2012) C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Scientific reports*, **2**.
46. Reddy, K., Zamiri, B., Stanley, S.Y.R., Macgregor, R.B. and Pearson, C.E. (2013) The Disease-associated r(GGGGCC)(n) Repeat from the C9orf72 Gene Forms Tract Length-dependent Uni- and Multimolecular RNA G-quadruplex Structures. *Journal of Biological Chemistry*, **288**, 9860-9866.
47. Kumari, S., Bugaut, A., Huppert, J.L. and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nature Chemical Biology*, **3**, 218-221.

48. Wieland, M. and Hartig, J.S. (2007) RNA Quadruplex-Based Modulation of Gene Expression. *Chemistry & Biology*, **14**, 757-763.
49. Brčić, J. and Plavec, J. (2015) Solution structure of a DNA quadruplex containing ALS and FTD related GGGGCC repeat stabilized by 8-bromodeoxyguanosine substitution. *Nucleic Acids Research*, **43**, 8590-8600.
50. Brčić, J. and Plavec, J. (2018) NMR structure of a G-quadruplex formed by four d(G4C2) repeats: insights into structural polymorphism. *Nucleic Acids Research*, gky886-gky886.
51. Reshetnikov, R.V., Kopylov, A.M. and Golovin, A.V. (2010) Classification of g-quadruplex DNA on the basis of the quadruplex twist angle and planarity of g-quartets. *Acta naturae*, **2**, 72-81.
52. Fadrna, E., Spackova, N., Sarzynska, J., Koca, J., Orozco, M., Cheatham, T.E., 3rd, Kulinski, T. and Sponer, J. (2009) Single Stranded Loops of Quadruplex DNA As Key Benchmark for Testing Nucleic Acids Force Fields. *Journal of chemical theory and computation*, **5**, 2514-2530.
53. Islam, B., Stadlbauer, P., Gil-Ley, A., Pérez-Hernández, G., Haider, S., Neidle, S., Bussi, G., Banas, P., Otyepka, M. and Sponer, J. (2017) Exploring the Dynamics of Propeller Loops in Human Telomeric DNA Quadruplexes Using Atomistic Simulations. *Journal of chemical theory and computation*, **13**, 2458-2480.
54. Parkinson, G.N., Ghosh, R. and Neidle, S. (2007) Structural basis for binding of porphyrin to human telomeres. *Biochemistry*, **46**, 2390-2397.
55. Martadinata, H. and Phan, A.T. (2009) Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K⁺ solution. *Journal of the American Chemical Society*, **131**, 2570-2578.
56. Havrila, M., Stadlbauer, P., Kührová, P., Banáš, P., Mergny, J.-L., Otyepka, M. and Šponer, J. (2018) Structural dynamics of propeller loop: towards folding of RNA G-quadruplex. *Nucleic Acids Res*, **46**, 8754-8771.
57. Morris, M.J., Wingate, K.L., Silwal, J., Leeper, T.C. and Basu, S. (2012) The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Res*, **40**, 4137-4145.
58. Wang, S.-R., Zhang, Q.-Y., Wang, J.-Q., Ge, X.-Y., Song, Y.-Y., Wang, Y.-F., Li, X.-D., Fu, B.-S., Xu, G.-H., Shu, B. *et al.* (2016) Chemical Targeting of a G-Quadruplex RNA in the Ebola Virus L Gene. *Cell Chemical Biology*, **23**, 1113-1122.
59. Ou, T.M., Lu, Y.J., Tan, J.H., Huang, Z.S., Wong, K.Y. and Gu, L.Q. (2008) G-quadruplexes: targets in anticancer drug design. *ChemMedChem*, **3**, 690-713.
60. Perrone, R., Nadai, M., Poe, J.A., Frasson, I., Palumbo, M., Palu, G., Smithgall, T.E. and Richter, S.N. (2013) Formation of a unique cluster of G-quadruplex structures in the HIV-1 Nef coding region: implications for antiviral activity. *PloS one*, **8**, e73121.
61. Wang, S.R., Min, Y.Q., Wang, J.Q., Liu, C.X., Fu, B.S., Wu, F., Wu, L.Y., Qiao, Z.X., Song, Y.Y., Xu, G.H. *et al.* (2016) A highly conserved G-rich consensus sequence in hepatitis C virus core gene represents a new anti-hepatitis C target. *Science advances*, **2**, e1501535.
62. Pande, V.S., Beauchamp, K. and Bowman, G.R. (2010) Everything you wanted to know about Markov State Models but were afraid to ask. *Methods*, **52**, 99-105.
63. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E., Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.*, **92**, 3817-3829.
64. Krepl, M., Zgarbova, M., Stadlbauer, P., Otyepka, M., Banas, P., Koca, J., Cheatham, T.E., Jurecka, P. and Sponer, J. (2012) Reference Simulations of Noncanonical Nucleic Acids with Different chi Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *Journal of chemical theory and computation*, **8**, 2506-2520.
65. Zgarbova, M., Luque, F.J., Sponer, J., Cheatham, T.E., Otyepka, M. and Jurecka, P. (2013) Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *Journal of chemical theory and computation*, **9**, 2339-2354.

66. Zgarbova, M., Sponer, J., Otyepka, M., Cheatham, T.E., Galindo-Murillo, R. and Jurecka, P. (2015) Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *Journal of chemical theory and computation*, **11**, 5723-5736.
67. Banas, P., Hollas, D., Zgarbova, M., Jurecka, P., Orozco, M., Cheatham, T.E., Sponer, J. and Otyepka, M. (2010) Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *Journal of chemical theory and computation*, **6**, 3836-3849.
68. Zgarbova, M., Otyepka, M., Sponer, J., Mladek, A., Banas, P., Cheatham, T.E. and Jurecka, P. (2011) Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *Journal of chemical theory and computation*, **7**, 2886-2902.
69. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparisons of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926-935.
70. Joung, I.S. and Cheatham, T.E. (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *Journal of Physical Chemistry B*, **112**, 9020-9041.
71. Bayly, C.I., Cieplak, P., Cornell, W.D. and Kollman, P.A. (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem.*, **97**, 10269-10280.
72. D.A. Case, D.S.C., T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman. (2016), University of California, San Francisco.
73. Lavery, R., Zakrzewska, K., Beveridge, D., Bishop, T.C., Case, D.A., Cheatham, T.E.I., Dixit, S., Jayaram, B., Lankas, F., Laughton, C. *et al.* (2010) A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA *Nucleic Acids Res.*, **38**, 299-313.
74. Cosconati, S., Marinelli, L., Trotta, R., Virno, A., De Tito, S., Romagnoli, R., Pagano, B., Limongelli, V., Giancola, C., Baraldi, P.G. *et al.* (2010) Structural and Conformational Requisites in DNA Quadruplex Groove Binding: Another Piece to the Puzzle. *J. Am. Chem. Soc.*, **132**, 6425-6433.
75. Fadrna, E., Spackova, N.a., Sarzynska, J., Koca, J., Orozco, M., Cheatham, T.E., III, Kulinski, T. and Sponer, J. (2009) Single Stranded Loops of Quadruplex DNA As Key Benchmark for Testing Nucleic Acids Force Fields. *J. Chem. Theory Comput.*, **5**, 2514-2530
76. Mukherjee, A., Lavery, R., Bagchi, B. and Hynes, J.T. (2008) On the Molecular Mechanism of Drug Intercalation into DNA: A Simulation Study of the Intercalation Pathway, Free Energy, and DNA Structural Changes. *J. Am. Chem. Soc.*, **130**, 9747-9755.
77. Gaillard, T. and Case, D.A. (2011) Evaluation of DNA Force Fields in Implicit Solvation. *J Chem Theory Comput*, **7**, 3181-3198.
78. Pasi, M., Maddocks, J.H., Beveridge, D., Bishop, T.C., Case, D.A., Cheatham, T.C., Dans, P.D., Jayaram, B., Lankas, F., Laughton, C. *et al.* (2014) mu ABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Research*, **42**, 12272-12283.
79. Da, L.-T., E, C., Shuai, Y., Wu, S., Su, X.-D. and Yu, J. (2017) T7 RNA polymerase translocation is facilitated by a helix opening on the fingers domain that may also prevent backtracking. *Nucleic acids research*, **45**, 7909-7921.
80. Jiang, H., Sheong, F.K., Zhu, L., Gao, X., Bernauer, J. and Huang, X. (2015) Markov State Models Reveal a Two-Step Mechanism of miRNA Loading into the Human Argonaute Protein: Selective Binding followed by Structural Re-arrangement. *PLOS Computational Biology*, **11**, e1004404.

81. Palermo, G., Casalino, L., Magistrato, A. and Andrew McCammon, J. (2019) Understanding the mechanistic basis of non-coding RNA through molecular dynamics simulations. *Journal of Structural Biology*.
82. Wang, J.M., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157-1174.
83. Shen, Z.H., Mulholland, K.A., Zheng, Y.J. and Wu, C. (2017) Binding of anticancer drug daunomycin to a TGGGGT G-quadruplex DNA probed by all-atom molecular dynamics simulations: additional pure groove binding mode and implications on designing more selective G-quadruplex ligands. *J Mol Model*, **23**.
84. Ryckaert, J., Ciccotti, G. and Berendsen, H.J.C. (1977) Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Chem. Phys.*, **23**, 327-341.
85. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T.A., Lee, H. and Pedersen, L.G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**, 8577-8593.
86. Procacci, P. and Berne, B.J. (1994) Multiple Time-Scale Methods For Constant-Pressure Molecular-Dynamics Simulations Of Molecular-Systems. *Mol. Phys.*, **83**, 255-272.
87. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling*, **14**, 33-38.
88. McGibbon, Robert T., Beauchamp, Kyle A., Harrigan, Matthew P., Klein, C., Swails, Jason M., Hernández, Carlos X., Schwantes, Christian R., Wang, L.-P., Lane, Thomas J. and Pande, Vijay S. (2015) MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal*, **109**, 1528-1532.
89. Pedregosa, F., Ga, #235, Varoquaux, I., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P. et al. (2011) Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, **12**, 2825-2830.
90. Rousseeuw, P.J. (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, **20**, 53-65.
91. Prinz, J.-H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., Chodera, J.D., Schütte, C. and Noé, F. (2011) Markov models of molecular kinetics: Generation and validation. *The Journal of Chemical Physics*, **134**, 174105.
92. Noé, F., Horenko, I., Schütte, C. and Smith, J.C. (2007) Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *The Journal of Chemical Physics*, **126**, 155102.
93. Csardi, G. and Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal, Complex Systems*, **1695**, 1--9.