Rowan University

Rowan Digital Works

6-25-2020

An in silico study of small molecule anti-cancer agents targeting DNA G-quadruplexes

Holli-Joi Sullivan
Rowan University

Follow this and additional works at: https://rdw.rowan.edu/etd

Part of the Medicinal and Pharmaceutical Chemistry Commons

**AN *IN SILICO* STUDY OF SMALL MOLECULE ANTI-CANCER AGENTS
TARGETING DNA G-QUADRUPLEXES**

by

Holli-Joi Sullivan

A Thesis

Submitted to the
Department of Chemistry and Biochemistry
College of Science and Mathematics
In partial fulfillment of the requirement
For the degree of
Master of Science in Pharmaceutical Sciences
at
Rowan University
June 10, 2020

Thesis Chair: Chun Wu, Ph.D.

Acknowledgments

Dr. Wu, thank you for your guidance over the past three years. Your faith in me has given me many opportunities to learn and grow both as a person and as an academic. The knowledge I have gained and the work ethic you have instilled in me will not be lost as I continue my PhD education and future career pursuits. To Dr. Caputo and Dr. Keck, thank you for being a part of my thesis committee and for your ongoing support and encouragement for my educational journey. Thank you to the funding sources for my research which was performed under the Rowan SEED grant and the High Performance Computing Facility at Rowan funded by the National Science Foundation under Grants RUI-1904797/MRI-1429467 and XSEDE MCB170088 and to Rowan Universities Campbell Library for granting me open access funding for several of my published works. Finally, to my lab mates, friends and family, it is no secret I have been engulfed in my studies since beginning my higher education training so I truly thank you for your support from afar throughout all these years.

Abstract

Holli-Joi Sullivan
AN *IN SILICO* STUDY OF SMALL MOLECULE ANTI-CANCER AGENTS
TARGETING DNA G-QUADRUPLEXES
2019-2020
Chun Wu, Ph.D.
Master of Science in Pharmaceutical Sciences

Free ligand binding molecular dynamic simulations are a powerful tool used to probe the ligand binding process, mechanism and pathway and the insight gained can help expedite the early stages of drug discovery. Using these methods, we model the binding of two small molecule anti-cancer agents BRACO19 and CX-5461 to a variety of DNA G-quadruplexes (G4s) and a DNA Duplex. The first study focuses on the binding of BRACO19 to three different topological folds (parallel, anti-parallel and hybrid) of the human telomeric G4s. Our detailed analysis identified the most stable binding modes were end stacking and groove binding for the G4s and duplex, respectively. With the parallel scaffold being most favorable, we suggest a conformation-selection mechanism where the relative population of the three scaffolds shifts to an increase of the parallel scaffold upon BRACO19 binding. The second study focuses on the binding of CX-5461 to human telomeric, c-KIT1, and c-Myc G4s. Our analysis was able to provide insight into a FRET-melting temperature increase assay measured the stabilizing effects of CX-5461 to each of these targets. The energetic and structural differences explained the different melting temperature between the G4s, while CX-5461's lack of intercalation to the duplex explained the difference between the G4s and duplex. Using our insight CX-5461 derivatives were deigned and docked with higher selectivity to the G4s over the duplex, which might aid in further optimization of CX-5461.

iv

Table of Contents

Table of Contents (Continued)

Table of Contents (Continued)

List of Figures

List of Figures (Continued)

List of Figures (Continued)

List of Tables

Chapter 1

Introduction to Free Ligand Molecular Dynamics Simulation and Its Application to Computer-Aided Drug Design

1.1 Motivations for a Computational Approach

Many questions exist regarding the role of biological molecules in human health and diseases. For example, why does one medication work better than another? How does a medication interact with the intended target? Can this medication interact with other targets? How can we quantify the binding of this medication to its target? For decades, the focus of many medicinal chemists has been attempting to provide answers to these fundamental questions. It is by answering these questions that researchers are able to identify and optimize small molecules for therapeutic use that selectively target pharmacologically relevant human diseases such as cancer, neurodegenerative disorders, and diabetes.

Notwithstanding the assiduous efforts of researchers, traditional development methods of a new medication is a costly and labor intensive process. From the initial identification phase until the approval for marketing is 14 years on average and costs between 2-3 billion dollars[1, 2]. Despite the unprecedented investments into pharmaceutical research and development from 1950 to 2008, the US Food and Drug Administration (FDA) approved just 1,222 new drugs averaging ~21 new drugs put on the market per year. Interestingly, around 2010 there were reports of large pharmaceutical companies beginning to use computational approaches in their discovery pipelines[3] and since then the number of new drugs put on the market has increased from 21 in 2008 to 59 in 2018[4].

1

By use of a computational approach, researchers are able to address the same fundamental questions regarding the role of biological molecules in human health and diseases on a much shorter timeframe and cost basis. The insight gained from these studies have a broad range of applications but has been specifically useful for pharmaceutical drug discovery. Computer-aided drug design (CADD) couples molecular modeling methods with molecular dynamics simulations for drug design. CADD tools are powerful and have had much success since the initial integration into the drug discovery pipeline. Visualization tool such as those implemented through VMD and Maestro are used to visualize high resolution coordinates solved by x-ray crystallography or nuclear magnetic resonance. The increased computational resources available enable molecular dynamics simulations programs such as Amber and Desmond to predict detailed atomic level insight into molecular process, mechanism and dynamics on the nanosecond to millisecond timescale[5, 6].

For example, an early computational approach to studying ligand binding was through the use of docking. Docking is an indispensable, fast and cost efficient method to predict the bound conformations of small molecule ligands to their pharmacologically relevant targets[7]. Understanding the binding conformation of a small molecule ligand is an important feature in structure-based drug design where the three-dimensional structural information is used to predict the key sites and interactions of the ligand and its target which can be used for rational drug design. Due to the quick and computational uncostly nature of docking, it can be extremely convenient for screening large libraries of molecules. However, docking does not capture the dynamics of the binding process and only provides a final binding pose.

2

This can be a disadvantage for studying ligand binding because the process is highly dynamic.

This limitation of docking was partially addressed through the introduction of molecular dynamics (MD) simulations to the field. MD is a computational technique used to simulate the dynamic behavior of molecular systems as over time. By using the docking pose or an experimentally solved binding pose, as the starting position, conventional MD simulations can be coupled with molecular docking to assess stability, refine the structures of the final complexes and rescore and calculate more detailed interaction energies of the docking poses[8]. This approach provides another level of dynamic insight, however it the result relies on the position of the ligand in the initial binding pose. Thus, the detailed and dynamic binding pathway and binding process still cannot be probed in full.

Free ligand molecular dynamics simulations are able to model the binding process of a small molecule starting from the unbound state without a rigid body assumption. Using a high temperature to randomize the initial position of the ligand, plus long time simulation, the phase space of a molecular system can be well sampled. Although this method is more costly in terms of time and computation, free ligand molecular dynamics binding simulations are more accurate and are able to predict the major binding site, ligand binding pose, detailed binding interactions, major driving force for ligand binding, binding affinity and energy, and even the binding pathway. Because neither the target or ligand are forced to be rigid during the ligand binding process, dynamic binding mechanisms are able to be observed like the induced fit and conformational-selection mechanisms. This is extremely important because ligands

3

www.manaraa.com

often bind to their targets using a dynamic mechanism rather than the lock-key mechanism. Even more, new ligand binding sites are often identified using this method, including high energy intermediate states, which are hard to study using experimental methods. Which is why the insights from the simulations can be very useful for structure based drug design, improve ligand binding and often helps guide lead optimization.

Free ligand molecular dynamics simulations are able to model the binding process of a small molecule starting from the unbound state without a rigid body assumption. Using a high temperature to randomize the initial position of the ligand, plus long time simulation, the phase space of a molecular system can be well sampled. Although this method is more costly in terms of time and computation, free ligand molecular dynamics binding simulations are more accurate and are able to predict the major binding site, ligand binding pose, detailed binding interactions, major driving force for ligand binding, binding affinity and energy, and even the binding pathway. Because neither the target or ligand are forced to be rigid during the ligand binding process, dynamic binding mechanisms are able to be observed like the induced fit and conformational-selection mechanisms. This is extremely important because ligands often bind to their targets using a dynamic mechanism rather than the lock-key mechanism. Even more, new ligand binding sites are often identified using this method, including high energy intermediate states, which are hard to study using experimental methods. Which is why the insights from the simulations can be very useful for structure based drug design, improve ligand binding and often helps guide lead optimization.

4

1.2 Motivations for Studying DNA G-Quadruplexes

DNA G-Quadruplexes (G4s) are secondary structures of nucleic acids formed by sequences that are rich in guanine. In terms of structure, when four of these guanine bases interact through hydrogen bonding they form a G-tetrad and when three g-tetrads stack on top of each other they form the G4 core which is stabilized by cations like potassium. Then to complete the structure, short 1-5 residue loops connect each g-quartet and it is the connection of these loops that defines the overall topology of the G4. The different structures are characterized into three major interconverting topologies which are parallel, anti-parallel, and hybrid. In the parallel structure the connecting loops are proceed in the same direction, where as the anti-parallel structure has each loop adopting the opposite configuration as the one that came before it, and in the hybrid structure there is a mixture of both.

In terms of locations in the body, these G4s are spontaneously formed in two major guanine rich regions. The first is at the end of the chromosome in the telomeres and the second is in the promoter region of oncogenes. As for the telomeric G4s, in normal cells, the telomeres shorten over time and cell division stops, however in cancerous cells, telomerase facilitates the process of replacing the shortened telomere. Essentially, telomerase is a reverse transcriptase enzyme that adds repeat segments to the 3' end of telomeric DNA, and because telomerase is highly expressed in almost all cancers, the segments added are often sequences that contain a cancerous mutation and the cells ultimately form a malignant tumor. Therefore, the major anticancer approach for targeting telomeric G4's is to stabilize the G4s using a ligand to prevent telomerase from functioning.

The promoter G4s are found is the promoter region of oncogenes throughout the chromosome. Some of these oncogenes include c-KIT, PDGF-A, c-Myc and VEGF. For example, the c-myc pathway plays an integral role in regulating c-Myc. This is a protein product and with this, the c-Myc protein functions in the processes of apoptosis, cell growth, and for transcriptional control on human telomerase reverse transcriptase. This is just one example of how the structures are formed in essential genes, which really highlights the importance of this secondary structure in the development and growth of cancerous cells. Much like the telomeric G4s, the therapeutic approach for promoter G4s is also to stabilize the G4 using a ligand, but in this case, stabilizing the G4 can help prevent the transcription of the gene that contains a cancerous mutation. While the formation of G-quadruplex structure vary to some extent for the different promoter regions of oncogenes, the consistent stabilization of these structures have been found in cancer development.

These structures are very prosing anti-cancer targets because they are more diverse and stable than the DNA duplex and over 300,000 sequence motifs have been identified within the human genome. However in order to target these structures, there are several major challenges that need to be overcome which is why there are no G4 stabilizers on the market to date. Specifically, the major challenge is that there is a low potency and a lack of selectivity to G4s over the duplex structure. Like many classes of medications, without the selectivity to the major target, adverse effects can be observed. Which is why improvements are needed. However, it has become difficult to make improvements to this class of drug without understanding the detailed binding mechanism. So in order to best utilize these promising structures for

6

their anticancer effects, the binding mechanism of the G4 stabilizers must first be understood.

1.3 Computational Methodologies

1.3.1 Molecular docking and structure based virtual screening. Molecular docking is an indispensable, fast and cost efficient method to predict the binding pose of small molecule ligands to their pharmacologically relevant targets and has been extremely convenient for screening large libraries of molecules. Molecular docking has been utilized in combination with structure-based virtual screening (VS) for drug discovery for over a decade to automatically dock of large libraries of ligands into a target in a relatively a short amount of time. The docking poses are achieved by using software that generates a molecular surface of the target structure based on high resolution structures. Using a known ligand binding site, the software generates spheres in the binding site and these spheres match with the new ligands to determine all possible orientations. Then the software will use a scoring function to rank each possible binding pose where a higher score indicates a more favorable binding. There are three major scoring functions that can be used. First, the force-field based approach uses physical-based functional forms based on experimental data to estimate binding affinity. Second, empirical scoring simplifies the parameters of the first approach to approximate favorable interactions, which is much faster but slightly less accurate. The third approach is knowledge-based scoring functions derived from known binding interactions and expected to be more common than by random distribution. Following the scoring functions, and especially useful in structure based virtual screenings, ADME (absorption, distribution, metabolism,

7

excretion) property prediction software can be utilized to provide insight into a wide range of pharmaceutically relevant properties about a molecule's suitability and druggability, such as octanol/water and water/gas log Ps, log S, and log BB. Because roughly 40% of all potential medications fail clinical trials due to poor ADME properties, early prediction is extremely cost and time effective and can prevent unnecessary testing of compounds that are likely to ultimately fail. In this study, we utilized Schrodinger's Glide docking program using extra precision which is a semi empirical scoring function utilizing water desolvation energy terms followed by their QikProp ADME prediction software.

$$\text{XP Glide Score} = \Delta E_{coul} + \Delta E_{vdW} + E_{bind} + E_{penalty}$$

$$E_{bind} = E_{hyd\_enclosure}^{4} + E_{hb\_nn\_motif}^{5} + E_{hb\_cc\_motif}^{6} + E_{pI}^{7} + E_{hb\_pair} + E_{phobic\_pair}$$

$$E_{penalty} = E_{desolv} + E_{ligand\_strain}$$

1.3.2 Molecular dynamics simulations. Molecular dynamics is a computational simulation method used to analyze the physical movements of atoms and molecules. Essentially, molecular dynamics simulations are able to mimic what atoms do in real life, based on a given potential energy function which is used to calculate the force experienced by any atom given the positions of the other atoms in the model system. The atoms and molecules are allowed to interact for a fixed period of time, giving a view of the dynamic evolution of the system. Then, by using Newton's Second Law, molecular dynamics simulations predict how the forces will affect the motions of the atoms. Thus probing the position and velocity of particles within a given system that when paired with

8

the structure gives a complete analysis of the interaction dynamics of the model system at high spatial and temporal resolution.

More specifically, the aforementioned forces calculated in a molecular dynamics simulation are based on what is a called a force field, a set of parameters to define the potential energy function of a molecular system. The parameters include both bonded and nonbonded interactions amongst the atoms in the model system. The bonded interactions are comprised of bond, angle and dihedral angles and the nonbonded interaction are the van der Waals and electrostatic interactions within the model system. Within this system, further coupling parameters are set to maintain a constant pressure and temperature. Two types of solvent models can be used to realistically model the interactions between the solvent and the solute molecules, which are called implicit and explicit solvent models. The implicit solvent model treats a solvent as one continuous medium using a dielectric constant, whereas the explicit solvent model uses thousands of solvent molecules. Although the explicit model can produce more accurate results when comparing systems that are highly variable from each other, due to the higher computational cost, an implicit model may be used as a fair alternative for systems that are more closely comparable.

Once the system is prepared, the molecular dynamic simulation is run on this system until an equilibrium state is reached. Using snap shots from the simulation a time progression is generated and used to generate the simulation trajectory. Each snapshot contained coordinates that specify the position of atoms in the system. From the trajectory, a number of post simulation analyses can be performed including a trajectory clustering. One method for this is a through a pairwise comparison which is done by first aligning each snapshot from the trajectory of the ligand-target complex. Once aligned, the

9

structures can be clustered into unique structural families based on the root mean square deviations (RMSD) of their atomic positions and the larger structural families can be identified as the most prevalent binding positions for the system.

1.3.3 Markovian state modeling. Markov State Models (MSMs) can be built from molecular dynamics simulation data and are a comprehensive statistical approach used to create understandable yet high-resolution models of the intrinsic kinetics of a system. In order to build the MSM microstates are built. To do this, the trajectories are first subjected to a clustering method, K-means clustering, which differs from the pairwise clustering method described above. Where pairwise clustering examines each pair of N structures, K-means defines a number of clusters and assigns structures to each cluster based on the relative similarity to the mean structure of each cluster. K-means clustering is a rapid clustering method able to exponentially scale with N, but the trade-off for speed is the level of accuracy delivered using this method. For the K-means clustering in this study, two features were used for the calculation, RMSD and center of mass. Center of mass was included because we observe the ligand bound in a near spherical cloud around the G-quadruplex, the RMSD values could overlap with RMSD values of ligands at the opposite ends of the cloud. By incorporating center of mass, the level of accuracy was maintained while reducing the time for the clustering to complete from days to minutes. Due to choosing to cluster into a handful of "macrostates" and directly and skipping over the experimentally unverifiable thousand "microstates", the expected convergence time of the implied timescales should be significantly greater than that of a model with a greater number of clusters. This results in a coarser grained model that trades finer detail for greater experimental testability and easier human

understanding. It is likely that directly clustering into "macrostates" still maintains the integrity of the MSM as verification through the Chapman-Kolmogorov test indicates that the model closely resembles the observed simulation data. Since a key feature of MSMs is that they are memoryless, the current state predicts the future state without knowing the past state. In terms of the kinetic insight this analysis can provide, the major result is the percentage of each state and the mean transition times between the major binding modes. The implied time method or the Chapman-Kolmogorov test is typically used to determine the lag time for counting the transitions. Once transition matrix is obtained, the mean passage time between states can be calculated.

1.3.4 MMGBSA/MMPBSA binding energy calculations. Quantifying the binding free energy of a ligand to its target is extremely useful in understanding its stability, and binding favorability to one binding site versus another. Of the binding free energy prediction methods, end point methods have been extremely useful when calculating the binding free energies of the unbound and bound states of a model system. Of the end point methods are the molecular mechanics generalized Born surface area (MM-GBSA) and molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) methods. These methods allow for energy decomposition into the following components: electrostatic, van der Waals, and hydrophobic interaction using surface area term. The calculations are performed on the receptor alone through minimization, on the ligand alone, and finally, the ligand target complex is calculated. The final output of these calculations contain the contribution of each component on the total binding energy which provides quantitative insight into the most favorable binding mode. Although these calculations are extremely useful in making general predictions and orders of stability, because the calculation does

11

not include a solute entropy, the reported energy could be an over-estimation of the true binding free energy. However, this is overcome relatively when the solute conformational entropies in different binding poses are comparable to one another and thus the relative binding free energy can be estimated relative to the MMGBSA binding energies. MM-PBSA[9] (Molecular Mechanics Poisson Boltzmann Surface Area) was used to analyze the energetics of the bound complexes. The MM-PBSA binding energy for a system was calculated based on three simulations: the ligand only, the DNA only and the DNA-ligand complex using equation 1. The equation is made of four components, equation 2: van der Waals interaction energy (VDW), hydrophobic interaction energy (SUR), electrostatic interaction (GBELE) and the change of the conformation energy for DNA and ligand which are calculated using equation 3 and 4.

Eq 1: $\Delta E = E_{complex} - E_{DNA\_free} - E_{lig\_free}$

Eq 2: $\Delta E = \Delta E_{vdw} + \Delta E_{SUR} + \Delta E_{GBELE} + \Delta E_{comformation}$

Eq 3: $\Delta E_x = E_{x\_complex} - E_{x\_DNA\_complex} - E_{x\_lig\_complex}$

$x= vdw, \; sur \; and \; gbele$

Eq 4: $\Delta E_{Comformation} = E_{DNA\_complex} + E_{lig\_complex} - E_{DNA\_free} - E_{lig\_free}$

1.4 Thesis Outline

In Chapter 2, the binding pathway of BRACO19 was probed by 27.5 µs molecular dynamics binding simulations with a free ligand to a DNA duplex and three different topological folds of the human telomeric DNA G-quadruplex (parallel, anti-parallel and hybrid). BRACO19 is a potent G-quadruplex binder, but its potential for clinical usage is hindered by its low selectivity towards DNA G-quadruplex over duplex. High-resolution

12

structures of BRACO19 in complex with neither single-stranded telomeric DNA G-quadruplexes nor B-DNA duplex are available, thus the detailed interactions of BRACO19 to these targets remain elusive. Our analyses probed these interactions and showed that the most stable binding modes were identified as end stacking and groove binding for the DNA G-quadruplexes and duplex, respectively. Among the three G-quadruplex topologies, the MM-GBSA binding energy analysis suggested that BRACO19′s binding to the parallel scaffold was most energetically favorable. We deciphered two lines of conflicting evidence, along with our binding energy data, to suggest a conformation-selection mechanism is used where the relative population shift of three scaffolds upon BRACO19 binding. Or in other words, there is an increase of the population of the parallel scaffold, and a decrease of populations of antiparallel and/or hybrid scaffolds. This hypothesis appears to be consistent with the fact that BRACO19 was specifically designed based on the structural requirements of the parallel scaffold and has since proven effective against a variety of cancer cell lines as well as toward a number of scaffolds. In addition, this binding mode is only slightly more favorable than BRACO19s binding to the duplex, explaining the low binding selectivity of BRACO19 to G-quadruplexes over duplex DNA. Our detailed analysis suggests that BRACO19′s groove binding mode may not be stable enough to maintain a prolonged binding event and that the groove binding mode may function as an intermediate state preceding a more energetically favorable end stacking pose. Finally we observed that base flipping played an important role in enhancing binding interactions which is an integral feature of an induced fit binding mechanism. In Chapter 3, we study the binding interactions of DNA G-quadruplex (G4) stabilizer, CX-5461, to human telomeric, c-KIT1, and c-Myc G4

structures as well as a DNA duplex. CX-5461 is in Phase I/II clinical trials for advanced hematologic malignancies and cancers with BRCA1/2 deficiencies. Recently, a FRET-melting temperature increase assay measured the stabilizing effects of 10 µM CX-5461 to canonical DNA duplex (~10 K) and three G4 forming sequences negatively implicated in BRCA1/2 mutations upon its binding: human telomeric (~30 K), c-KIT1 (~27 K) and c-Myc (~25 K). Without experimentally solved structures of these CX-5461-G4 complexes, CX-5461's binding interactions remain elusive. To probe these interactions, we performed 16.5 µs free ligand molecular dynamics binding simulations of CX-5461 to the DNA duplex and three G4s. Our analyses detailed their thermodynamic, kinetic, and structural nature at the molecular level. CX-5461's average MM/PBSA of the major poses were calculated for the human telomeric (-28.6 kcal/mol), c-KIT1 (-23.9 kcal/mol), c-Myc (-22.0 kcal/mol) G4s, and DNA duplex (-15 kcal/mol) systems. These energetic differences coupled with structural differences at the 3' site explained the different melting temperature between the G4s, while CX-5461's lack of intercalation to the duplex explained the difference between the G4s and duplex. CX-5461 derivatives were deigned and docked with higher selectivity to the G4s over the duplex, which might aid in further optimization of CX-5461.

Chapter 2

Binding of BRACO19 to a Telomeric G-Quadruplex DNA Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent

2.1 Introduction

2.1.1 Human telomeric G-quadruplex DNA. The first therapeutically important G-quadruplex sequence was located in the single stranded 3' overhang of human telomeric DNA [10, 11], and contains numerous repeats of $d(TTAGGG)_n$ sequences capped by Shelterin complexes [12-15]. The Shelterin complexes provide protection against nuclease attacks, chromosomal end-to-end fusion and gene erosion at cell divisions [16]. After each cell replication the telomere truncates by 50-200 base pairs, when the telomere is exhausted and the Hayflick limit is reached. This hayflick limit essentially states that a normal human cell can only replicate and divide forty to sixty times before it cannot divide anymore, and will break down by programmed cell death or apoptosis thus cell senescence and apoptosis are triggered [17, 18]. In cancer cells, a reverse transcriptase telomerase adds nucleotides to the telomere, immortalizing the cells [19, 20]. Telomerase is found to be overexpressed in 80-85% of tumor cells underscoring why telomerase inhibition is a logical therapeutic approach in cancer treatment. Despite the potential of this approach challenges include: (i) a time delay in which the telomere length needs to be established for the ultimate apoptosis trigger [17, 21, 22] and (ii) that studies suggest an alternate mechanism for telomerase maintenance might be activated upon telomerase inhibition [23-25]. However, it has been reported that the telomere cannot be hybridized by telomerase when the single stranded 3' overhang folds into a G-quadruplex [26], instead the

15

telomeric G-quadruplex is perceived as DNA damage and stimulates cell level apoptosis [11, 27]. Accentuating how a ligand that stabilizes the telomeric G-quadruplex can be an efficacious anti-cancer therapy.

2.1.2 Topological folds of the human telomeric G-quadruplex DNA. Extensive research has been performed that show G-quadruplexes are highly polymorphic and their topological fold depends on factors such as nucleic acid sequence, ions and the presence of small molecules [28-32]. Though studied for decades, the most biologically relevant topological fold of human telomeric G-quadruplex remains an elusive and controversial debate. In 1993, Wang and Patel published the first solution structure of the human telomeric sequence $d[AG_3(T_2AG_3)_3]$ in Na^+ containing solution which adopts anti-parallel topology (PDB ID: 143D); formally referred to as 3(-lwd+ln)[33]. In 2002, Parkinson and coworkers published a K^+-containing crystal structure of the human telomeric DNA in a parallel topological fold, referred to as 3(-p-p-p) using the nomenclature recently described by Dvorkin et al.[34] (PDB ID: 1KF1)[35]. The parallel crystal structure published by Parkinson and coworkers was different from the preceding studies which reported the DNA in an anti-parallel topological fold in Na^+ containing solution [33, 36, 37]. In the years following, experiments providing evidence for both topological folds continued to publish. The parallel topological fold was suggested the most biologically relevant form in K^+ containing crystal because the polymorphism of the G-quadruplex structure was lost in 40% PEG or 50% ethanol solutions, i.e., dehydrated solutions[38]. Heddi and Phan studied the human telomeric sequence under crowded conditions with NMR, using the same dehydrating crowding agents used in X-ray crystallography, and found that the parallel conformation predominated [39]. In contrast, solution studies using NMR and [125]I-

16

radioprobing were also published providing evidence for the anti-parallel topology in both K⁺- and Na⁺-containing solution, several of which reported the parallel and anti-parallel topologies coexist under both ionic conditions [40-43]. Not long after Parkinson and coworker's parallel crystal structure was published solution studies began to identify additional topological folds for the human telomeric DNA[44-48]. Yang and coworkers [49] showed the same sequence with two additional naturally occurring adenine residues at the 5' and 3' termini in K⁺ solution folds into a hybrid scaffold in 2007 (PDB ID: 2HY9), formally referred to as 3(-p-lw-ln). The skepticism regarding the predominating topology in physiological conditions has led to studies suggesting that rather than the parallel topology, both the anti-parallel [43, 50, 51] and hybrid [45, 49] forms are most physiologically relevant.

The contradicting results being published gave rise to discussion regarding the experimental accuracy of using dehydrating crowding agents like PEG to report the structure of the highly flexible and dynamic DNA G-quadruplex. In 2005, Li and coworkers published work that directly refuted the biological relevance of the parallel stranded crystal structure of the human telomeric DNA G-quadruplex reported by Parkinson and coworkers delineating that by using certain dehydrating solvents, like PEG, crowding conditions are not being mimicked but rather conditions of extreme water depletion that is misrepresentative of physiological conditions. [28, 50] Using acetylnitrile, a non-dehydrating crowding agent, Miller et al. proposed the structure of the human telomeric sequence was not identical to the structure published in crystalline state, supporting the role of hydration in the stability and conformation of this human telomeric sequence. Using Ficoll and *Xenopus laevis* egg extract compared to PEG, Hansel and

17

coworkers suggested the parallel scaffold is not the preferred topology under physiological conditions, but rather the parallel, anti-parallel, and hybrid topologies co-exist under native conditions. Stating that in Ficoll or cellular extracts, the more predominate conformations in the co-existing equilibrium mixture are likely the anti-parallel and/or the hybrid topologies. Analyzing the studies overall, evidence suggests that this sequence forms multiple intramolecular G-quadruplex scaffolds in K^+ solution and the intramolecular parallel G-quadruplex observed in the K^+-containing crystal appears unlikely to be the major form in K^+- containing solution[41, 42, 52-56]. Given the broad range of evidence to support each of the three scaffolds and without evidence to rule out the predominance of any of the reported scaffolds, one of each the parallel, anti-parallel, and hybrid topological folds were modeled in this study.

2.1.3 Anti-cancer agent BRACO19. Based on structural requirements of the parallel-stranded telomeric G-quadruplex binding site, BRACO19, a tri-substituted acridine, was rationally designed with computer modelling [57, 58] and has since been one of the most widely studied G-quadruplex binders (Figure 1). BRACO19 has been reported to inhibit telomerase which causes telomere shortening [59]; its experimental *in vivo* activity against a variety of cell lines is reported (Table 1). Furthermore, BRACO19 have been shown effective in anti-viral, and anti-parasitic treatments. BRACO19 also demonstrates broad anti-viral activity by stabilizing the G-quadruplexes found in pro-viral genomes [60] such as the viruses HIV-1, HSV-1, EBV, HHV-6, and HBV [61]. BRACO19 also showed *in vitro* anti-parasitic activity in bloodstream forms of *T. brucei brucei*, against promastigotes of *L. major*, against *P. falciparum* [62]*, as well as against a human non-tumoral lung cell line (MRC-5)* [63]. A major factor that has prevented BRACO19 from clinical usage is a low

18

selectivity towards G-quadruplex over duplex DNA (K_quad/K_dup=40 fold, K: ligand binding constant) [58], which has the potential to cause reverse effects. To achieve higher selectivity (e.g. 10^5 fold), better understanding of the binding nature of BRACO19 with DNA G-quadruplex and duplex DNA is required.
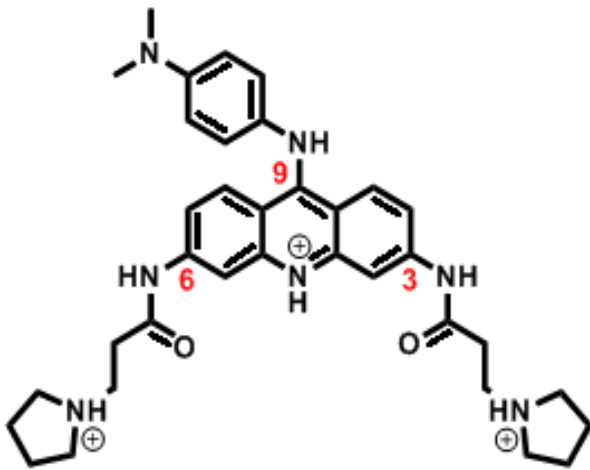
Figure 1. Chemical Structure of BRACO19.

Table 1
In vivo activity of BRACO19 against various cancer cell lines.

Cell lines	Tissue type	IC$_{50}$	References
MCF7	Breast cancer (human)	2.5μM	64, 65
A549	Lung cancer (human)	2.4μM	64, 66
DU145	Prostate cancer (human)	2.3μM	64, 67
HT-29	Colon cancer (human)	2.7μM	64, 68
HGC-27	Gastric carcinoma	2.6μM	64, 69
A2780	Ovarian cancer (human)	2.5μM	64, 70
WI-38	Lung fibroblast (human)	10.7μM	64, 71
IMR90	Lung fibroblast (human)	>25μM	64, 72
U87	Glioblastoma (human)	1.45μM	73, 74
U251	Glioblastoma (human)	1.55μM	73
SHG-44	Glioma (human)	2.5μM	73
UXF1138L	Uterus carcinoma (human)	2.5μM	75
CH1	Lymphoma (mouse)	10.1μM	76
SKOV3	Ovarian cancer (human)	13.0μM	76, 77
CLL	Chronic lymphocytic leukemia	80μM	78, 79
AML	Acute myeloid leukemia	80μM	79
--	Prolymphocytic leukemia	80μM	79

Despite the high interest of BRACO19 in complex with biologically relevant single stranded intramolecular DNA G-quadruplex formed by one chain (e.g., d(AGGGTTAGGGTTAGGGTTAGGG)), there are no high-resolution structures of BRACO19 binding to the antiparallel or the hybrid topology. The only high-resolution structure available is a bimolecular parallel G-quadruplex in complex with BRACO19 (PDB ID: 3CE5), where the intermolecular G-quadruplex is formed from two 12-residue chains (i.e., d(TAGGGTTAGGGT)). [80] Because bimolecular (12mer) or intramolecular (22mer) adopt the same parallel topology, suggested by Parkinson et al. [35] and later confirmed by Phan et al. [81] in both Na+ or K+ in solution under crowded conditions, this crystal structure provides the following invaluable interaction insights: BRACO19 interacts asymmetrically with the guanine bases of the intermolecular G-quadruplex

through π–π interactions and the nitrogen atom of the acridine ring aligns with the K$^+$ cations within the ion pore. Nonetheless the additional 5' and 3' residues at the two ends could introduce artifacts when comparing the BRACO19 binding modes on the biologically relevant unimolecular parallel scaffold formed the single stranded telomeric sequence. To remove the artifacts, we used another crystal structure (PDB ID: 1KF1) containing the apo form of the parallel intramolecular telomeric G-quadruplex in our BRACO19 binding studies. Furthermore, because the most biologically relevant form may not be the parallel form, the binding of BRACO19 to the antiparallel and the hybrid form are required to better understand its biology relevant binding mode leading to its anti-cancer properties. [28, 50]

2.1.4 Experimental overview. In this work, by using free ligand MD binding simulations with AMBER OL15 DNA and GAFF2 ligand force fields [82] (Table 2), we were able to predict a binding mode of BRACO19 to the double stranded parallel telomeric G-quadruplex that is consistent with the crystal complex structure (PDB ID: 3CE5). Furthermore, the binding modes and the ligand binding pathways were characterized in detail. We extended our free ligand MD simulations to characterize the binding pathway of BRACO19 to the parallel, anti-parallel, hybrid DNA G-quadruplexes and duplex DNA (Figure 2). Major binding poses, (top binding, bottom binding and groove binding) were identified and detailed binding pathways were characterized. The dynamic and energetic properties of the three major binding modes were thoroughly studied. We suggest that the similar binding energy of the groove binding pose to the duplex and of the top stacking pose to the parallel G-quadruplexes may be responsible for the low selectivity (40 fold) of

21

BRACO19. This study may provide insight into the further modification of BRACO19 and

other G-quadruplex binders to enhance overall selectivity and efficacy.

Table 2

Molecular dynamics simulations systems

ID	DNA	No. of ligand	No. run	Drug Initial State	NPT eq. (ns)	NVT (ns)	Total time (μs)
1	n/a	1	2	Free	1	500	1
2	Duplex(d([GC]$_{10}$)$_2$)	0	2	Free	1	500	1
3	3(-p-p-p) (1KF1)	0	2	Free	1	500	1
4	3(-lwd+ln) (143D)	0	2	Free	1	500	1
5	3(-p-lw-ln) (2HY9)	0	2	Free	1	500	1
6	Duplex(d([GC]$_{10}$)$_2$)	1	9+1	Free	1	500+2000	6.5
7	3(-p-p-p) (1KF1)	1	8+2	Free	1	500+2000	8.0
8	3(-lwd+ln) (143D)	1	9+1	Free	1	500+2000	6.5
9	3(-p-lw-ln) (2HY9)	1	9+1	Free	1	500+2000	6.5

22
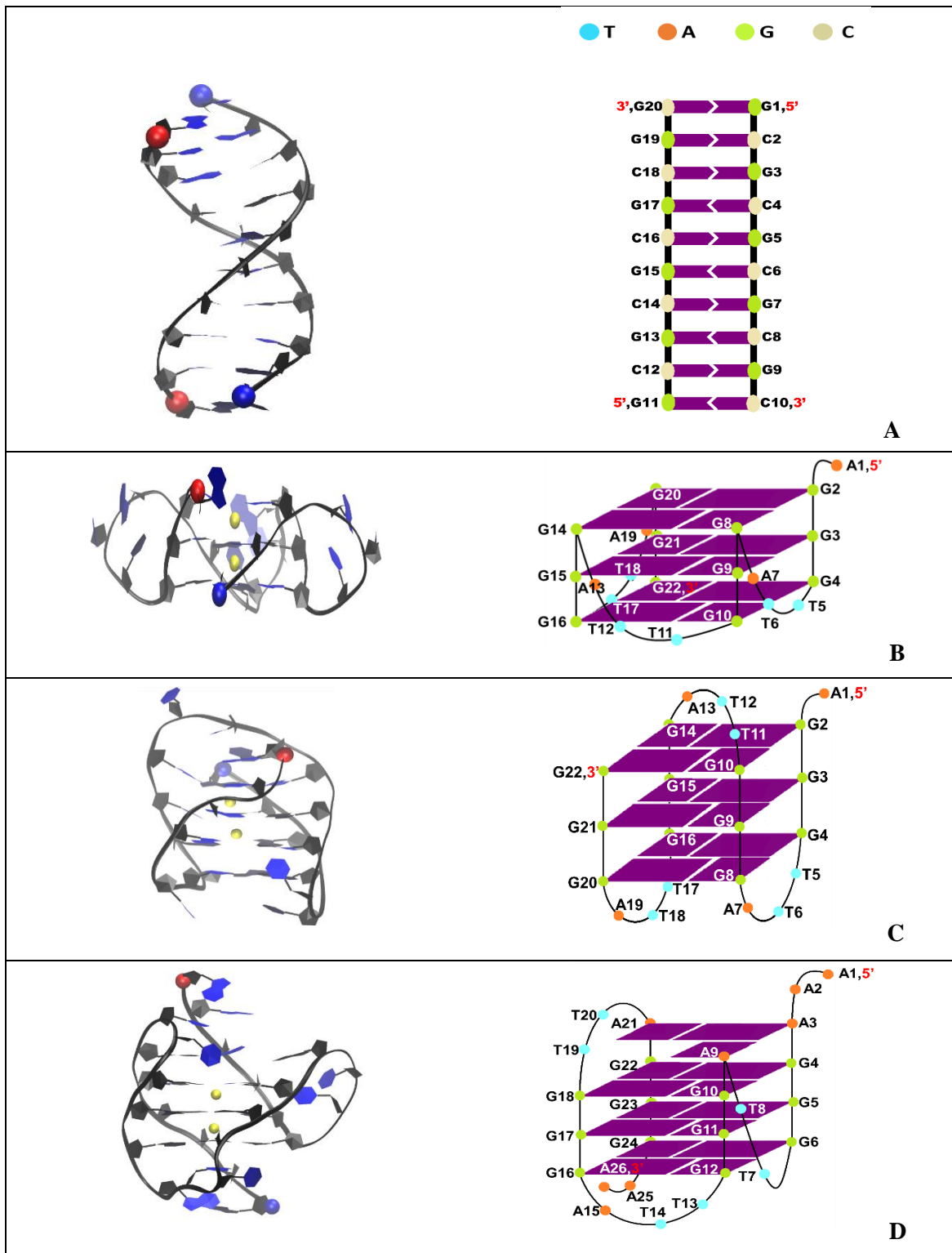
Figure 2. Structure of human telomeric duplex DNA (**A**), human telomeric parallel DNA G-quadruplex (PDB ID: 1KF1) 3(-p-p-p) (**B**), human telomeric anti-parallel DNA G-quadruplex (PDB ID: 143D) 3(-lwd+ln) (**C**), and human telomeric hybrid DNA quadruplex (PDB ID: 2HY9) 3(-p-lw-ln) (**D**). 5' and 3' ends of the DNA chain are indicated by red and blue spheres, respectively.

23

2.2 Methods

2.2.1 Molecular dynamic simulation system setup. A total of 9 systems were constructed: a BRACO19 only system, a B-DNA duplex structure of $d([GC]_{10})_2$, the X-ray crystal structure of the parallel telomeric DNA G-quadruplex, the NMR solved anti-parallel telomeric DNA G-quadruplex and the NMR-solved hybrid telomeric DNA G-quadruplex and four DNA-ligand systems (Table 2). The B-DNA duplex structure of $d([GC]_{10})_2$ was built using Maestro program. The four free ligand-DNA systems were constructed with a free BRACO19 molecule that was 10 Å away from the DNA (Figure A1). A water box of truncated octahedron with 10 Å water buffer was used to solvate the unbound system and was neutralized by K^+ and 0.15 M KCl was added. The DNA structures were represented by a refined version of the AMBER DNA OL15 (i.e., parm99bsc0 [83] $+\chi_{OL4}$ [84] $+ \varepsilon/\zeta_{OL1}$ [85] $+ \beta_{OL1}$ [86] updates), water was represented by TIP3P model [87] and the K^+ ions were represented by the K^+ model developed by Cheatham group [88]. The standard AMBER protocol was used to obtain the force field for the BRACO19 molecule: after the geometry optimization of BRACO19 at the HF/6-31G* level, the molecular electrostatic potential (MEP) of the BRACO19 molecule was calculated at the same theory level; then the partial charges of BRACO19 atoms were determined by MEP using Restrained Electrostatic Potential/RESP method with two stage fitting [89]; and the AMBER GAFF2 [90] force field provided the rest of the force field parameters. The supporting document of our previous paper [82] provides the BRACO19 force field in Mol2 format. The nucleic acid simulations have been widely practiced in AMBER DNA force fields [91-95]. In our previous studies, the binding pathways of

24

doxorubicin [96] and Telomestatin [97], anti-cancer drugs to the B-DNA fragment [96] and to the human telomeric hybrid G-quadruplex [97], respectively, have been simulated.

2.2.2 Molecular dynamics simulation protocols. The ten production runs for all complex systems were conducted using the AMBER 16 simulation package [90]. The detailed protocol followed our previous studies [96] [97]. After minimizing the energy, the Maxwell-Boltzmann distribution was used to conduct all simulation runs with different initial velocities, which were assigned based on random seeds. For the free binding system, an extra 500 ps pre-run at high temperature (500 K) was carried out to randomize the position and orientation of the free ligand, while the receptor was fixed. Better sampling of binding poses and pathway was enabled by multiple independent simulations. To equilibrate the system density, a short 1.0 ns MD simulation in the NPT ensemble mode (constant pressure and temperature) was conducted, where the DNA and ligand were subjected to Cartesian restraints (1.0 kcal/mol/Å). For the 500 ns production run, the NVT ensemble mode (constant volume and temperature) was used to enhance the simulation code stability. The representative trajectory for major binding modes was picked from each system and extended to 2000 ns. All bonds connecting hydrogen atoms were constrained by SHAKE [98] which enabled a 2.0 fs time step in the simulations. Long-range electrostatic interactions under periodic boundary conditions were treated using the particle-mesh Ewald method [99] (the fourth order of the B-spline charge interpolation, charge grid spacing of ~1.0 Å; and direct sum tolerance of 10^{-5}). 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 cost, a two-stage RESPA approach [100] was used to calculate non-bonded forces where the short range forces was

25

updated once per time step and the long range forces was updated twice per time step. The Langevin thermostat with a coupling constant of 2.0 ps was used to control the temperature. The trajectories were saved at 50.0 ps intervals for analysis.

2.2.3 Checking the convergence of the simulations. The initial structure was used as a reference to calculate the root mean square deviation (RMSD) of DNA backbone. The stability of the DNA structures was indicated by the flat and small RMSDs (Figures A2, A4, A6 and A8). An atom-to-atom distance cutoff of 3.0 Å was used to calculate atom contacts between the DNA structure and BRACO19. The stable contact number indicated the steady state of the simulation systems (Figures A3, A5, A7 and A9). We defined a stable complex as one with greater than 10 atom contacts.

2.2.4 Binding mode identification. Accounting to the stability of the DNA backbone in the binding process, the DNA backbone of the stable complexes was aligned by a least square fitting. Daura's algorithm [101] was used to cluster the aligned complexes into different structural families based on the 2 Å pair-wise RMSD cutoff of the BRACO19 only without ligand fit. The centroid structure was defined as a structure with the largest number of neighbors in the structural family and was used to represent that structural family. Based on visual inspection, super-families corresponding to major binding modes were formed by merging the centroid structures (Figure 3; Figure A14-A17).

26
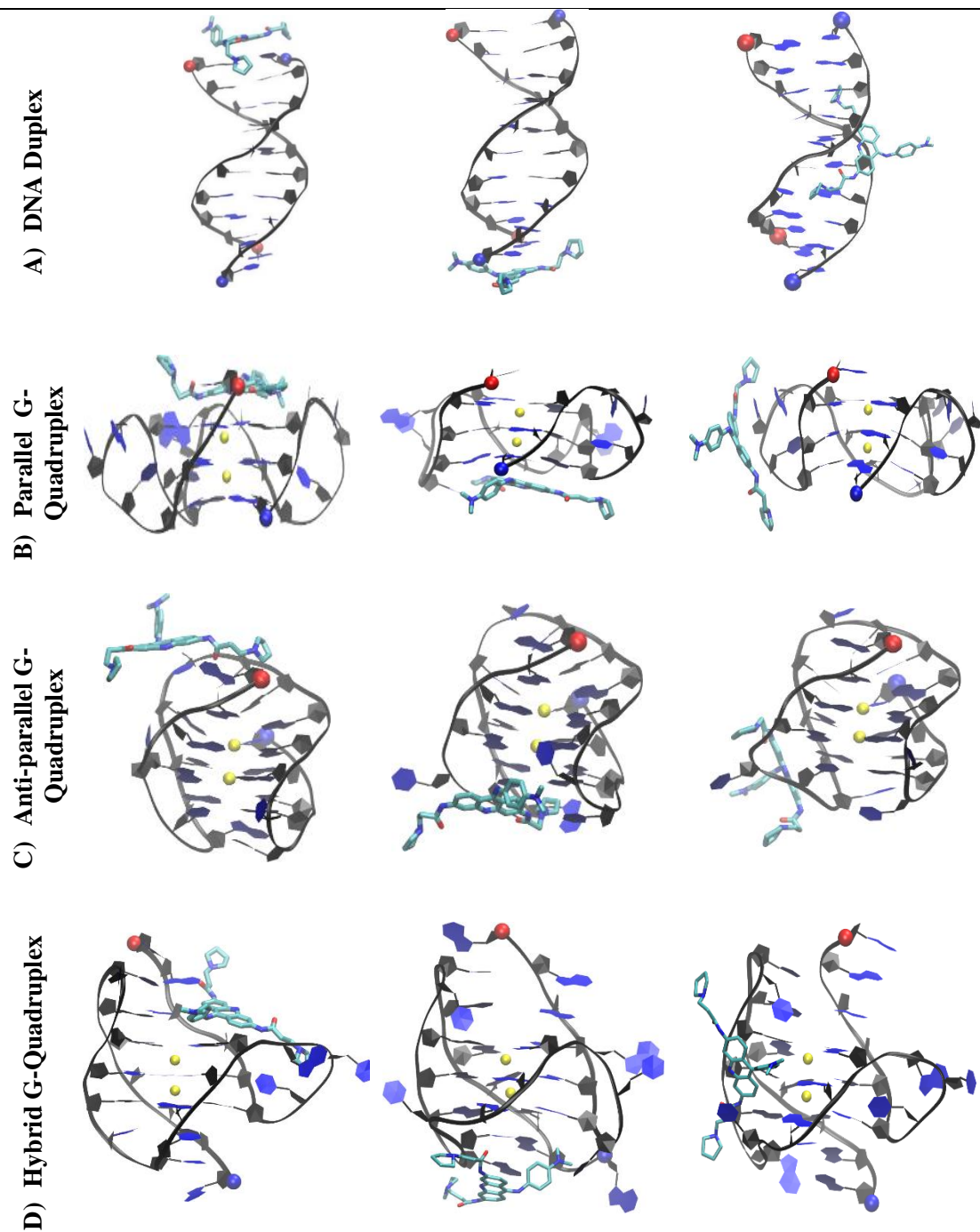
Figure 3. The major binding modes of BRACO19 in complex with the human telomeric DNA. **A:** Duplex; **B:** Parallel G-quadruplex; **C**: Anti-parallel human telomeric G-quadruplex; **D**: Hybrid human telomeric G-quadruplex. **A-D**: Top binding (**left**), Bottom binding (**middle**) and groove binding (**right**); 5' end and 3' end are represented by the red and blue ball respectively.

27

2.2.5 Order parameters to characterize DNA-drug binding pathway. The DNA-drug binding process was characterized by using five order parameters: hydrogen bond analysis, drug-base dihedral angle, DNA/ligand RMSD, center-to-center and K^+-K^+ distance (R) and MM-GBSA binding energy (ΔE). A hydrogen bond was defined by 3.5Å distance cutoff between H-bond donor and H-bond acceptor and 120° donor-H-acceptor angle cutoff. The hydrogen bonds were calculated for the top/first, middle/second and bottom/third base tetrads. For the three G-quadruplexes, the three G-tetrads were defined so that 5' is close to the first G-tetrad. The dihedral angle between the plane of the stable G-tetrad of the DNA that is close to drug binding site and the BRACO19's ring plane was defined as the dihedral angle. After aligning the DNA, the ligand RMSD was calculated with reference to the first frame of the trajectory. The length from the DNA center to the drug molecule center was defined as the center-to-center distance (R). The distance between the K^+ ions present in the DNA G-quadruplex was defined as K^+-K^+ distance. The energetics of the bound complexes were analyzed using MM-GBSA[54] (Molecular Mechanics Generalized Born-Surface Area) module in the AMBER package (GB1 model with salt concentration of 0.15M, mBondi radii set, and surface tension of 0.0072 kcal/$Å^2$) to avoid the large energy fluctuation of the explicit solvent.

It was reported that even when considering the relative solvation free energy, good predictions can be made for charged molecules by the GB models on the hydration free energy [102]. Under this assumption, ions were removed from charged DNA systems in this study. This approach was validated in our previous study, in which this MM-GBSA protocol successfully assessed the binding energy of doxorubicin, an anti-cancer drug, to a B-DNA fragment (d(CGATCG)$_2$) [103]. Under comparable entropic terms, the relative

28

binding free energy estimated by the MM-GBSA binding energies can be used to rank drugs or their binding poses if a single molecule is considered [104]. It has been established by systematic benchmarking studies up to 1864 crystal complexes that ranking of the ligand binding affinity can be achieved by relative MM-GBSA binding energy calculations [105-107]. In a previous work we studied the use of MM-GBSA versus MM-PBSA as a predictor of BRACO19's relative binding energy ($\Delta\Delta E$) over a range of ionic strengths [82]. The highly comparable relative binding energies in both the MM-GBSA and MM-PBSA calculations supports the use of MM-GBSA in ranking the binding poses of BRACO19 in this study. The MM-GBSA binding energy for each system was calculated from three simulations[108]: ligand only, DNA only and DNA-ligand complex using equation 1. Equation 2 is made of four components: van der Waals interaction energy (VDW), hydrophobic interaction energy (SUR), electrostatic interaction (GBELE) and the change of the conformation energy for DNA and ligand. These terms were calculated using equation 3 and 4.

Eq 1: $\Delta E = E_{complex} - E_{DNA\_free} - E_{lig\_free}$

Eq 2: $\Delta E = \Delta E_{vdw} + \Delta E_{SUR} + \Delta E_{GBELE} + \Delta E_{comformation}$

Eq 3: $\Delta E_x = E_{x\_complex} - E_{x\_DNA\_complex} - E_{x\_lig\_complex}$, $x=$ VDW, SUR and GBELE

Eq 4: $\Delta E_{Comformation} = E_{DNA\_from\_complex} + E_{lig\_from\_complex} - E_{DNA\_free} - E_{lig\_free}$

The standard backbone dihedral angles (α, β, γ, δ, ϵ and ζ) around the covalent bonds of the deoxyribose and χ about the glycosidic bond were defined (Figure A27) to characterize the conformational changes.

29

2.3 Results

2.3.1 Multiple drug binding modes were observed in free ligand binding simulations. Starting from an unbound state, we simulated ten 500-ns production runs for each system. The convergence of the binding simulations was confirmed (see the method section), a sampling plot was generated to trace the position of one atom of BRACO19 through the length of the entire simulation period for each system (Figure A31). The last snapshots from each simulated trajectory of the duplex-BRACO19 system is listed in Figure A10 and indicate the stability of the DNA structures where the base pairing was maintained. The last snapshots of all the simulated trajectories of the G-quadruplex-ligand systems are listed in Figures A11, A12 and A13 and indicate the stability of the G-quadruplex structures where the G-tetrads were maintained. Multiple binding sites were observed in the ten duplex DNA-BRACO19 trajectories. The clustering analysis described in the methods section was employed to categorize the stable complexes that were extracted from the trajectories into structural families. By setting a threshold of 1% population, 14 structural families of complexes were identified (Figure A14). These 8 structural families were further merged into three binding modes: groove binding, top stacking and bottom stacking. Binding to the groove of the duplex accounted for 81% of the total population. Additionally, end stacking to the top of the duplex accounted for 4% and end stacking to the bottom of the duplex made up 2% of the total population (Figure A14). Three binding modes were observed in the ten parallel G-quadruplex DNA-BRACO19 trajectories. The clustering analysis was employed to categorize the stable complexes that were extracted from the trajectories into 11 structural families (Figure A15). The three binding modes observed were: top stacking, bottom stacking and groove binding. Top stacking to the

30

parallel G-quadruplex DNA accounted for 28%, bottom stacking accounted for 41% and groove binding for 29% of the total population (Figure A15). Multiple binding sites were observed in the ten anti-parallel DNA G-quadruplex-BRACO19 trajectories. The clustering analysis was employed to categorize the stable complexes that were extracted from these trajectories into 9 structural families (Figure A16). Three binding modes were observed: top, bottom and groove binding. Bottom binding to the anti-parallel G-quadruplex DNA accounted for 46%, top binding for 25% and groove binding for 40% of the total population (Figure A16). Multiple binding sites were observed in the ten hybrid G-quadruplex DNA-BRACO19 trajectories. The same clustering analysis was employed to categorize the stable complexes that were extracted from these trajectories into 11 structural families (Figure A17). Three binding modes were observed: top, groove and bottom binding. Groove binding to the hybrid G-quadruplex DNA accounted for 43%, top binding for 33%, and bottom binding for 20% of the total population (Figure A17). Two dimensional interaction diagrams of BRACO19 in complex with each DNA system, in each major binding pose, is available in the supporting document (Figure A18).

2.3.2 MM-GBSA binding energy calculations. MM-GBSA binding energy calculations were carried out, as depicted in methods section, to examine the relative binding affinities of the major binding modes of BRACO19 with respect to the DNA (Table 3). From this it was clear the VDW interaction contributes most to the total binding energy and in ranking the binding poses for each DNA-ligand system. The most favorable binding energy for the duplex-BRACO19 complex was the groove binding mode (-61.7±8.0 kcal/mol), followed by the bottom stacking mode (-34.6±5.7 kcal/mol) and the top stacking mode (-33.7±5.3 kcal/mol). VDW packing, responsible for the VDW

31

energy contribution, was the primary contributor to the binding energy of the three

modes. BRACO19 bound to parallel G-quadruplex DNA in three binding pose where top

stacking (-62.3±4.5kcal/mol) was the most energetically favorable, closely followed by

the bottom binding mode (-61.8±1.5 kcal/mol), and groove binding (-37.6±7.2 kcal/mol)

was the least stable of the three. BRACO19 bound to the anti-parallel G-quadruplex DNA

in three binding poses where bottom binding (-53.9±5.8 kcal/mol) was the most stable of

the three, groove binding exhibiting a binding energy of -43.1±7.2 kcal/mol and top

binding had the lowest binding energy (-42.8±4.1kcal/mol). BRACO19 bound to the

hybrid G-quadruplex DNA in three binding poses as well. Top binding (-40.5±5.4

kcal/mol) was the most stable of the three, followed by groove binding (-35.7±5.1

kcal/mol) and bottom binding (-29.0±12.9 kcal/mol).

Table 3

MM-GBSA binding Energy (kcal/mol) of BRACO19 to human telomeric DNA duplex and Quadruplexes.

System	Pose	$^1\Delta E_{VDW}$	$^2\Delta E_{SUR}$	$^3\Delta E_{GBELE}$	$^4\Delta E_{CONF}$	$^5\Delta E_{TOT}$	$^6\Delta\Delta E_{TOT}$
DNA Duplex	Top	-28.9±4.0	-2.3±0.5	-5.2±3.1	2.7±1.2	-33.7±5.3	28.6
	Bottom	-28.4±4.1	-2.2±0.5	-4.8±3.2	0.8±3.8	-34.6±5.7	27.7
	Groove	-57.9±9.5	-5.2±0.7	-3.9±4.5	5.2±2.8	**-61.7±8.0**	**0.6**
Parallel Quad	Top	-63.1±5.7	-5.2±0.6	-1.9±4.4	7.9±5.2	**-62.3±4.5**	**0**
	Bottom	-44.5±3.3	-3.79±0.2	-11.35±2.0	-2.2±3.9	**-61.8±1.5**	0.5
	Groove	-37.0±6.4	-3.1±0.4	-8.9±4.5	11.4±4.5	-37.6±7.2	24.7
Anti-Parallel Quad	Top	-29.1±9.0	-2.5±0.8	-9.2±3.0	-2.1±4.4	-42.8±4.1	19.5
	Bottom	-41.5±11.4	-4.0±1.1	-8.5±4.3	0.1±4.6	**-53.9±5.8**	**8.4**
	Groove	-43.0±6.0	-3.4±0.5	-7.5±2.6	10.9±2.3	-43.1±7.2	19.2
Hybrid Quad	Top	-44.2±11.4	-4.3±1.0	-12.1±5.0	20.0±9.2	**-40.5±5.4**	**21.8**
	Bottom	-25.7±5.8	-2.8±0.7	-16.3±6.0	15.8±8.5	-29.0±12.9	33.3
	Groove	-40.5±6.6	-4.0±0.5	-14.9±5.2	23.7±3.6	-35.7±5.1	26.6

2.3.3 Binding of BRACO19 to the duplex DNA. BRACO19 binds to the groove of the duplex DNA, without inducing appreciable DNA structure fluctuation. The representative trajectories for the three major binding modes of BRACO19 to the human telomeric duplex DNA (Figure 4 and Figure A19). In all ten binding trajectories, the DNA showed low structural fluctuation with a RMSD of 2.4 Å (Figure A2) where the hydrogen bonds between the base pairs were maintained. In the representative trajectory of BRACO19 binding to the groove of the human telomeric duplex DNA in Figure 4, an initial

interaction was observed as early as 3 ns and the final binding pose was achieved at ~14 ns which was maintained throughout the remainder of the trajectory. The limited fluctuation in the five order parameters explained the limited structural dynamics. The other representative trajectory of BRACO19 groove binding (Figure A19) also exhibited rapid binding and limited dynamics.

34
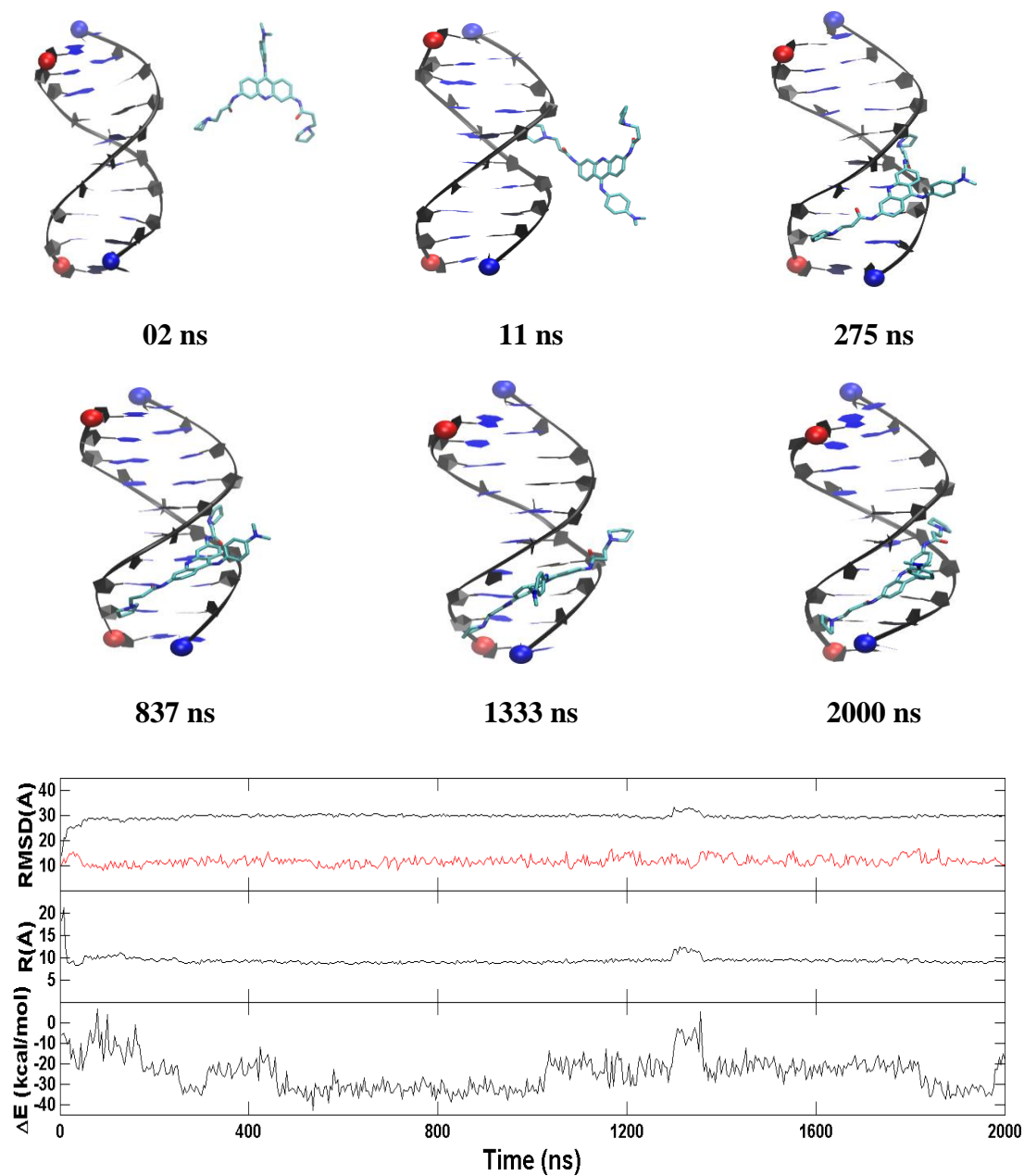
Figure 4. A representative groove binding trajectory of the duplex DNA. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. **Bottom:** receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

2.3.4 Binding of BRACO19 to the parallel G-quadruplex. The representative trajectories for the two major binding modes of BRACO19 to the parallel human telomeric G-quadruplex DNA (Figure 5 and A20) (the top stacking mode) and Figure A21 (the bottom stacking mode). In all ten binding trajectories, the DNA showed low structural fluctuation with RMSD of 2.4 Å (Figure A4) and the hydrogen bonds in the three G-tetrads were maintained. In the representative trajectory of BRACO19 binding to the top of the human telomeric parallel G-quadruplex DNA in Figure 5, an initial interaction was observed as early as 2 ns. BRACO19 underwent several top pose adjustments until ~750 ns when the final binding pose was achieved and maintained throughout the remainder of the trajectory. Despite the repositions of BRACO19 in the early portion of the simulation, there were limited fluctuations in the order parameters. The other representative trajectory of the top stacking mode of BRACO19 binding to parallel G-quadruplex (Figure A20) also exhibited quick binding and limited dynamics, with the early interaction to the complex at 4 ns and attaining the stable top binding pose at 25 ns where it maintained the top staking pose with minor repositions until 1391 ns where it remained for the length of the trajectory. The representative trajectory for the bottom binding pose (Figure A21) achieved the final binding pose within 10 ns and displayed high stability as indicated by the limited fluctuations in the order parameter plot. The binding energy for top/bottom stacking fluctuated between -60 and -75 kcal/mol while groove stacking varied between -35 and -45 kcal/mol after attaining a stable binding pose.
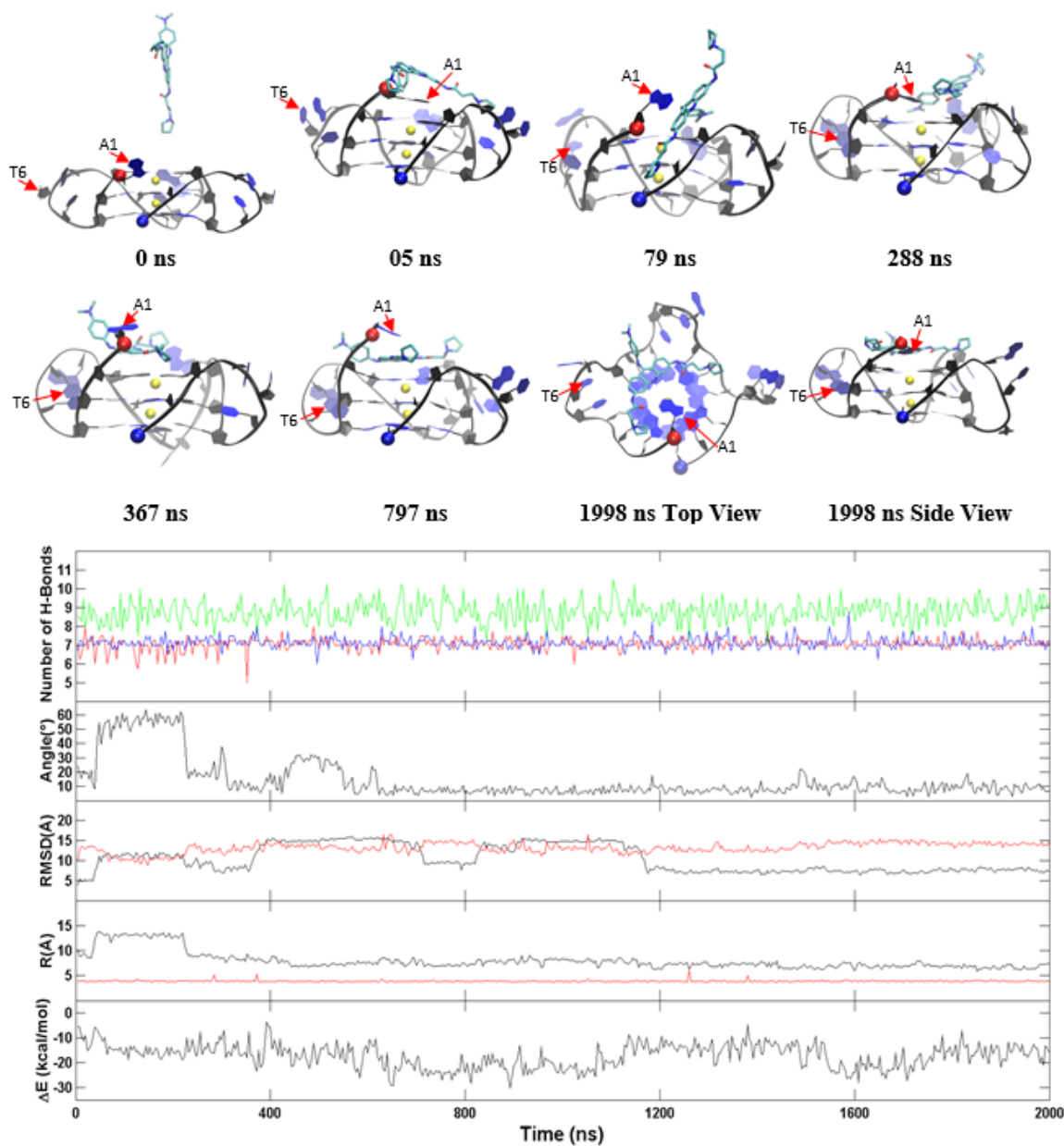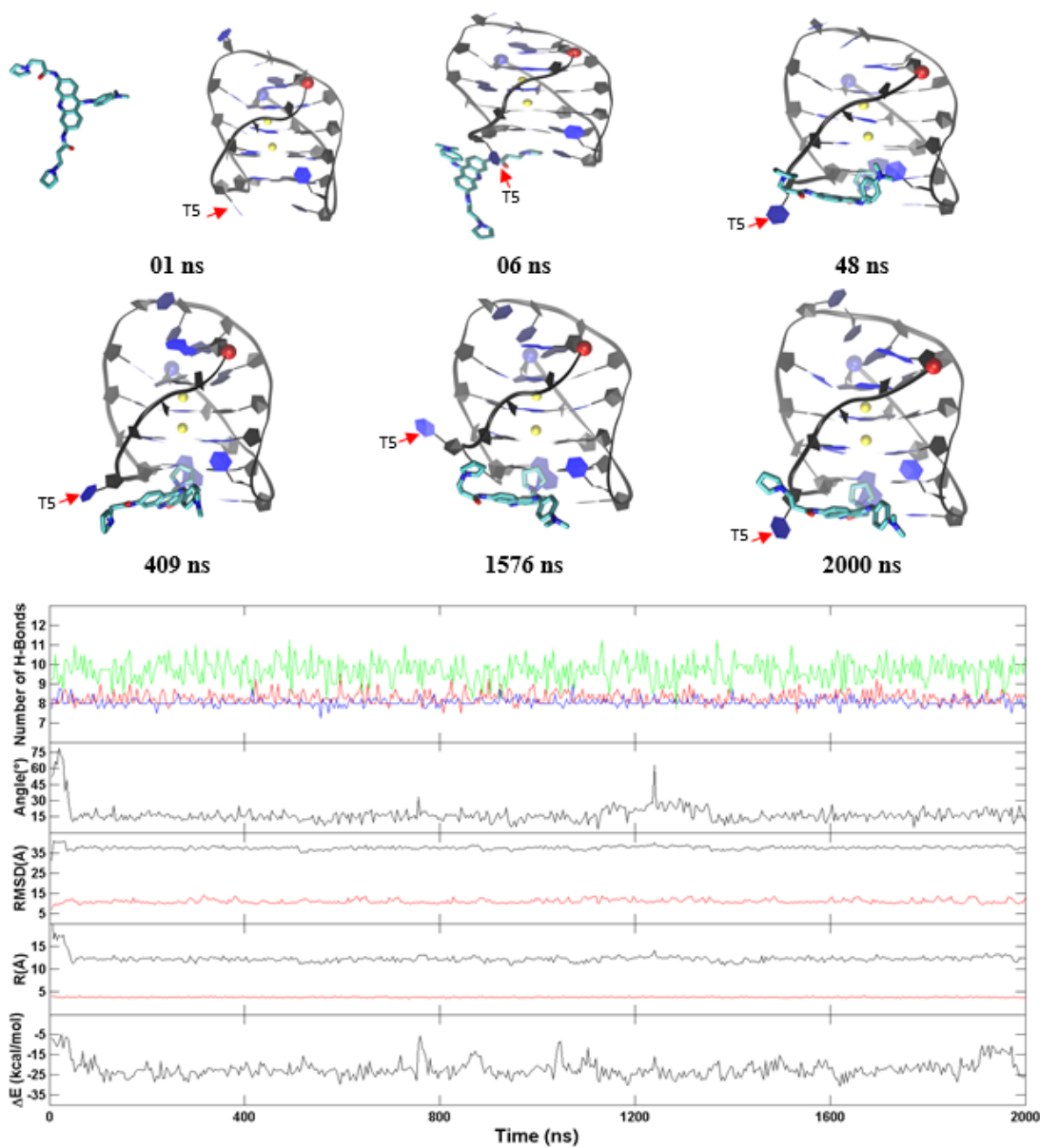
Figure 5. A representative top stacking trajectory of the parallel G-quadruplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first G4 (green), second G4 (red) and third G4 (blue) tetrads of the DNA structure (Figure 2), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

Major fluctuations were observed in the terminal residues, however T6 in particular is discussed here as it demonstrates highest fluctuation. T6 flipped out at 15 ns and flipped back at 45 ns, flipped out at 69 ns and flipped in at 100 ns and it finally flipped out at 114 ns and remained same throughout the rest of the trajectory. This flipping out of the base is mainly characterized by α, β, γ and χ (Figure 6). Another example of BRACO19 binding to the parallel scaffold facilitated by base flipping is illustrated in Figure A30A which shows the terminal residue A1 clearly flipping outward which provided adequate space for BRACO19 to bind to the top G-quartet, closest to the 5' end.
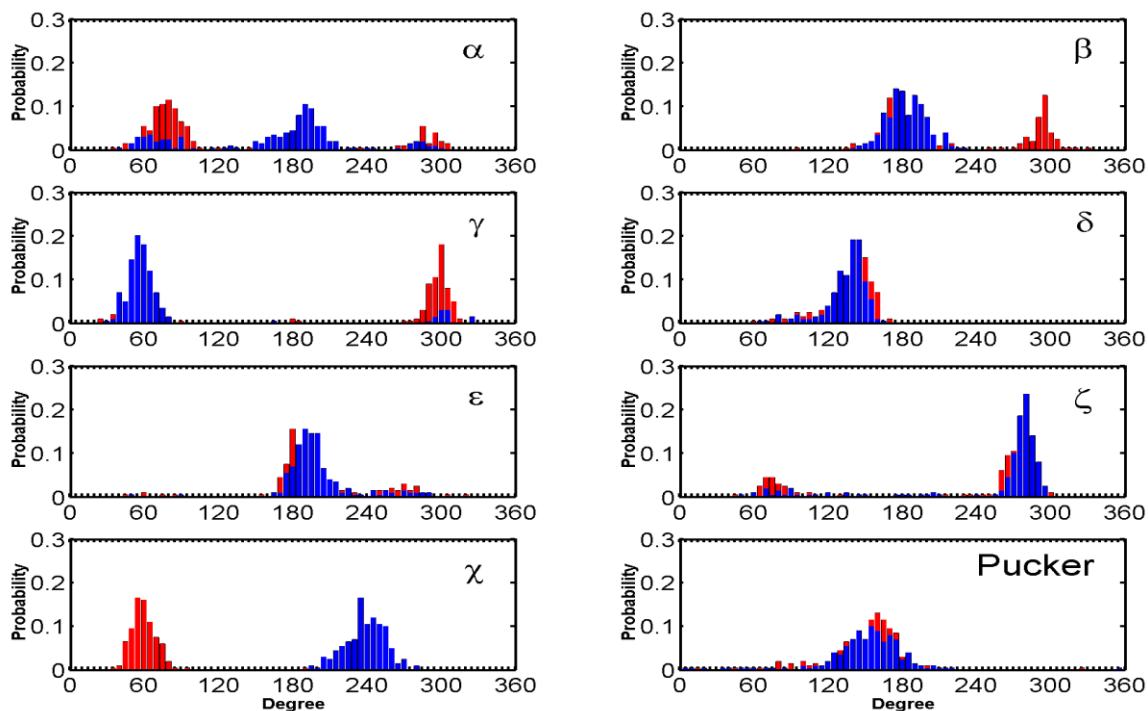

Figure 6. Histograms comparing the backbone torsion angles of residue T6 between the free ligand binding simulation (red) of the top stacking mode of the parallel G-quadruplex and the stability simulation of the crystal structure (blue) of the parallel G-quadruplex within last 200 ns.

2.3.5 Binding of BRACO19 to the anti-parallel G-quadruplex. The representative trajectories for the three major binding modes of BRACO19 to the anti-parallel human telomeric G-quadruplex DNA are characterized in Figure 7 and A23 (the bottom stacking mode) and Figure A22 (the top binding mode). In all ten binding trajectories, the DNA showed high structural fluctuation in four trajectories with RMSD of 3.2 Å (Figure A6), the hydrogen bonds in the three G-tetrads were maintained and the distance between K^+ ions remained stable in all trajectories. The representative trajectory of the bottom stacking mode of BRACO19 (Figure 7) on the human telomeric anti-parallel DNA G-quadruplex showed an initial interaction at 5 ns. The final binding pose was achieved within 48 ns and was maintained throughout the rest of the trajectory. The limited structural dynamics were explained by the limited fluctuation in the five order parameters. The representative trajectory for the top binding mode (Figure A22) is similar to the bottom binding trajectories with a rapid binding and limited fluctuation of order parameters with first interaction at 5 ns and attainment of the final binding pose by 55 ns. The binding energy for bottom stacking and groove binding fluctuated between -55 and -65 kcal/mol while top stacking varied between -40 and -50 kcal/mol after attaining a steady binding pose.

39

Figure 7. A representative bottom stacking trajectory of the anti-parallel G-quadruplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first (red), second G4 (cyan), third G4 (blue), fourth G4 (black) and fifth (green) layers of the DNA structure , the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

The dihedral angles of the G-tetrads in free ligand binding simulations indicate low fluctuations and are consistent through the binding process. Major fluctuations were observed in the terminal residues, T5 in particular is discussed here as it demonstrates highest fluctuation. Through the binding process BRACO19s major interaction was with T5, which opened up as BRACO19 approached and at 29 ns flipped out to let BRACO19 in and flipped back at 40 ns and staying open after interacting with BRACO19. This flipping out of the base is mainly characterized by ε and ζ (Figure 8). Another example of this base flipping for the anti-parallel topology is illustrated in Figure A30B, where BRACO19's major interaction is with base A7. As a result of BRACO19's interaction with base A7, the base T5 flips upward allowing base A7 to flip to the outside of BRACO19 maximizing the binding interactions between the G-quadruplex and BRACO19. The base T5 remains flipped up for the remainder of the trajectory and the interaction where A7 is partially intercalating BRACO19 onto the G-quadruplex is also maintained.

41
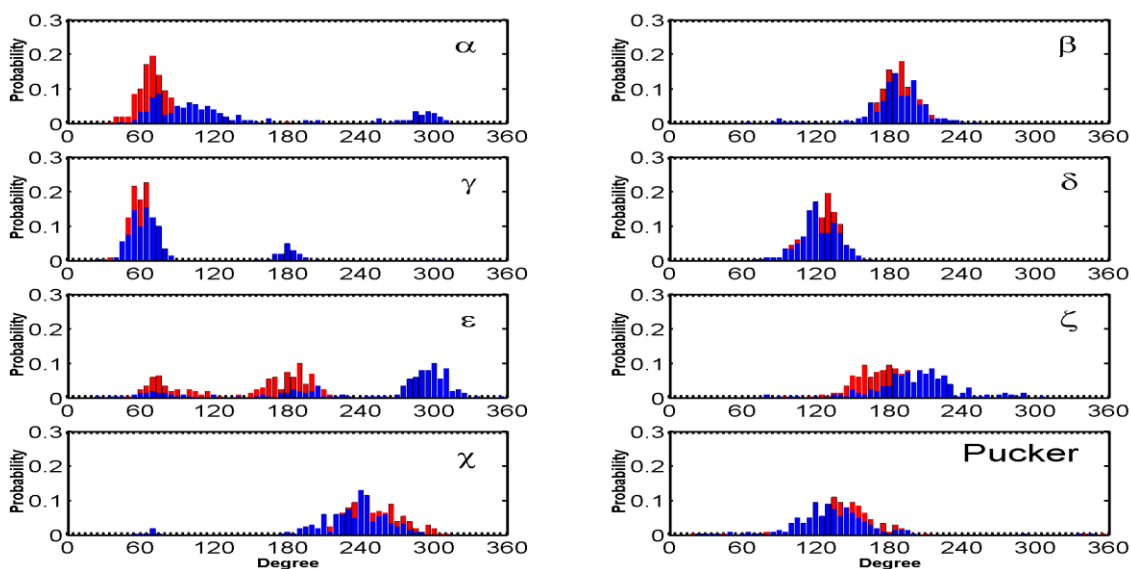
Figure 8. Histograms comparing the backbone torsion angles of residue T05 between the free ligand binding simulation (red) of the top stacking trajectory of the anti-parallel G-quadruplex and the stability simulation of the crystal structure (blue) of the anti-parallel G-quadruplex.

2.3.6 Binding of BRACO19 to the hybrid G-quadruplex. The representative trajectories for the three-major binding modes of BRACO19 with respect to the hybrid human telomeric G-quadruplex DNA are characterized in Figure 9 and A24 (the top binding mode), Figure A25 (the bottom binding mode) and Figure A26 (the groove binding mode). Out of the ten binding trajectories, the DNA showed high structural fluctuation in five trajectories with RMSD of 2.9 Å (Figure A8), the hydrogen bonds in the three G-tetrads were maintained and the distance between K^+ ions remained stable in all trajectories. The representative trajectory of BRACO19 top stacking onto the hybrid G-quadruplex DNA showed an initial interaction at 3 ns, the final binding pose was attained as early as 30 ns and was maintained throughout the rest of the trajectory. The limited structural dynamics were explained by the limited fluctuation in the five order parameters.

The representative trajectories of the bottom (Figure A25) and groove binding (Figure A26) are similar to the top binding trajectories in rapid binding and limited fluctuation of order parameters. Early interaction of BRACO19 with the quadruplex was at 9 ns and 2 ns respectively and the final binding pose was attained by 51 and 13 ns respectively. The binding energy for all binding modes varied between -55 and -65 kcal/mol after attaining the steady binding pose.

43
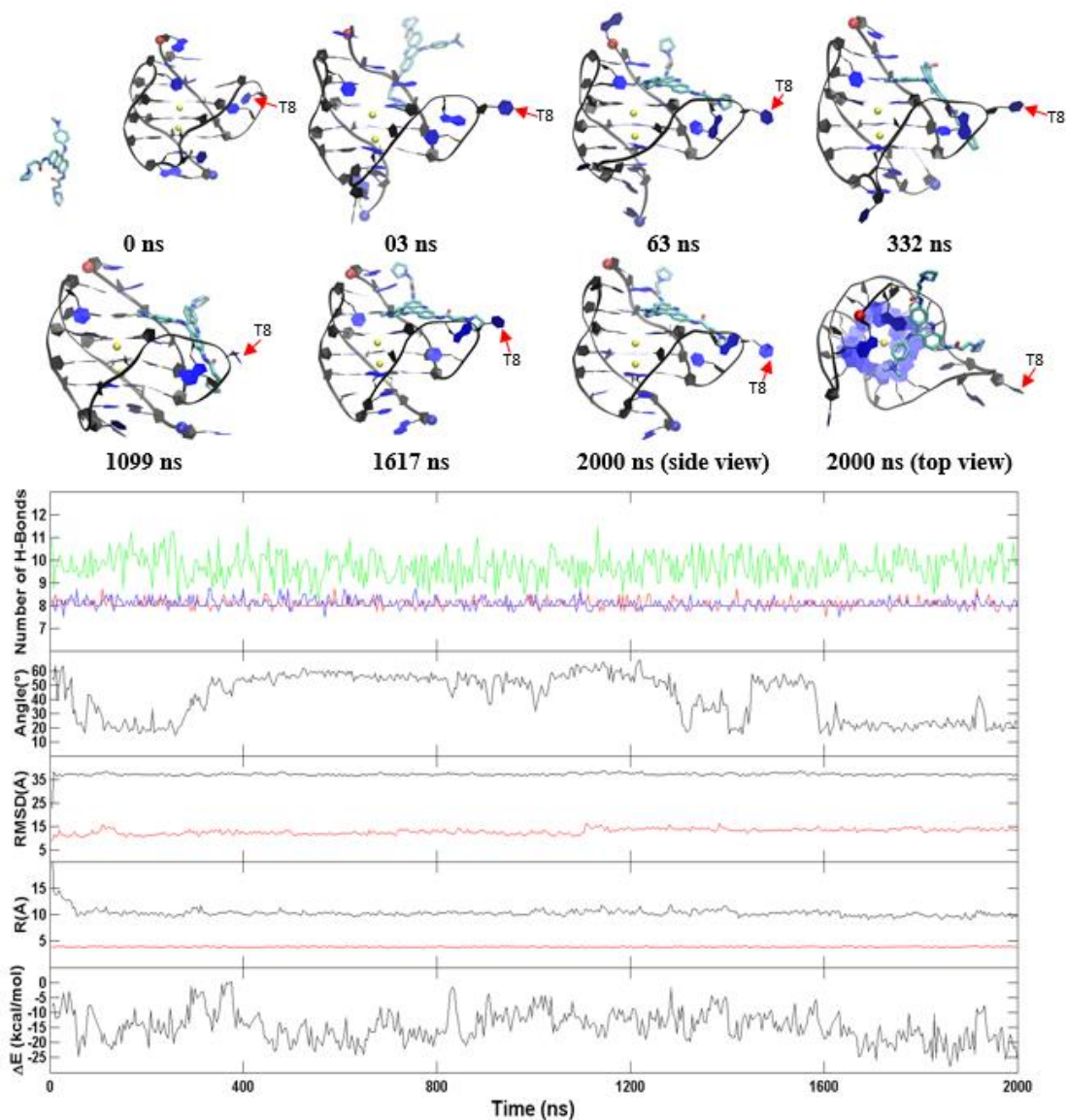
Figure 9. A representative top binding trajectory of the hybrid G-quadruplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. K$^+$ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first (red), third G4 (blue), and fifth (green) layers of the DNA structure, the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K$^+$-K$^+$ distance (R/red) and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

The dihedral angles of the G-tetrads in free ligand binding simulations indicate

low fluctuations and are consistent through the binding process. Major fluctuations were

observed in the terminal residues, T8 in particular is discussed here as it demonstrates highest fluctuation. T8 flipped out upon simulation and remained flipped through a majority of the simulation. This flipping out of the base is mainly characterized by α, δ, ε and ζ (Figure 10). Another example for the hybrid scaffold is presented in Figure A30C. The initial binding of BRACO19 was to the groove of the G-quadruplex before interacting with the 3' terminal residue which appeared to steer BRACO19 toward a bottom binding interaction. In this trajectory, both the 3' terminal residue A23 as well as loop residue T13 flip upward (clear in the 48 ns snapshot) which allowed BRACO19 to bind to the bottom of the G-quadruplex. Bases A23 and T13 made slight adjustments in their position for the remainder of the trajectory, whereas BRACO19 remained stably bound to the bottom of the G-quadruplex.
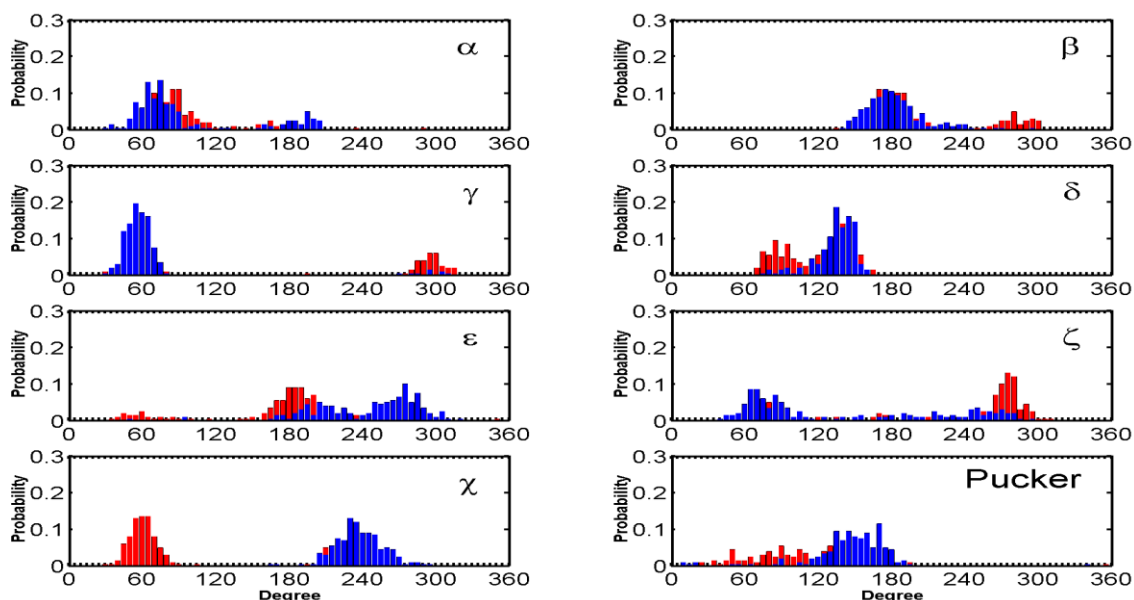
Figure 10. Histograms comparing the backbone torsion angles of residue T8 between the free ligand binding simulation (red) of the top binding trajectory of the hybrid G-quadruplex and the stability simulation of the crystal structure (blue) of the hybrid G-quadruplex within the last 200 ns.

2.4 Discussion

After the discovery of the greater existence of G-quadruplexes in malignant tumors than in normal tissues, the interest in G-quadruplex DNA as a target for cancer therapeutics has increased. BRACO19, an effective G-quadruplex stabilizing ligand, is a promising anticancer drug candidate, yet its low preferential binding affinity (about ~40-fold) to the telomeric single-stranded G-quadruplex DNA over duplex DNA remains to be enhanced. For better molecular insights, the binding of BRACO19 to a duplex 20mer DNA $(d([GC]_{10})_2)$ and to the parallel, anti-parallel and hybrid telomeric G-quadruplexes were investigated in this study using free ligand binding molecular dynamics simulations. Out of various binding modes for each system, the MM-GBSA binding energy calculations showed that the most stable binding pose was the groove binding mode for the duplex, the top/bottom stacking mode for the parallel G-quadruplex, the bottom stacking mode of the anti-parallel G-quadruplex and the top stacking mode of the hybrid G-quadruplex (Table 3). The order of the relative binding energy of BRACO19 to these DNA forms are as follows: -62.3 ± 4.5 kcal/mol of the top stacking to the parallel G-quadruplex ($\Delta\Delta E=0$ kcal/mol), -61.7 ± 8.0 kcal/mol of the groove binding to the duplex DNA ($\Delta\Delta E=0.6$ kcal/mol), $-53.9.4\pm5.8$ kcal/mol of the bottom stacking to the anti-parallel G-quadruplex ($\Delta\Delta E=8.4$ kcal/mol) and -40.5 ± 5.4 kcal/mol of the top stacking to the hybrid G-quadruplex ($\Delta\Delta E=21.8$ kcal/mol). For all the systems, breaking down the binding energy indicated that the VDW term makes the biggest contribution to the total binding energy (Table 3). This indication suggests introducing target or drug specific packing optimization as a prospect for further stabilization of the G-quadruplex. A limitation of the MMPBSA binding energy calculations are that they do not include the conformational changes

46

involved in the folding process of the G-quadruplexes nor do they consider the relative stability of the different scaffolds. Because of this, MMPBSA calculations alone may not be sufficient enough to predict the most favorable scaffold under physiological conditions.

There are the two lines of conflicting evidences on major target form of BRACO19: 1). under the solution condition with cellular extracts as crowding agents, the more predominate conformations are likely the anti-parallel and/or the hybrid topologies. 2). There is no high-resolution complex structures of BRACO19 binding to antiparallel or the hybrid scaffold, except for parallel stranded. Our binding energy data suggest a hypothesis that reconcile the conflict: the relative population shift of three scaffolds upon BRACO19 binding (i.e., an increase of population of parallel scaffold, a decrease of populations of antiparallel and/or hybrid scaffold). This hypothesis appears to be consistent with the facts that BRACO19 was specifically designed based on the structural requirements of the parallel scaffold and has since proven effective against a variety of cancer cell lines as well as toward a number of scaffolds.

With groove binding predicted to be the least energetically favorable, and based on our visual inspection of each trajectory, our data suggests that BRACO19's groove binding pose is likely not stable enough to maintain a prolonged binding event and that under a more extended timeline the groove binding mode may function as an intermediate state preceding a more energetically favorable end stacking pose. To support this, Figure A29 A-C provides representative snapshots of three simulation runs from each G-quadruplex system are presented. As for the anti-parallel system, we attribute the comparable top and groove binding poses to the anti-parallel topology. Based on our observations the diagonal loop (T11, T12, A13) atop the G-quartet, closest to the 5'

47

terminal, obstructs BRACO19's ability to achieve a stable stacking pose on the top G-quartet. Therefore BRACO19's major interaction with the top of the anti-parallel G-quadruplex is with the TTA diagonal loop, which offers no benefit over the groove binding pose in terms of binding interaction.

If these binding modes have comparable entropic energies then our relative binding energies suggest that BRACO19 binds preferentially to the parallel G-quadruplex over the anti-parallel and the hybrid G-quadruplexes if assuming equally abundant conformations in physiological condition. In the same way, our relative binding energy data shows that BRACO19 binds preferentially to the telomeric parallel G-quadruplexes over the DNA duplex. This qualitatively explains the experimental observation of weak preferential binding affinity difference of BRACO19 on the two DNA forms (40 fold of the selectivity). For that reason, it can be suggested that a ligand modification that destabilizes the duplex groove binding mode but stabilizes the G-quadruplex top stacking mode will enhance the binding selectivity of the ligand. For example, adding a rigid planar ring fragment to the acridine may facilitate top stacking rather than groove binding and increase the van der Waals interactions in turn increasing selectivity and binding affinity of the prospective drug towards the G-quadruplex. This suggestion is consistent with the original SAR data in the development of BRACO19 from prototype BSU6048 in which the addition of the ring at position 9 (makings of BRACO19) increased the drug selectivity from 10-fold to 40 fold towards human telomeric G-quadruplexes over duplex DNA [57, 109, 110]. The addition of the methylated anilino group, as opposed to the hydrogenated aniline at the 9th position slightly decreased binding to the duplex, while maintaining binding to the G-quadruplex [111]. It is also to be noted that the sidechains on 3 and 6 contribute to the groove binding of both

48

DNA duplex and G-quadruplex which could be the reason behind low selectivity. We observed the side chains of the 3rd and 6th position to play a role in BRACO19's ability to fully bind to the groove of the duplex DNA, but the side chains appear to play less of a role in G-quadruplex binding; which we primarily observed as an interaction with the acridine core. Thus, suggestions can be made to reduce the length of these side chains. These side chains exist in protonated form at physiological pH however, Table 3 indicates that the contribution of electrostatic interactions to the binding affinity is very low and therefore modifications can be suggested to the substituents at 3rd and 6th position of the acridine. Modifications such as loss of positive charge which would increase the hydrophobicity which could in fact increase the van der Waals interactions and reduction of the length of the side chains.

Encouragingly, the binding pose of BRACO19 to the parallel human telomeric G-quadruplex is consistent with to the only available crystal structure of BRACO19 (PDB ID: 3CE5) (Figure 11). In both the crystal structure and the structure from our study, the acridine core binds to the G-quartet closest to the 3' terminal with the nitrogen from the acridine core facing inward in-line with the K^+ cations. The anilino group at the 9th position faces away from the G4 core and the two substituents at the 3rd and 6th position are also positioned outward, allowing the drug to remain planar and stack onto the G-quartet. Although a top binding pose was not reported in the crystal structure of the double stranded parallel topology, our study suggests due to the symmetry of the single stranded parallel G-quadruplex topology both the top and bottom of the G-quadruplex core offer comparable binding opportunities for BRACO19. In support of this, our MM-GBSA analysis (Table 3) showed the most energetically favorable top and bottom

binding modes were within 0.5 kcal/mol; where the top binding pose (-62.3±4.5

kcal/mol) was slightly more favorable than the bottom binding pose (-61.8±1.5 kal/mol).

Our study also clearly showed BRACO19 in a top binding pose closely matching the

description published for the bottom stacking in the crystal structure (Figure 11).

Together this provides evidence to support that both end stacking modes could offer

equal binding for BRACO19 in the single stranded parallel scaffold of the human

telomeric DNA G-quadruplex. In addition to this, a crystal structure of a Pt-tripod in

complex with the hybrid DNA G-quadruplex sequence was recently solved (PDB ID:

5Z80), which shows binding to the top of the G4 similar to the binding pose observed in

our study (Figure A28).
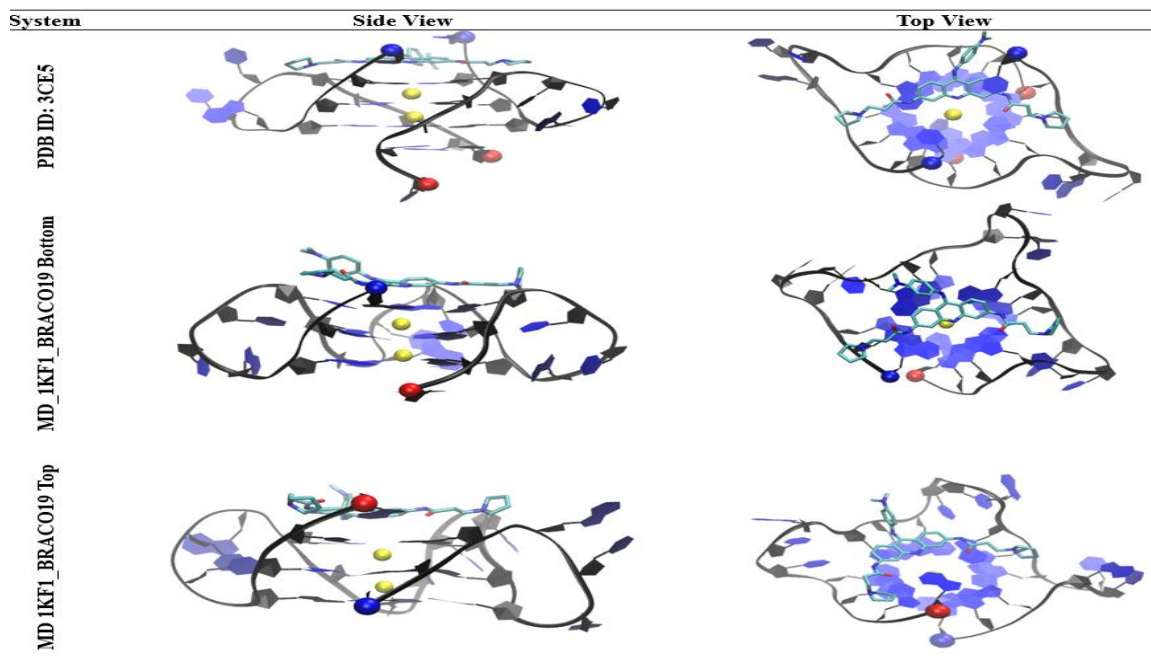
Figure 11. Comparison of the Double Stranded parallel G-quadruplex-BRACO19 complex (PDB ID: 3CE5) and both the Bottom and Top Binding Modes of the Single Stranded parallel G-quadruplex-BRACO19 complex. The 5' residues are represented by a red ball and the 3' residues in blue.

50

Although longer simulation periods are required to confirm, evidence of an induced fit binding mechanism was observed in each of the BRACO19-G-quadruplex complexes. The representative trajectories in Figures 5, 7 and 9 show one example per system where a base flipping mechanism worked to enhance the binding of BRACO19 to the receptor. The base flipping mechanism was ongoing and recurrent through conformational changes that occurred during the binding event. Despite an absolute equilibrium not being reached under the restricted simulation period, the persistent use of the base flipping mechanism and the resultant beneficial binding interactions, as observed over the timeline of the MMGBSA energy analysis, suggest the potential use of an induced fit binding mechanism facilitated by base flipping. We also provide a detailed analysis of the dihedral angles of the residue showing the largest fluctuation in each system compared to the apo form. The dihedral angles helped to characterize the changes of the bases that may contribute to an induced fit binding mechanism. Including a description of the binding events of a second example from each system illustrated in Figure A30. This figure demonstrates three important characteristics that suggest the use an induced fit binding mechanism used by BRACO19. For the parallel topology (Figure A30A), the flipping out of the 5' terminal base A1 led to the repositioning of BRACO19 on top of the top G-quartet. We observed two possible outcomes for the mechanism involving 5' terminal base flipping: (i) the 5' base will flip back on top of BRACO19 intercalating it onto the top G-quartet or (ii) the 5' terminal base will position itself in plane with BRACO19 and base pair; both mechanisms enhance the binding interactions between BRACO19 and the DNA G-quadruplex. The anti-parallel DNA G-quadruplex (Figure A30B) provides an example of two bases from the same loop changing position in order to enhance the binding of BRACO19 to the DNA G-quadruplex.

51

In this case, the flipping upward of base T5 during the simulation run allowed base A7 to flip outward and reposition itself around the outside of BRACO19 so that BRACO19 was partially intercalated to the groove of the DNA G-quadruplex which maximized its binding interactions. The hybrid topology (Figure A30C) provides an example of both a terminal and loop residues flipping outward to allow BRACO19 to reposition into a binding pose that enhances its binding interactions. In this case, the 3' terminal residue A23 and loop residue T13 both flip outward allowing BRACO19 to stack to the bottom of the G-quadruplex. Residue A23 flips back on top of BRACO19 intercalating it while T13 remains flipped outward to provide sufficient room for BRACO19. Together with the analysis of the dihedral angles, the flexibility of both the terminal and loop residues -which through their conformational changes allow BRACO19 to positon itself in a more favorable binding pose and enhance its binding interactions- show characteristics of an induced fit binding mechanism. It was by use of the free ligand MD binding simulations, as opposed to rigid body docking, that we were able to observe the flipping of the terminal and loop bases during the binding process which we suggest are integral for BRACO19 to achieve the most favorable binding pose.

Chapter 3

To Probe the Binding of CX-5461, an Anti-Cancer DNA G-Quadruplex Stabilizer, to Human Telomeric, c-KIT-1, and c-Myc G-Quadruplexes and a DNA Duplex Using Molecular Dynamics Binding Simulations

3.1 Introduction

3.1.1 DNA G-quadruplexes as anti-cancer targets. The use of DNA G-Quadruplexes (G4s) as novel therapeutic targets has been a rapidly developing field over the last decade because compounds targeting the G4s have demonstrated a high potential against a variety of cancer cell lines. DNA G-quadruplexes demonstrate very useful characteristics as drug targets including high diversity, stability and much slower dissociation when compared to DNA duplexes. [112, 113] With well over 300,000 sequence motifs identified within the human genome, the design of small molecules targeting G4s as anti-cancer agents has become a primary focus of many researchers. [114-116]

3.1.2 CX-5461 as an anti-cancer agent. Small molecule CX-5461 (Figure 12) is a DNA G-quadruplex stabilizer whose structure contains a rigid benzothiazole-based core and two flexible side chains: methyl diazepane based (R1) and methylpyrazine based (R2). CX-5461 was designed for superior *in vivo* stability and pharmacokinetics currently in Phase I/II clinical trials for advanced hematologic malignancies and cancers with BRCA1/2 deficiencies[117]. CX-5461's first working mechanism was identified as G4 binder directly inhibiting the binding of RNA Polymerase I, which has implications for cancer therapeutics.[118] More specifically, by binding to ribosomal DNA (rDNA) G4s formed within the rDNA promoter, CX-5461 prevents the binding of the transcription factor, SL1,

53

and subsequently RNA Polymerase I, to the rDNA promoter which prevents the transcription of rDNA.[119, 120] Recently a second, unexpected, mechanism was identified for CX-5461 whereby it disrupts the cells replication fork by binding to and stabilizing chromosomal DNA G4 structures in cancer cells. Although experiments have yet to identify specific G4 targets for CX-5461 in the human genome, experimental evidence discovered CX-5461's specific roles at DNA G4s include the ability to selectively bind to and stabilize G4 structures of human cells lines *in vitro*, and increase the number of *in vivo* G4 structures. [117] These properties are extremely advantageous for cancer therapeutics, and evident from recent work, CX-5461 is a promising therapeutic agent for a variety of targets. In fact, as research expands, so do the number of potential targets for CX-5461 including solid tumors[121], acute myeloid leukemia[122, 123], multiple myeloma[124, 125], neuroblastoma tumors[126], prostate cancer[127], osteosarcoma[128], acute lymphoblastic leukemia[129, 130], epithelial ovarian cancer[131-133], arterial injury-induced neointimal hyperplasia[134], and even non-cancerous diseases such as cytomegalovirus[135, 136], Herpes Simplex type I virus[135], and African trypanosomiasis[137]. However, without an experimentally solved structure of CX-5461 in complex with any G4 structure the specific interactions associated with the binding of CX-5461 and ultimate stabilization of the G4 remains to be fully understood.
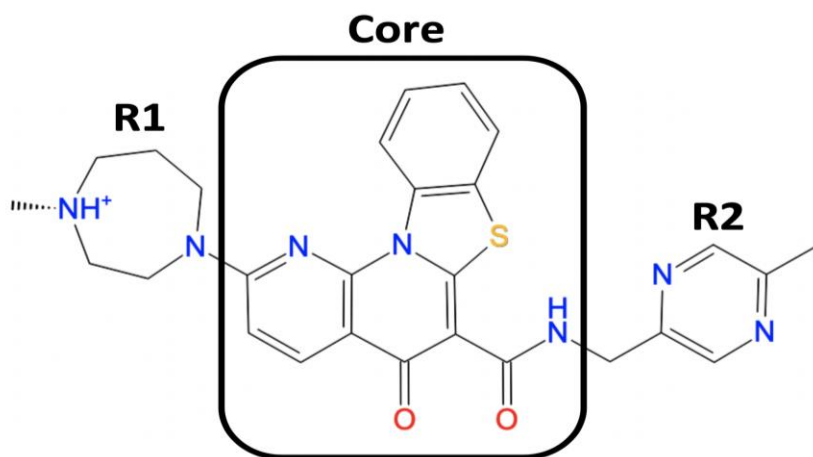
Figure 12. Chemical Structure of CX-5461. Two distinct regions are defined here where the region outlined in black represents the rigid core of CX-5461 and the two side chains (R1 and R2) represent flexible regions of CX-5461.

3.1.3 Experimental motivation. It is critical for G4 stabilizers to have a high binding affinity to G4's and demonstrate high selectivity over DNA duplexes to reduce the reverse effects. Experimentally, this has been shown using a DNA duplex as a negative control compared to several G4 targeting ligands which have effectively demonstrated a higher affinity and selectivity towards G4s over DNA duplexes.[112, 138] In one study, Xu et. al performed a FRET-melting temperature increase assay to test CX-5461's stabilizing effects to the canonical DNA duplex structure and three different G4 forming sequences which have been implicated in the cancerous complications resulting from BRCA1/2 mutations (human telomeric, c-KIT1, and c-Myc) [139-144]. Using the double stranded DNA duplex as a negative control to the G4 systems[117], the melting temperatures of each system was measured in the apo form. Then CX-5461 was added to each system to measure the increase in melting temperature upon CX-5461 binding to each DNA fragment, where a higher increase in melting temperature indicates a higher stabilizing effects thus higher binding affinity. The results of the FRET melting temperature assay indicated that with 10

55

µM CX-5461 the highest melting temperature increase was demonstrated by the human

telomeric system (~30 K) followed by the c-KIT1 (~27 K) and c-Myc (~25 K) G4s and the

DNA duplex (~10 K). Thus, these results show that the stabilizing effect due to CX-5461

binding was highest in the human telomeric complex followed by the c-KIT1 and c-Myc

G4s and then the DNA duplex. The difference in melting temperature increase between the

G4s and duplex complex systems (15+ K) suggest that CX-5461 can selectively bind to

and stabilize G4 structures over duplex DNA[117]. Along with the three G4 systems having

significantly higher melting temperatures than the DNA duplex, they also varied from each

other. Due to this, it is essential to compare the binding modes and mechanisms of CX-

5461 in complex with the G4s versus the DNA duplex to identify specific differences that

may help explain the higher binding affinity and selectivity demonstrated by CX-5641 to

the G4s over the duplex and the variance among three G4s.

3.1.4 Experiment design overview. Based on the three G4 forming sequences

used in the FRET melting temperature assay performed by Xu and coworkers[117], the

solved G4 scaffolds were obtained from the Protein Data Bank and used in this study.

These include a human telomeric G4 (PDB ID: 1KF1), c-KIT1 G4 (PDB ID: 4WO3), and

c-MYC promoter G4 (PDB ID: 2MGN)[145-147] (Figure 13). Due to the sequence difference

these G4s also vary in structure. The human telomeric DNA G4 (Figure 13A;E) is made

of four parallel DNA strands with three linking TTA trinucleotide loops which connect

the top of one stand to the bottom of another forcing the strands into a parallel

configuration which is highly symmetric. This parallel scaffold was chosen based on the

understanding gained in our previous work which showed that the human telomeric G4

forming sequences use a conformation-selection mechanism where the relative

56

population shift of the three possible scaffolds shift upon binding, resulting in an increase in the population of the parallel scaffold, and a decrease of populations of the antiparallel and/or hybrid scaffolds. [148] On the other hand, the c-KIT1 DNA G4 (Figure 13B;F) has an anti-parallel scaffold with double chain-reversal and a long lateral stem loop at the 3' region made of five nucleotides, two of which (A16 and G20) are capable of pairing. There is also one non-G-tract guanine that is part of the core of stacked G-quartets and the short single and dinucleotide loops of this c-KIT1 G4 are extremely flexible and show extensive base flipping. Whereas the c-Myc DNA G4 (Figure 13C;G) has a hybrid scaffold with a snapback motif that is adopted by the 3′-end GAAGG segment that forms a stable diagonal loop containing a G(A-G) triad and caps the 3' side of the G-tetrad. For our DNA duplex system we use a GC rich DNA duplex (Figure 13D;G), rather than using an oligonucleotide comprised of a polyethylene glycol linker able to fold into a hairpin as used in the FRET melting temperature assay by Xu and coworkers, which we feel is a more suitable comparison under physiological conditions.
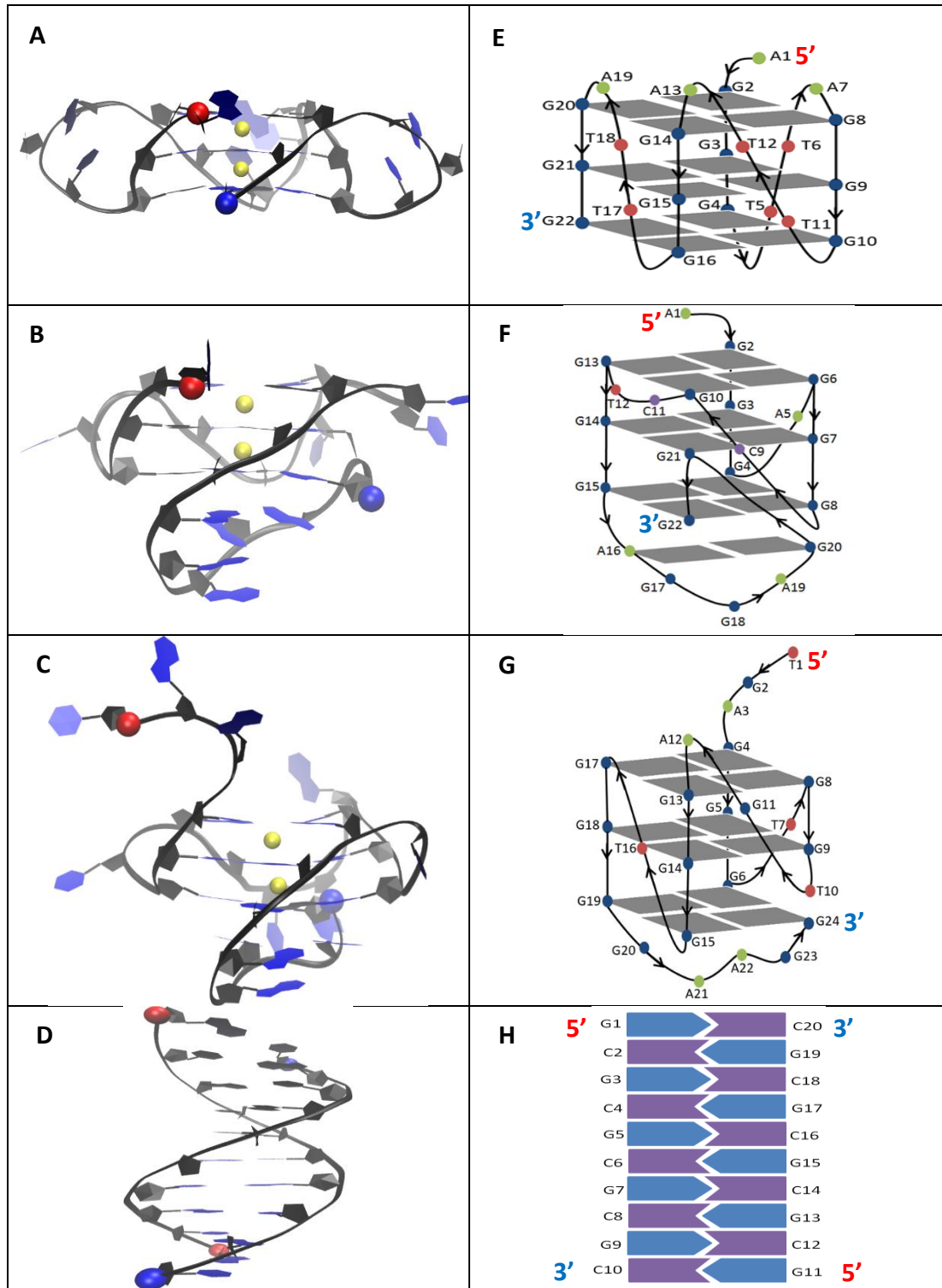
Figure 13. The initial configuration (A-D) and topology models of the simulation systems (E-H). **A&E:** Duplex DNA, **B&F:** Human telomeric quadruplex G4 (PDB ID: 1KF1), **C&G:** c-KIT1 quadruplex (PDB ID: 4WO3), and **D&H:** c-Myc quadruplex (PDB ID: 2MGN). 5' and 3' are indicated by red and blue spheres, respectively. K⁺ ions are represented in yellow.

58

3.1.5 Relevant studies. To date, a variety of studies have demonstrated using molecular modeling and simulations as a powerful approach to identify structural details at a molecular level. Hou et al. used this approach to probe the stability of six ligand-G-quadruplex DNA complexes structurally determined by experimental approaches [149]. Many of MD simulations studies are ligand binding studies that effectively provide mechanistic insight into the binding of small molecules to G4 DNA [150-155]. Information such as DNA-ligand binding free energy calculations, identification of ligand/G4 binding sites, and ligand binding modes were successfully determined using a modeling system that utilized the standard parm99 Amber force field using parmbsc0 parameters and a K^+ cation in the center of the G-tetrads to neutralize the system [152, 153, 156]. Deng et. al resolved ligand-binding specificity using absolute free binding energy calculations for both c-MYC [157] and human telomeric [158] G-quadruplex DNA . The work of Luo and Mu studied the binding of small molecules to human telomeric G-quadruplex using all-atomic molecular dynamic simulations [159]. Kumar and coworkers studied the binding of small molecules to G4 formed by the c-MYC promoter [147]. Some studies, like that performed by Chatterjee and coworkers have also had success performing an in silico screening on G4 structures formed by the c-MYC oncogene [160]. Further, the Lemkuls group has performed extensive work on the binding of small molecules to the c-KIT1 promoter G4 using molecular dynamics simulations [161-163]. In addition to small molecules targeting single G-quadruplexes, Praadeepkumar and coworkers studied the binding of small molecules that stabilize multiple G-quadruplex forming sequences including the c-MYC and c-KIT1 promoter G4s [164]. Modeling studies have also produced insight into a variety of G-quadruplex forming sequences. Research done on telomeric G4s have successfully calculated realistic

intermolecular and relative binding energies as well as determined binding modes and pathways. [155, 165, 166] Research performed by Liu and others extensively studied potassium binding with human telomeric intra-molecular G-quadruplex using molecular dynamics[167]. These studies provide invaluable insight into the model systems and strongly suggests the importance of using atomistic simulations to rationalize biologically relevant phenomena.[152, 165] Even more, many biological studies have been performed beyond computation that highlight the potential of targeting DNA G-quadruplexes with small molecules like CX-5461, however there are very limited complex structures available.

3.1.6. Experiment goals. With these facts in mind, the goal of this study is to use free ligand all atom molecular dynamics binding simulations to study the binding of CX-5461 to the human telomeric, c-KIT1 and c-Myc G4s and a DNA duplex. Our post simulation analysis will identify the major binding pose, binding mechanisms and binding pathways of the CX-5461 complexes and provide novel insight as to how CX-5461 has been experimentally shown to selectively bind to and stabilize these G4s. Through our analysis we also address the order of stability of each system and features that differentiate the binding of CX-5461 to the G4's and the DNA Duplex which help to understand the experimentally determined binding affinity and selectivity of CX-5461 to the G4 structures over the duplex. With the interaction insights, we propose optimizations to CX-5461 that may increase its interactions with G4s but decrease its interactions with the DNA duplex, which may improve its anti-cancer capability with less reversed effects.

60

3.2 Methods

3.2.1 Simulation methods overview. A full description of the methods used in this study is provided in the Supporting Information. In brief, a total of nine systems were constructed: three free ligand quadruplex-ligand complex systems (PDB ID's: 1KF1, 4WO3, 2MGN), one free ligand DNA duplex-ligand complex system, the DNA only form of each respective system, as well as a CX-5461 only system (Table 4). The DNA duplex-ligand system was constructed using a B-DNA duplex structure, sequence: $d([GC]_{10})_2$, using the Maestro program. The three DNA quadruplex-ligand systems were solvated inside a water box of truncated octahedron with 10 Å water buffer. Cl^- or K^+ counter ions were used to neutralize the system. The DNA fragments were represented using a refined version of the AMBER DNA OL15 (i.e., parm99bsc0[168] $+\chi_{OL4}$[169]$+ \varepsilon/\zeta_{OL1}$[170]$+ \beta_{OL1}$[171] updates). The water was represented using the TIP3P and the K^+/Na^+ model developed by Cheatham group was used to represent the K^+ ions.[172] The standard AMBER protocol was used to create the force field for CX-5461. This procedure included calculating the molecular electrostatic potential (MEP) of the CX-5461 molecule at the HF/6-31G* level after geometry optimization at the same theory level. Along with other parameters collected from the AMBERGAFF2[90] force field, the MEP was used to identify the partial charges of the CX-5461 atoms using Restrained Electrostatic Potential/RESP method with two stage fitting[173]. Using AMBER DNA force fields are a highly effective and widely used in nucleic acid simulations.[174-177] This experiment was able to simulate the binding process of the DNA G-Quadruplex (G4) stabilizer CX-5461, to a human telomeric G4 (PBD ID: 1KF1), a human c-KIT1 G4 (PBD ID: 4WO3), and a c-MYC promoter G4 (PBD ID: 2MGN),[145-147] as well as a B-DNA fragment. [178] The AMBER GAFF2 force field of CX-

5461 in Mol2 format is provided at the end of the supporting document. The detailed protocol for these simulations follow an earlier study;[179] where the AMBER 16 simulation package was used for the production runs of all four systems.[90] Following the Maxwell-Boltzmann distribution, atoms of the system were assigned different initial velocities by use of random seeds after the energy minimization. Thirty independent trajectories were run for each of the four complex systems to allow for better sampling of binding poses and pathway. In order to equilibrate the system density, a 500 ns production run at 300 K included a 1.0 ns molecular dynamics in the NPT ensemble mode (constant pressure and temperature). During this production run the DNA and the ligand were under Cartesian restraints (1.0 kcal/mol/Å), and 499.0 ns molecular dynamics in the NVT ensemble mode (constant volume and temperature). Two or three representative trajectories for each of four complex system were further extended into 1999.0 ns. A 2.0 fs time step in the simulations was created using SHAKE[180], which was able to constrain any bond connecting hydrogen atoms. 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}) were treated with the particle-mesh Ewald method.[181] The long range van der Waals interactions were based on a uniform density approximation; the cutoff distance for short-range non-bonded interactions was 10 Å. Non-bonded forces were calculated using a two-stage RESPA approach.[182] During this approach, the short-range forces were updated every step whereas the long range forces were updated every two steps. Using a Langevin thermostat with a coupling constant of 2.0 ps, the temperature was controlled and the trajectories were saved at 50.0 ps intervals for analysis.

62

Table 4

Molecular dynamics simulation runs.

ID[1]	DNA	No. Ligand	No. run	No. Water molecules	ions	Box Size (Å)*	Drug Initial Pose	NPT eq. (ns)	NVT (ns)	Total time (μs)
1	0	1	2	1420	1 Cl-	40.6	Free	1	1999	4
2	1(h-Tel)	0	2	4671	21 K+	60.9	N/A	1	1999	4
3	1(c-KIT1)	0	2	3954	21 K+	58.1	N/A	1	1999	4
4	1(c-Myc)	0	2	4843	23 K+	61.9	N/A	1	1999	4
5	1(Duplex)	0	2	4515	18 K+	55.4	N/A	1	1999	4
6	1(h-Tel)	1	27/3	8261	20 K+	72.4	Free	1	499 /1999	19.5
7	1(c-KIT1)	1	28/2	6371	20 K+	67.0	Free	1	499 /1999	18.0
8	1(c-Myc)	1	28/2	5958	22 K+	65.8	Free	1	499 /1999	18.0
9	1(Duplex)	1	28/2	5282	17 K+	57.0	Free	1	499 /1999	18.0

[1]Systems 1-4 refer to the free DNA-only systems, system 5 refers to the CX-5461 free ligand simulation, systems 6-9 refer to the free DNA plus free ligand simulations (6:9: Human telomeric (PDB ID: 1KF1), c-KIT1 (PDB ID: 4WO3), c-Myc (PDB ID: 2MGN) and Duplex complexes, respectively).
*Triclinic box equivalent to the true truncated octahedral box

3.3 Results

3.3.1. Convergence analysis for the apo form systems. To validate the force field used in our simulations, independent 2μs stability simulation runs carried out for each apo form. We performed a root mean square deviation (RMSD) analysis which compared the deviation of the DNA backbone in each snapshot to the initial structure. The flat RMSD values showed that the ligand only or the apo DNA remained stable in each simulation run and after taking the average of each run per system a figure is

63

presented in Figure B1. Nonetheless, the three G4 systems showed a larger structure deviation from the initial conformation (i.e., ~4Å of human telomeric G-quadruplex, ~3Å of c-KIT1 G-quadruplex and ~5Å of c-Myc G-quadruplex) due to the fluctuation of non-G-tetrad parts. Next, we analyzed the last snapshots of each apo form simulation run (Figure B2) and found that each the scaffold of each system was visually maintained when compared to the initial structure. A deeper understanding of the RMSD values were obtained through visual inspection of the apo form trajectories. For the human telomeric system, there appears to be high flexibility at the 5' terminal residue as well as for each of the three, three residue, connecting loops. The c-KIT1 system showed high flexibility at the 5' terminal residues as well as the four connecting loops, ranging from 2 to 5 residues long. The c-Myc system showed the greatest flexibility at the longer 5' terminal segment and also at the varied length connecting loops. The most notable difference of the last snap shots was for the c-Myc system, where some residues of the longer and highly flexible 5' terminal segment stabilized on the G4 core which was consistent in both runs, which is likely because the original PDB structure of this G4 contained a ligand and our simulation runs allowed the DNA to relax into an apo form. Following this analysis we took representative snapshots (Figure B3) for each of the apo DNA simulation systems: human telomeric (A), c-KIT1 (B), c-Myc (C) G-quadruplexes, and DNA duplex (D). Along with each representative trajectory is an order plot which shows that each system maintained a backbone RMSD, the potassium ions in each G4 system maintained their initial positions, and each systems MMPBSA energy relative to the initial snapshot showed small fluctuations but the average remained the same overall.

3.3.2. Convergence analysis for the holo form systems. A variety of post simulation analyses were performed to ensure proper sampling and convergence was reached in our simulations. First, we generated a plot showing the position of a single atom of CX-5461 in each trajectory (Figure B4). Due to the clear distribution of binding around the DNA in each system, we concluded there was a good sampling in each system. Following this, the root mean square deviation (RMSD) of the DNA backbone was calculated against the initial structure for the complex systems and the average of each system was calculated and presented (Figure B5). The flat and small receptor RMSDs in each system indicated the stability of the DNA structures during the simulation period. Next, atom contacts between the DNA structure and CX-5461 were calculated using a 3.0 Å cutoff (Figure B6). Here we defined a stable complex as a complex with a number of atom contacts between the DNA and ligand greater than 30. The stable contact number in this figure indicates the simulation systems reached a steady state in all runs. With our analyses suggesting proper sampling and convergence, we started looking at the binding poses. The last snap shots and a table summarizing each system's final binding poses are provided in the supporting document (Figures B7-B10; Table B1). An additional stability analysis characterizing each systems geometry was performed which will be discussed later (Table B2-B6).

3.3.3. Clustering analysis of the holo forms. Following the simulation runs, we performed a k-means clustering to extract the major binding modes of each system. Each system was grouped into three major structural families and represented by a centroid structure for which each snapshot was most similar. The centroid structures along with the structural families overall percentage is presented in Figure 14. In addition, two

65

dimensional interaction diagrams along with top and side views of the most abundant clusters of each system are presented in the supporting document (Figure B11 A-D) and Figures B12-B13 compare the top and groove clusters for each system, respectively.

For the human telomeric complex system the top binding mode accounted for 37.8%, the bottom binding mode 20.8%, and the side binding mode 36.9%. Due to the symmetry of the human telomeric G4, binding opportunities were roughly equal for the top and bottom binding modes. This is clear when comparing the similar major binding mode for the top and bottom modes and from the specific details gained from the two dimensional interaction diagrams in Figure B11A. In the top binding mode CX-5461's rigid core (Figure 12) is parallel to the 5' G-quartet and the entire CX-5461 molecule interacts with each residue of the 5' G-quartet (G^2, G^8, G^{14}, and G^{20}). In addition, it is further stabilized by a full intercalation by terminal reside A^1 and a hydrogen bond is formed between the nitrogen of flexible side chain R1 and residue G^8. In the bottom binding mode, CX-5461 also stacks parallel with the 3' G-quartet and interacts with each of the residues of this G-quartet (G^4, G^{10}, G^{16}, G^{22}). In addition to these interactions, flexible side chain R2 is further stabilized by two loops residues. First of which is residue T^5 which forms a partial intercalation around R2 and the second is residue A^7 which folds down toward the binding site and forms a hydrogen bond with a nitrogen of R2. In the groove binding mode, each loop also provides equal binding opportunity due to the identical sequence (TTA) and connections to the G4 core. From the most abundant pose you can see the $T^{11}T^{12}A^{13}$ loop is the primary interaction area where T^{12} forms a hydrogen bond with a nitrogen of R1. Although part of the rigid core of CX-5461 is able to stack inside a groove on the G4 core, it is evident that in this mode most of the core

66

and all of R2 is exposed to solvent and forms less interactions than in the top and bottom binding modes.

For the c-KIT1 complex system (Figure 14; Figure B11B) the top binding mode accounted for 53.5%, the bottom binding mode 24.0%, and the side binding mode 15.7%. The topology of the c-KIT1 G4 contains a short 5' segment and a flexible 5' face loop ($C^{11}T^{12}$) which often stack on top of the 5' g-tetrad. In the most abundant top biding pose for the c-KIT1 system, the rigid core of CX-5461 stacks on top the 5' terminal and interacts with all four residues of the 5' G-quartet (G^2, G^6, G^{10}, G^{13}). Further stabilization is provided through interactions with terminal residue A^1 which fully intercalates CX-5641 onto the G4 core. It also interacts with both residues of the 5' face loop where C^{11} flips upward to stabilize the interaction. The most abundant bottom binding mode shows a much different binding pose. The c-KIT1 G4 has a AGGAG 3' terminal loop for which A^{16} and G^{20} base pair directly under the 3' G-quartet and a smaller GGA loop is formed. In the most abundant bottom binding pose, the $A^{16}G^{20}$ base pair blocks CX-5461 ability to interact with the 3' G-quartet. Instead, CX-5461 interacts with each residue of the $G^{17}G^{18}A^{19}$ 3' loop where flexible side chain R1 folds upward to interact with the residue G^{18} through hydrogen bonding. In this mode, R1 also interacts with residue G^{20}, whereas R2 is entirely exposed to solvent and unbound. For the most abundant groove binding mode for the c-KIT1 system we observe a pose similar to the bottom binding mode. Here we see that one residue, C^{11}, of the 5' face loop flips outward and interacts with CX-5461 and T^{12} residue of this same loop also interact with CX-5461. In this mode we also see that while R2 is flipped outward and entirely exposed to solvent, R1 stacks along the G4 core, interacting with core residues (G^{10}, G^{13}, G^{21}, G^{22}).

67

For the c-Myc complex system (Figure 14; Figure B11C) the top binding mode accounted for 59.6%, the bottom binding mode 9.5%, and the side binding mode 26.7%. The topology of the c-Myc system has a several unique features. One of which is the highly flexible three residue ($T^1G^2A^3$) 5' terminal segment which adopts several positions including being flipped completely outward and stacked on top of the 5' G-quartet. The dynamic between these two positions is critical for the most abundant top binding pose for the c-Myc system. In this mode, CX-5461 is fully intercalated onto the 5' G-quartet due to the direct stacking of the G^2 residue of the 5' segment. In more detail we see that T1 is stacked on top of G^2 and that A^3 is flipped outward allowing its initial position to be occupied by CX-5461 while both residues still interact with CX-5461. Further, as mentioned this interaction allows for a full intercalation where CX-5461 also interacts with all four residues of the 5' G-quartet (G^4, G^8, G^{13}, G^{17}). One additional interaction occurs between loop residue A12 which flips upward to enhance the binding interactions of this pose. A second unique feature of the c-Myc topology is the four residue $G^{20}A^{21}A^{22}G^{23}$ 3' face loop which connects residues G^{19} and G^{24} of the 3' G-quartet, where G^{24} is the 5' terminal residue. Evident from the most abundant bottom binding pose, this 4 residue loop directly blocks CX-5461's access to the 3' G-tetrad which prevents bottom intercalation. Due to this, in the most abundant pose the flexible side chain R1 of CX-5461 is a stacking onto the DNA backbone of this loop where its positive charge is able to interact with residues G^{20}, A^{21}, A^{22} and G^{23}. However, R2 is completely exposed to solvent in this mode. In the groove binding mode, a similar binding mode to the bottom binding mode is observed where flexible side chain R1 folds upward and the

positively charged nitrogen hydrogen bonds with base T^7 while maintaining interactions with bases G^8 and G^9.

For the duplex complex system (Figure 14; Figure B11D) the top binding mode accounted for 3.8%, the bottom binding mode 4.5%, and the groove binding mode 25.7%. Although the top and bottom binding modes are reported to consistency, terminals of DNA are rare in the human body so we chose to focus our analyses on the groove binding mode due to its physiological relevance. Here we see the flexible side chain R1 of CX-5461 binding inside the groove of the duplex and forms a hydrogen bond with residue C^6, while R2 and the core are exposed to solvent.

Overall, it was clear that for each G4 system the top binding mode was most abundant and that in these poses, the DNA's interaction with rigid core of CX-5461 was most a key factor in stable binding. In the duplex system, groove binding was most abundant and the flexible side chains of CX-5461 were most involved in stable binding for this system. However, we feel the most significant differences in the major binding modes are attributed to structural differences and their impact on CX-5461 ability to interact with the G4 core. This was made clear when comparing the binding modes of the bottom binding poses to the more favorable top binding pose. Moreover, we observed very similar pattern of binding interactions between our G4 side binding and the duplex groove binding. With this in mind we wanted to explore energetic implications of these differences in binding and determine if our predictions about the binding interactions matched the binding energies.

69
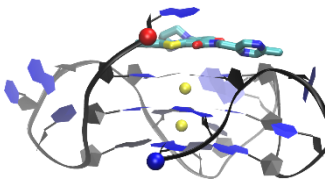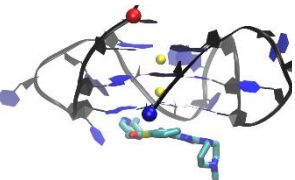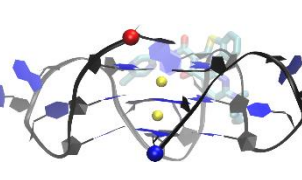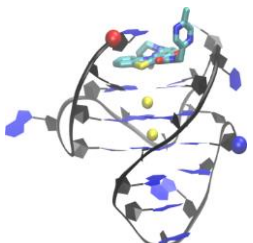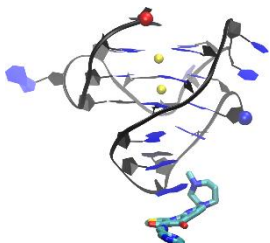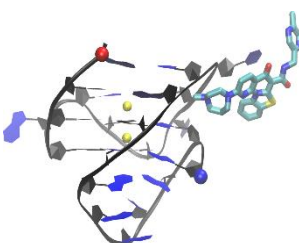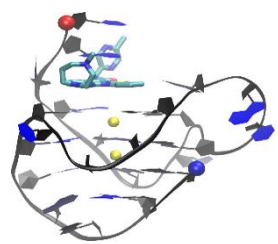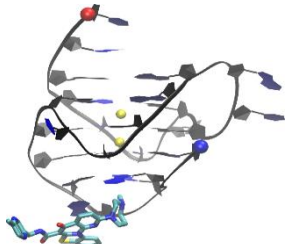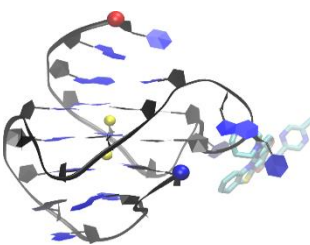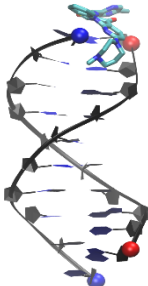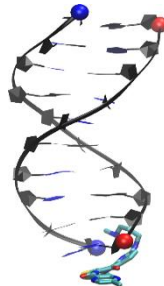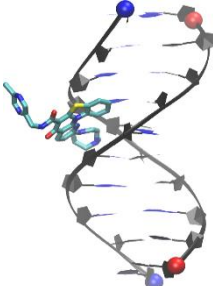
System	Top Stacking	Bottom Stacking	Side Binding
h-Tel			
	37.8%	20.8%	36.9%
c-KIT1			
	53.5%	24.0%	15.7%
c-Myc			
	59.6%	9.5%	26.7%
Duplex			
	3.8%	4.5%	25.7%

Figure 14. Major binding modes of CX-5461 binding to the human telomeric G4 (PDB ID: 1KF1), c-KIT1 G4 (PDB ID: 4WO3), c-Myc G4 (PDB ID: 2MGN) and duplex DNA.

70

3.3.4. MM/PBSA binding energy analysis. Of the four systems, the top intercalating mode of the human telomeric DNA G-quadruplex (1KF1) was shown to be the most energetically favorable binding mode (Table 5). Calculated to be -37.0 kcal/mol, this binding mode was -3.4 kcal/mol more favorable than the bottom binding mode (-33.6 kcal/mol) for the human telomeric G4 system, and -21.7 kcal/mol more favorable than the side binding mode (-15.3 kcal/mol).

For the c-KIT1 G4, the top binding mode, calculated to be -33.1 kcal/mol, was most energetically favorable of this system. This binding mode was -11.1 kcal/mol more favorable than the CX-5461's bottom binding pose (-22.0 kcal/mol) in the c-KIT1 G4 and -16.6 kcal/mol more favorable than the side binding pose (-16.5 kcal/mol). However when comparing c-KIT1's top binding pose to the human telomeric G4's top binding pose, the human telomeric G4 has a more favorable binding energy by -3.9 kcal/mol.

For the c-Myc G4 system, the top binding mode was also most energetically favorable calculated at -32.6 kcal/mol. The top binding mode was more favorable than the bottom binding mode (-15.5 kcal/mol) by -17.1 kcal and more favorable than the side binding mode by -14.8 kcal/mol. When comparing c-Myc's top binding pose to the human telomeric G4's top binding pose, the human telomeric G4 has a more favorable binding energy by -4.4 kcal/mol.

The duplex groove binding mode was the least energetically favorable, measuring -15.0 kcal/mol. This binding energy of the duplex groove mode was comparable to the side binding modes of the G4 systems: human telomeric (-15.3 kcal/mol), c-KIT1 (-16.5 kcal/mol), and c-Myc (-17.8 kcal/mol). The bottom binding mode of the c-Myc system (-

15.5 kcal/mol) was also comparable to the duplex systems groove binding mode. These results suggest that CX-5461 is selective to DNA G-quadruplex systems over DNA duplex because each of the most favorable binding poses for the G4 systems is at least twice as energetically favorable as the duplex groove binding mode. Further the less favorable MM-PBSA binding energy of the G4 systems suggests that side binding may be an intermediate state.

CX-5461's average MM/PBSA binding energy over the major poses were calculated for the human telomeric (-28.6 kcal/mol), c-KIT1 (-23.9 kcal/mol), c-Myc (-22.0 kcal/mol) G4s, and DNA duplex (-15.0 kcal/mol) systems. The binding energy was broken down into van der Waals (VDW), hydrophobic interactions (SUR), electrostatic interactions (GBELE), and the conformational energy change induced from the complex formation (CONF) (Table 5). From the table it is clear that the hydrophobic interactions (Δ_{SUR}) contribute the most to the total energy. As expected, the top intercalation binding modes have the most energetically favorable hydrophobic interactions and are much more favorable than the duplex systems groove binding mode.

72

Table 5

MM-PBSA binding energy of CX-5461 binding to the human telomeric G-quadruplex (1KF1), c-KIT1 G-quadruplex (4WO3), c-Myc G-quadruplex (2MGN), and duplex DNA in each of the major binding modes.

Sys.	BS	Δ_{VDW}[1]	Δ_{SUR}	Δ_{PBELE}	Δ_{CONF}	Δ_{TOT}	$\Delta\Delta E$[2]	ΔT_m[3]
h-Tel	T	-9.3±6.3	-32.9±3.1	10.5±1.3	-5.4±3.8	**-37.0±3.9**	0	30
	B	-3.5±5.9	-22.3±2.4	6.0±3.8	13.8±4.6	-33.6±0.7	3.4	
	S	-4.5±3.6	-23.7±2.0	11.3±2.3	1.5±8.0	-15.3±1.5	21.7	
c-KIT	T	-4.4±0.9	-25.4±0.5	8.2±0.6	-11.5±2.7	-33.1±0.7	3.9	27
	B	-2.0±1.6	-11.0±0.7	5.1±1.0	-14.1±2.3	-22.0±0.4	15	
	S	-3.0±1.9	-12.8±0.7	7.1±1.4	-7.8±11	-16.5±0.6	20.5	
c-Myc	T	-13.7±4.3	-35.2±1.5	16.4±0.7	-0.1±3.9	-32.6±4.7	4.4	25
	B	-4.3±0.8	-16.0±0.2	9.0±0.5	-4.3±4.7	-15.5±1.0	21.5	
	S	-5.3±6.2	-25.7±3.0	15.9±1.6	-2.7±1.4	-17.8±2.5	19.2	
Dup.	G	-2.1±41.2	-23.1±2.1	8.2±2.3	3.2±3.9	-13.8±3.9	23.2	10
	T	-4.3±0.8	-15.4±0.7	3.8±0.9	0.1±0.9	-16.1±2.7	20.9	
	B	-3.8±1.2	-15.9±1.8	4.6±2.3	0.02±0.8	-15.1±1.8	21.9	

[1] The parameters in this table are reported in units of kcal/mol.

ΔVDW = Change of VDW energy in gas phase upon complex formation

ΔSUR = Change of energy due to surface area change upon complex formation

ΔEBELE = Change of GB reaction field energy + gas phase Elec. energy upon complex formation

ΔCONF = Change of conormational energy upon complex formation

ΔTOT = ΔVDW + ΔSUR + ΔEBELE + ΔCONF Change of potential energy in water upon complex formation

Sys. refers to system and BS refers to binding site: Top (T), Bottom (B), Side (S), or Groove (G).

[2] $\Delta\Delta E$[2] = (Δ_{TOT} - (**-37.0**))

[3] Experimental melting temperature increase (in Kelvin) of each system with 10 µM CX-5461 from a FRET melting temperature assay[1]. Values are estimated based on the figure in the literature.

3.3.5. Markovian state model overview. To characterize the binding of CX-5461

to the DNA G4s and duplex we performed a Markovian State Model (MSM) analysis.

We present the MSM along with representative trajectories and order plots for each of the

73

thermodynamically favorable state in each system: human telomeric (Figures 15-17), c-KIT1 (Figures 18-19), c-Myc (Figures 20-21), and duplex (Figure 22;B14). MSMs can be built from MD simulation data and are a comprehensive statistical approach used to create understandable yet high-resolution models of the intrinsic kinetics of a system.[183] Our MSM analysis follows a similar procedure to that of our previous work[184] which also examined the binding pathways and kinetic information of G-quadruplex structures. Our implied timescales of each cluster for all lag times of each system are presented in Figure B15. Due to choosing to cluster into a handful of "macrostates" and directly and skipping over the experimentally unverifiable thousand "microstates", the expected convergence time of the implied timescales should be significantly greater than that of a model with a greater number of clusters. This results in a coarser grained model that trades finer detail for greater experimental testability and easier human understanding [183, 185]. It is likely that directly clustering into "macrostates" still maintains the integrity of the MSM as verification through the Chapman-Kolmogorov test (Figure B16A-D) indicates that the model closely resembles the observed simulation data. Two dimensional network models are also presented in the supporting document for each system (Figure B17).

3.3.6. Markovian state model for the human telomeric system. For the human telomeric system, there were four major parallel pathways were observed for CX-5461: unbound to top binding, unbound to bottom binding, and unbound to side binding as an intermediate state before transitioning to either a top or bottom binding pose. The mean first passage times between the four states are shown in Figure 15 where green arrows indicate the more likely transition while blue arrows indicate a less likely transition. The top (37.8%) and bottom (20.8%) binding poses are the thermodynamically favorable

binding states which collectively make up approximately 58.5% of the simulation. Our

calculated first mean passage times indicated that the pathway from unbound directly to

the top binding state is slightly faster (3.3 µs) than unbound directly to the bottom

binding state (4 µs) and both the transition states starts unbound and going from the side

to top (1.2 µs+3.1 µs=4.3 µs) and side to bottom (1.2 µs+5.7 µs=6.9 µs) transition states.

The approximate interstate flux for unbound to top binding was 1:15, unbound to side

binding was 3:5, unbound to bottom binding was 2:3, side binding to top binding was

unidirectional from side binding to top binding, and side binding to bottom binding was

20:1.
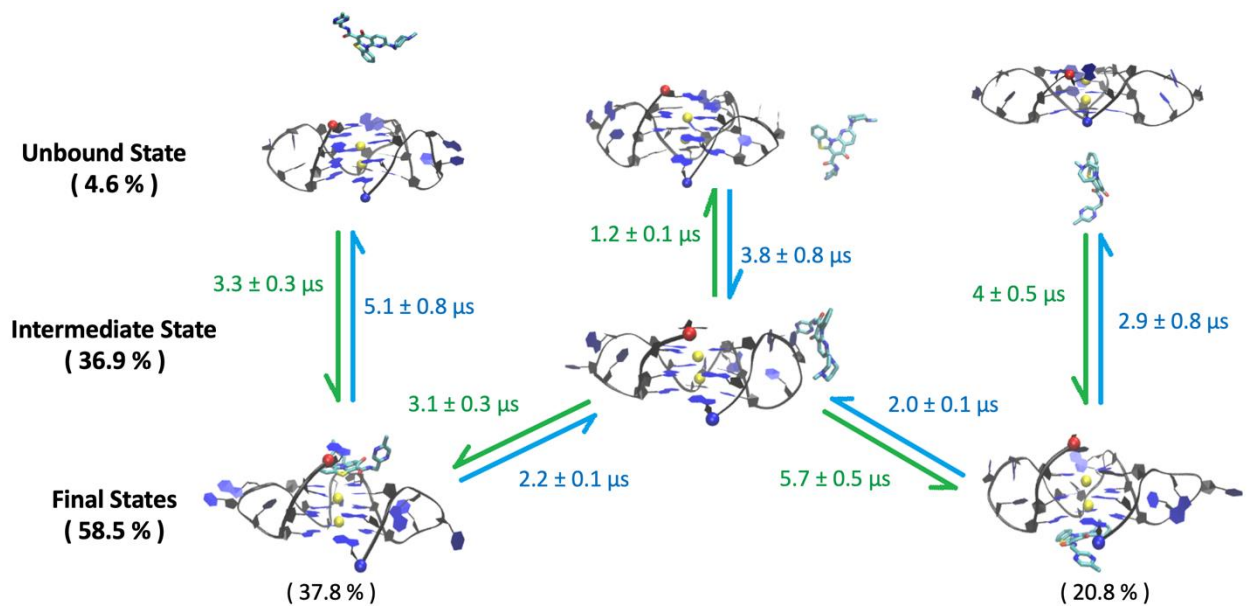
Figure 15. The mean first passage times between the four states (unbound, side transition, top, and bottom) of the human telomeric DNA G-quadruplex and CX-5461 complex system.

We present two representative trajectories which characterize the binding of the most thermodynamically favorable states, one trajectory showing the top unbound to top binding mode and one showing the unbound to side binding transition state ending in a top binding mode. We do not include an unbound to bottom or transition state to bottom binding mode trajectory in the main text due to the structural symmetry of the human telomeric system. However, an additional trajectory of CX-4561 in the bottom binding mode is reported in the supporting document (Figure B18). The representative top binding mode of CX-5461 shows a clear induced fit binding mechanism (Figure 16). At 2 ns the central chromophore of CX-5461 makes initial contact with loop residue A13. At 19 ns, CX-5461 repositions into a top staking mode on top of the first G4 layer. At 101 ns, A1 flips outward making room for CX-5461 to better reposition on the top G4 layer. By 128 ns residue A1 flips back and binds to the top of CX-5461, allowing CX-5461 to intercalate between A1 and the top G4 tetrad where it remained until the end of the 2000 ns simulation. The order parameters indicate the system was stable throughout the trajectory. This quadruplex was able to maintain ~8, ~10, ~8 hydrogens in the top, middle, and bottom layers, respectively. The drug-base dihedral angle remained small after the binding of CX-5461. The ligand RMSD indicated it was stable by the small and flat RMSD of ~2.5 Å. The DNA backbone was generally maintained with slight changes concurrent with the most prominent binding changes of the complex during the trajectory, but remains stable overall maintaining a RMSD of ~15 Å after the final binding pose is achieved. The K^+ cations remained stable, keeping ~4 Å between them. The MM-PBSA binding energy revealed slight fluctuations but a binding energy of ~-30 kcal/mol was maintained for the majority of the trajectory. In the transition state

trajectory (Figure 17), the final binding pose is similar to the representative top binding pose however until ~88 ns CX-5461 was bound to the $T^{11}T^{12}A^{13}$ loop before intercalating into the top of the 5' G-quartet. Similar order parameters were observed for this trajectory, however the time to stability was slightly longer in this case due to the transition.
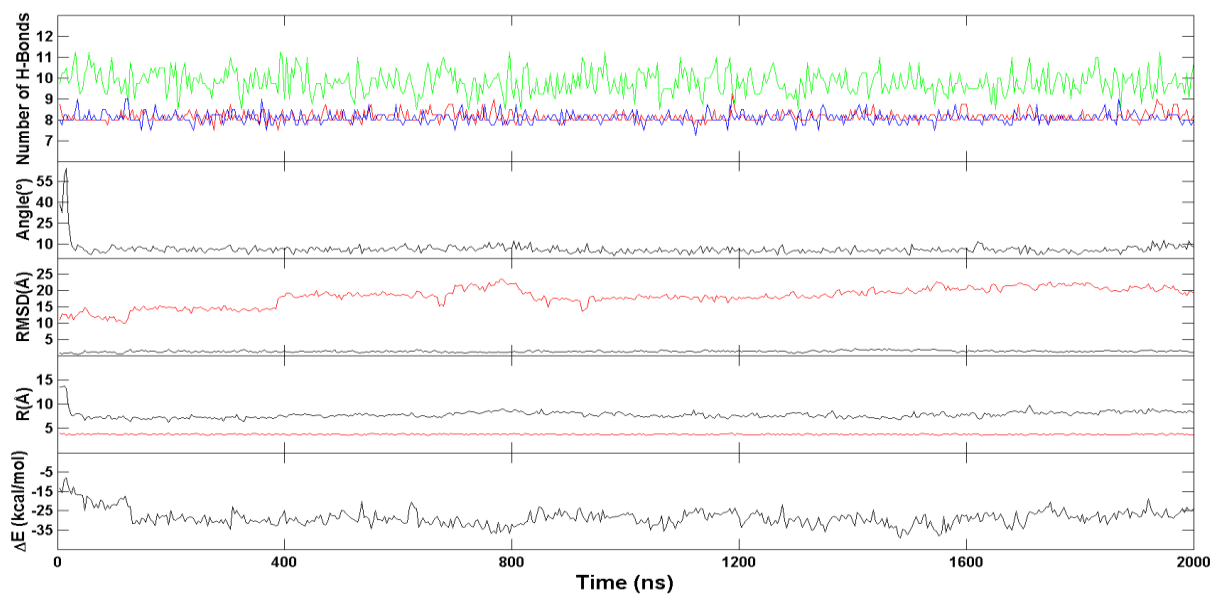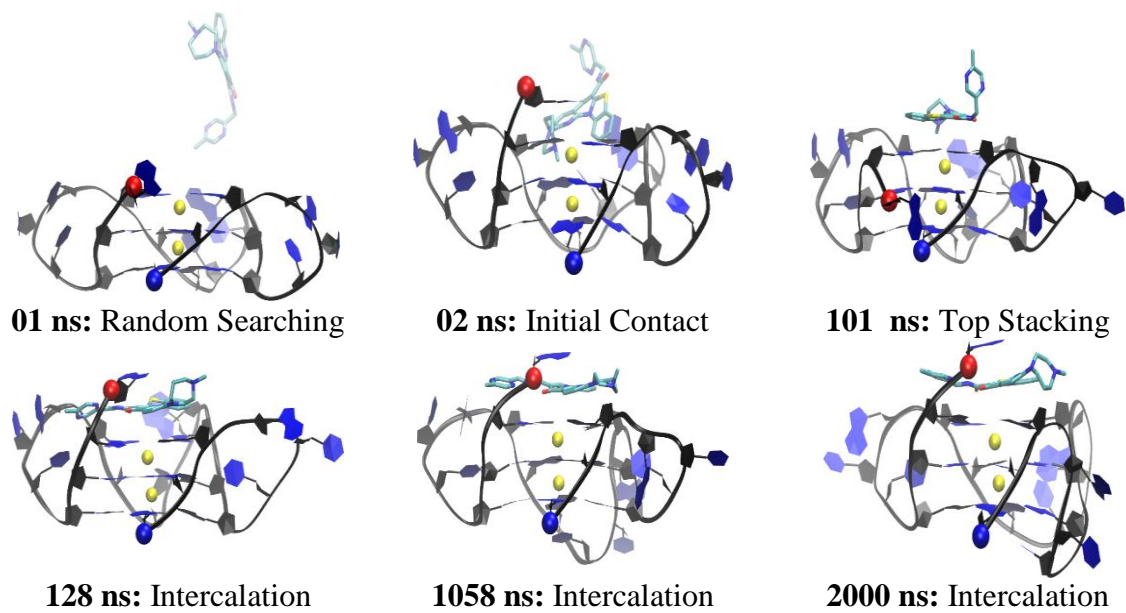
Figure 16. Snapshots from the representative trajectory of the top binding mode of the human telomeric G4, run 13, including an order parameter plot. Illustrated in the plot is the breaking and reforming of hydrogen bonds per G4 layer (Top/Red, Middle/Green and Bottom/Blue), drug-base dihedral angle, ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and K^+ to K^+ distance(red), and the MM-GBSA binding energy (ΔE in kcal/mol). 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K^+ ions are indicated by yellow spheres.

1 ns: Random Searching   **28 ns:** Side Binding   **89 ns:** Intercalation

301 ns: Full Intercalation   **391 ns:** Intercalation   **1943 ns:** Top Stacking
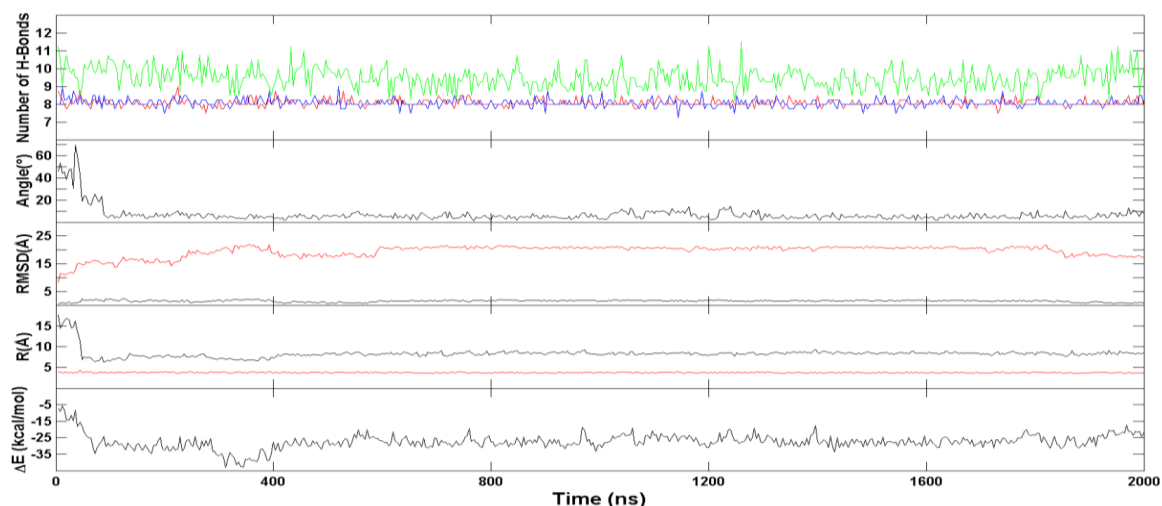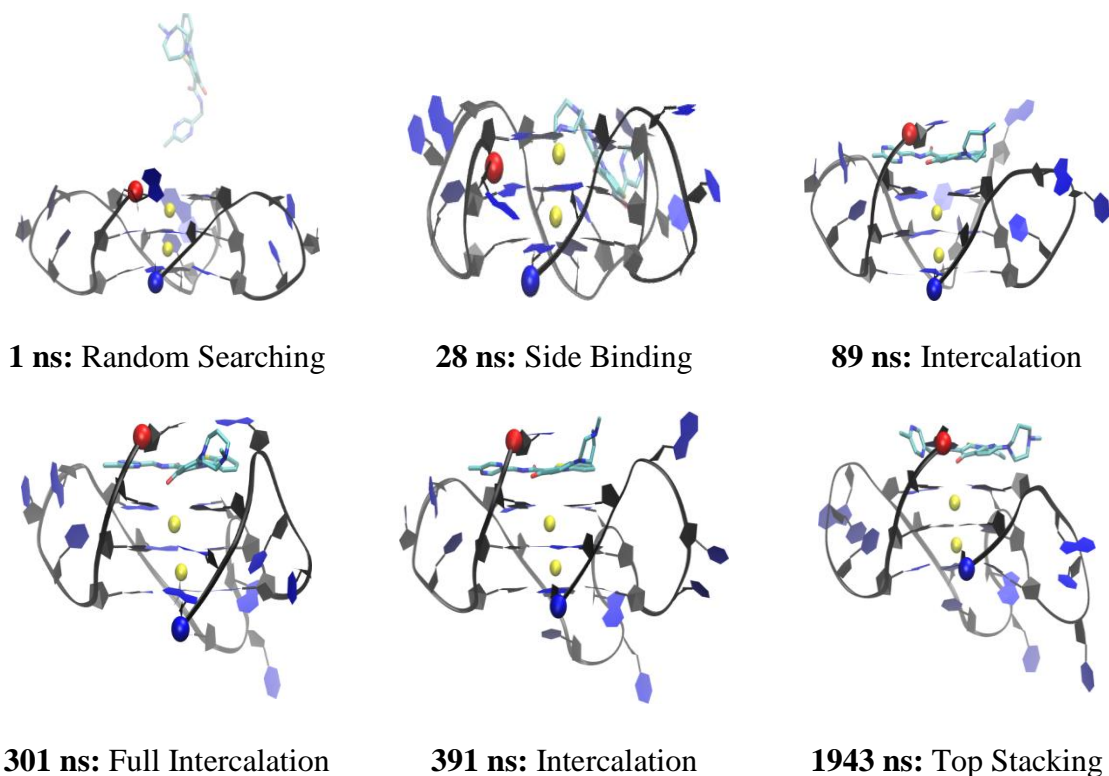
Figure 17. Snapshots from the representative trajectory of the top binding mode of the human telomeric G4, run 05, including an order parameter plot. Illustrated in the plot is the breaking and reforming of hydrogen bonds per layer(Top/Red, Middle/Green and Bottom/Blue), drug-base dihedral angle, ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and K^+ to K^+ distance(red), and the MM-GBSA binding energy (ΔE in kcal/mol). 5' and 3' ends of the DNA chain are indicated by a red and blue ball, respectively. K^+ ions are indicated by yellow spheres.

3.3.7 Markovian state model for the c-KIT1 system. The c-KIT1 G4 system shows three major parallel pathways leading to one thermodynamically favorable top binding state: unbound to top binding, and unbound to side binding as an intermediate state before transitioning to a top binding pose, and unbound to bottom binding before transitioning to a side binding pose and finally transitioning to a top binding pose. The mean first passage times between the three states are shown in Figure 18 where green arrows indicate the more likely transition while blue arrows indicate a less likely transition. Each of the pathways led to a thermodynamically favorable top binding state which accounted for 53.5 % of the simulation period and occurred in 1.2 µs. The unbound to side to top pathway was the next fastest totaling 2.7 µs. The slowest pathway was from the unbound state to the side and finally ending in a top binding mode which totaled 6.8 µs. Important to note is that we believe the MSM determined that the bottom binding mode is not a thermodynamically favorable state in this system because of the limited simulation period, however we expect that if the simulations were extended further, a thermodynamically favorable bottom binding site would be seen. The approximate interstate flux for unbound to top binding was 3:4, unbound to side binding was 3:4, unbound to bottom binding was 1:10, side binding to top binding was 1:30, and side binding to bottom binding was 1:50, and top binding to bottom binding was 1:4.
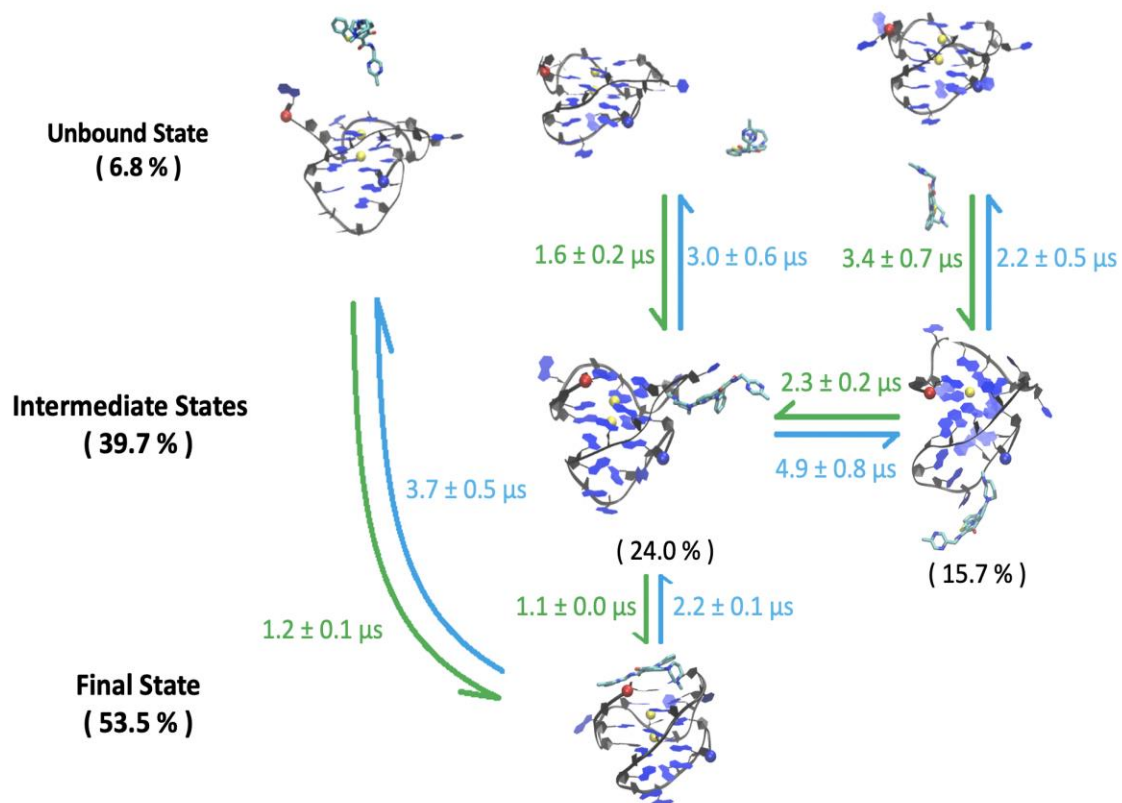
Figure 18. The mean first passage times between the four states (unbound, side transition, top, and bottom) of the c-KIT1 DNA G-quadruplex and CX-5461 complex system.

Since the only thermodynamically favorable state is the top binding mode, we present one representative trajectory of this mode (Figure 19) in the main text. For the top binding mode of the c-KIT1 G4 the simulation begins with random searching of CX-5461. By 4 ns, CX-5461 makes initial contact with the c-KIT1 G4 by interacting with residue A5. This interaction continues until CX-5461 moves to the top of the G4 at 230 ns, where the central chromophore of CX-5461 binds to residue G6 of the top G4 quartet. At 230 ns the 2-methylpyrazine side chain of CX-5461 makes the first interaction with the G4 side loop residue C11 and this interaction continues until 383 ns when residue C11 flips upward

81

and guides the interaction of CX-5461 deeper into the top binding pocket of the G4, allowing it to become sandwiched between bases A1 and G6. At 413 ns the reside C11 flips outward and base T12 is able to bind to flip over and bind on top of CX-5461, along with residue A1 which has maintained its sandwiching interaction with CX-5461 since 383 ns. This binding pose and interactions are maintained for the remainder of the 2000 ns trajectory. The A1 and A12 pair maintained ~2 hydrogen bonds, the top, middle and bottom G4 layers maintained 8, 10, and 5 hydrogen bonds, respectively. The G17 and A19 base pair did not form hydrogen bonds in this run. Around 400 ns the drug-base dihedral angle as well as MMPBSA binding energy stabilizes, with very minimal fluctuations; this finding is consistent with the timing of CX-5461 positioning into the final binding pose for this trajectory. The ligand RMSD and ligand center to DNA center distance stabilize around 150 ns which is concurrent with the ligand repositioning from the side to the top of the quadruplex where it remains for the length of the trajectory. The DNA RMSD and K^+ cation distance remain stable throughout the trajectory. There additional trajectories presented in the supporting document for the top (Figure B19) and bottom binding modes (Figure B20-21) which show CX-5461 binding to both the outside of the bottom loop as well as inside of the bottom loop, but not actually interacting with the G4 core. In addition to this, a trajectory is presented in the supporting document to show the side to top binding mode (Figure B22).

82

1 ns: Random Searching   **4 ns:** Side Loop Interaction   **144 ns:** Top Stacking

383 ns: Sandwiched   **413 ns:** Full Intercalation   **1993 ns:** Intercalated
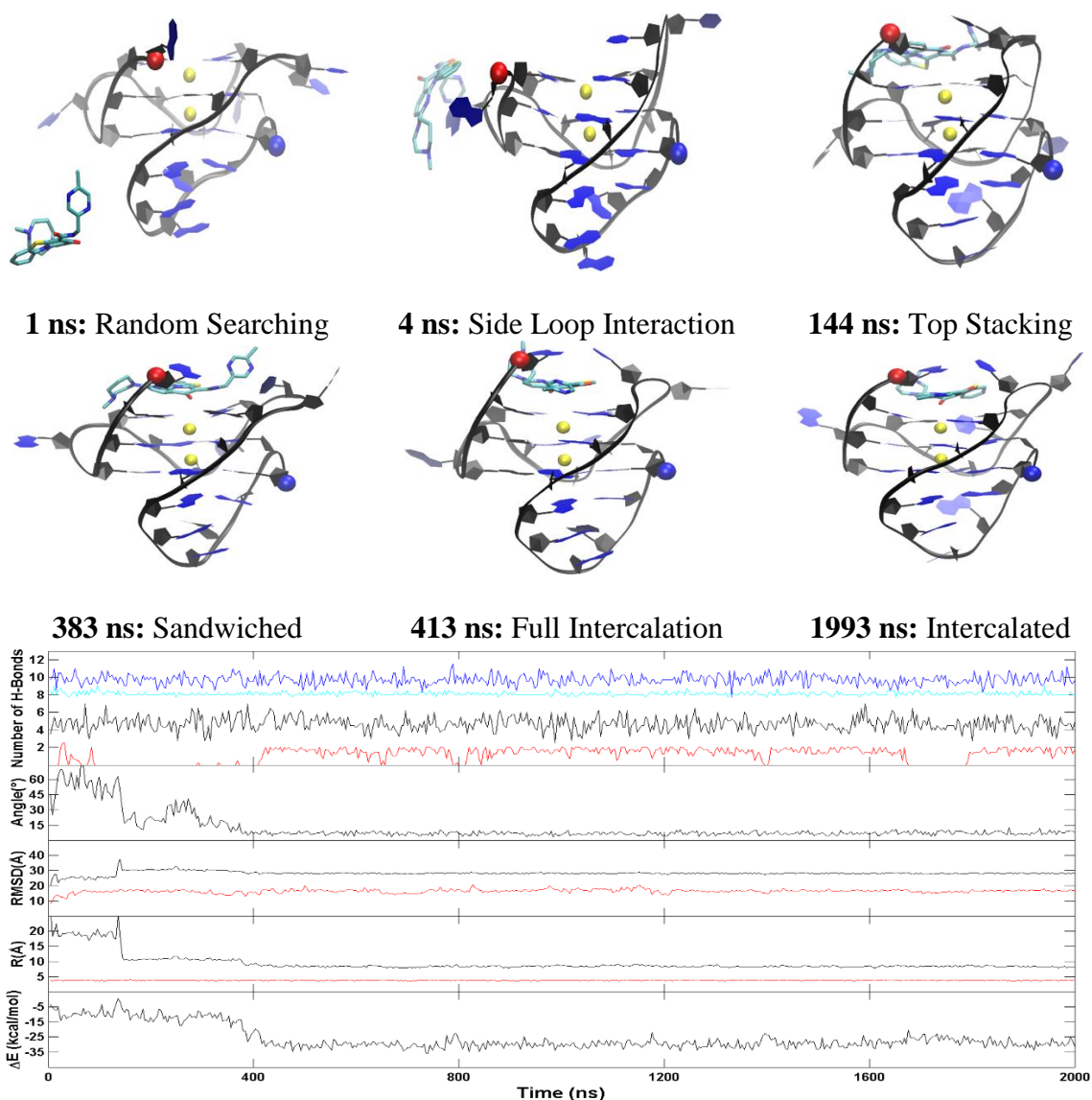
Figure 19. Snapshots from the representative trajectory of the top intercalating binding mode of the c-KIT1 G4, run 19, including an order parameter plot. Illustrated in the plot is the breaking and reforming of hydrogen bonds per layer (5' Face/Red, Top G-Tetrad/Cyan, Middle G-Tetrad /Blue, Bottom G-Tetrad /Black, and 3' Face/Green), drug-base dihedral angle, ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and K^+ to K^+ distance(red), and the MM-GBSA binding energy (ΔE in kcal/mol). 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K^+ ions are indicated by yellow spheres.

3.3.8. Markovian state model for the c-Myc system. For the 2MGN system, three major parallel pathways were observed: unbound to top, unbound to bottom, and unbound to side transition and ending in a top binding mode. Unique to this system, the bottom binding pose appears to be highly unstable and likely acts as an off pathway intermediate state where CX-5461 binds to the bottom from an unbound state and once again goes back to the unbound state and follows one of the other pathways leading to the thermodynamically favorable top binding mode. The mean first passage times between the states are shown in Figure 20 where green arrows indicate the more likely transition while blue arrows indicate a less likely transition. The top (59.6%) and bottom (9.5%) binding poses collectively make up approximately 69% of the simulation. The transition from unbound directly to the top binding state (1.4 µs) is slightly faster than from unbound to the top binding state through the side transition state (2.4 µs). Transition from the unbound to the bottom binding pose is significantly slower calculated to be 16.7 µs. The approximate interstate flux for unbound to top binding was 1:5, unbound to side binding was 1:1, unbound to bottom binding was 1:4, and side binding to top binding was 1:3.
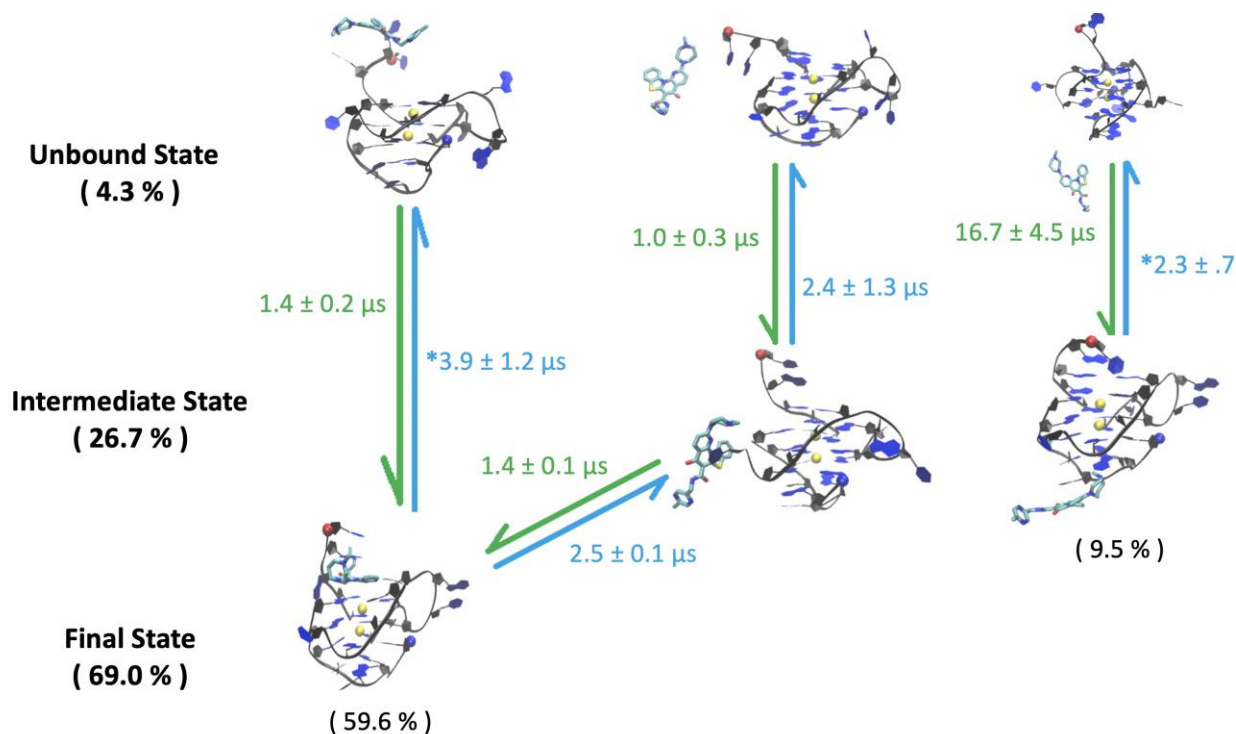
Figure 20. The mean first passage times between the four states (unbound, side transition, top, and bottom) of the c-MYC DNA G-quadruplex and CX-5461 complex system.

The interaction between the c-MYC G4 and CX-5461 is particularly interesting because of the clear use of a base flipping mechanism by both the terminal residues, T1, G2, and A3, but also the loop residue A12 (Figure 21). The first interaction of CX-5461 to the c-MYC promoter G4 was at 1 ns where it made contact with the residues of the 5' terminal. At 2 ns, the terminal residue A3 had already flipped on top of CX-5461 and sandwiched it onto the top G-quartet; simultaneously, the methyl diazepane side chain of CX-5461 made interaction with the loop residue A12. At 6 ns loop residue A12 has flipped on top of CX-5461, which then fully intercalated into the top binding pocket of the c-MYC G4, being stabilized by both residues A3 and A12. This interaction remained until 1136 ns

85

when terminal residue T1 flipped over and joined residues A3 and A12 binding on top of CX-5461. At 1742 ns terminal residue T1 flipped up again, and the ligand remained intercalated into the interface for the remainder of the trajectory. The final binding pose that CX-5461 maintained for the majority of the trajectory involved residues A3 and A12 intercalating the drug between residues G4, G17, and G8. The A3 and A12 base pair maintains ~2 hydrogen bonds, the top, middle and bottom G-quartet maintain ~8, ~ 9, and ~5 hydrogen bonds, respectively; the G20, A22 and G23 base pairing also maintained ~2 hydrogen bonds. The drug-base dihedral angle and MMPSBA binding energy is roughly maintained over the trajectory with a slight decrease after 1200 ns, consistent with the drugs final binding pose of full intercalation. The DNA and ligand RMSD, drug center to ligand center, and K^+ cation distances are maintained throughout the majority of the trajectory. There additional trajectories presented in the supporting document for the top (Figure B23) and bottom binding modes (Figure B24-25) which show CX-5461 binding to both the outside of the bottom loop as well as inside of the bottom loop, but not actually interacting with the G4 core. In addition to this, a trajectory is presented in the supporting document to show the side to top binding mode (Figure B26).
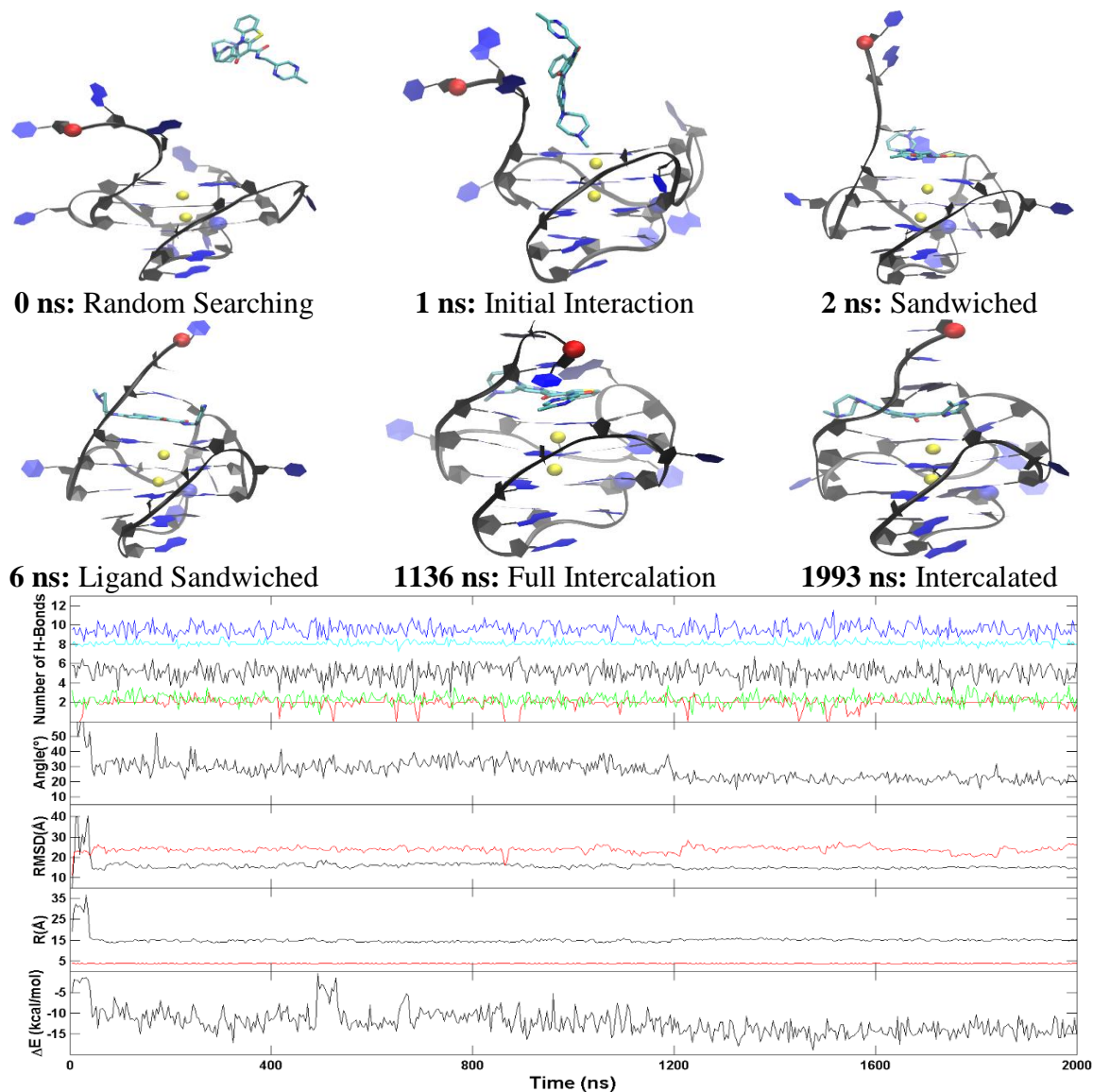
Figure 21. Snapshots from the representative trajectory of the top intercalating binding mode onto the c-Myc G4, run 14, including an order parameter plot. Illustrated in the plot is the breaking and reforming of hydrogen bonds per layer (5' Face/Red, Top G-Tetrad/Cyan, Middle G-Tetrad /Blue, Bottom G-Tetrad /Black, and 3' Face/Green), drug-base dihedral angle, ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and K+ to K+ distance(red), and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow spheres.

3.3.9. Markovian state model for the duplex system. The MSM plot of the duplex system and mean first passage times are presented in Figure B14 where green arrows indicate the more likely transition while blue arrows indicate a less likely transition. This plot shows the major pathways for the duplex system include one thermodynamically favorable groove biding state (25.7%) as well top and bottom states that end up in a groove binding mode. Since the groove binding mode is the only one of physiological relevance for a long chromosome DNA, pathways leading to this mode are discussed here. The pathway from unbound directly to the groove binding state is slightly faster (2.0 µs) than from unbound to the transition states (top: 4.7 µs or bottom: 5.0 µs) leading to the groove binding and is significantly more abundant. The representative trajectory for the groove binding mode of CX-5461 to the DNA duplex is shown in Figure 22. Clear from the representative snap shots, CX-5461 binds to the groove around 35 ns with slight repositions until fully relocating to another groove binding site at 845 ns. Although minor adjustments are seen, CX-5461 remains in this binding pose for the remainder of the 2000 ns trajectory. The order plot shows changes concurrent with the repositioning of CX-5461, where each parameter shows fluctuations until ~850 ns where the final, stable binding pose is achieved. After 850 ns the ligand and DNA RMSD maintains 30 Å and 10 Å, respectively. The drug center to DNA center distance is ~15, and the MMPBSA binding energy is roughly -15 kcal/mol. Interestingly, in that simulation run we observed CX-5461 scaling the surface of the DNA duplex, unable to find a stable binding pose, repositioning around the grooves of the DNA duplex while maintaining interaction during the entire binding process. Three additional trajectories for this binding mode is presented in the supporting document (Figure B27-B29).

88
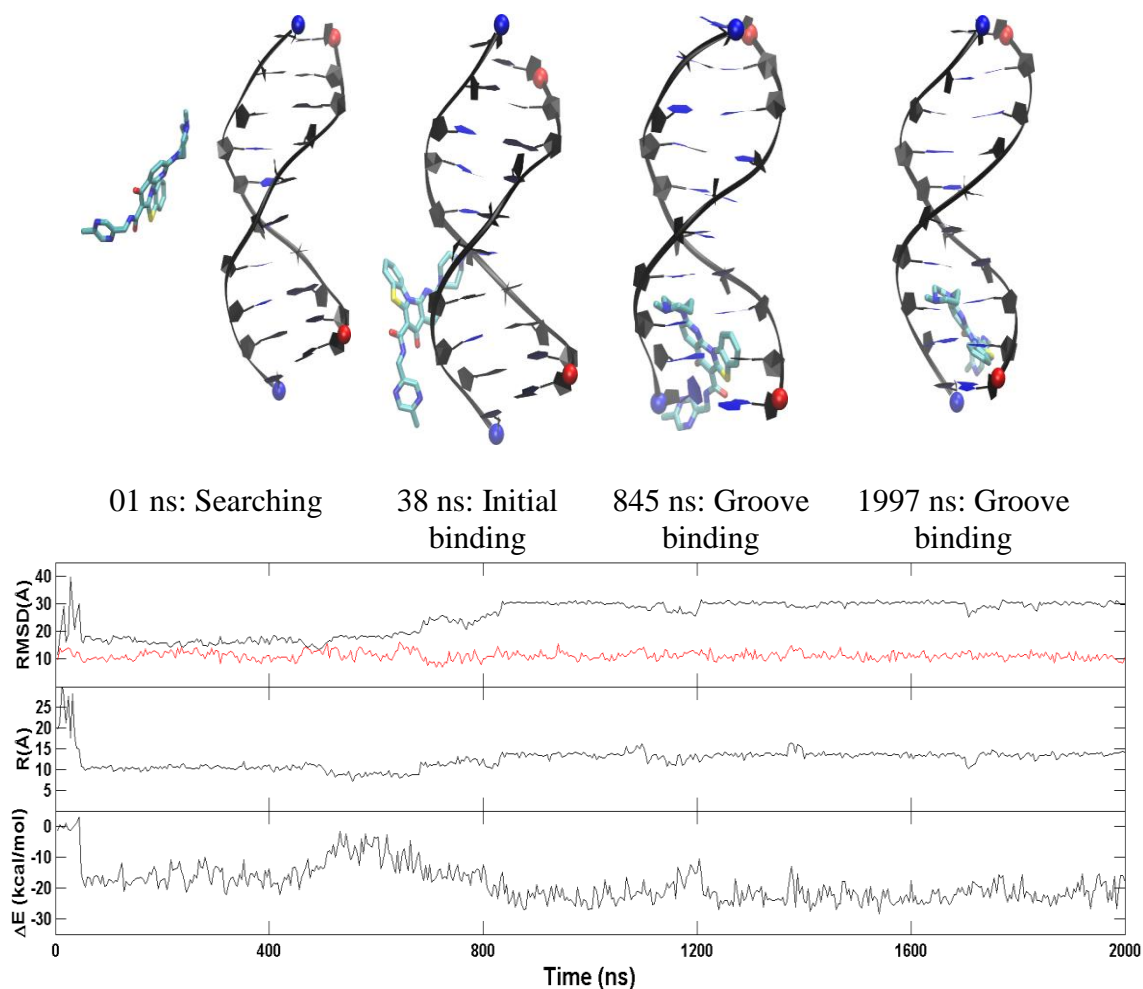
Figure 22. Snapshots from the representative trajectory of the bottom binding mode of the DNA duplex system, run 12, including an order parameter plot. Illustrated in the plot is the ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow balls.

3.3.10. Geometry characterization. The geometry of the inter G4 layers from the MD simulations is an area that has not yet been thoroughly studied despite literature suggesting that the H-rise and H-twist geometry parameters were the most important parameters to analyze the G4s overall structure. Characterizing these parameters is

important because despite a number of works defining duplex DNA, there has been little research into the helical rise and rotation of the quadruplex structure from molecular dynamic simulations. We first performed a geometry analysis on both the apo and holo forms to validate the parameters set for our simulation systems and to gain insight into the overall geometry of the system (Table B2), which is a technique that has been successfully implemented into our previous works[184, 186]. Helical B-DNA is commonly defined by maintaining average values of 3.32 Å and 34.3° for H-rise and H-twist, respectively. The results of the H-rise and H-twist helical geometry parameter analysis revealed comparable values for each of the systems being analyzed, providing quantitative support that the helical structure was maintained throughout the simulations. Specifically, we identified the average values for the H-rise (~2.6 Å) and H-twist (~20 Å) of the bases within the three quartets of each G4 system and reproduced an average value for the rise parameter (~3.5 Å). In this work, it is clear that the H-rise and H-twist in the G-quadruplex are smaller than those in standard B-DNA (3.32 Å and 34.3° for H-rise and H-twist). The G4 geometry analysis in this study is supported by a recent bioinformatics study on 74 G-quadruplex structures in the PDB databank which showed the distribution of the twist angle between the two adjacent G-quartets for bimolecular parallel G-quadruplexes is ranged from 10º to 35º with a maximum at 31º±3º.[187] By comparing the apo and holo forms, it is clear that the binding of CX-5461 on the top and bottom of the each G-quadruplex did not change these parameters much. Thus, we conclude that ligand binding can slightly modulate the G-quadruplex inter-layer geometry. We also calculated base pair-axis, base pair-step, and paired base-base parameters, as detailed in the method section, to characterize the duplex geometry in the DNA only system versus in the most favorable groove binding pose (Table

B3-B5). The values from the groove binding system generally agree with the values from the DNA only simulation. For example, the DNA only simulation has average values of ~3.3 Å and ~35° for H-rise and H-twist, respectively. Whereas the groove binding mode has average values of ~3.3 Å and ~35° for H-rise and H-twist, respectively. This is also supported from the DNA backbone dihedral angle and pucker calculations (Table B5): although there are some subtle differences, the majority of values are close to each other. The hydrogen bond networks of the major binding modes of the G4 systems are presented in Figures B30-B32 which show subtle differences.

3.3.11. Root mean square fluctuation analysis. The RMSF plots for each system as well as a topology model which shows the position of each residue for which the RMSF was calculated and presented in Figure 23. In each data set three large (~5 Å) peaks were identified which correspond to three loops connecting the G-quartets: $T^5T^6A^7$, $T^{11}T^{12}A^{13}$, and $T^{17}T^{18}A^{19}$. Evident form the data here, the apo form has higher overall fluctuations when compared to the complex systems. Overall, there are two pronounced differences in the data series. The first is the lower fluctuations of the $T^{11}T^{12}A^{13}$ loop in the apo form simulation run, which may be due to this loop frequently flipping outward to clear room for CX-5461 binding. The second is in the bottom binding pose where the fluctuation of the $T^{17}T^{18}A^{19}$ loop has significantly decreased fluctuation in residues T^{17} and T^{18}, which may be due to a stabilizing effect because of CX-5461 directly interacting with residue G^{16} in this binding mode. In addition we see here that the terminal residues have roughly equal fluctuations due the structure symmetry of the G4. In the c-KIT1 systems the major peaks corresponded to residues of the connecting loops: A^5, C^9, C^{11}, T^{12} and G^{17}, G^{18} and A^{19}. Overall, the apo form trajectory had higher fluctuations than the

91

complex systems indicating that CX-5461 binding slightly reduced the overall fluctuations at these peaks. The once exception was at loop C^{11}, T^{12} where the top binding mode has this region with higher fluctuations. This is due to this loop flipping upward and forming a hydrogen bond between R1 of CX-5461. Here we clearly see a difference in the terminal residues where the 5' terminal is significantly higher than the 3' terminal, which is likely due to the fact that the 3' terminal is actually a residue of the 3' G-tetrad whereas the 5' terminal is a free residue. There were 5 major peaks identified for the c-Myc systems corresponding to residues T^{1-A3}, T^7, T^{10-A12}, T^{16}, and A^{21}. Each residue showing significant fluctuation was either a part of the large 5' terminal segment (T^1, G^2, A^3), or a part of a connecting loop. In the c-Myc systems, the RMSF values were comparable in each peak except for the T^{10}, G^{11} and A^{12} region. The top binding mode had a high fluctuation at residue T^{10} but lower fluctuations at G^{11} and A^{12}, this is likely due to this loop flipping upward to stabilize the binding of CX-5461. Where the higher fluctuations of T^{10} are a result of the further distance travelled to flip upward compared to G^{11} which flips outward and A^{12} which directly binds to CX-5461. The bottom system had the largest fluctuation at T^{10} and G^{11} which flip outward during the simulation and the receptor only system had equal fluctuations at residues T^{10}, G^{11} and A^{12} which are also flipped outward. The terminal residues are most notably different here where the long 5' segment is highly flexible (5-6 Å) and the 3' terminal, which is a part of the 3' G-quartet has very low fluctuations (~1 Å). The duplex DNA systems (grey and black) showed peaks at residues G^1, C^{10}, G^{11}, and C^{20}, which are the terminal residues of the double helix. Overall, we see that the human telomeric G4 loops have the highest fluctuations of all systems which may help to explain the less favorable binding energy of the side

binding modes compared to the other systems and the fluctuations of the apo G4 forms

are larger than the complex systems, suggesting that CX-5461 stabilizes the G4 DNA

upon binding.
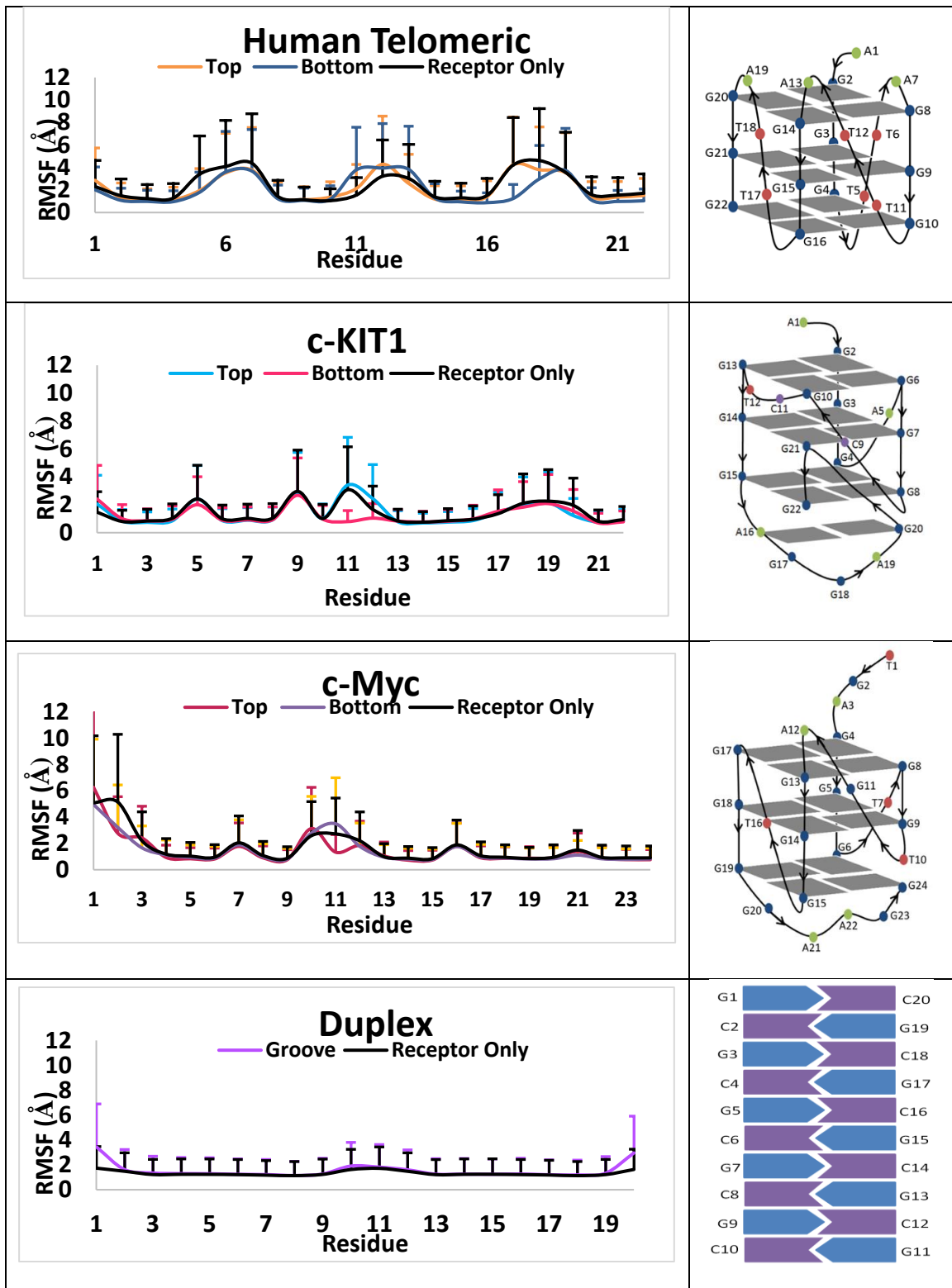
Figure 23. The root mean square fluctuation (RMSF) plot of the 11 systems (receptor only, plus top and bottom binding modes of 1KF1, 2MGN, and 4WO3, groove binding mode of the duplex system) with topology include on the right for reference. Standard deviations are represented by error bars.

3.4 Discussion

DNA G-quadruplexes (G4s) are over represented in a wide variety of cancers making them a prime therapeutic target. [188] Stabilizing G4s of cancerous cells has become a novel therapeutic technique that has been shown to inhibit cancerous cell growth and replication leading to the cancerous cells death. [189] A promising new anti-cancer medication, CX-5461 is in in Phase I/II clinical trials for advanced hematologic malignancies and cancers with BRCA1/2 deficiencies [117]. Experimental evidence identified CX-5461's specific roles at G4s as the ability to bind to and stabilize G4 structures in human cancer cell lines, increase the number of G4 sites, as well as selectivity bind to G4 structures on the human genome.[117] These properties are extremely advantageous for cancer therapeutics and evident from recent work CX-5461 is a promising therapeutic agent for a variety of targets, however the detailed binding mode and mechanism for which CX-5461 interacts with the G4 structures remains elusive. Our analyses sought to characterize the binding of CX-5461 to the human telomeric, c-KIT1, c-Myc G4s as well as a DNA duplex to provide insight into the binding process and help to explain the experimentally reported order of binding induced stability.

Our simulation analysis identified three major binding features conserved in each G4 system. First, we observed that 5' intercalation was the most thermodynamically favorable binding mode. This was evident from the MM/PBSA binding energy (Table 5) and trajectory clustering (Figure 14). Second, the Markovian State Model (MSM) of each G4 system revealed that there were multiple parallel pathways all leading to the thermodynamically favorable top intercalation mode (Figures 15, 18, 20). Each of which include a direct pathway from the unbound state to top binding as well as indirect

95

pathways which could be using side binding as an on-pathway intermediate state. The lower MM/PBSA binding energies of the side binding modes (Table 5) provide thermodynamic support of the side binding intermediate states. The third common binding feature was that the top binding mode of each system clearly were not achieved through the use of a lock-key mechanism. Instead, each G4 system demonstrated the use of a combination of the induced fit and conformational selection mechanisms. In more detail, our apo form simulations (Figure B2-3) show that because of the short flexible segments at the 5' face of the G4's there are some instances of base flipping observed for both the 5' terminal and 5' face loop residues. However when CX-5461 binds, the population is shifted toward the population which these bases stabilize and stack onto the G4 core or CX-5461 which clearly shows evidence of a conformational-selection mechanism. In addition, the most thermodynamically stable state is only achieved by a base flipping insertion mechanism where terminal and side loop residues flip outward to make room for CX-5461 insertion followed by the bases flipping back to make contact with CX-5461, which demonstrated the use of an induced fit binding mechanism not observed in the apo form. A clear example is the interaction between the c-Myc G4 and CX-5461 which shows the use of a base flipping mechanism by both the terminal residues, T1, G2, and A3, and also the loop residue A12 (Figure 21), which is supported by the RMSF (Figure 23) and 2D interaction diagrams (Figure B11). Therefore, although there are structural differences between the G4s, these same general binding characteristics of CX-5461 are conserved which we suggest could be applicable to other targets not studied here.

Three notable differences were also extracted from our simulation analysis. It was evident that the structural features of each system caused subtle differences in binding modes (Table B7), binding energies (Table 5), and MSMs (Figures 15, 18, 20). First, the human telomeric system's structural symmetry on the 5' and 3' faces allowed equal binding opportunities for both sites. The human telomeric system was the only MSM to have parallel pathways leading to the bottom binding pose as a thermodynamically favorable state (Figure 15), and the only system to have the binding energy of the bottom binding mode be comparable to the top binding mode (Table 5). The conclusion made by comparing the bottom binding modes for each G4 system leads to the second difference: each 3' site is different and as a result has a different effect on CX-5461 binding, which we discuss in more detail further in the discussion. The third major difference is the trend of the MSMs. As mentioned, the structural symmetry of the human telomeric system clearly differentiated it from the c-KIT1 and c-Myc, but the two non-symmetric G4s also had stark differences from each other. Most notably, the c-Myc systems bottom binding pose was determined to be the most thermodynamically unstable by the MM/PBSA binding energy analysis. This is demonstrated in the MSM where the bottom binding mode is a lowly abundant state which we suggest likely acts as an off pathway intermediate state. We believe this happens by CX-5461 binding to the bottom of the G4 from an unbound state and once again goes back to the unbound state and follows one of the other pathways which leads to the thermodynamically favorable top binding mode. In contrast to the c-KIT1 G4 system which utilizes the bottom binding mode as one of its on-pathway intermediates before leading to the more thermodynamically favorable top binding mode. Together, understanding these

97

similarities and differences allowed us to provide insight into the binding mode and mechanism of CX-5461 and make sense of the experimental phenomenon published on CX-5461's affinity and selectivity to these targets.

Our simulation protocols and final binding poses were validated by comparing the binding of CX-5461 to the c-Myc promoter G4 in our study to the only experimentally solved structure of these G4s in complex with a small molecule to date. This was the binding pose of Phen-DC$_3$ in complex with Pu24T (PDB ID: 2MGN) which was solved using NMR spectroscopy [190]. Our study used Pu24T, the intramolecular G-quadruplex formed from the c-Myc promoter, from the Phen-DC3-Pu24T complex and simulated this G4s interactions with CX-5461. The major binding mode of CX-5461 identified in our study closely matches the binding pose of Phen-DC$_3$ solved using NMR spectroscopy (Figure 24). The highly flexible 5' terminal segment binds to and stabilizes both Phen-DC$_3$ and CX-5461 into an intercalation to the top G-quartet where the 5' terminal residue is bound to the center aromatic core of both ligands. This comparison provides support that the key interactions for this G4 are π-stacking with the guanine bases of the top G-tetrad, as well as validation for the simulation methods used in our study. Our study also supports the prediction made by Xu and coworkers[117] which suggests that CX-5461 may also be an end stacking G4 stabilizer due the structural relationship between CX-5461 and QQ58 (Figure B33), which was biophysically determined to be an end stacking G4 stabilizer in one study.

98

Side	Top

Cx-5461-Pu24T complex

Phen-DC₃-Pu24T Complex (2MGN)
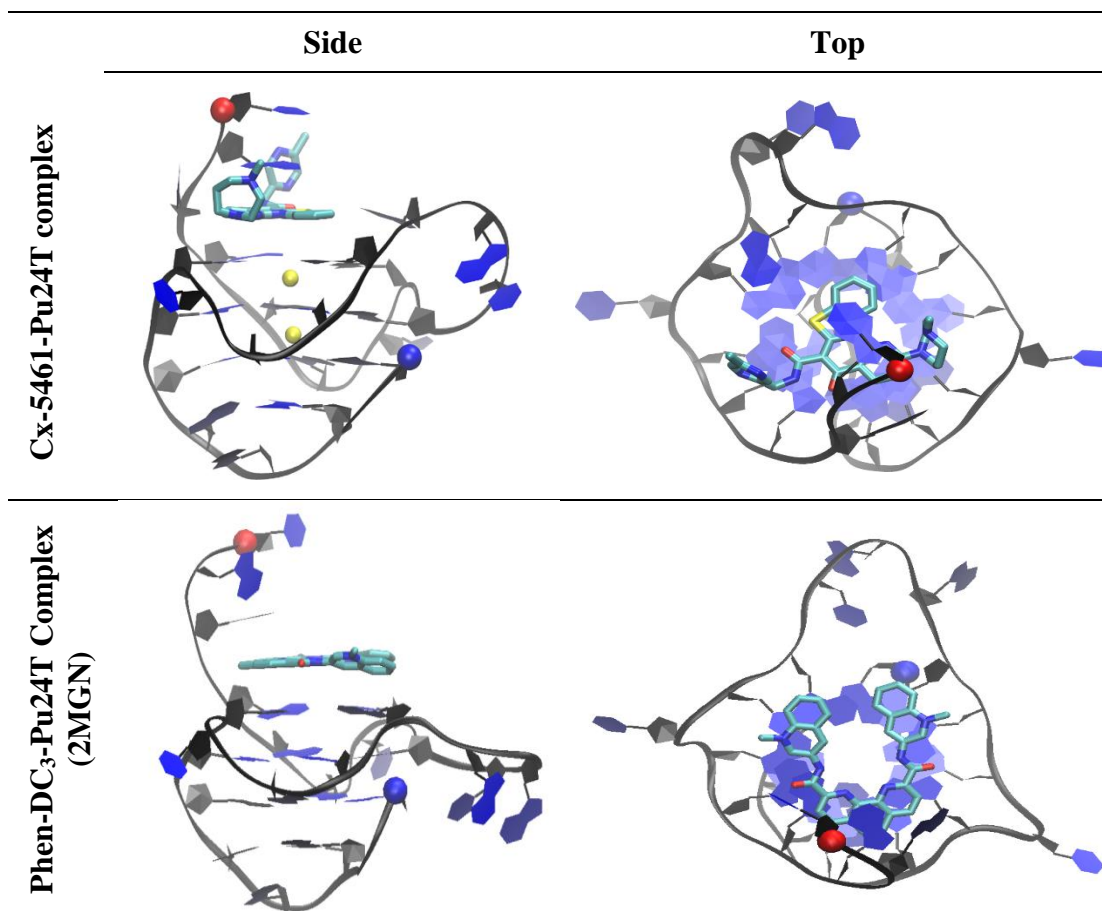
Figure 24. Major binding mode comparison of the CX-5461-Pu24T complex and Phen-DC₃-Pu24T Complex (PDB ID: 2MGN). 5' and 3' of the DNA chain are indicated by a red and blue sphere, respectively. K^+ cations are represented by yellow spheres.

As mentioned in the introduction, Xu and co-workers FRET-melting temperature increase assay [117] was performed to determine the affinity of CX-5461 binding to the G4s and duplex structure. A higher binding affinity was correlated to a higher melting temperature increase which indicated more energy was required to break the bonds formed upon CX-5461 binding, and was essentially a measurement of enthalpy. The results of their assay indicated that with 10 μM CX-5461 the highest melting temperature increase, and thus the stabilizing effect due to CX-5461 binding, was demonstrated by the human telomeric system (~30 K) followed by the c-KIT1 G4 (~27 K), the c-Myc G4 (~25 K) and

99

the DNA duplex (~10 K). Evident from this data, CX-5461 stabilizes the three G4 systems much more than the DNA duplex (~15-20 K). But what was causing marginally different melting temperature increases (~2-5 K) for the three G4 systems? In our study, each system's MM-PBSA binding energy calculations, which are also a measurement of binding enthalpy, were consistent with the order of stabilizing effect upon CX-5461 binding determined in the FRET-melting temperature assay (Table 5). CX-5461's average MM/PBSA over the major poses were calculated for the human telomeric (-28.6 kcal/mol), c-KIT1 (-23.9 kcal/mol), c-Myc (-22.0 kcal/mol) G4s, and DNA duplex (-15.0 kcal/mol) systems. Evident from this, the three G4 systems much higher MM-PBSA binding energies as compared to the DNA duplex. Thus, it appears that the binding energy differences lead to the observed difference in the melting temperature change for these systems.

The unique structural differences at the 3' region provide insight into the binding modes and the binding energy differences. Due to the structural symmetry of the human telomeric G4, the binding poses and binding energies of the top (-37 kcal/mol) and bottom (-33 kcal/mol) modes were very comparable. However, this was not true for the c-KIT1 and c-Myc G4s, which contain 3 (GGA) and 4 (GAAG) residue long diagonal loops, respectively, on the 3' end. Instead, the bottom binding mode of c-Myc had a very unfavorable MMPBSA binding energy (-15.5 kcal/mol) which was comparable to the duplex system (~-15 kcal/mol), and the c-KIT1 system was not much more favorable (-22 kcal/mol). In these systems we observed a decreased ability for CX-5461 to interact directly with the third G-quartet due to hindrance from the diagonal loops which led to unequal binding opportunities when compared to the top binding mode. From our analysis we suggest that the more ordered diagonal loop of c-KIT1, which contains an A^{16}-G^{20} base

100

pair, may contribute to the marginally higher stability determined from the FRET-melting temperature assay, when compared to the c-Myc G4 which contains the 4 residue diagonal loop (GAAG). Although CX-5461 was not able to intercalate onto the 3' G-quartet for the c-KIT1 and c-Myc G4's there was one case for both the c-KIT1 and c-Myc G4s where CX-4561 was able to intercalate into the 3' diagonal loop connecting the 3' G-tetrad, but our MMPBSA binding energy analysis revealed this binding mode was even less favorable in both the c-KIT1 (-11.66 kcal/mol) and c-Myc (-13.17 kcal/mol) systems. These facts help to explain why both of these G4s have a lower binding energy and melting temperature increase than the human telomeric G4, but also why the c-KIT1 G4 has a marginally higher affinity and binding energy when compared to the c-Myc G4. Thus, we propose that the structural differences at the 3' region of each G4 causes unequal CX-5461 binding opportunities which made for differences in our observed binding modes and may also play a critical role in the experimentally reported affinities from the melting temperature increase assay.

Along with characterizing the binding features of CX-5461's in each system to understand CX-5461s major role as a G4 binder and stabilizer observed in experiments, our study aimed to provide insight into the selectivity of CX-5461 to G4 structures over DNA duplex. The motivation for this is based on the major finding from the FRET-melting temperature assay performed by Xu and co-workers research showed CX-5461 selectively binds to and stabilizes G4 structures over duplex DNA[117]. This is an extremely desirable feature for a G4 stabilizer, since a major limiting factor for the therapeutic use of this class of drugs is the lack of selectivity to G4s over DNA duplex structures. Encouragingly, our trajectory analysis supports the selectivity of CX-5461 to G4s. The major finding used to

101

conclude this was that for each of our G4 systems, an intercalation mode was observed for CX-5461, however for our duplex DNA system no groove intercalation mode was observed. The effects of this lack of highly stable binding pose in the duplex system was supported by MMPBSA binding energy analysis which showed that the duplex system had the lowest binding energy. The most energetically favorable binding pose observed in our duplex DNA system (-15.0 kcal/mol) was only comparable to the least favorable binding poses of the G4 systems, supporting that CX-5461 is selective to G4 structures over duplex DNA. In addition, our simulation analysis which supports that CX-5461 is not a DNA duplex intercalator is consistent with an intercalation assay comparing CX-5461's ability to intercalate into calf thymus DNA to a known DNA intercalator Actinomycin-D[118]. This assay even showed that CX-5461 was even a weak minor groove binder at concentrations as high as 50 µM, which is consistent with our less favorable binding energy calculations for CX-5461 to the duplex when compared to the G4s.

To further compare the binding of CX-5461 to a known intercalator, we compared the binding pose of CX-5461 to RHPS4, another G4 ligand with an aromatic core (Figure B34) from our previous work. [191] Evident from comparing the chemical structures, CX-5461 is longer and contains two flexible side chains whereas RHPS4 maintains a very planar aromatic structure. In the previous work, RHPS4 was modeled with the same duplex DNA and human telomeric G4 structure (1KF1) used in this work, among other G4s. RHPS4 was able to fully intercalate into the duplex DNA with a binding energy of -46.8±4.6 kcal/mol and in the human telomeric simulations the most favorable binding mode was bottom stacking with a MMPBSA binding energy of -48.9±2.4 kcal/mol. RHPS4's lack of selectivity toward a DNA structure indicated modifications were needed

for it to be a successful G4 stabilizer. In this study, CX-5461 demonstrates an advantage in G4 selectivity over RHPS4 which we attribute to the two aromatic side chains of CX-5461. These two side chains limit CX-5461's planarity and therefore ability to fit into the narrow grooves of the DNA duplex structure. From our observations we believe that the positive charge on the methyl diazepane side chain, coupled with the flexibility of the methylpyrazine side chain, contributes to the selectivity of CX-5461 to G4 structures over duplex DNA.

Taking a deeper look at the overall binding pattern of CX-5461 to the DNA duplex, since it was clear CX-5461 was not a DNA duplex intercalator, we observed trajectories which showed an interesting mechanism comparable to a recent study.[192] It is conventional that a ligand may completely unbind and begin researching for a binding site in cases where ligand binding is not favorable. In our duplex system however, in cases when ligand binding was not favorable, we observed CX-5461 diffusing the surface of the DNA duplex searching for a stable binding site, maintaining a partial interaction throughout the entire binding process. This was made even clearer when the trajectory was extended from 500 ns to 2000 ns and provides unique insight into the binding interactions of CX-5461 to the duplex. We see that the drugs side chains (R1 and R2) are able to maintain interactions with the DNA, however a favorable binding pose is not achieved due to the lack of intercalation into the grooves of the duplex. Since the major factor limiting the therapeutic use of DNA G4 stabilizers to date has been a lack of selectivity to the DNA duplex, we kept this mechanism in mind when suggesting possible optimizations to this drug.

We developed a hypothesis to increase the binding of CX-5461 to the G4 structures and decrease the binding of CX-5461 to the duplex structure based on our analysis of the detailed interactions of each binding mode at the G4 and duplex structures. The detailed interactions (Figures B11-B13) of the most thermodynamically favorable binding poses (Figures 14; Table 5) indicate that CX-5461's rigid core (Figure 12) binding to the G4s is essential in stable binding. One level of support for this conclusion is that the core of CX-5461 is not exposed to solvent in our two dimensional interaction diagrams of the most thermodynamically favorable top binding poses (Figure B12). Further, the second most thermodynamically favorable pose, human telomeric G4 bottom binding (Figure B11A), also shows that the rigid core shows very little exposure to solvent. However in the c-KIT1 and c-Myc bottom mode and all three G4 systems groove binding mode, the amount of solvent the rigid core of CX-5461 is exposed to is far greater (Figure B11 A-D). With these facts in mind we were able to suggest minor modifications be made to the benzothiazole-based scaffold of CX-5461 with the hypothesis that increasing the length of the core could increase the binding energy and stability. At each possible point of substitution, we picked 2 function groups (fluorine or chlorine) that could be substituted for hydrogen (Figure 25; Figure B35) leaving a total 56 compounds generated in our combinatorial library. Each new compound was docked to the same orthosteric binding site as the most abundant cluster for the most thermodynamically favorable top binding mode, and showed similar hydrogen bonds, π-π and hydrophobic interactions. Of these compounds, we chose four derivates (Figure 25) based on their synthesizability as determined by a minimal number of functional groups

104

added, and their more negative XP docking scores for the G4 but less negative XP

docking scores for the Duplex DNA.
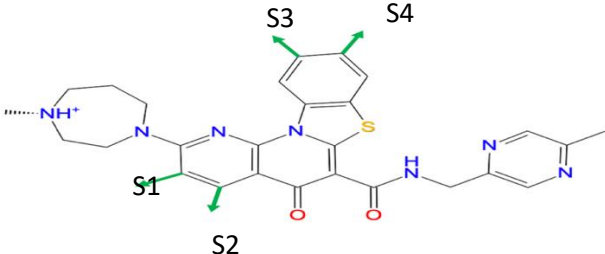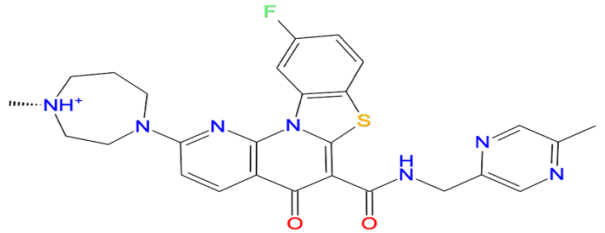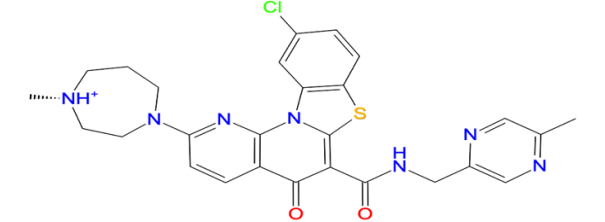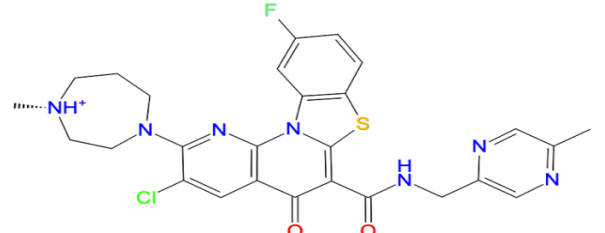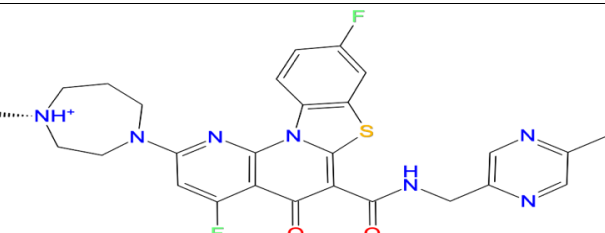
Chemical Structure of CX-5461: Substitution Sites Marked	Docking Score	
	h-Tel: -8.386 CKIT1: -8.112 c-Myc: -8.567 Duplex: -5.944	
Chemical Structure of New Ligand	**Docking Score**	**△CX-5461**
	h-Tel: -8.930 CKIT1: -8.552 c-Myc: -9.165 Duplex: -4.241	h-Tel: -0.544 CKIT1: -0.440 c-Myc: -0.598 Duplex: +2.703
	h-Tel: -9.272 CKIT1: -8.771 c-Myc: -8.818 Duplex: -3.326	h-Tel: -0.886 CKIT1: -0.659 c-Myc: -0.251 Duplex: +2.618
	h-Tel: -8.913 CKIT1: -8.179 c-Myc: -8.808 Duplex: -4.334	h-Tel: -0.527 CKIT1: -0.067 c-Myc: -0.241 Duplex: +1.610
	h-tel: -9.179 c-KIT1: -8.175 c-Myc: -8.667 Duplex: -5.273	h-Tel: -0.793 CKIT1: -0.063 c-Myc: -0.100 Duplex: +0.671

Figure 25. Chemical structure of CX-5461 and derivates identified through virtual screening, including docking scores. For CX-5461 green arrows indicate substitution sites for the derivates on this table. For the derivates, docking scores are provided as well as the difference between their docking scores and CX-5461.

106

While picking possible substitution sites, we closely compared our work to the original structure activity relationship study for which CX-5461 was identified[118]. Of note, their work found that the addition of a F at our position S3 significantly increased the cellular activity of the compound, to a level even higher than reported for CX-5461. However, this addition was not made to CX-5461 itself. Rather than using the methylpyrazine based flexible side chain R2 (Figure 12) for their trial at this site they used a pyrrolidinoethylamine side chain. They discovered that this combination, although highly beneficial in terms of cellular activity, had an extremely low oral absorption. They later discovered that the addition of the methylpyrazine based side chain (R2) drastically increased the oral availability (~40-fold increase: ~2,300 to ~92,000 ng h/mL), however failed to revisit the possibility of using F at substitution site S3 with the addition of R2 to the aromatic core. Which led us to believe that because R2 provides such a drastic increase in oral availability, and the F at site S3 leads to a drastic increase in cellular activity, this combination could provide the resolve for both important factors. Our docking results further supported the promise of this combination where F at site S3 increased the binding to all G4's and provided the largest decrease in duplex binding (+2.703 kcal/mol) of all our derivatives. A closely following second was a Cl substitution at site S3 which also increased G4 binding and decreased duplex binding (+2.618 kcal/mol). A third ligand was identified which substituted F at site S3 and Cl at site S1 which increased G4 binding and decrease duplex binding (+1.610 kcal/mol). In addition to these ligands, we identified a number of other derivates which could be promising. This includes a ligand with F substitutions at sites S2 and S4 which increased binding for each the human telomeric, c-KIT1 and c-Myc G4s and decreased duplex binding (Figure

107

25) and a ligand which had a Cl substitution at site S2 which increased binding for each the human telomeric and c-Myc G4s but decreased binding for both the c-KIT1 G4 and duplex (Figure B35). All together suggesting substitutions at site S3 may address both the necessary biological properties as well as well as the intended decrease in duplex binding.

Furthermore, an interesting trend was observed for the human telomeric system. First, of the 56 ligands the intended effect (i.e. increase G4 binding or decrease duplex binding) was most observed for the human telomeric G4 system (48%) followed by the duplex (44%). Looking closer at the docking results, there were many new derivatives which produced the intended effect in just the human telomeric and duplex systems, six of which are exemplified in the supporting document (Figure B35). Putting together the most thermodynamically favorable system in our study was the human telomeric G4, the stability results of the melting temperature assay, and the results of our docking study, we suggest that using CX-5461 derivatives to specifically target the human telomeric G4 could be a promising therapeutic approach. This hypothesis is consistent with limited experimental testing such as a telomere FISH assay[117] which showed an increased frequency of telomere defects in BRCA -/- HCT116 cells after exposure to CX-5461 which they used to provide support of CX-5461's G4 stabilizing and ability to induce genome instability specifically at G4s in human cells. However, only targeting human telomeric G4 provides a limited solution that ties back into a major limiting factor of G4 stabilizers to date, whereby, the structure of the G4s overall do not provide a large binding pocket, as seen in proteins and other targets. This has implications on the reported efficacies where a G4 stabilizer has yet to bind in the nM concentration range.

One approach to remedy this is to generate a hybrid G4 binder that has both the pharmacophores of a G4 intercalator and groove binder. The intention of this ligand design is that the intercalator core of this ligand will intercalate onto a G-quartet and, that when connected by a flexible linker, the groove binder like side chains may wrap around the G4 core like arms and further stabilize the interactions. As far as we know, there has been no such ligand developed with this binding pose to date. Of course, further experimentation is required to support this hypothesis.

References

1.	DiMasi, J. A.; Feldman, L.; Seckler, A.; Wilson, A., Trends in Risks Associated With New Drug Development: Success Rates for Investigational Drugs. *Clinical Pharmacology & Therapeutics* **2010**, 87, 272-277.

2.	Yella, J. K.; Yaddanapudi, S.; Wang, Y.; Jegga, A. G., Changing Trends in Computational Drug Repositioning. *Pharmaceuticals (Basel)* **2018**, 11, 57.

3.	Yao, L.; Evans, J. A.; Rzhetsky, A., Novel opportunities for computational biology and sociology in drug discovery. *Trends Biotechnol* **2010**, 28, 161-170.

4.	FDA, In Products, N. D. a. F. C. s. N. M. E. a. N. T. B., Ed.; 2018.

5.	Huggins, D. J.; Biggin, P. C.; Dämgen, M. A.; Essex, J. W.; Harris, S. A.; Henchman, R. H.; Khalid, S.; Kuzmanic, A.; Laughton, C. A.; Michel, J.; Mulholland, A. J.; Rosta, E.; Sansom, M. S. P.; van der Kamp, M. W., Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity. *WIREs Computational Molecular Science* **2019**, 9, e1393.

6.	Cole, D. J.; Horton, J. T.; Nelson, L.; Kurdekar, V., The future of force fields in computer-aided drug design. *Future Medicinal Chemistry* **2019**, 11, 2359-2363.

7.	Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J., Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols* **2016**, 11, 905-919.

8.	Salmaso, V.; Moro, S., Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Front Pharmacol* **2018**, 9, 923-923.

9.	Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. I., Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Model. *Acc. Chem. Res* **2000**, 33, 889-897.

10.	Debray, J.; Zeghida, W.; Jourdan, M.; Monchaud, D.; Dheu-Andries, M. L.; Dumy, P.; Teulade-Fichou, M. P.; Demeunynck, M., Synthesis and Evaluation of Fused Bispyrimidinoacridines as Novel Pentacyclic Analogues of Quadruplex-Binder BRACO-19. *Org. Biomol. Chem.* **2009**, 7, 5219-5228.

11.	Doluca, O.; Withers, J. M.; Filichev, V. V., Molecular Engineering of Guanine-Rich Sequences: Z-DNA, DNA Triplexes, and G-Quadruplexes. *Chem. Rev. (Washington, DC, U. S.)* **2013**, 113, 3044-3083.

12.   Chung, W. J.; Heddi, B.; Tera, M.; Iida, K.; Nagasawa, K.; Phan, A. T., Solution Structure of an Intramolecular (3+1) Human Telomeric G-Quadruplex Bound to a Telomestatin Derivative. *J. Am. Chem. Soc.* **2013**, 135, 13495-13501.

13.   Moyzis, R. K.; Buckingham, J. M.; Cram, L. S.; Dani, M.; Deaven, L. L.; Jones, M. D.; Meyne, J.; Ratliff, R. L.; Wu, J. R., A Highly Conserved Repetitive DNA Sequence, (TTAGGG)n, Present at the Telomeres of Human Chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, 85, 6622-6626.

14.   Wright, W. E.; Tesmer, V. M.; Huffman, K. E.; Levene, S. D.; Shay, J. W., Normal Human Chromosomes have Long G-Rich Telomeric Overhangs at One End. *Genes Dev.* **1997**, 11, 2801-2809.

15.   de Lange, T., Shelterin: the Protein Complex that Shapes and Safeguards Human Telomeres. *Genes & development* **2005**, 19, 2100-10.

16.   Palm, W.; de Lange, T., How Shelterin Protects Mammalian Telomeres. *Annual review of genetics* **2008**, 42, 301-34.

17.   Harley, C. B.; Futcher, A. B.; Greider, C. W., Telomeres Shorten During Ageing of Human Fibroblasts. *Nature (London, U. K.)* **1990**, 345, 458-460.

18.   Zakian, V. A., Telomeres: Beginning to Understand the End. *Science* **1995**, 270, 1601.

19.   Moorhouse, A. D.; Santos, A. M.; Gunaratnam, M.; Moore, M.; Neidle, S.; Moses, J. E., Stabilization of G-Quadruplex DNA by Highly Selective Ligands Via Click Chemistry. *J. Am. Chem. Soc.* **2006**, 128, 15972-15973.

20.   Greider, C. W.; Blackburn, E. H., A Telomeric Sequence in the RNA of Tetrahymena Telomerase Required for Telomere Repeat Synthesis. *Nature (London, U. K.)* **1989**, 337, 331-337.

21.   Shay, J. W.; Wright, W. E., Telomerase Therapeutics for Cancer: Challenges and New Directions. *Nat. Rev. Drug Discovery* **2006**, 5, 577-584.

22.   Asai, A.; Oshima, Y.; Yamamoto, Y.; Uochi, T.-a.; Kusaka, H.; Akinaga, S.; Yamashita, Y.; Pongracz, K.; Pruzan, R.; Wunder, E.; Piatyszek, M.; Li, S.; Chin, A. C.; Harley, C. B.; Gryaznov, S., A Novel Telomerase Template Antagonist (GRN163) as a Potential Anticancer Agent. *Cancer Res.* **2003**, 63, 3931.

23.   Dunham, M. A.; Neumann, A. A.; Fasching, C. L.; Reddel, R. R., Telomere Maintenance by Recombination in Human Cells. *Nature genetics* **2000**, 26, 447-50.

24.   Bechter, O. E.; Zou, Y.; Walker, W.; Wright, W. E.; Shay, J. W., Telomeric Recombination in Mismatch Repair Deficient Human Colon Cancer Cells after Telomerase Inhibition. *Cancer Res.* **2004**, 64, 3444.

25.   Hu, J.; Hwang, S. S.; Liesa, M.; Gan, B.; Sahin, E.; Jaskelioff, M.; Ding, Z.; Ying, H.; Boutin, A. T.; Zhang, H.; Johnson, S.; Ivanova, E.; Kost-Alimova, M.; Protopopov, A.; Wang, Y. A.; Shirihai, O. S.; Chin, L.; DePinho, R. A., Anti-Telomerase Therapy Provokes ALT and Mitochondrial Adaptive Mechanisms in Cancer. *Cell (Cambridge, MA, U. S.)* **2012**, 148, 651-663.

26.   Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M., Inhibition of Telomerase by G-Quartet DNA Structures. *Nature (London, U. K.)* **1991**, 350, 718-720.

27.   Denchi, E. L.; de Lange, T., Protection of Telomeres Through Independent Control of ATM and ATR by TRF2 and POT1. *Nature (London, U. K.)* **2007**, 448, 1068-1071.

28.   Hänsel, R.; Löhr, F.; Foldynová-Trantírková, S.; Bamberg, E.; Trantírek, L.; Dötsch, V., The Parallel G-Quadruplex Structure of Vertebrate Telomeric Repeat Sequences is not the Preferred Folding Topology under Physiological Conditions. *Nucleic Acids Res.* **2011**, 39, 5768-5775.

29.   Davis, J. T., G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angewandte Chemie (International ed. in English)* **2004**, 43, 668-98.

30.   Patel, D. J.; Phan, A. T.; Kuryavyi, V., Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic acids research* **2007**, 35, 7429-55.

31.   Neidle, S., The structures of quadruplex nucleic acids and their drug complexes. *Current opinion in structural biology* **2009**, 19, 239-50.

32.   Phan, A. T., Human telomeric G-quadruplex: structures of DNA and RNA sequences. *The FEBS journal* **2010**, 277, 1107-17.

33.   Wang, Y.; Patel, D. J., SOLUTION STRUCTURE OF THE HUMAN TELOMERIC REPEAT D AG(3)(T(2)AG(3))3 G-TETRAPLEX. *Structure* **1993**, 1, 263-282.

34.   Dvorkin, S. A.; Karsisiotis, A. I.; da Silva, M. W., Encoding canonical DNA quadruplex structure. *Science Advances* **2018**, 4.

35.   Parkinson, G. N.; Lee, M. P. H.; Neidle, S., Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002**, 417, 876-880.

36.   Smith, F. W.; Feigon, J., Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. *Nature* **1992**, 356, 164-168.

37.   Horvath, M. P.; Schultz, S. C., DNA G-quartets in a 1.86 A resolution structure of an Oxytricha nova telomeric protein-DNA complex. *Journal of molecular biology* **2001**, 310, 367-77.

38.   Miller, M. C.; Buscaglia, R.; Chaires, J. B.; Lane, A. N.; Trent, J. O., Hydration Is a Major Determinant of the G-Quadruplex Stability and Conformation of the Human Telomere 3 ' Sequence of d(AG(3)(TTAG(3))(3)). *Journal of the American Chemical Society* **2010**, 132, 17105-17107.

39.   Heddi, B.; Phan, A. T., Structure of Human Telomeric DNA in Crowded Solution. *Journal of the American Chemical Society* **2011**, 133, 9824-9833.

40.   Redon, S.; Bombard, S.; Elizondo-Riojas, M.-A.; Chottard, J.-C., Platinum cross-linking of adenines and guanines on the quadruplex structures of the AG3(T2AG3)3 and (T2AG3)4 human telomere sequences in Na+ and K+ solutions. *Nucleic acids research* **2003**, 31, 1605-1613.

41.   Phan, A. T.; Patel, D. J., Two-Repeat Human Telomeric d(TAGGGTTAGGGT) Sequence Forms Interconverting Parallel and Antiparallel G-Quadruplexes in Solution: Distinct Topologies, Thermodynamic Properties, and Folding/Unfolding Kinetics. *Journal of the American Chemical Society* **2003**, 125, 15021-15027.

42.   Ying, L.; Green, J. J.; Li, H.; Klenerman, D.; Balasubramanian, S., Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences* **2003**, 100, 14629.

43.   He, Y. J.; Neumann, R. D.; Panyutin, I. G., Intramolecular quadruplex conformation of human telomeric DNA assessed with I-125-radioprobing. *Nucleic acids research* **2004**, 32, 5359-5367.

44.   Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D., Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic acids research* **2006**, 34, 2723-35.

45.   Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J., Structure of the human telomere in K+ solution: An intramolecular (3+1) G-quadruplex scaffold. *Journal of the American Chemical Society* **2006**, 128, 9963-9970.

46.   Phan, A. T.; Luu, K. N.; Patel, D. J., Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K+ solution. *Nucleic acids research* **2006**, 34, 5715-9.

113

47.   Xu, Y.; Noguchi, Y.; Sugiyama, H., The new models of the human telomere d[AGGG(TTAGGG)3] in K+ solution. *Bioorganic & medicinal chemistry* **2006**, 14, 5584-91.

48.   Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T., Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J Am Chem Soc* **2009**, 131, 4301-9.

49.   Dai, J. X.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. Z., Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic acids research* **2007**, 35, 2440-2450.

50.   Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B., Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic acids research* **2005**, 33, 4649-4659.

51.   Qi, J. Y.; Shafer, R. H., Covalent ligation studies on the human telomere quadruplex. *Biophysical Journal* **2005**, 88, 57A-57A.

52.   Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T., Extreme conformational diversity in human telomeric DNA. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, 102, 18938.

53.   Xu, Y.; Sugiyama, H., Highly Efficient Photochemical 2'-Deoxyribonolactone Formation at the Diagonal Loop of a 5-Iodouracil-Containing Antiparallel G-Quartet. *Journal of the American Chemical Society* **2004**, 126, 6274-6279.

54.   Ourliac-Garnier, I.; Elizondo-Riojas, M.-A.; Redon, S.; Farrell, N. P.; Bombard, S., Cross-Links of Quadruplex Structures from Human Telomeric DNA by Dinuclear Platinum Complexes Show the Flexibility of Both Structures. *Biochemistry* **2005**, 44, 10620-10634.

55.   Vorlíčková, M.; Chládková, J.; Kejnovská, I.; Fialová, M.; Kypr, J., Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic acids research* **2005**, 33, 5851-5860.

56.   Rujan, I. N.; Meleney, J. C.; Bolton, P. H., Vertebrate telomere repeat DNAs favor external loop propeller quadruplex structures in the presence of high concentrations of potassium. *Nucleic acids research* **2005**, 33, 2022-2031.

57.   Yang, D.; Okamoto, K., Structural Insights into G-Quadruplexes: Towards New Anticancer Drugs. *Future Med. Chem.* **2010**, 2, 619-646.

58.   Read, M.; Harrison, R. J.; Romagnoli, B.; Tanious, F. A.; Gowan, S. H.; Reszka, A. P.; Wilson, W. D.; Kelland, L. R.; Neidle, S., Structure-based design of selective and

potent G quadruplex-mediated telomerase inhibitors. *Proceedings of the National Academy of Sciences* **2001**, 98, 4844.

59.   Incles, C. M.; Schultes, C. M.; Kempski, H.; Koehler, H.; Kelland, L. R.; Neidle, S., A G-Quadruplex Telomere Targeting Agent Produces p16-Associated Senescence and Chromosomal Fusions in Human Prostate Cancer Cells. *Mol. Cancer Ther.* **2004**, 3, 1201-1206.

60.   Perrone, R.; Butovskaya, E.; Daelemans, D.; Palù, G.; Pannecouque, C.; Richter, S. N., Anti-HIV-1 Activity of the G-Quadruplex Ligand BRACO-19. *J. Antimicrob. Chemother.* **2014**, 69, 3248-3258.

61.   Ruggiero, E.; Richter, S. N., G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic acids research* **2018**, 46, 3270-3283.

62.   Belmonte-Reche, E.; Martinez-Garcia, M.; Guedin, A.; Zuffo, M.; Arevalo-Ruiz, M.; Doria, F.; Campos-Salinas, J.; Maynadier, M.; Lopez-Rubio, J. J.; Freccero, M.; Mergny, J. L.; Perez-Victoria, J. M.; Morales, J. C., G-Quadruplex Identification in the Genome of Protozoan Parasites Points to Naphthalene Diimide Ligands as New Antiparasitic Agents. *Journal of Medicinal Chemistry* **2018**, 61, 1231-1240.

63.   Arevalo-Ruiz, M.; Doria, F.; Belmonte-Reche, E.; De Rache, A.; Campos-Salinas, J.; Lucas, R.; Falomir, E.; Carda, M.; Perez-Victoria, J. M.; Mergny, J. L.; Freccero, M.; Morales, J. C., Synthesis, Binding Properties, and Differences in Cell Uptake of G-Quadruplex Ligands Based on Carbohydrate Naphthalene Diimide Conjugates. *Chemistry-a European Journal* **2017**, 23, 2157-2164.

64.   Gunaratnam, M.; Greciano, O.; Martins, C.; Reszka, A. P.; Schultes, C. M.; Morjani, H.; Riou, J.-F.; Neidle, S., Mechanism of acridine-based telomerase inhibition and telomere shortening. *Biochemical Pharmacology* **2007**, 74, 679-689.

65.   Brandes, L. J.; Hermonat, M. W., Receptor Status and Subsequent Sensitivity of Subclones of MCF-7 Human Breast Cancer Cells Surviving Exposure to Diethylstilbestrol. *Cancer Research* **1983**, 43, 2831.

66.   Fang, R.; Aust, A. E., Induction of Ferritin Synthesis in Human Lung Epithelial Cells Treated with Crocidolite Asbestos. *Archives of Biochemistry and Biophysics* **1997**, 340, 369-375.

67.   Mickey, D. D.; Stone, K. R.; Wunderli, H.; Mickey, G. H.; Vollmer, R. T.; Paulson, D. F., Heterotransplantation of a Human Prostatic Adenocarcinoma Cell Line in Nude Mice. *Cancer Research* **1977**, 37, 4049.

68.   Chen, T. R.; Drabkowski, D.; Hay, R. J.; Macy, M.; Peterson Jr, W., WiDr is a derivative of another colon adenocarcinoma cell line, HT-29. *Cancer Genetics and Cytogenetics* **1987**, 27, 125-134.

115

69.   Akagi, T.; Kimoto, T., HUMAN CELL LINE (HGC-27) DERIVED FROM METASTATIC LYMPH-NODE OF GASTRIC CANCER. *Acta Medica Okayama* **1976**, 30, 215-219.

70.   Alizadehnohi, M.; Nabiuni, M.; Nazari, Z.; Safaeinejad, Z.; Irian, S., The synergistic cytotoxic effect of cisplatin and honey bee venom on human ovarian cancer cell line A2780cp. *Journal of Venom Research* **2012**, 3, 22-27.

71.   Landers, J. E.; Cassel, S. L.; George, D. L., Translational Enhancement of Oncogene Expression in Human Tumor Cells Containing a Stabilized Wild-Type p53 Protein. *Cancer Research* **1997**, 57, 3562.

72.   Nichols, W. W.; Murphy, D. G.; Cristofalo, V. J.; Toji, L. H.; Greene, A. E.; Dwight, S. A., CHARACTERIZATION OF A NEW HUMAN DIPLOID CELL STRAIN, IMR-90. *Science* **1977**, 196, 60-63.

73.   Zhou, G. T.; Liu, X. R.; Li, Y. Q.; Xu, S. B.; Ma, C. Y.; Wu, X. M.; Cheng, Y.; Yu, Z. Y.; Zhao, G.; Chen, Y., Telomere targeting with a novel G-quadruplex-interactive ligand BRACO-19 induces T-loop disassembly and telomerase displacement in human glioblastoma cells. *Oncotarget* **2016**, 7, 14925-14939.

74.   Olopade, O. I.; Jenkins, R. B.; Ransom, D. T.; Malik, K.; Pomykala, H.; Nobori, T.; Cowan, J. M.; Rowley, J. D.; Diaz, M. O., Molecular Analysis of Deletions of the Short Arm of Chromosome 9 in Human Gliomas. *Cancer Research* **1992**, 52, 2523.

75.   Burger, A. M.; Dai, F. P.; Schultes, C. M.; Reszka, A. P.; Moore, M. J.; Double, J. A.; Neidle, S., The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Research* **2005**, 65, 1489-1496.

76.   Harrison, R. J.; Reszka, A. P.; Haider, S. M.; Romagnoli, B.; Morrell, J.; Read, M. A.; Gowan, S. M.; Incles, C. M.; Kelland, L. R.; Neidle, S., Evaluation of by disubstituted acridone derivatives as telomerase inhibitors: the importance of G-quadruplex binding. *Bioorganic & Medicinal Chemistry Letters* **2004**, 14, 5845-5849.

77.   Morimoto, H.; Safrit, J. T.; Bonavida, B., Synergistic effect of tumor necrosis factor-alpha- and diphtheria toxin-mediated cytotoxicity in sensitive and resistant human ovarian tumor cell lines. *The Journal of Immunology* **1991**, 147, 2609.

78.   Kellner, J.; Wierda, W.; Shpall, E.; Keating, M.; McNiece, I., Isolation of a novel chronic lymphocytic leukemic (CLL) cell line and development of an in vivo mouse model of CLL. *Leukemia Research* **2016**, 40, 54-59.

79.   Rankin, A. M.; Faller, D. V.; Spanjaard, R. A., Telomerase inhibitors and 'T-oligo' as cancer therapeutics: contrasting molecular mechanisms of cytotoxicity. *Anti-Cancer Drugs* **2008**, 19, 329-338.

80.   Campbell, N. H.; Parkinson, G. N.; Reszka, A. P.; Neidle, S., Structural Basis of DNA Quadruplex Recognition by an Acridine Drug. *Journal of the American Chemical Society* **2008**, 130, 6722-6724.

81.   Heddi, B.; Phan, A. T., Structure of human telomeric DNA in crowded solution. *Journal of the American Chemical Society* **2011**, 133, 9824-33.

82.   Machireddy, B.; Kalra, G.; Jonnalagadda, S.; Ramanujachary, K.; Wu, C., 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* **2017**, 57, 2846-2864.

83.   Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E. r.; Laughton, C. A.; Orozco, M., Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, 92, 3817–3829.

84.   Krepl, M.; Zgarbova, M.; Stadlbauer, P.; Otyepka, M.; Banas, P.; Koca, J.; Cheatham, T. E.; Jurecka, P.; Sponer, J., Reference Simulations of Noncanonical Nucleic Acids with Different chi Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J Chem Theory Comput* **2012**, 8, 2506-2520.

85.   Zgarbova, M.; Luque, F. J.; Sponer, J.; Cheatham, T. E.; Otyepka, M.; Jurecka, P., Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J Chem Theory Comput* **2013**, 9, 2339-2354.

86.   Zgarbova, M.; Sponer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurecka, P., Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J Chem Theory Comput* **2015**, 11, 5723-5736.

87.   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparisons of simple potential functions for simulating liquid water. *J. Comp. Phys.* **1983**, 79, 926-935.

88.   Joung, I. S.; Cheatham, T. E., Determination of Alkali and Halide Monovalent ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B.* **2008**, 112, 9020-9041.

89.   Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem.* **1993**, 97, 10269-10280.

90.   Case, D. A.; Betz, R. M.; Botello-Smith, W.; Cerutti, D. S.; Cheatham, I., T.E. ; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B. *AMBER 2016*, University of California, San Francisco, 2016.

91.   Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T. E. I.; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J., 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.* **2010**, 38, 299-313.

92.   Cosconati, S.; Marinelli, L.; Trotta, R.; Virno, A.; De Tito, S.; Romagnoli, R.; Pagano, B.; Limongelli, V.; Giancola, C.; Baraldi, P. G.; Mayol, L.; Novellino, E.; Randazzo, A., Structural and Conformational Requisites in DNA Quadruplex Groove Binding: Another Piece to the Puzzle. *J. Am. Chem. Soc.* **2010**, 132, 6425-6433.
93.   Fadrna, E.; Spackova, N. a.; Sarzynska, J.; Koca, J.; Orozco, M.; Cheatham, T. E., III;; Kulinski, T.; Sponer, J., Single Stranded Loops of Quadruplex DNA As Key Benchmark for Testing Nucleic Acids Force Fields. *J Chem Theory Comput* **2009**, 5, 2514-2530

94.   Mukherjee, A.; Lavery, R.; Bagchi, B.; Hynes, J. T., 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.* **2008**, 130, 9747-9755.

95.   Deng, N. J.; Wickstrom, L.; Cieplak, P.; Lin, C.; Yang, D. Z., Resolving the Ligand-Binding Specificity in c-MYC G-Quadruplex DNA: Absolute Binding Free Energy Calculations and SPR Experiment. *Journal of Physical Chemistry B* **2017**, 121, 10484-10497.

96.   Lei, H.; Wang, X.; Wu, C., Early Stage Intercalation of Doxorubicin to DNA Fragments Observed in Molecular Dynamics Binding Simulations. *Journal of molecular graphics & modelling* **2012**, 38, 279-89.

97.   Mulholland, K.; Wu, C., Binding of Telomestatin to a Telomeric G-Quadruplex DNA Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent. *J. Chem. Inf. Model* **2016**, 56, 2093-2102.

98.   Ryckaert, J.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comp. Phys.* **1977**, 23, 327-341.

99.   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T. A.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Comp. Phys.* **1995**, 103, 8577-8593.

100.   Procacci, P.; Berne, B. J., Multiple Time-Scale Methods For Constant-Pressure Molecular-Dynamics Simulations Of Molecular-Systems. *Molec. Phys.* **1994**, 83, 255-272.

101.   Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E., Peptide Folding: When Simulation Meets Experiment. *Angew Chem Int Edit* **1999**, 38, 236-240.

102.   Kongsted, J.; Soderhjelm, P.; Ryde, U., How Accurate are Continuum Solvation Models for Drug-Like Molecules? *J. Comput. Aided Mol. Des.* **2009**, 23, 395-409.

103.   Lei, H.; Wang, X.; Wu, C., Early stage intercalation of doxorubicin to DNA fragments observed in molecular dynamics binding simulations. *Journal of molecular graphics & modelling* **2012**, 38, 279-89.

104.   Kongsted, J.; Soderhjelm, P.; Ryde, U., How accurate are continuum solvation models for drug-like molecules? *Journal of Computer-Aided Molecular Design* **2009**, 23, 395-409.

105.   Kollman, P.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.; Cheatham, T., Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, 33, 889-897.

106.   Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, 51, 69-82.

107.   Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the Performance of the Molecular Mechanics/Poisson Boltzmann Surface Area and Molecular Mechanics/Generalized Born Surface Area Methods. II. The Accuracy of Ranking Poses Generated from Docking. *J. Comput. Chem.* **2010**, 32, 866-877.

108.   Mulholland, K.; Siddiquei, F.; Wu, C., 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* **2017**, 18685-18694

109.   White, E. W.; Tanious, F.; Ismail, M. A.; Reszka, A. P.; Neidle, S.; Boykin, D. W.; Wilson, W. D., Structure-Specific Recognition of Quadruplex DNA by Organic Cations: Influence of Shape, Substituents and Charge. *Biophys. Chem.* **2007**, 126, 140-153.

119

110.   Harrison, R. J.; Gowan, S. M.; Kelland, L. R.; Neidle, S., Human Telomerase Inhibition by Substituted Acridine Derivatives. *Bioorg. Med. Chem. Lett.* **1999**, 9, 2463-8.

111.   Harrison, R. J.; Cuesta, J.; Chessari, G.; Read, M. A.; Basra, S. K.; Reszka, A. P.; Morrell, J.; Gowan, S. M.; Incles, C. M.; Tanious, F. A.; Wilson, W. D.; Kelland, L. R.; Neidle, S., Trisubstituted acridine derivatives as potent and selective telomerase inhibitors. *Journal of Medicinal Chemistry* **2003**, 46, 4463-4476.

112.   Neidle, S., Quadruplex Nucleic Acids as Novel Therapeutic Targets. *Journal of medicinal chemistry* **2016**, 59, 5987-6011.

113.   Risitano, A.; Fox, K. R., Stability of intramolecular DNA quadruplexes: Comparison with DNA duplexes. *Biochemistry* **2003**, 42, 6507-6513.

114.   Hansel-Hertsch, R.; Di Antonio, M.; Balasubramanian, S., DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nature reviews. Molecular cell biology* **2017**, 18, 279-284.

115.   Todd, A. K.; Johnston, M.; Neidle, S., Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic acids research* **2005**, 33, 2901-7.

116.   Huppert, J. L.; Balasubramanian, S., Prevalence of quadruplexes in the human genome. *Nucleic acids research* **2005**, 33, 2908-16.

117.   Xu, H.; Di Antonio, M.; McKinney, S.; Mathew, V.; Ho, B.; O'Neil, N. J.; Dos Santos, N.; Silvester, J.; Wei, V.; Garcia, J.; Kabeer, F.; Lai, D.; Soriano, P.; Banath, J.; Chiu, D. S.; Yap, D.; Le, D. D.; Ye, F. B.; Zhang, A. N.; Thu, K.; Soong, J.; Lin, S. C.; Tsai, A. H. C.; Osako, T.; Algara, T.; Saunders, D. N.; Wong, J.; Xian, J.; Bally, M. B.; Brenton, J. D.; Brown, G. W.; Shah, S. P.; Cescon, D.; Mak, T. W.; Caldas, C.; Stirling, P. C.; Hieter, P.; Balasubramanian, S.; Aparicio, S., CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nature Communications* **2017**, 8, 14432

118.   Haddach, M.; Schwaebe, M. K.; Michaux, J.; Nagasawa, J.; O'Brien, S. E.; Whitten, J. P.; Pierre, F.; Kerdoncuff, P.; Darjania, L.; Stansfield, R.; Drygin, D.; Anderes, K.; Proffitt, C.; Bliesath, J.; Siddiqui-Jain, A.; Omori, M.; Huser, N.; Rice, W. G.; Ryckman, D. M., Discovery of CX-5461, the First Direct and Selective Inhibitor of RNA Polymerase I, for Cancer Therapeutics. *Acs Medicinal Chemistry Letters* **2012**, 3, 602-606.

119.   Andrews, W. J.; Panova, T.; Normand, C.; Gadal, O.; Tikhonova, I. G.; Panov, K. I., Old Drug, New Target: Ellipticines Selectively Inhibit RNA Polymerase I Transcription. *Journal of Biological Chemistry* **2013**, 288, 4567-4582.

120.   Ye, F. B.; Hamza, A.; Singh, T.; Flibotte, S.; Hieter, P.; O'Neil, N. J., Characterization of the phototoxicity, chemigenetic profile, and mutational signatures of the chemotherapeutic CX-5461 in Caenorhabditis elegans. *bioRxiv* **2019**, 2019.12.20.884981.

121.   Hilton, J.; Cescon, D. W.; Bedard, P.; Ritter, H.; Tu, D.; Soong, J.; Gelmon, K.; Aparicio, S.; Seymour, L., CCTG IND. 231: A phase 1 trial evaluating CX-5461 in patients with advanced solid tumors. *Annals of Oncology* **2018**, 29, III8.

122.   Wang, T. T.; Shatara, M.; Liu, F. B.; Knight, T.; Edwards, H.; Wang, G.; Lin, H.; Wang, Y.; Taub, J. W.; Ge, Y. B., Simultaneous cotargeting of ATR and RNA Polymerase I transcription demonstrates synergistic antileukemic effects on acute myeloid leukemia. *Signal Transduction and Targeted Therapy* **2019**, 4, 44.

123.   Leung, A. W. Y.; Anantha, M.; Prosser, K. E.; Wehbe, M.; Walsby, C. J.; Bally, M. B., A novel formulation of CX-5461, a small-molecule inhibitor of rRNA synthesis, and its use for treatment of acute myeloid leukemia models. *Cancer Research* **2016**, 76.

124.   Lee, H.; Wang, H.; Lin, H.; Baladandayuthapani, V.; He, J.; Jones, R. J.; Kuiatse, I.; Gu, D. M.; Wang, Z. Q.; O'Brien, S.; Keats, J. J.; Yang, J.; Davis, R. E.; Orlowski, R. Z., RNA Polymerase I Inhibition with CX-5461 As a Novel Strategy to Target MYC in Multiple Myeloma. *Blood* **2015**, 126.

125.   Maclachlan, K. H.; Cuddihy, A.; Hein, N.; Cullinane, C.; Harrison, S. J.; Hannan, R.; Poortinga, G., Novel Combination Therapies with the RNA Polymerase I Inhibitor CX-5461 Significantly Improve Efficacy in Multiple Myeloma. *Blood* **2017**, 130.

126.   Taylor, J. S.; Zeki, J.; Ornell, K.; Coburn, J.; Shimada, H.; Ikegaki, N.; Chiu, B., Down-regulation of MYCN protein by CX-5461 leads to neuroblastoma tumor growth suppression. *Journal of Pediatric Surgery* **2019**, 54, 1192-1197.

127.   Rebello, R. J.; Kusnadi, E.; Cameron, D. P.; Pearson, H. B.; Lesmana, A.; Devlin, J. R.; Drygin, D.; Clark, A. K.; Porter, L.; Pedersen, J.; Sandhu, S.; Risbridger, G. P.; Pearson, R. B.; Hannan, R. D.; Furic, L., The Dual Inhibition of RNA Pol I Transcription and PIM Kinase as a New Therapeutic Approach to Treat Advanced Prostate Cancer. *Clinical Cancer Research* **2016**, 22, 5539-5552.

128.   Li, L. M.; Li, Y.; Zhao, J. S.; Fan, S. L.; Wang, L. G.; Li, X., CX-5461 induces autophagy and inhibits tumor growth via mammalian target of rapamycin-related signaling pathways in osteosarcoma. *Oncotargets and Therapy* **2016**, 9, 5985-5997.

129.   Negi, S. S.; Brown, P., rRNA synthesis inhibitor, CX-5461, activates ATM/ATR pathway in acute lymphoblastic leukemia, arrests cells in G2 phase and induces apoptosis. *Oncotarget* **2015**, 6, 18094-18104.

130.   Negi, S. S.; Brown, P., Transient rRNA synthesis inhibition with CX-5461 is sufficient to elicit growth arrest and cell death in acute lymphoblastic leukemia cells. *Oncotarget* **2015**, 6, 34846-34858.

131.   Cornelison, R.; Dobbin, Z. C.; Katre, A. A.; Jeong, D. H.; Zhang, Y. F.; Chen, D. Q.; Petrova, Y.; Llaneza, D. C.; Steg, A. D.; Parsons, L.; Schneider, D. A.; Landen, C. N., Targeting RNA-Polymerase I in Both Chemosensitive and Chemoresistant Populations in Epithelial Ovarian Cancer. *Clinical Cancer Research* **2017**, 23, 6529-6540.

132.   Cornelison, R.; Llaneza, D. C.; Petrova, Y.; Dobbin, Z. C.; Schneider, D. A.; Landen, C. N., Inhibition of autophagy potentiates cytotoxicity of CX-5461 treatment in chemoresistant epithelial ovarian cancer. *Cancer Research* **2017**, 77.

133.   Cornelison, R.; Dobbin, Z. C.; Katie, A. A.; Jeong, D. H.; Petrova, Y.; Llaneza, D. C.; Steg, A. D.; Zhang, Y.; Schneider, D. A.; Landen, C. N., Targeting RNA-polymerase I using CX-5461 as a mechanism for treating chemotherapy resistant epithelial ovarian cancer. *Cancer Research* **2016**, 76.

134.   Ye, Q.; Pang, S.; Zhang, W. J.; Guo, X. T.; Wang, J. L.; Zhang, Y. T.; Liu, Y.; Wu, X.; Jiang, F., Therapeutic Targeting of RNA Polymerase I With the Small-Molecule CX-5461 for Prevention of Arterial Injury-Induced Neointimal Hyperplasia. *Arteriosclerosis Thrombosis and Vascular Biology* **2017**, 37, 476-484.

135.   Kostopoulou, O. N.; Wilhelmi, V.; Raiss, S.; Ananthaseshan, S.; Lindstrom, M. S.; Bartek, J.; Soderberg-Naucler, C., Human cytomegalovirus and Herpes Simplex type I virus can engage RNA polymerase I for transcription of immediate early genes. *Oncotarget* **2017**, 8, 96536-96552.

136.   Westdorp, K. N.; Terhune, S. S., Impact of RNA polymerase I inhibitor CX-5461 on viral kinase-dependent and -independent cytomegalovirus replication. *Antiviral Research* **2018**, 153, 33-38.

137.   Kerry, L. E.; Pegg, E. E.; Cameron, D. P.; Budzak, J.; Poortinga, G.; Hannan, K. M.; Hannan, R. D.; Rudenko, G., Selective inhibition of RNA polymerase I transcription as a potential approach to treat African trypanosomiasis. *Plos Neglected Tropical Diseases* **2017**, 11.

138.   Cosconati, S.; Rizzo, A.; Trotta, R.; Pagano, B.; Iachettini, S.; De Tito, S.; Lauri, I.; Fotticchia, I.; Giustiniano, M.; Marinelli, L.; Giancola, C.; Novellino, E.; Biroccio, A.; Randazzo, A., Shooting for selective druglike G-quadruplex binders: evidence for telomeric DNA damage and tumor cell death. *Journal of medicinal chemistry* **2012**, 55, 9785-92.

139.   Uziel, O.; Yerushalmi, R.; Zuriano, L.; Naser, S.; Beery, E.; Nordenberg, J.; Lubin, I.; Adel, Y.; Shepshelovich, D.; Yavin, H.; Ben Aharon, I.; Pery, S.; Rizel, S.; Pasmanik-Chor, M.; Frumkin, D.; Lahav, M., BRCA1/2 mutations perturb telomere

122

biology: characterization of structural and functional abnormalities in vitro and in vivo. *Oncotarget* **2016**, 7, 2433-2454.

140.   Rosen, E. M., BRCA1 in the DNA damage response and at telomeres. *Front Genet* **2013**, 4, 85-85.

141.   Sekido, Y.; Ohigashi, S.; Takahashi, T.; Hayashi, N.; Suzuki, K.; Hirota, S., Familial Gastrointestinal Stromal Tumor with Germline KIT Mutations Accompanying Hereditary Breast and Ovarian Cancer Syndrome. *Anticancer Research* **2017**, 37, 1425-1431.

142.   Waisbren, J.; Uthe, R.; Siziopikou, K.; Kaklamani, V., BRCA 1/2 gene mutation and gastrointestinal stromal tumours: a potential association. *BMJ Case Rep* **2015**, 2015, bcr2014208830.

143.   Grushko, T. A.; Dignam, J. J.; Das, S.; Blackwood, A. M.; Perou, C. M.; Ridderstråle, K. K.; Anderson, K. N.; Wei, M.-J.; Adams, A. J.; Hagos, F. G.; Sveen, L.; Lynch, H. T.; Weber, B. L.; Olopade, O. I., MYC Is Amplified in BRCA1-Associated Breast Cancers. *Clinical Cancer Research* **2004**, 10, 499.

144.   Chen, Y.; Xu, J.; Borowicz, S.; Collins, C.; Huo, D.; Olopade, O. I., c-Myc activates BRCA1 gene expression through distal promoter elements in breast cancer cells. *BMC Cancer* **2011**, 11, 246-246.

145.   Zhang, X.; Xu, C. X.; Di Felice, R.; Sponer, J.; Islam, B.; Stadlbauer, P.; Ding, Y.; Mao, L.; Mao, Z. W.; Qin, P. Z., Conformations of Human Telomeric G-Quadruplex Studied Using a Nucleotide-Independent Nitroxide Label. *Biochemistry* **2016**, 55, 360-72.

146.   Collie, G. W.; Campbell, N. H.; Neidle, S., Loop flexibility in human telomeric quadruplex small-molecule complexes. *Nucleic acids research* **2015**, 43, 4785-99.
147.   Tawani, A.; Mishra, S. K.; Kumar, A., Structural insight for the recognition of G-quadruplex structure at human c-myc promoter sequence by flavonoid Quercetin. *Scientific reports* **2017**, 7, 3600.

148.   Machireddy, B.; Sullivan, H.-J.; Wu, C., Binding of BRACO19 to a Telomeric G-Quadruplex DNA Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent. *Molecules* **2019**, 24, 1010.

149.   Hou, J. Q.; Chen, S. B.; Tan, J. H.; Luo, H. B.; Li, D.; Gu, L. Q.; Huang, Z. S., New insights from molecular dynamic simulation studies of the multiple binding modes of a ligand with G-quadruplex DNA. *J Comput Aided Mol Des* **2012**, 26, 1355-68.
150.   Xuan-Yu M., H.-X. Z., Mihaly M., and Meng C., Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Current Computer Aided Drug Design* **2011**, 7, 12.

151.   Machireddy, B.; Kalra, G.; Jonnalagadda, S.; Ramanujachary, K.; Wu, C., 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* **2017**, 57, 2846-2864.

152.   Di Leva, F. S.; Novellino, E.; Cavalli, A.; Parrinello, M.; Limongelli, V., Mechanistic insight into ligand binding to G-quadruplex DNA. *Nucleic acids research* **2014**, 42, 5447-55.

153.   Federica Moraca, J. A., Francesco Ortuso, Anna Artese, Bruno Pagano, Ettore Novellino,; Stefano Alcaro, M. P. a. V. L., <PNAS-2017-Moraca-E2136-45(2).pdf>.

154.   Fadrna, E.; Spackova, N.; Stefl, R.; Koca, J.; Cheatham, T. E., 3rd; Sponer, J., Molecular dynamics simulations of Guanine quadruplex loops: advances and force field limitations. *Biophysical journal* **2004**, 87, 227-42.

155.   Hong Z., S. X., Haojun L., Structural Dynamics of Human Telomeric G-Quadruplex Loops Studied by Molecular Dynamics Simulations. *PloS one* **2013**, 8, e71380.

156.   Rebic, M.; Laaksonen, A.; Sponer, J.; Ulicny, J.; Mocci, F., Molecular Dynamics Simulation Study of Parallel Telomeric DNA Quadruplexes at Different Ionic Strengths: Evaluation of Water and Ion Models. *The journal of physical chemistry. B* **2016**, 120, 7380-91.

157.   Deng, N.; Wickstrom, L.; Cieplak, P.; Lin, C.; Yang, D., 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* **2017**, 121, 10484-10497.

158.   Deng, N.; Xia, J.; Wickstrom, L.; Lin, C.; Wang, K.; He, P.; Yin, Y.; Yang, D., 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* **2019**, 24, 1574.

159.   Luo, D.; Mu, Y., All-Atomic Simulations on Human Telomeric G-Quadruplex DNA Binding with Thioflavin T. *The Journal of Physical Chemistry B* **2015**, 119, 4955-4967.

160.   Bhat, J.; Mondal, S.; Sengupta, P.; Chatterjee, S., In Silico Screening and Binding Characterization of Small Molecules toward a G-Quadruplex Structure Formed in the Promoter Region of c-MYC Oncogene. *ACS Omega* **2017**, 2, 4382-4397.

161.   Salsbury, A. M.; Lemkul, J. A., Molecular Dynamics Simulations of the c-kit1 Promoter G-Quadruplex: Importance of Electronic Polarization on Stability and Cooperative Ion Binding. *The Journal of Physical Chemistry B* **2019**, 123, 148-159.

162.   Salsbury, A. M.; Lemkul, J. A., Influence of Monovalent Cations on the Dynamics of the C-KIT1 Promoter G-Quadruplex using Polarizable Molecular Dynamics Simulations. *Biophysical journal* **2020**, 118, 220a.

163.   Dean, T.; Salsbury, A. M.; Lemkul, J. A., Dynamics of the 1:2:1 and 1:6:1 C-Myc G-Quadruplexes with the Drude Polarizable Force Field. *Biophysical journal* **2020**, 118, 65a.

164.   Pany, S. P. P.; Bommisetti, P.; Diveshkumar, K. V.; Pradeepkumar, P. I., Benzothiazole hydrazones of furylbenzamides preferentially stabilize c-MYC and c-KIT1 promoter G-quadruplex DNAs. *Organic & Biomolecular Chemistry* **2016**, 14, 5779-5793.

165.   Read M., W. A., Harrison J., Gowan S., Kelland L., Dosanjh H., and Neidle S., Molecular Modeling Studies on G-Quadruplex Complexes of Telomerase Inhibitors: Structure−Activity Relationships. *Journal of medicinal chemistry* **1999**, 42, 4538-4546.

166.   Mulholland, K.; Wu, C., 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* **2016**, 56, 2093-2102.

167.   Wang, Z.; Chen, R.; Hou, L.; Li, J.; Liu, J.-P., Molecular dynamics and principal components of potassium binding with human telomeric intra-molecular G-quadruplex. *Protein & Cell* **2015**, 6, 423-433.

168.   Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E. r.; Laughton, C. A.; Orozco, M., Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophysical journal* **2007**, 92, 3817–3829.

169.   Krepl, M.; Zgarbova, M.; Stadlbauer, P.; Otyepka, M.; Banas, P.; Koca, J.; Cheatham, T. E.; Jurecka, P.; Sponer, J., 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* **2012**, 8, 2506-2520.

170.   Zgarbova, M.; Luque, F. J.; Sponer, J.; Cheatham, T. E.; Otyepka, M.; Jurecka, P., Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *Journal of Chemical Theory and Computation* **2013**, 9, 2339-2354.

171.   Zgarbova, M.; Sponer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurecka, P., 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* **2015**, 11, 5723-5736.

172.   Joung, I. S.; Cheatham, T. E., Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *Journal of Physical Chemistry B* **2008**, 112, 9020-9041.

173.   Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *Journal of Physical Chemistry* **1993**, 97, 10269-10280.

174.   Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T. E. I.; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J., A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA *Nucleic acids research* **2010**, 38, 299-313.

175.   Cosconati, S.; Marinelli, L.; Trotta, R.; Virno, A.; De Tito, S.; Romagnoli, R.; Pagano, B.; Limongelli, V.; Giancola, C.; Baraldi, P. G.; Mayol, L.; Novellino, E.; Randazzo, A., Structural and Conformational Requisites in DNA Quadruplex Groove Binding: Another Piece to the Puzzle. *Journal of the American Chemical Society* **2010**, 132, 6425-6433.

176.   Fadrna, E.; Spackova, N. a.; Sarzynska, J.; Koca, J.; Orozco, M.; Cheatham, T. E., III;; Kulinski, T.; Sponer, J., Single Stranded Loops of Quadruplex DNA As Key Benchmark for Testing Nucleic Acids Force Fields. *Journal of Chemical Theory and Computation* **2009**, 5, 2514-2530

177.   Mukherjee, A.; Lavery, R.; Bagchi, B.; Hynes, J. T., On the Molecular Mechanism of Drug Intercalation into DNA: A Simulation Study of the Intercalation Pathway, Free Energy, and DNA Structural Changes. *Journal of the American Chemical Society* **2008**, 130, 9747-9755.

178.   Lei, H.; Wang, X.; Wu, C., Early stage intercalation of doxorubicin to DNA fragments observed in molecular dynamics binding simulations. *Journal of molecular graphics & modelling* **2012**, 38, 279-89.

179.   Mulholland, K.; Wu, C., 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* **2016**, 56, 2093-2102.

180.   Ryckaert, J.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *Journal of Computational Physics* **1977**, 23, 327-341.

181.   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T. A.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *Journal of Computational Physics* **1995**, 103, 8577-8593.

182.   Procacci, P.; Berne, B. J., Multiple Time-Scale Methods For Constant-Pressure Molecular-Dynamics Simulations Of Molecular-Systems. *Molecular Physics* **1994**, 83, 255-272.

183.   Pande, V. S.; Beauchamp, K.; Bowman, G. R., Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **2010**, 52, 99-105.

184.   Mulholland, K.; Sullivan, H.-J.; Garner, J.; Cai, J.; Chen, B.; Wu, C., Three-Dimensional Structure of RNA Monomeric G-Quadruplex Containing ALS and FTD Related G4C2 Repeat and Its Binding with TMPyP4 Probed by Homology Modeling based on Experimental Constraints and Molecular Dynamics Simulations. *ACS Chem Neurosci* **2020**, 11, 57-75.

185.   Jiang, H.; Sheong, F. K.; Zhu, L.; Gao, X.; Bernauer, J.; Huang, X., 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* **2015**, 11, e1004404.

186.   Sullivan, H.-J.; Readmond, C.; Radicella, C.; Persad, V.; Fasano, T. J.; Wu, C., 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* **2018**, 3, 14788-14806.

187.   Reshetnikov, R. V.; Kopylov, A. M.; Golovin, A. V., Classification of G-Quadruplex DNA on the Basis of the Quadruplex Twist Angle and Planarity of G-Quartets. *Acta Naturae* **2010**, 2, 72-81.

188.   Biffi, G.; Tannahill, D.; Miller, J.; Howat, W. J.; Balasubramanian, S., Elevated Levels of G-Quadruplex Formation in Human Stomach and Liver Cancer Tissues. *PloS one* **2014**, 9, e102711.

189.   Neidle, S., Quadruplex Nucleic Acids as Novel Therapeutic Targets. *Journal of Medicinal Chemistry* **2016**, 59, 5987-6011.

190.   Chung, W. J.; Heddi, B.; Hamon, F.; Teulade-Fichou, M. P.; Phan, A. T., Solution Structure of a G-quadruplex Bound to the Bisquinolinium Compound Phen-DC3. *Angewandte Chemie-International Edition* **2014**, 53, 999-1002.

191.   Mulholland, K.; Siddiquei, F.; Wu, C., 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* **2017**, 19, 18685-18694.

127

192.   Shan, Y. B.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E., How Does a Drug Molecule Find Its Target Binding Site? *Journal of the American Chemical Society* **2011**, 133, 9181-9183.

193.   Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K., Conformational analysis of nucleic acids revisited: Curves. *Nucleic Acids Research* **2009**, 37, 5917-5929.

194.   Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling* **1996**, 14, 33-38.

195.   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.; Pande, Vijay S., MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **2015**, 109, 1528-1532.

196.   Pedregosa, F.; Ga; #235; Varoquaux, l.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; #201; Duchesnay, d., Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, 12, 2825-2830.

197.   Rousseeuw, P. J., Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **1987**, 20, 53-65.

198.   Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F., Markov models of molecular kinetics: Generation and validation. *The Journal of Chemical Physics* **2011**, 134, 174105.

199.   Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C., Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *The Journal of Chemical Physics* **2007**, 126, 155102.

200.   Csardi, G.; Nepusz, T., The igraph software package for complex network research. *InterJournal, Complex Systems* **2006**, 1695, 1--9.

201.   Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. I., Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Model. *Accounts of Chemical Research* **2000**, 33, 889-897.

202.   Kollman, P.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.; Cheatham, T., Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Accounts of Chemical Research* **2000**, 33, 889-897.

128

203.   Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the Performance of the Molecular Mechanics/Poisson Boltzmann Surface Area and Molecular Mechanics/Generalized Born Surface Area Methods. II. The Accuracy of Ranking Poses Generated from Docking. *Journal of Computational Chemistry* **2010**, 32, 866-877.

204.   Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* **2011**, 51, 69-82.

205.   Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T., Assessing the Performance of MM/PBSA and MM/GBSA Methods. 3. The Impact of Force Fields and Ligand Charge Models. *Journal of Physical Chemistry B* **2013**, 117, 8408-8421.

206.   Sun, H.; Li, Y.; Tian, S.; Xub, L.; Hou, T., Assessing the Performance of MM/PBSA and MM/GBSA Methods. 4. Accuracies of MM/PBSA and MM/GBSA Methodologies Evaluated by Various Simulation Protocols Using PDBbind Data Set. *Physical Chemistry and Chemical Physics* **2014**, 16, 16719-16729.

207.   Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **2004**, 47, 1739-1749.

208.   Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W., Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *Journal of Computer Aided Molecular Design* **2013**, 27, 221-234.

209.   Ioakimidis, L.; Thoukydidis, L.; Mirza, A.; Naeem, S.; Reynisson, J., Benchmarking the Reliability of QikProp. Correlation Between Experimental and Predicted Values. *QSAR Combinatorial Sciences* **2008**, 27, 445-456.

210.   Ho, P.; Carter, M., *DNA Structure: Alphabet Soup for the Cellular Soul*. 2011.

129

Appendix A

Binding of BRACO19 to a Telomeric G-Quadruplex DNA Probed by All-Atom

Molecular Dynamics Simulations with Explicit Solvent

DNA Duplex + Unbound BRACO19	3(-p-p-p) DNA G-Quadruplex + Unbound BRACO19
3(-lwd+ln) DNA Quadruplex + Unbound BRACO19	3(-p-lw-ln) DNA Quadruplex + Unbound BRACO19
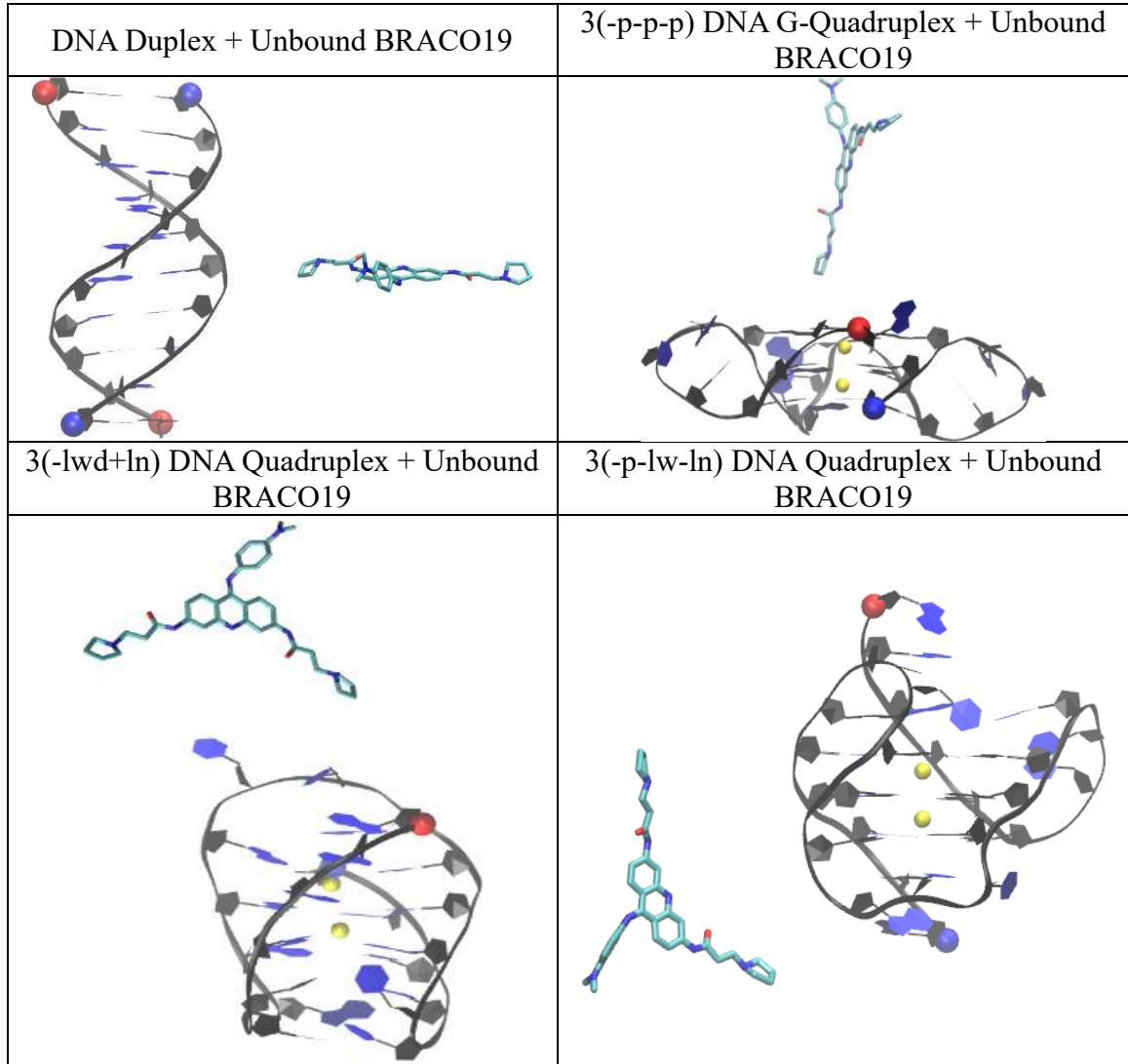
Figure A1. Initial structures of the simulation systems. 5' and 3' are indicated by a red and blue ball, respectively.
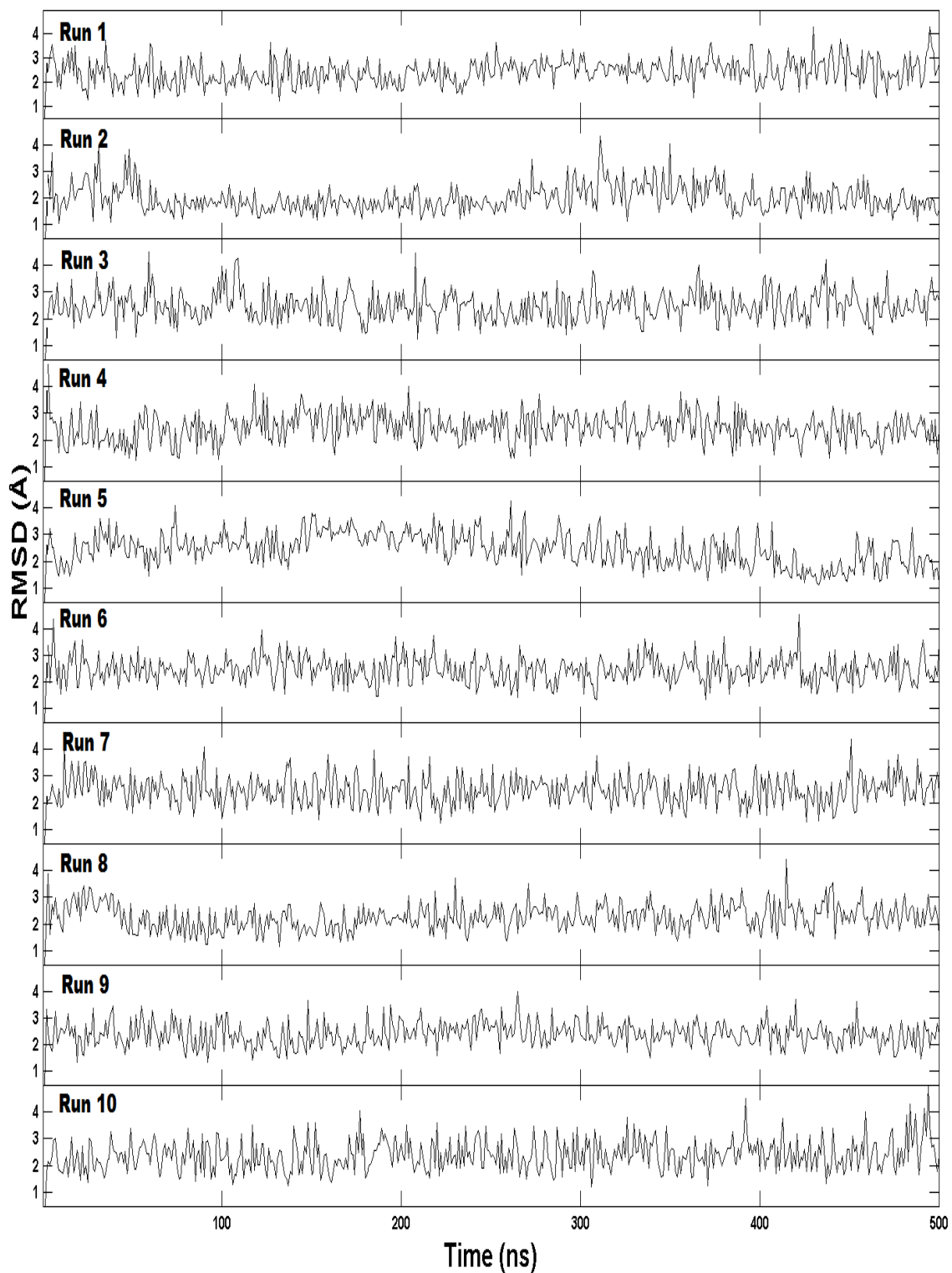
Figure A2. RMSD of the BRACO19-duplex DNA complex in each trajectory.

131
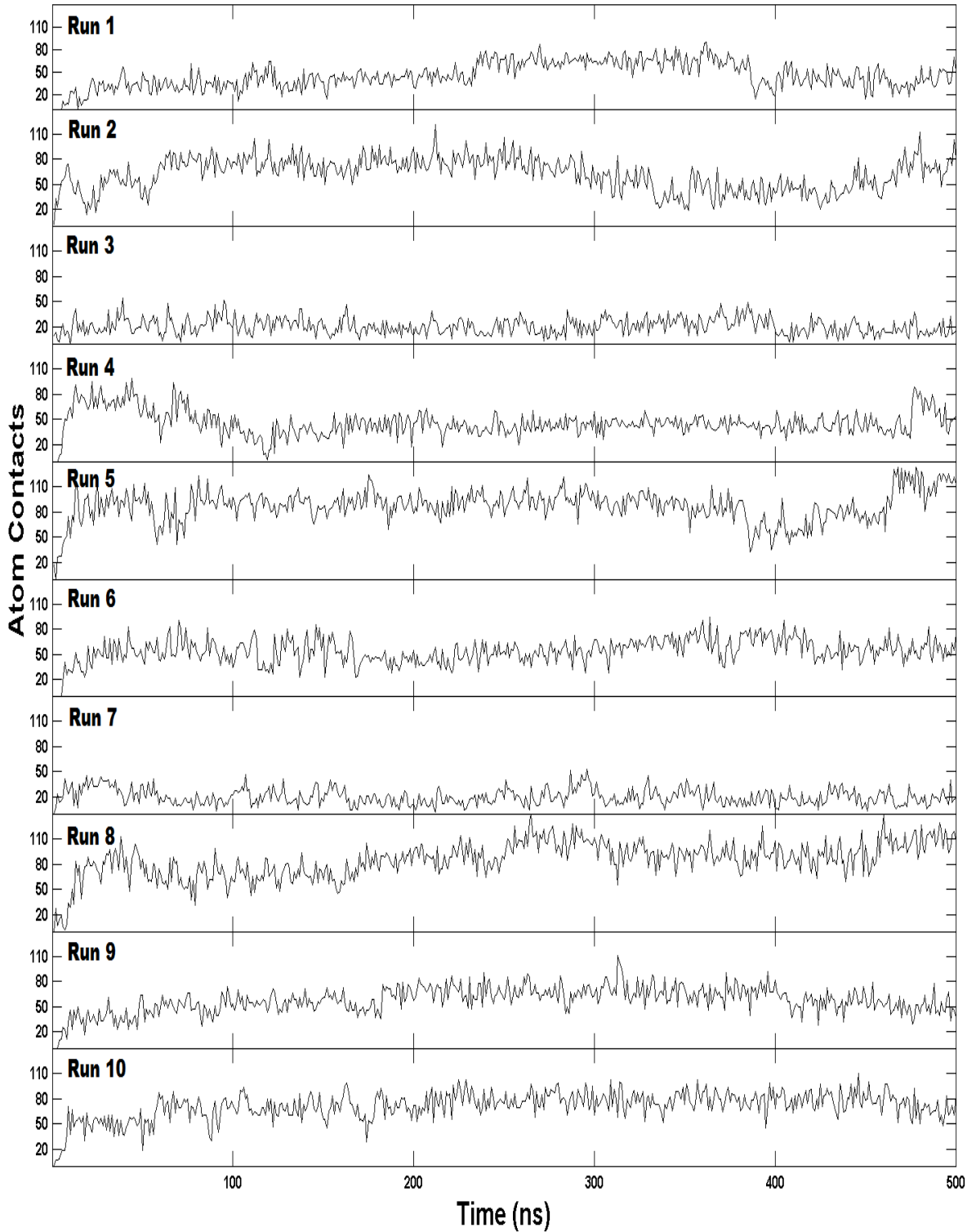
Figure A3. The contact number between BRACO19 and the duplex DNA in each trajectory.
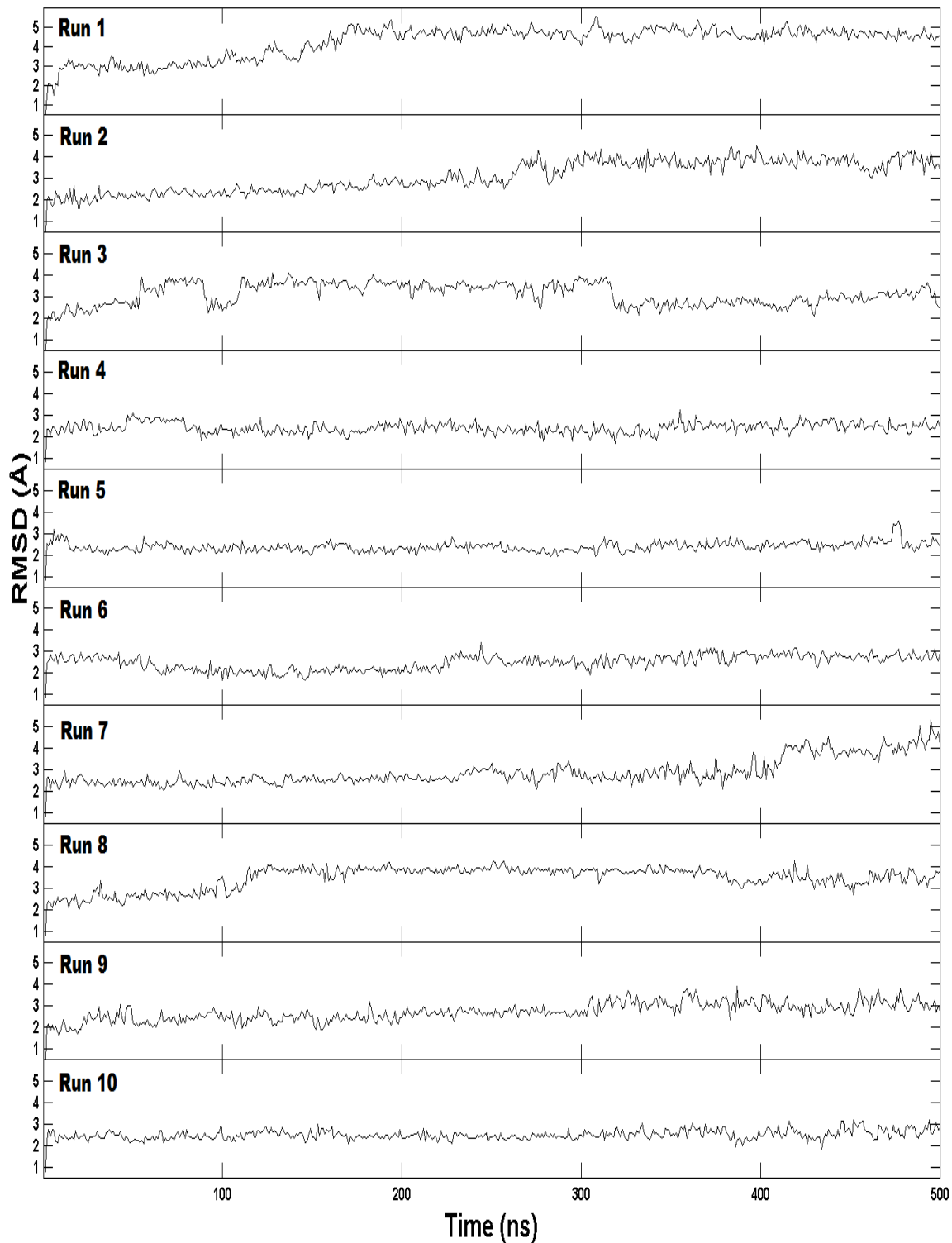
Figure A4. RMSD of the parallel topological fold of DNA in complex with BRACO19 in each trajectory.
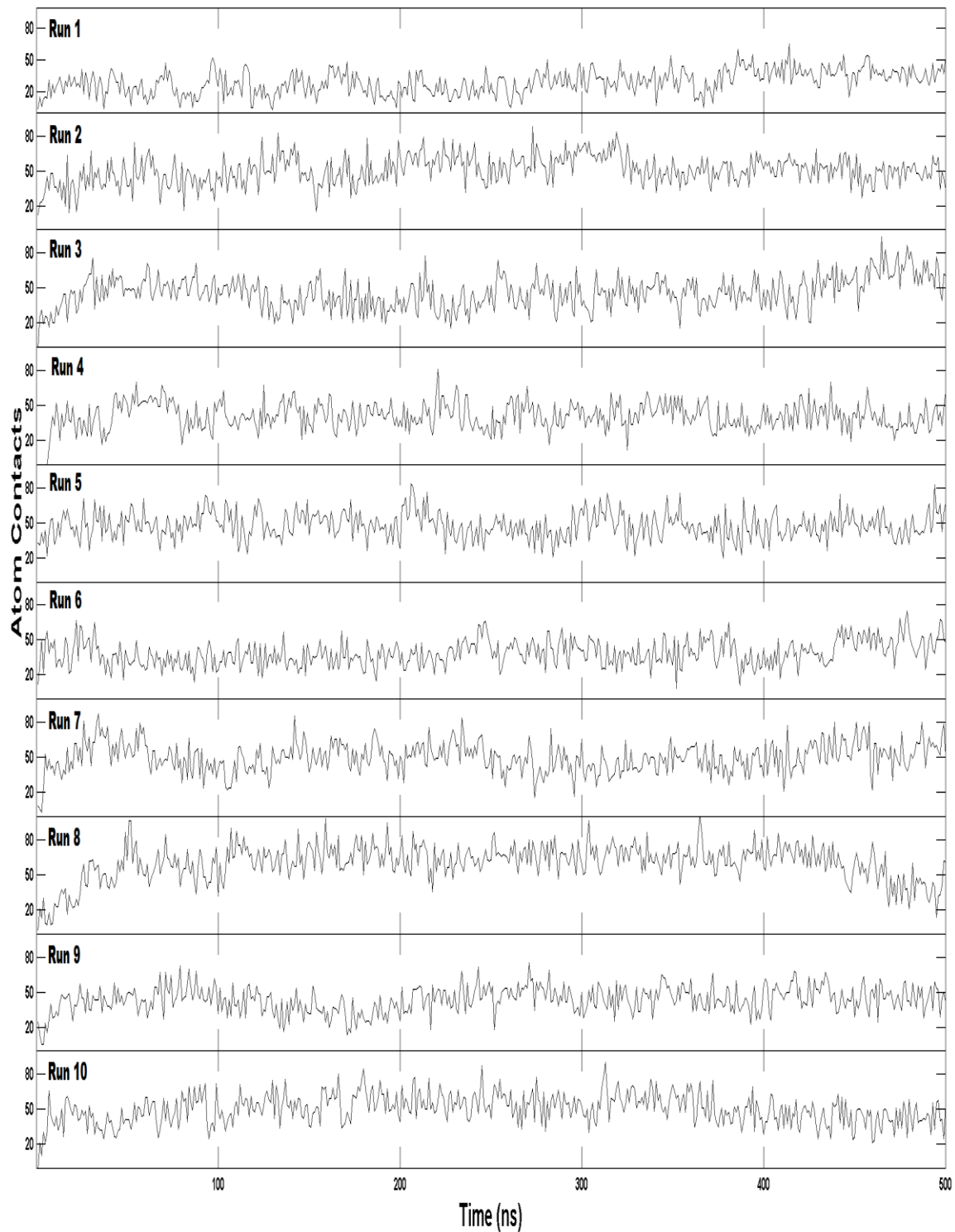
Figure A5. The contact number between the parallel topological fold of DNA and BRACO19 in each trajectory.
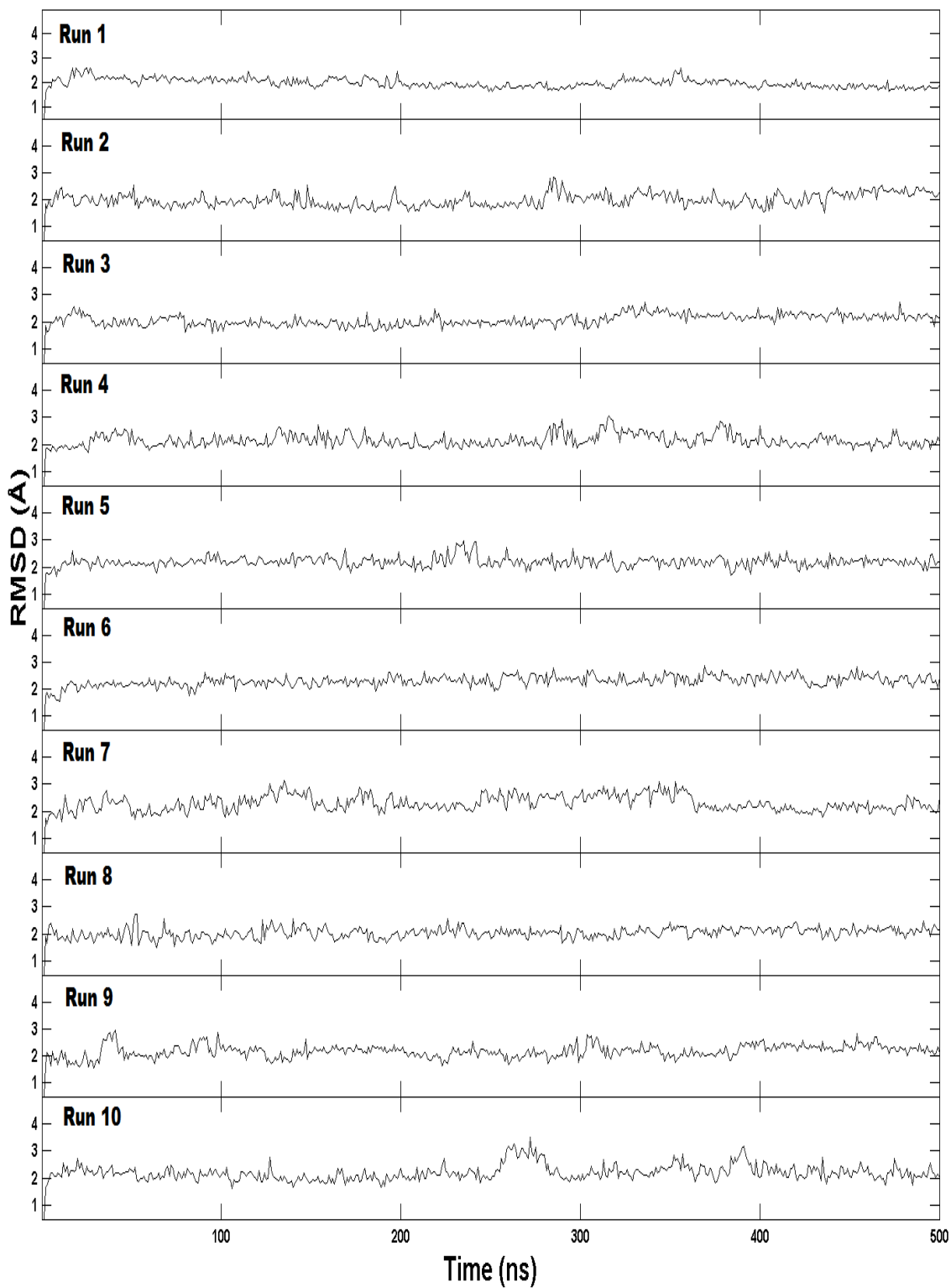
Figure A6. RMSD of the anti-parallel topological fold of DNA in complex with BRACO19 in each trajectory.

135
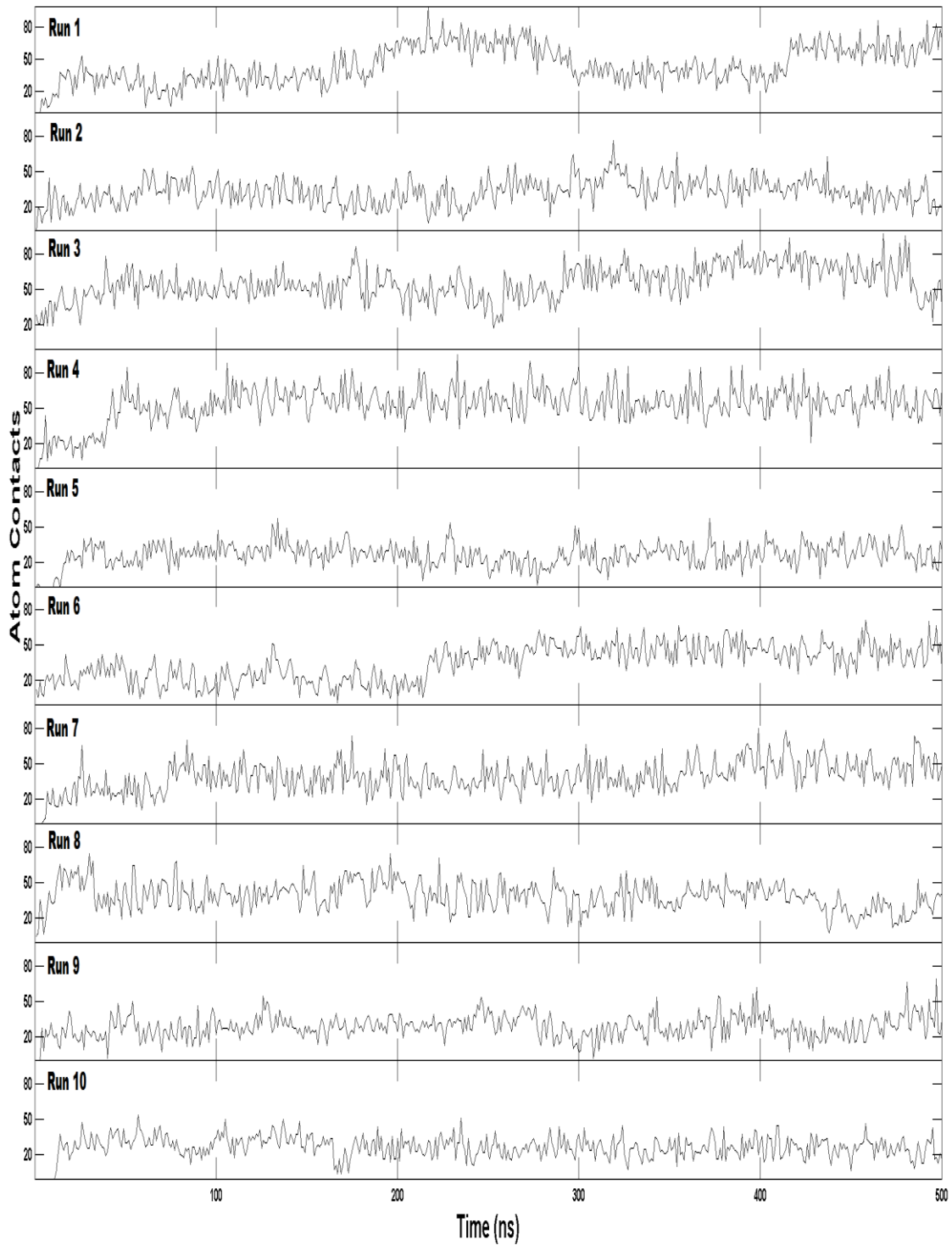
Figure A7. The contact number between the anti-parallel topological fold of DNA and BRACO19 in each trajectory.
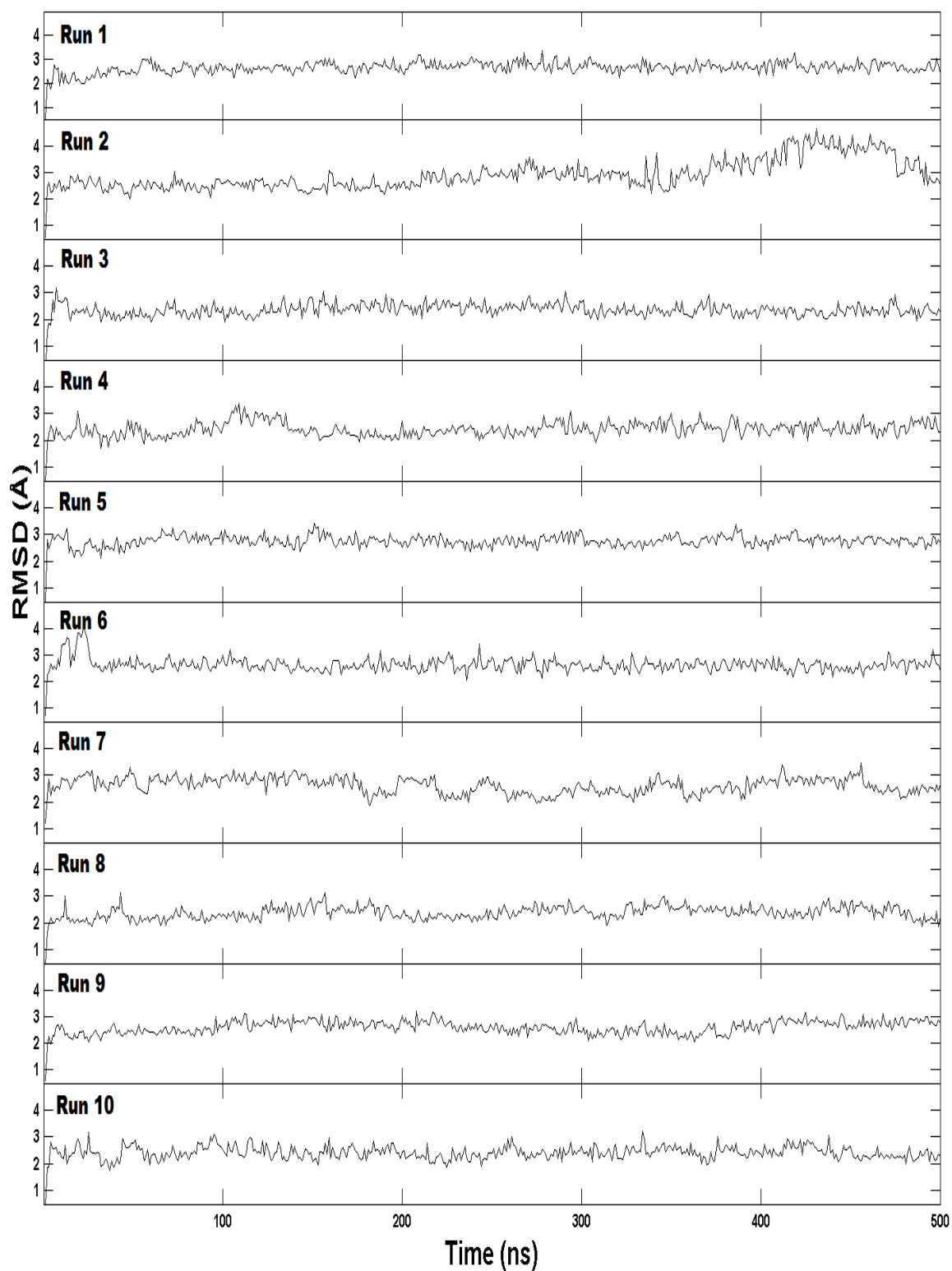
Figure A8. RMSD of the hybrid topological fold of DNA in complex with BRACO19 in each trajectory.

www.manaraa.com
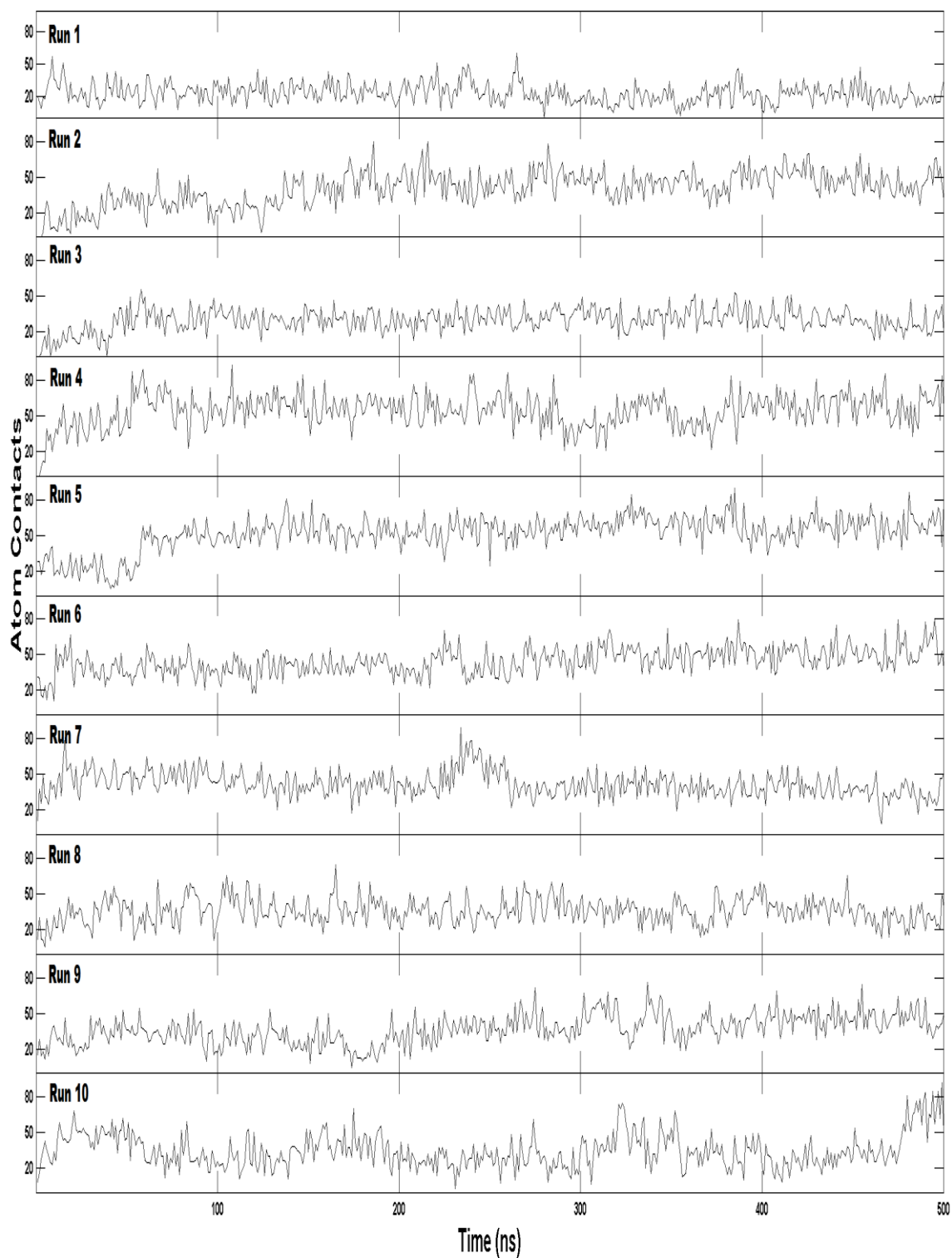
Figure A9. The contact number between the hybrid topological fold of DNA and BRACO19 in each trajectory.
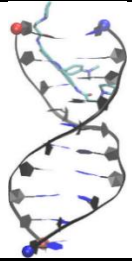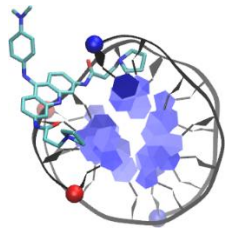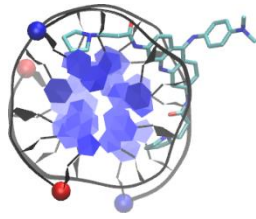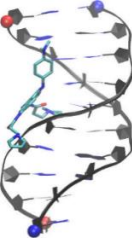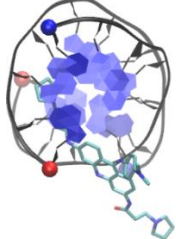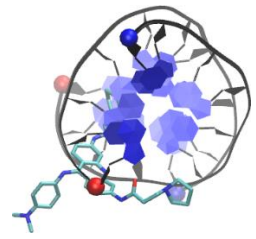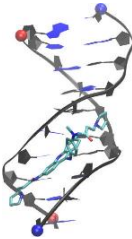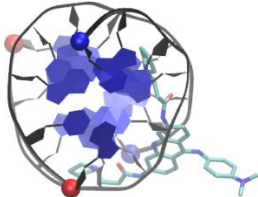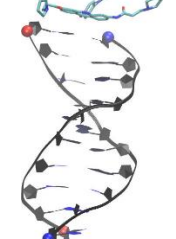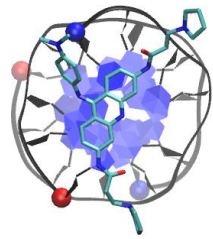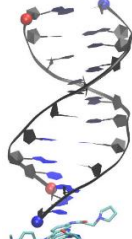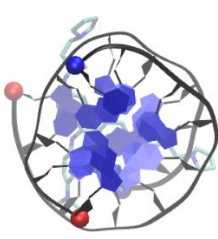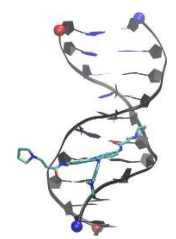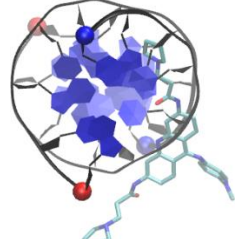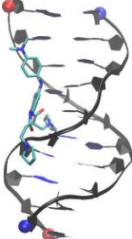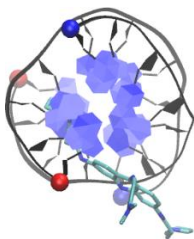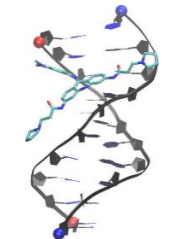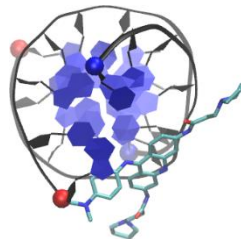
Front View	Description	Front View	Description
Run 01 (Groove)		Run 06 (Groove)	
Run 02 (Groove)		Run 07 (Groove)	
Run 03 (Groove)		Run 08 (Top)	
Run 04 (Bottom)		Run 09 (Groove)	
Run 05 (Groove)		Run 10 (Groove)	

Figure A10. Last snapshots of the 10 BRACO19-duplex DNA complex simulations. 5' and 3' arc indicated by a red and blue ball, respectively.

Front View	Description	Front View	Description
Run 01 (Top)		Run 06 (Top)	
Run 02 (Top)		Run 07 (Bottom)	
Run 03 (Bottom)		Run 08 (Top)	
Run 04 (Bottom)		Run 09 (Bottom)	
Run 05 (Groove)		Run 10 (Groove)	

Figure A11. Last snapshots of the 10 3(-p-p-p)-BRACO19 complex simulations. 5' and 3' arc indicated by a red and blue ball, respectively.
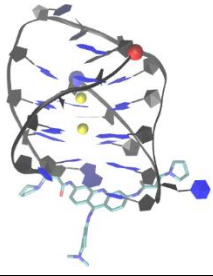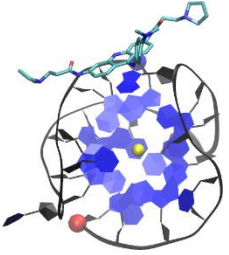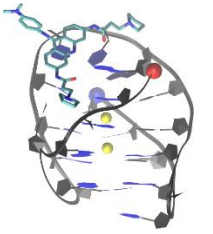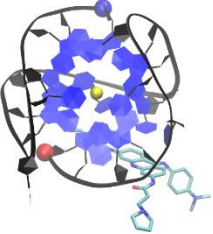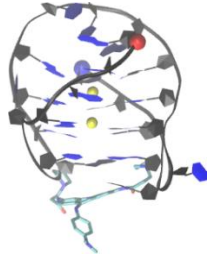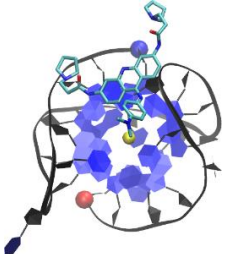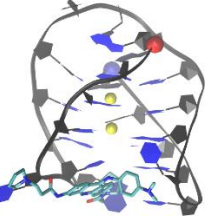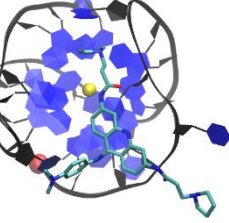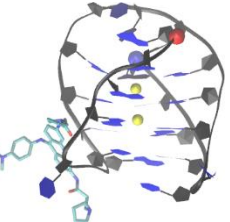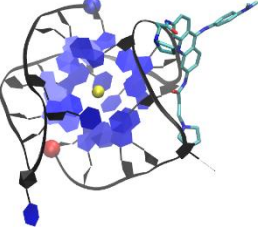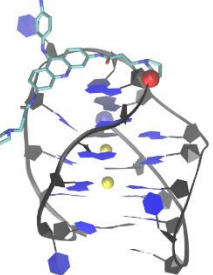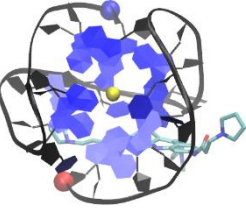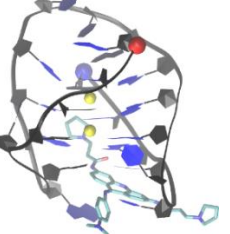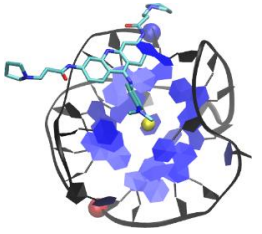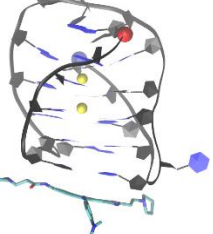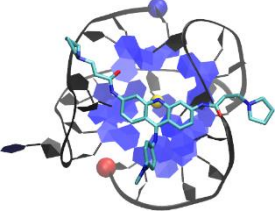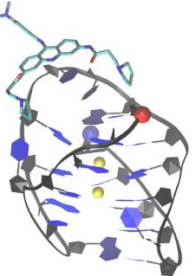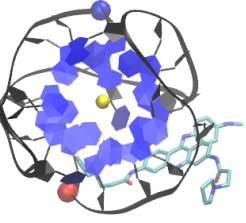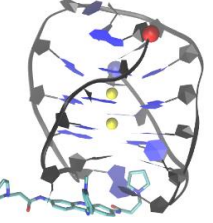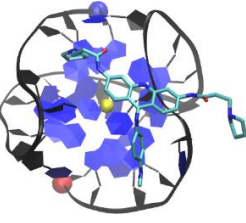
Front View	Description	Front View	Description
Run 01 (Bottom Groove)		Run 06 (Top groove)	
Run 02 (Bottom)		Run 07 (Bottom)	
Run 03 (Groove)		Run 08(Top groove)	
Run 04 (Bottom)		Run 09 (Bottom)	
Run 05 (Top groove)		Run 10 (Bottom)	

Figure A12. Last snapshots of the 10 3(-lwd+ln)-BRACO19 complex simulations. 5' and 3' arc indicated by a red and blue ball, respectively.

141
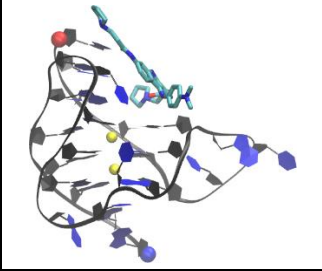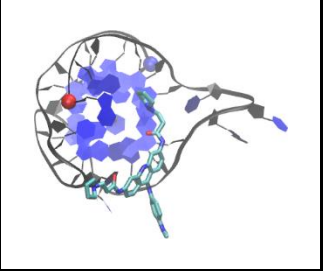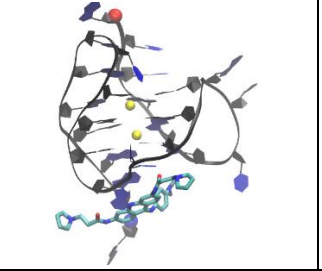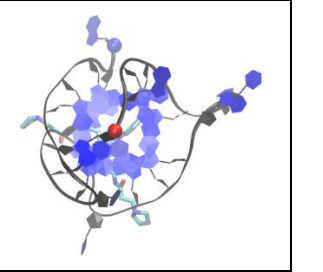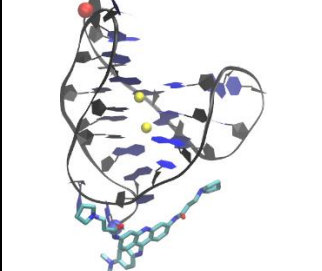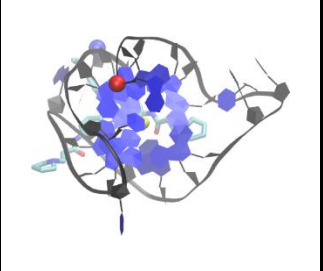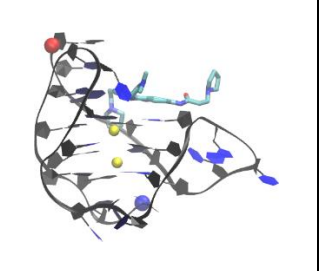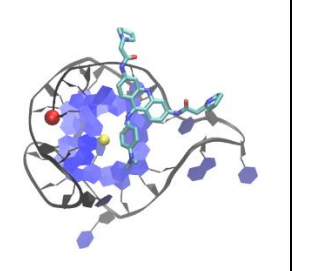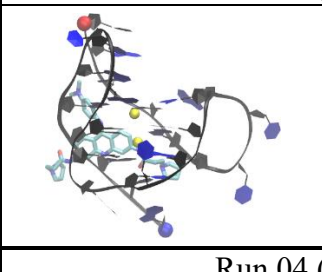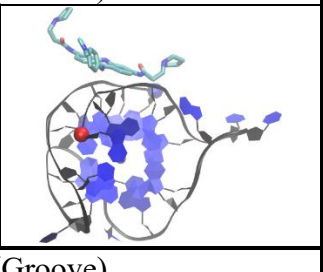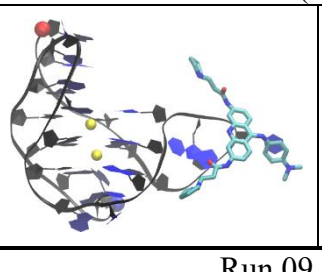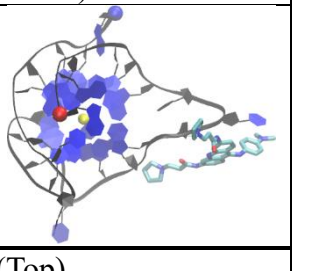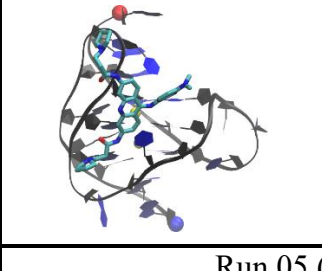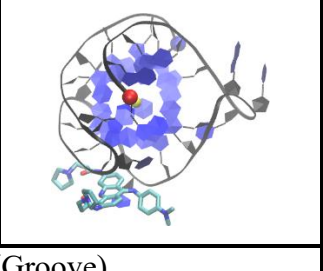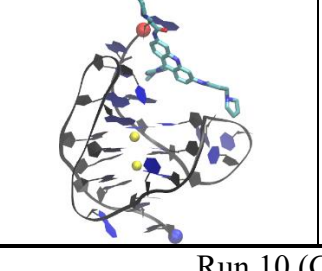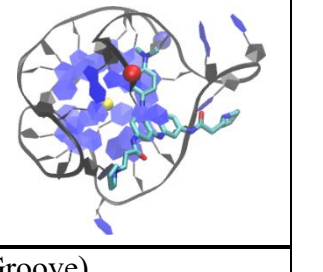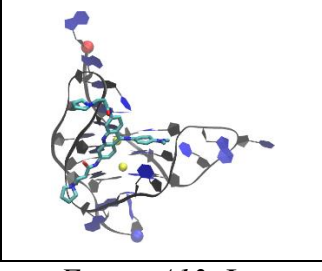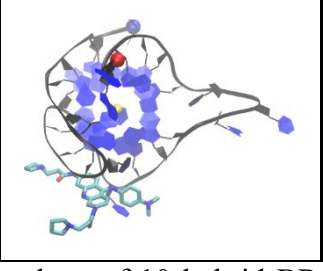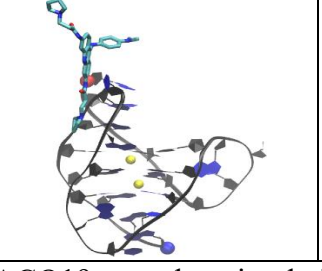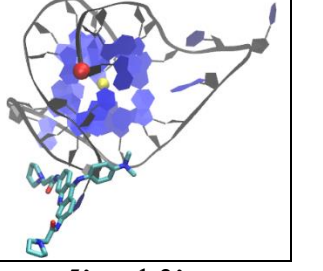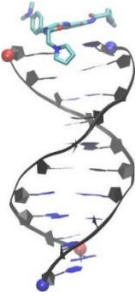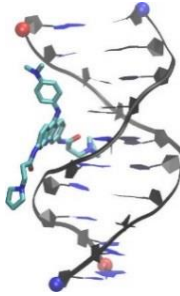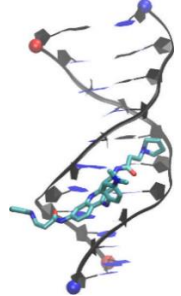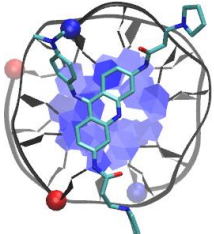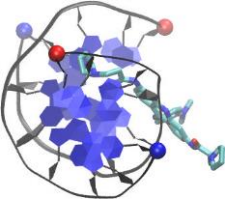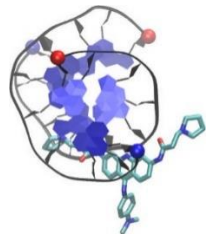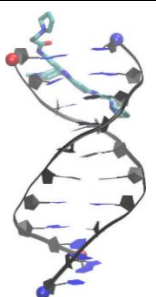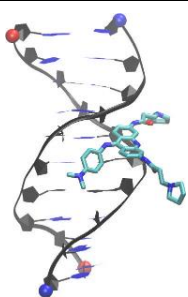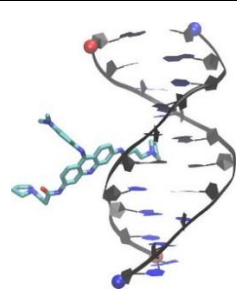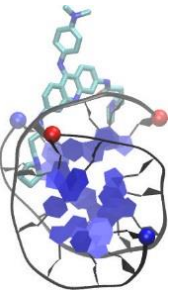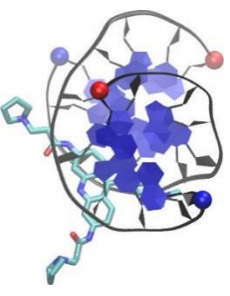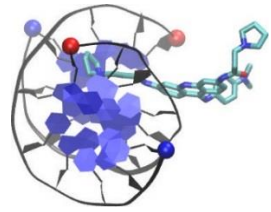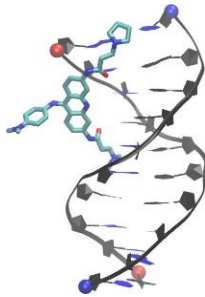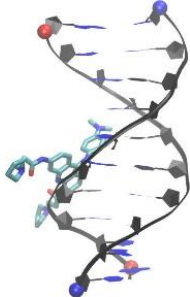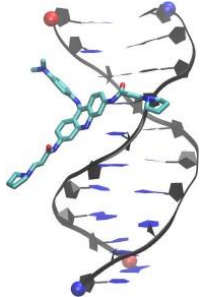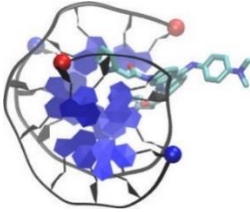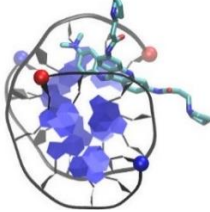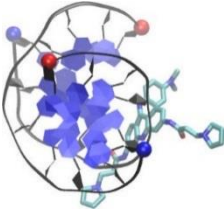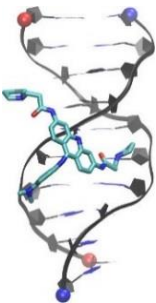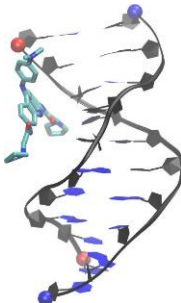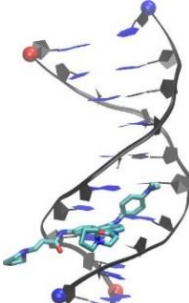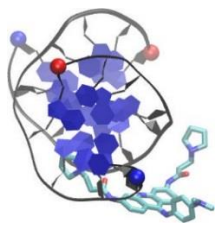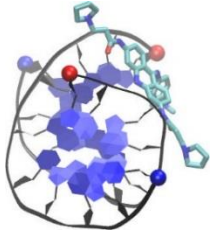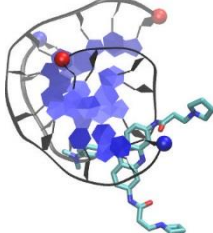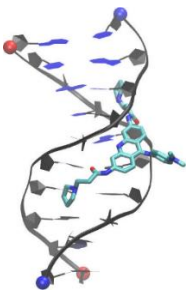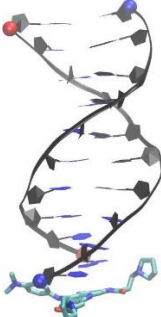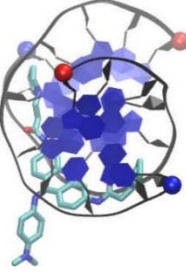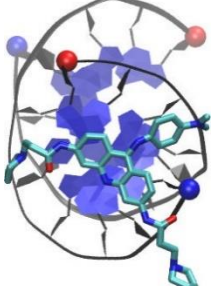
Front View	Description	Front View	Description
Run 01 (Top)		Run 06 (Bottom)	
Run 02 (Bottom)		Run 07 (Top)	
Run 03 (Groove)		Run 08 (Groove)	
Run 04 (Groove)		Run 09 (Top)	
Run 05 (Groove)		Run 10 (Groove)	
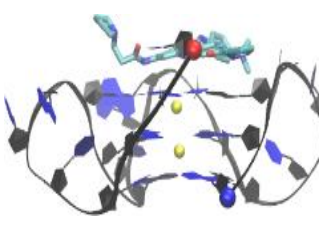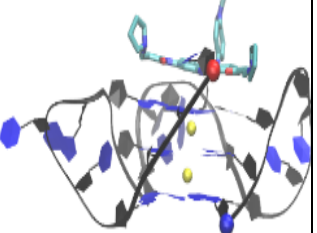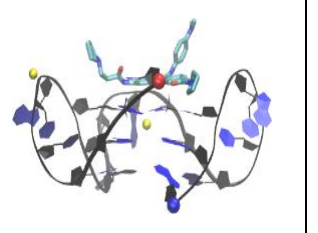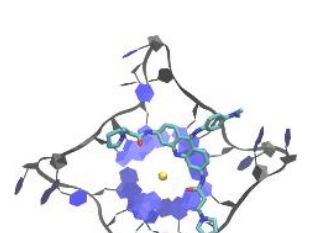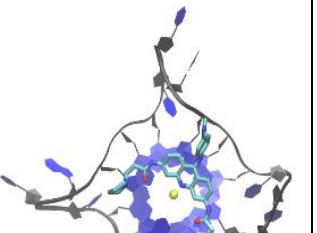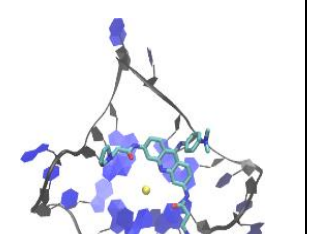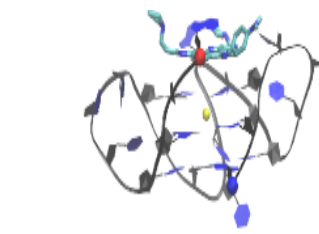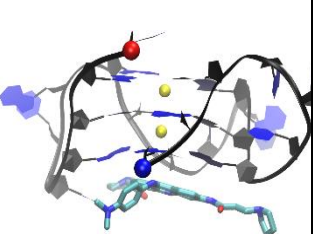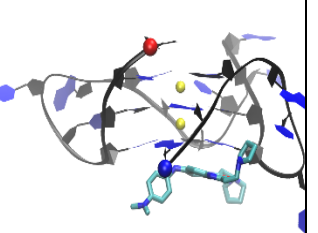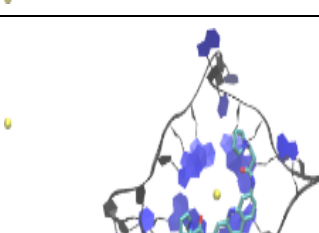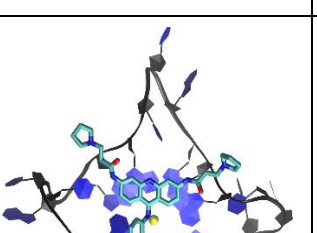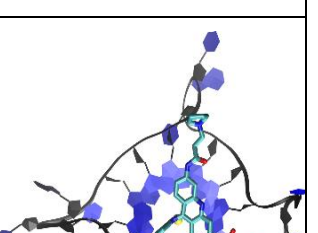
Figure A13. Last snapshots of 10 hybrid BRACO19 complex simulations. 5' and 3' arc indicated by a red and blue ball, respectively.

Binding model	Top Stacking	Groove Binding	Groove Binding
Cluster ID	A1	B1	B2
Representative Structure (Front View)			
Representative Structure (Top/Bottom View)			
Population	4%	23%	19%
Binding Model	Groove Binding	Groove Binding	Groove Binding
Cluster ID	B3	B4	B5
Representative Structure (Front View)			
Representative Structure (Top/Bottom View)			

143

Population	10%	6%	5%
Binding Model	Groove Binding	Groove Binding	Groove Binding
Cluster ID	B6	B7	B8
Representative Structure (Front View)			
Representative Structure (Top/Bottom View)			
Population	4%	3%	3%
Binding Model	Groove Binding	Groove Binding	Groove Binding
Cluster ID	B9	B10	B11
Representative Structure (Front View)			

144

Representative Structure (Top/Bottom View)			
Population	2%	2%	2%
Binding model	Groove Binding	Bottom Binding	
Cluster ID	B12	C1	
Representative Structure (Front View)			
Representative Structure (Top/Bottom View)			
Population	2%	2%	

*Figure A*14. Representative structures of the most populated complex structure families (population $\geqslant$ 1 %) from the clustering analysis of the combined binding trajectories. 5' and 3' are indicated by a red and blue ball, respectively for the DNA duplex-BRACO19 complex system.

Binding model	Top Stacking	Top Stacking	Top Stacking
Cluster ID	A1	A2	A3
Representative Structure (Front View)			
Representative Structure (Top/Bottom View)			
Population	16%	5%	4%
Binding Model	Top Stacking	Bottom Stacking	Bottom Stacking
Cluster ID	A4	B1	B2
Representative Structure (Front View)			
Representative Structure (Top/Bottom View)			

146
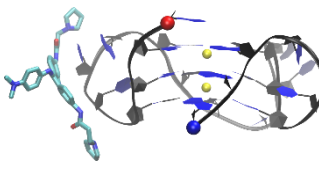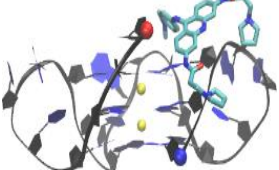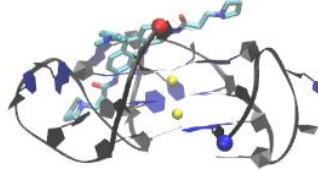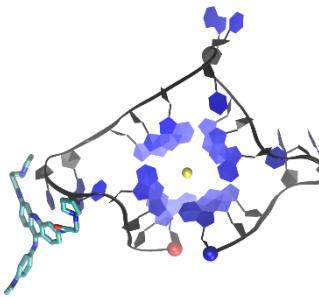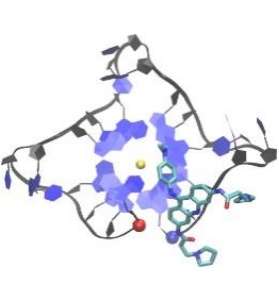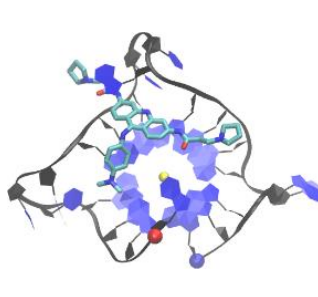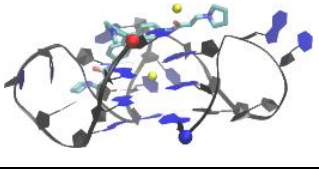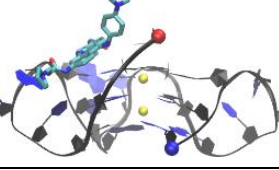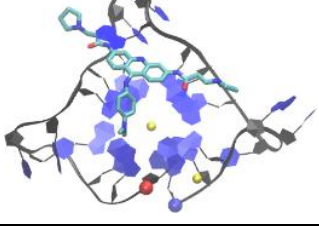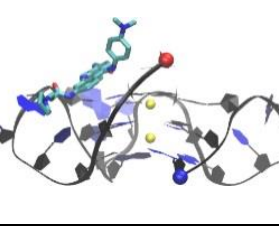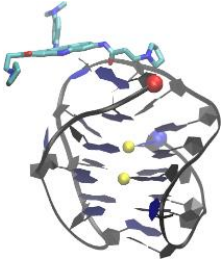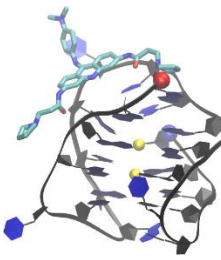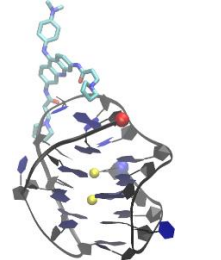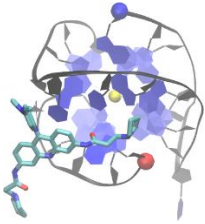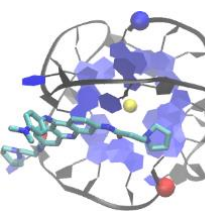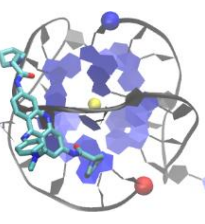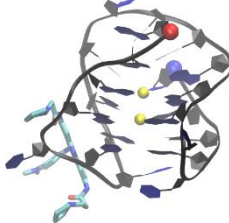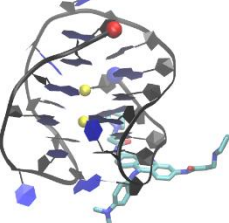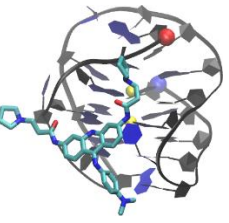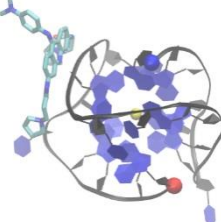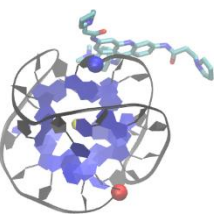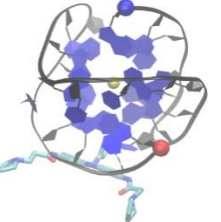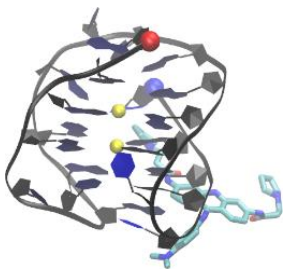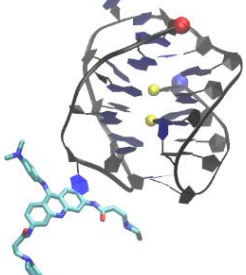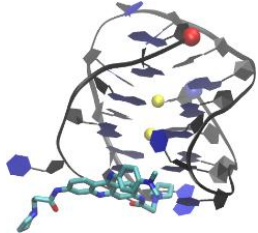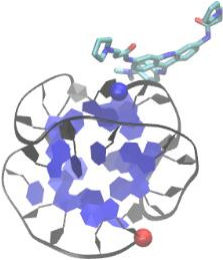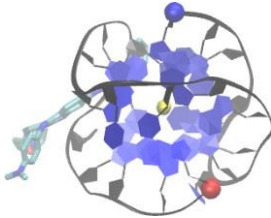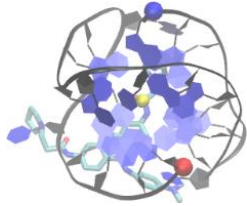
Population	3%	31%	10%
Binding Model	Groove Binding	Groove Binding	Groove Binding
Cluster ID	C1	C2	C3
Representative Structure (Front View)			
Representative Structure (Top/Bottom View)			
Population	9%	8%	7%
Binding Model	Groove Binding	Groove Binding	
Cluster ID	C4	B5	
Representative Structure (Front View)			
Representative Structure (Top/Bottom View)			
Population	3%	2%	

*Figure A*15. Representative structures of the most populated complex structure families (population $\geq$ 1 %) from the clustering analysis of the combined binding trajectories. 5'

147
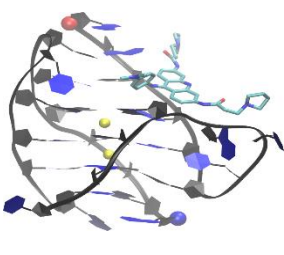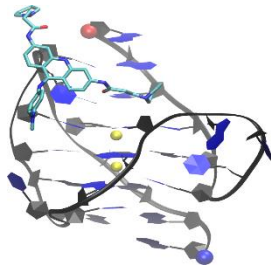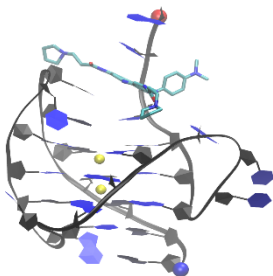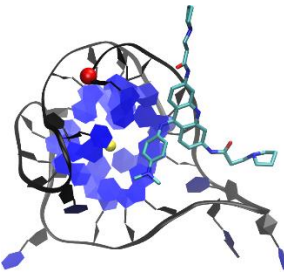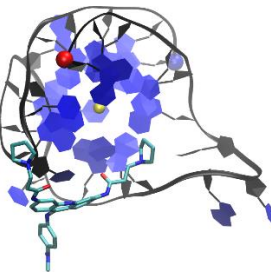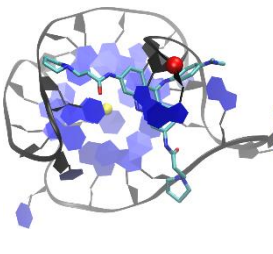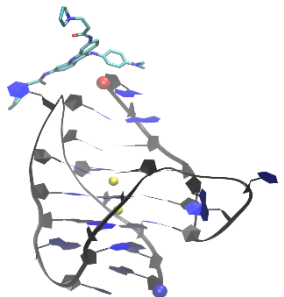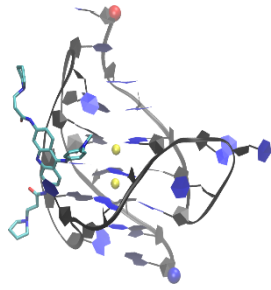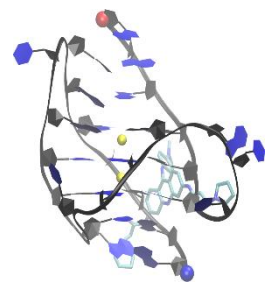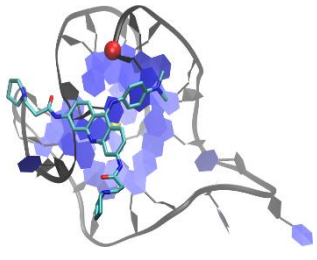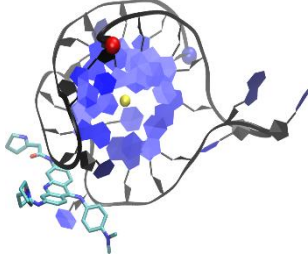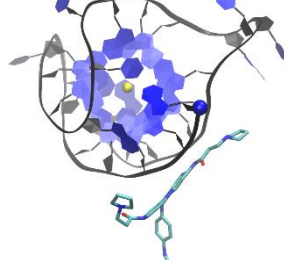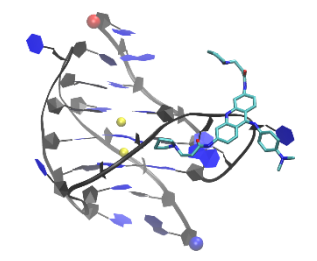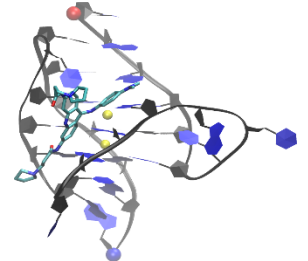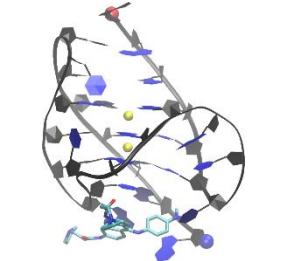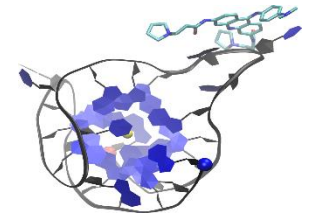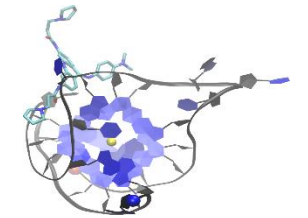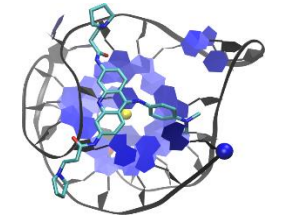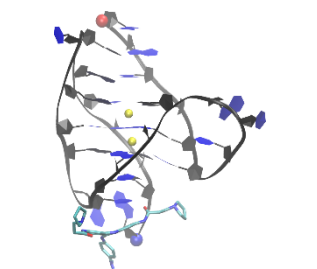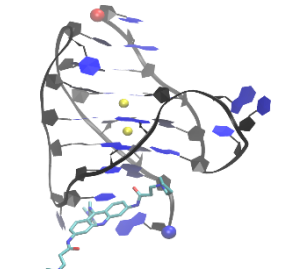
and 3' are indicated by a red and blue ball, respectively for the 3(-p-p-p)-BRACO19 complex system.
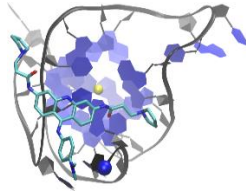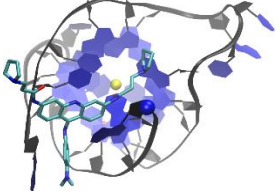
Binding model	Top Stacking	Top Stacking	Top Stacking
Cluster ID	A1	A2	A3
Representative Structure (Front View)			
Representative Structure (Top/Bottom View)			
Population	19%	4%	2%
Binding Model	Groove Binding	Groove Binding	Groove Binding
Cluster ID	B1	B2	B4
Representative Structure (Front View)			
Representative Structure (Top/Bottom View)			
Population	15%	10%	3%

149

Binding Model	Groove Binding	Groove Binding	Bottom Binding
Cluster ID	B5	B6	C1
Representative Structure (Front View)			
Representative Structure (Top/Bottom View)			
Population	1%	1%	45%

*Figure A*16. Representative structures of the most populated complex structure families (population $\geq$ 1 %) from the clustering analysis of the combined binding trajectories. 5' and 3' are indicated by a red and blue ball, respectively for the 3(-lwd+ln)-BRACO19 complex system.

Binding model	Top Stacking	Top Stacking	Top Stacking
Cluster ID	A1	A2	A3
Representative Structure (Front View)			
Representative Structure (Top/Bottom View)			
Population	12%	9%	7%
Binding Model	Top stacking	Groove Binding	Groove Binding
Cluster ID	A4	B1	B2
Representative Structure (Front View)			

151

Representative Structure (Top/Bottom View)			
Population	5%	22%	10%
Binding Model	Groove Binding	Groove Binding	Bottom Binding
Cluster ID	B3	B4	C1
Representative Structure (Front View)			
Representative Structure (Top/Bottom View)			
Population	9%	2%	10%
Binding Model	Bottom Binding	Bottom Binding	
Cluster ID	C2	C3	
Representative Structure (Front View)			

Representative Structure (Top/Bottom View)			
Population	9%	1%	

*Figure A*17. Representative structures of the most populated complex structure families (population ≥ 1 %) from the clustering analysis of the combined binding trajectories. 5' and 3' are indicated by a red and blue ball, respectively for the 3(-p-lw-ln)-BRACO19 complex system.

	2D interactions	2D interactions with expanded interacting residues
(A) Groove Binding Mode of Duplex		
(B) Top Stacking mode of parallel G-Quadruplex		

(C) Groove Binding Mode of 3(-p-p-p) G-Quadruplex

(D) Top Stacking Mode of anti-parallel G-Quadruplex

(E) Groove Binding Mode of anti-parallel G-Quadruplex		
(F) Bottom Stacking Mode of anti-parallel G-Quadruplex		

156

(G) Top Stacking mode of hybrid G-Quadruplex

(H) Groove Binding Mode of anti-parallel G-Quadruplex
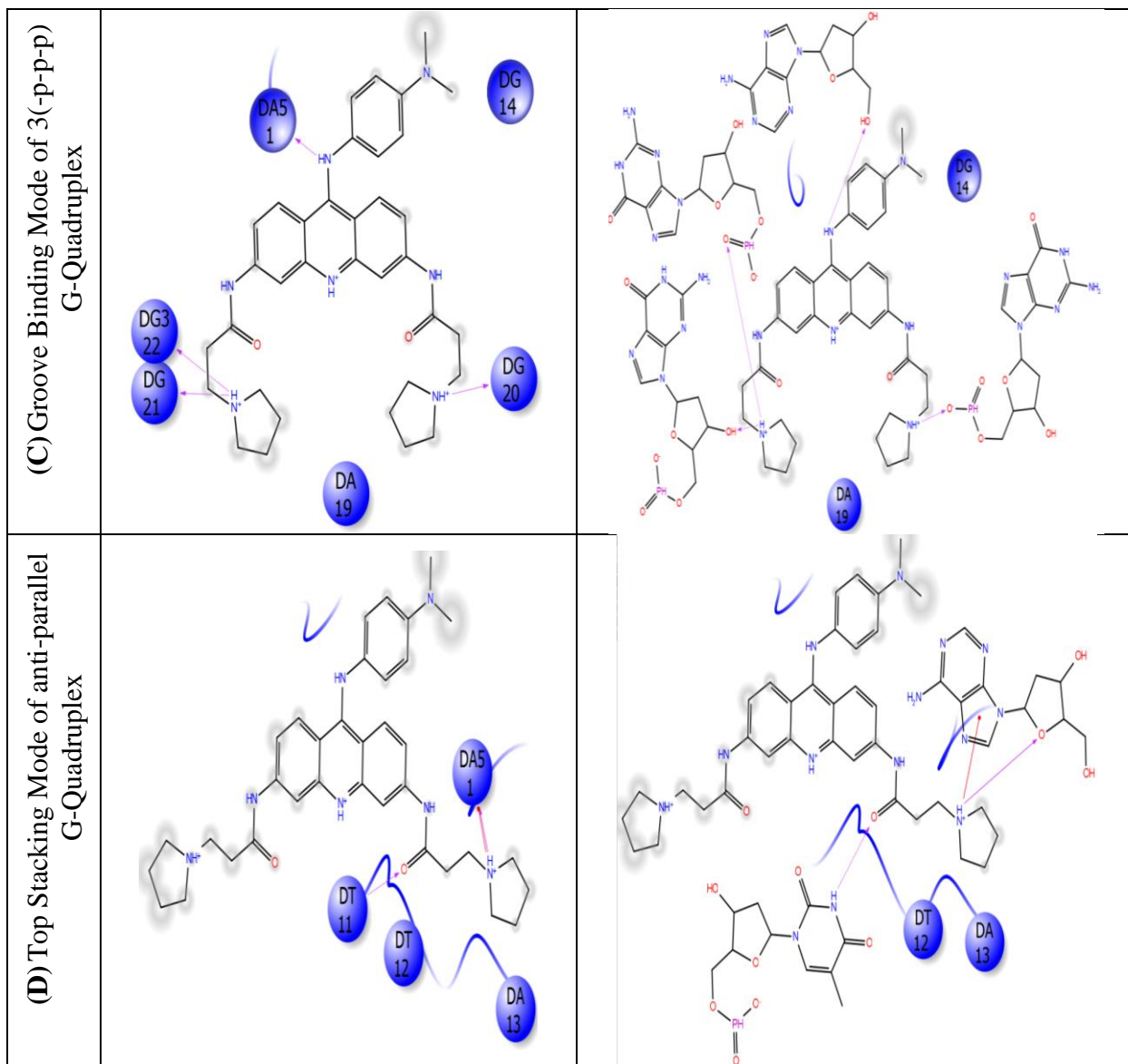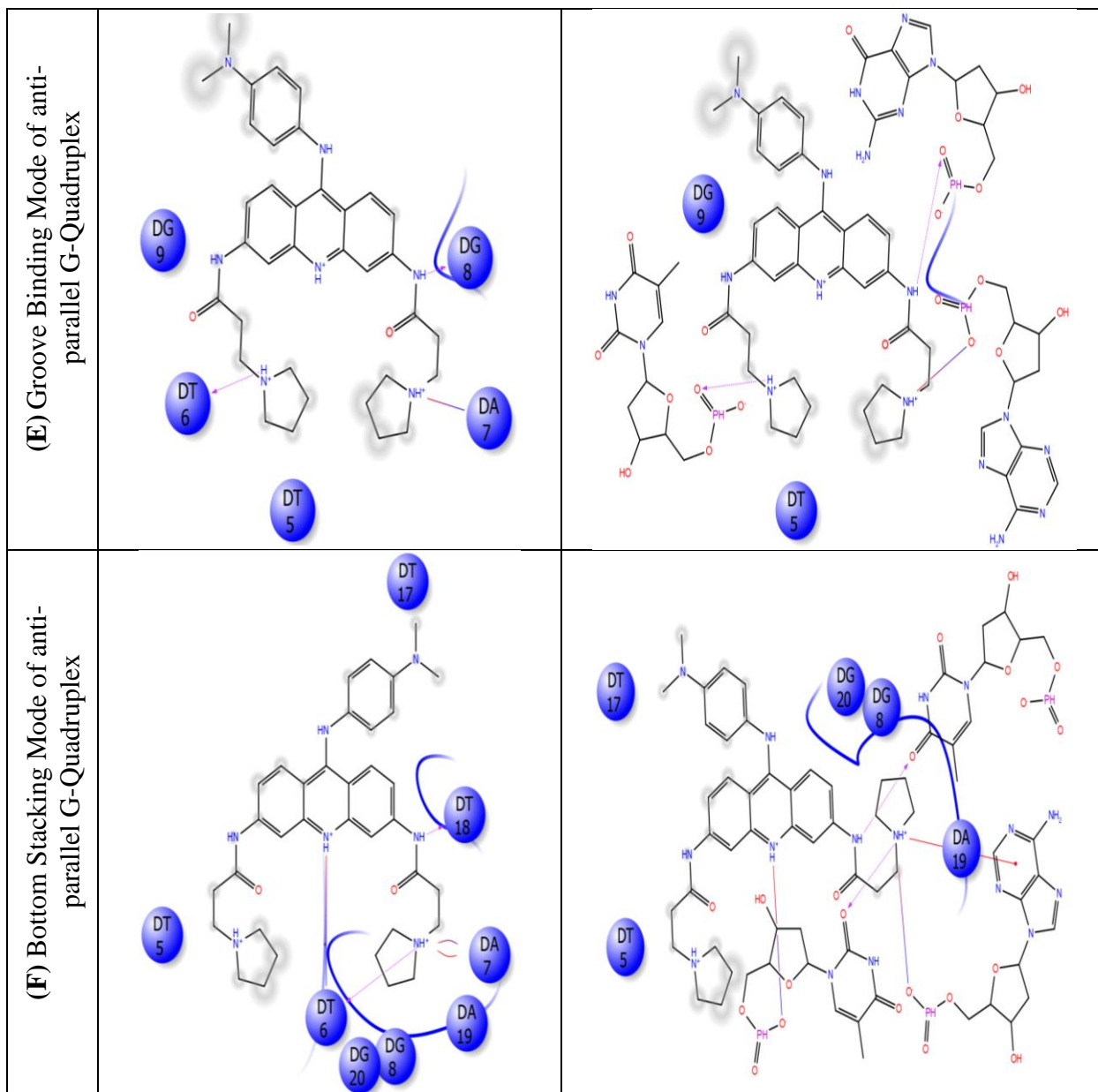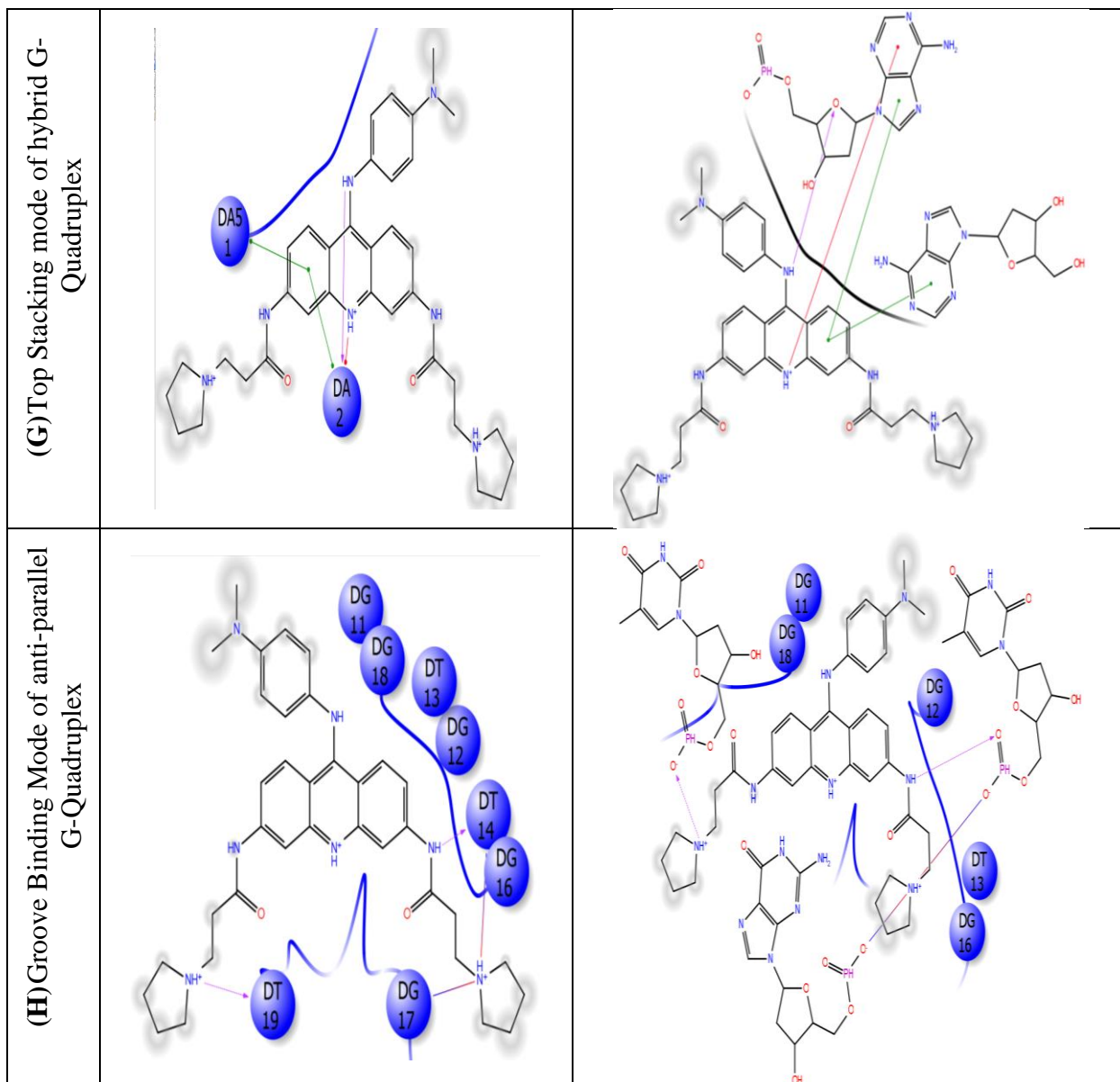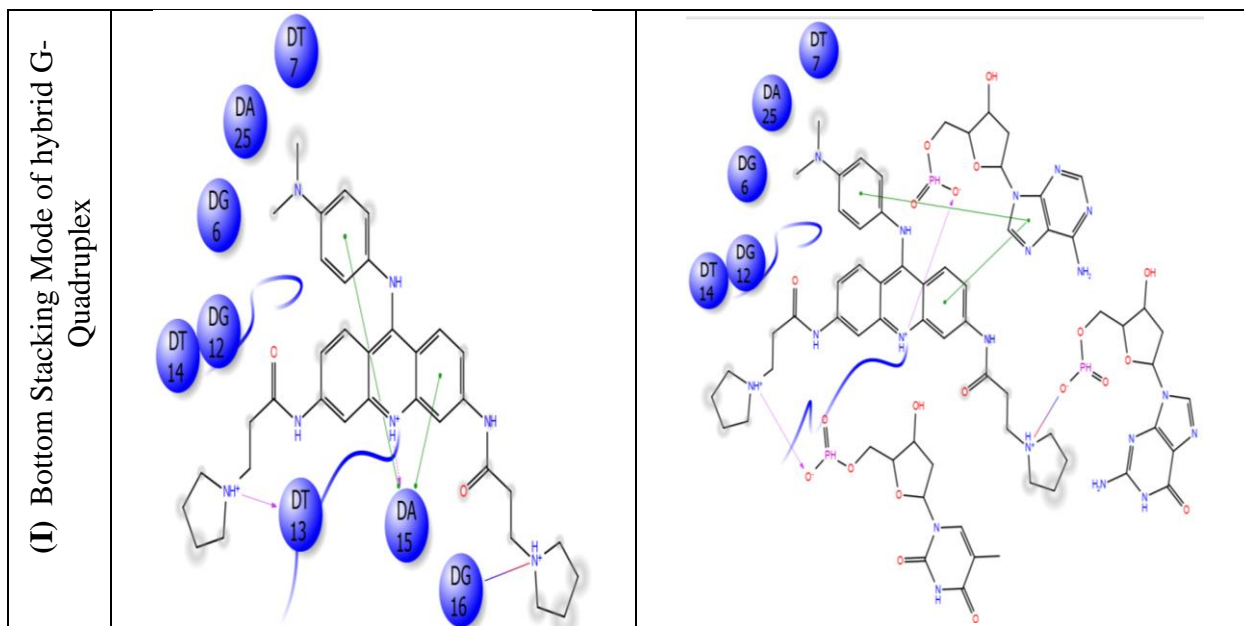
Figure A18. 2D interactions of BRACO19 in representative structures of (A) Groove Binding Mode of the Duplex DNA, (B) Top Stacking mode of the parallel G-Quadruplex, (C) Groove Binding Mode of the parallel G-Quadruplex, (D) Top Stacking mode of the anti-parallel G-Quadruplex, (E) Groove Binding Mode of the anti-parallel G-Quadruplex, (F) Bottom Stacking Mode of the anti-parallel G-Quadruplex, (G) Top Stacking mode of the hybrid G-Quadruplex, (H) Groove Binding Mode of the anti-parallel G-Quadruplex and (I) Bottom Stacking Mode of the hybrid G-Quadruplex.

158
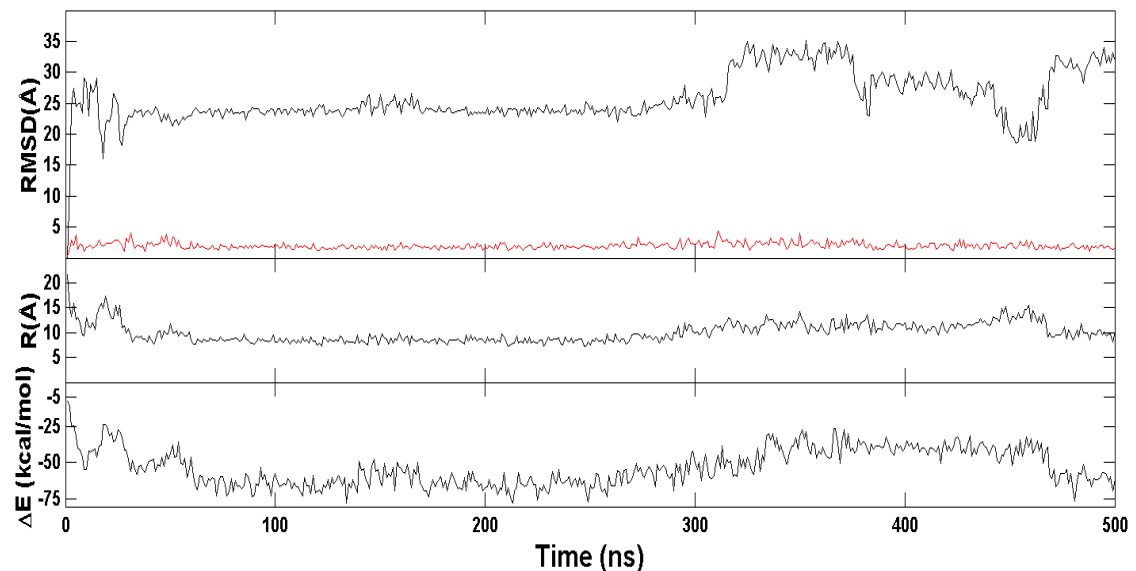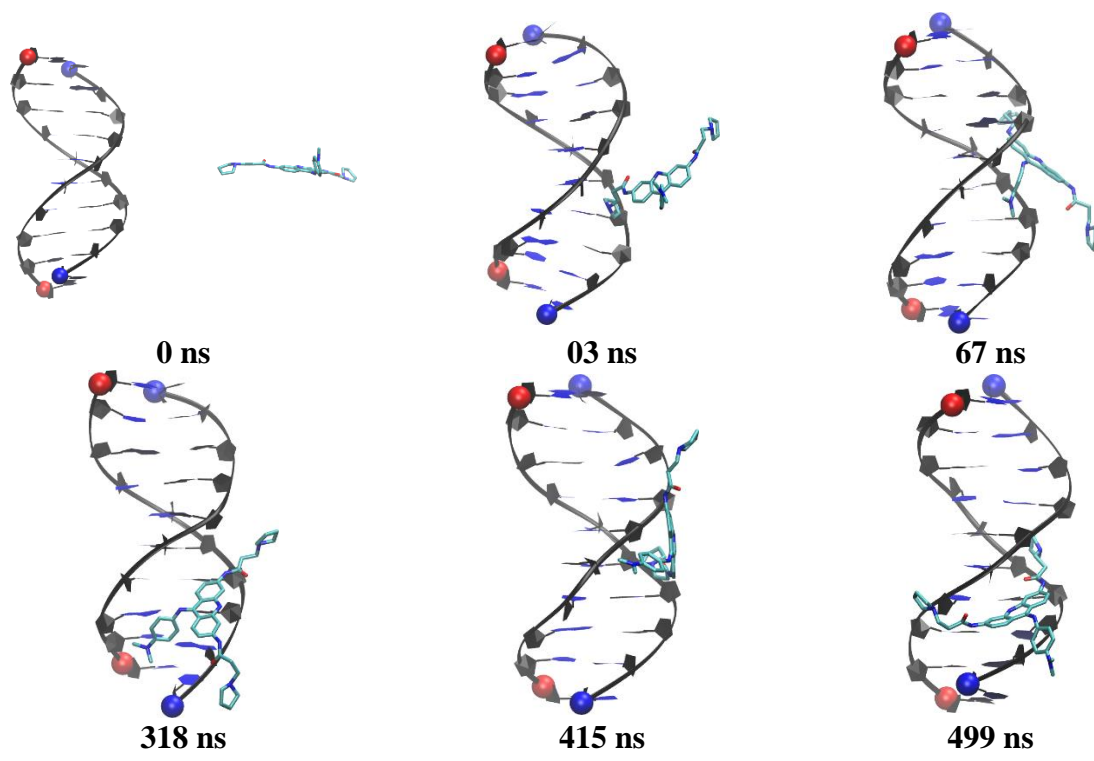
Figure A19. Another representative groove binding trajectory of the duplex DNA. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first base pair (green), second base pair (red) and third base pair (blue) tetrads of the DNA structure (figure 2), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

159
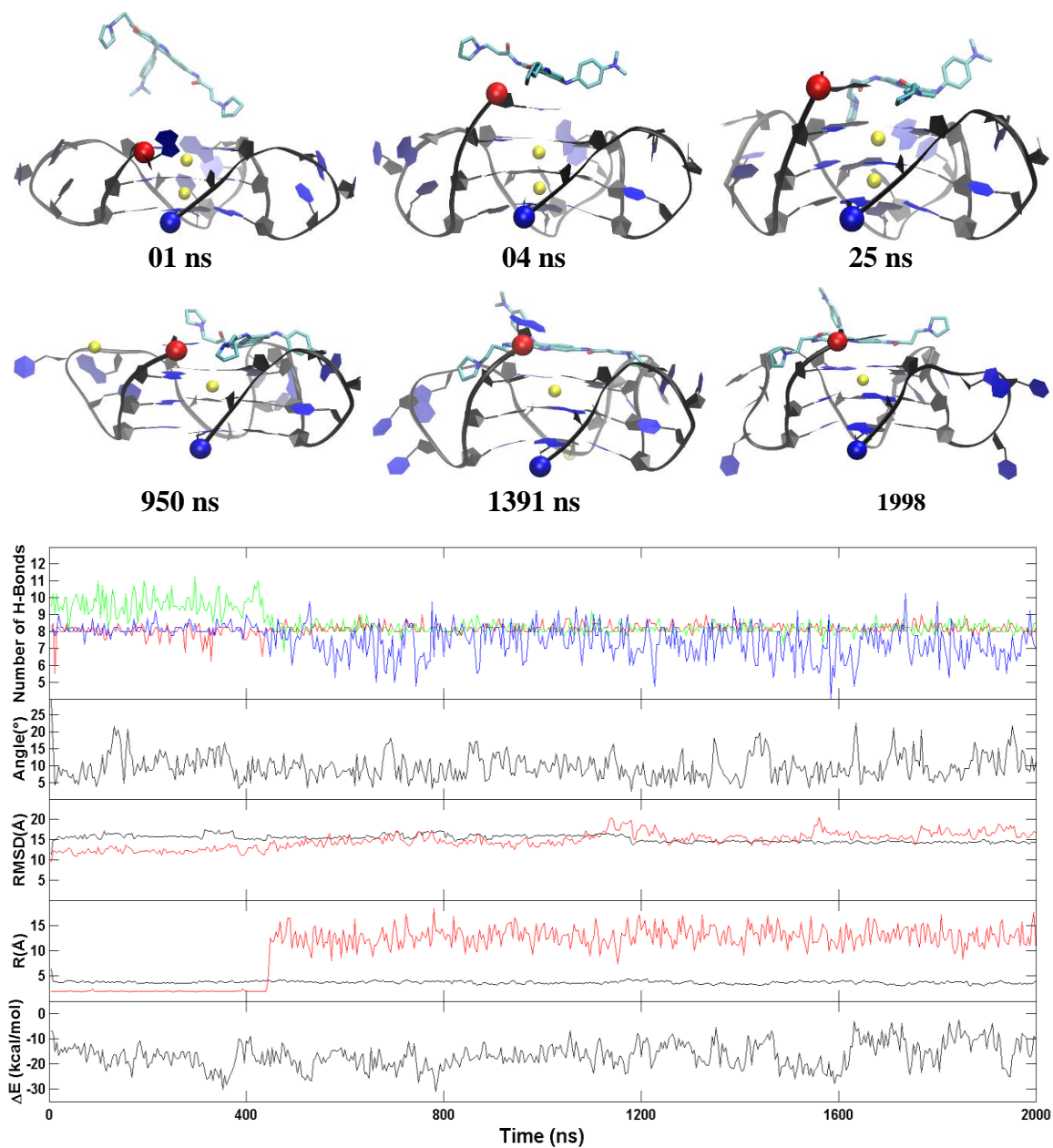
Figure A20. Another representative top stacking trajectory of the parallel G-quadruplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K$^+$ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first G4 (green), second G4 (red) and third G4 (blue) tetrads of the DNA structure (Figure 2), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K$^+$-K$^+$ distance (R/red) and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

160

Figure A21. A representative bottom binding trajectory of the parallel G-quadruplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first G4 (green), second G4 (red) and third G4 (blue) tetrads of the DNA structure (Figure 2), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

161

Figure A22. A representative top groove binding trajectory of the anti-parallel G-quadruplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first (red), second G4 (cyan), third G4 (blue), fourth G4 (black) and fifth (green) layers of the DNA structure (Figure S9), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-GBSA binding energy (ΔE) (cf. methods section for definition).
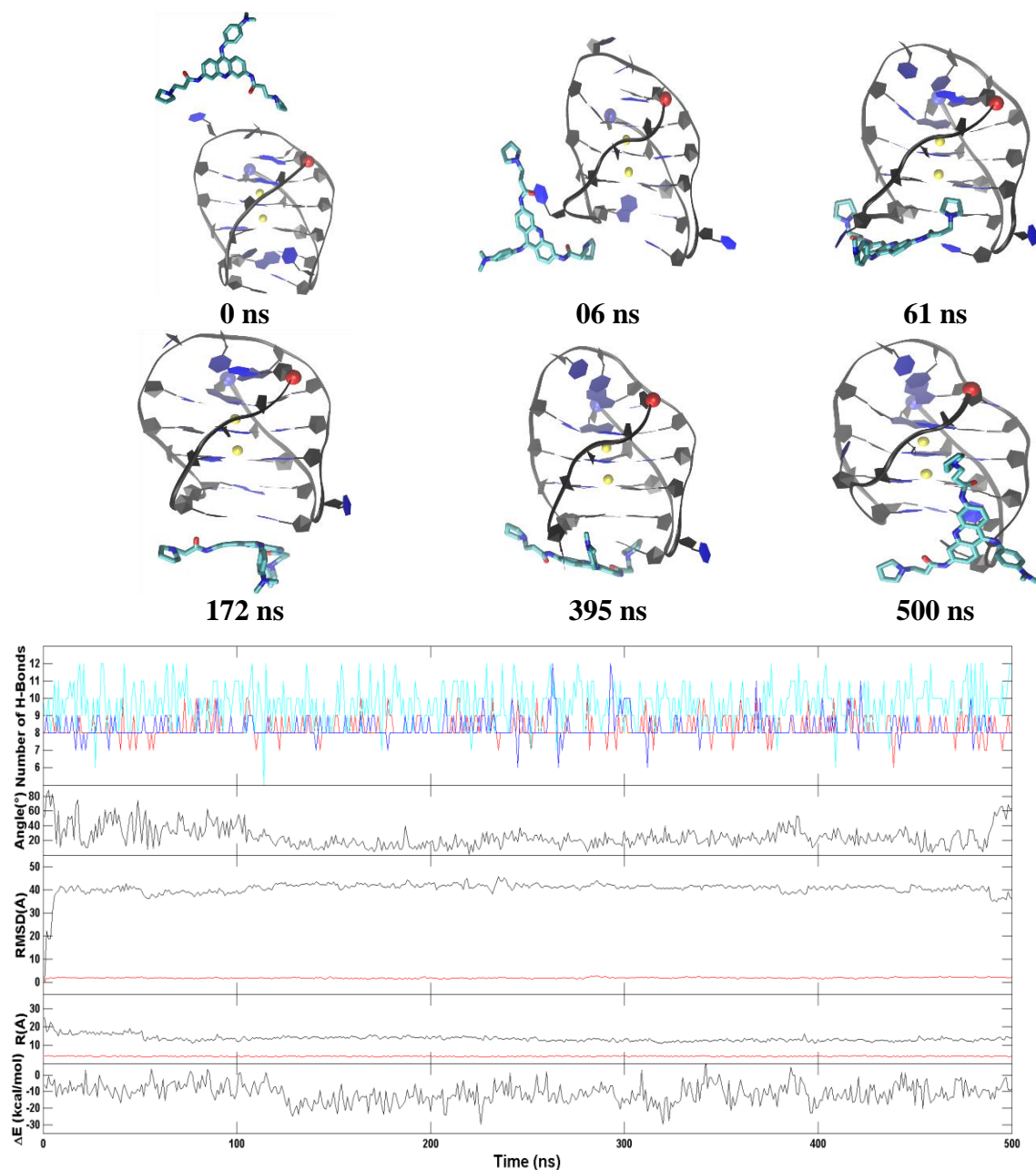
Figure A23. Another representative bottom stacking trajectory of the anti-parallel G-quadruplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first G4 (green), second G4 (red) and third G4 (blue) tetrads of the DNA structure (Figure 2), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

Figure A24. Another representative top binding trajectory of the hybrid G-quadruplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first G4 (green), second G4 (red) and third G4 (blue) tetrads of the DNA structure (Figure 2), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

Figure A25. Another representative bottom binding trajectory of the hybrid G-quadruplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first G4 (green), second G4 (red) and third G4 (blue) tetrads of the DNA structure (Figure 2), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

165

Figure A26. Another representative groove binding trajectory of the hybrid G-quadruplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first G4 (green), second G4 (red) and third G4 (blue) tetrads of the DNA structure (Figure 2), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

166

Figure A27. Backbone Torsion Angles of DNA.

System	Side View	Top View	Ligand Chemical Structure
Tel26_PtTripod (PDB ID: 5Z80)			
MD_Tel26_BRACO19			
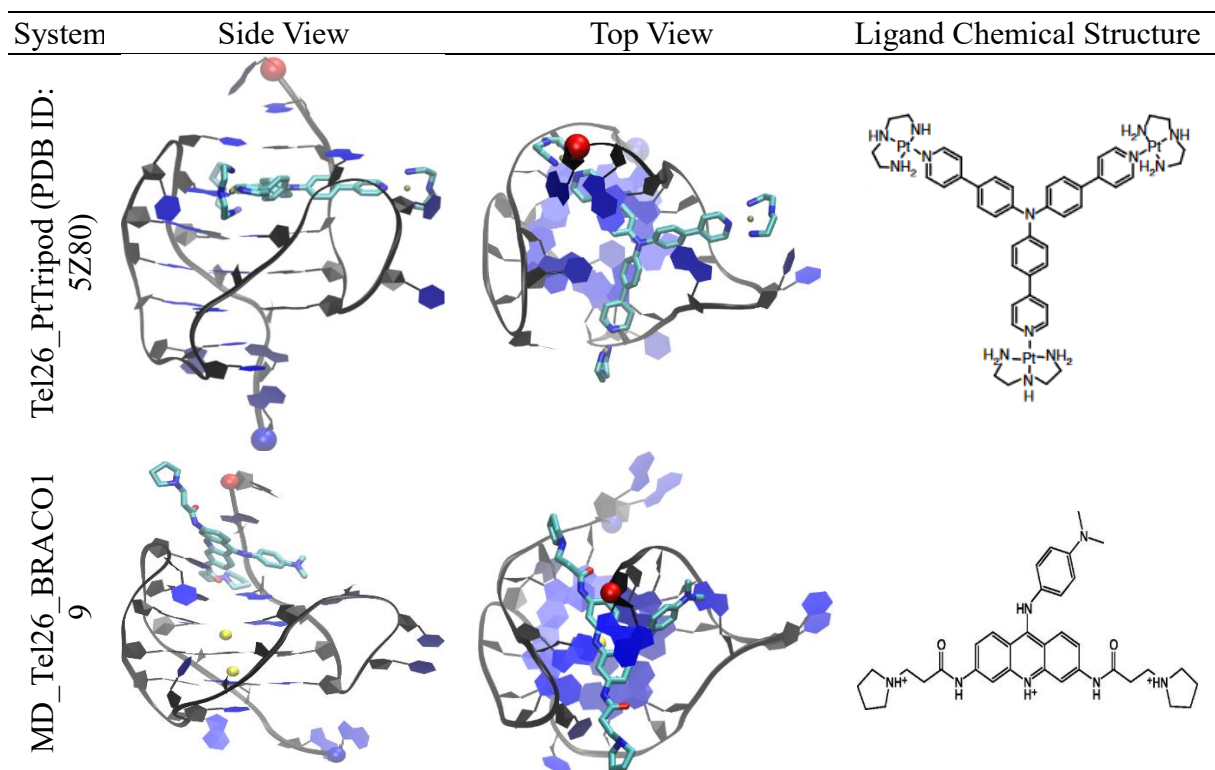
Figure A28. Comparison of Tel26 hybrid G-quadruplex-Pt Tripod Complex Crystal Structure (PDB ID: 5Z80) to Tel26 hybrid G-quadruplex-BRACO19 Complex. The 5' residues are represented by a red ball and the 3' residues in blue.

A.

01 ns   03 ns   17 ns   118 ns   200 ns   400 ns

00 ns   19 ns   26 ns   56 ns   102 ns   463 ns

00 ns   05 ns   79 ns   288 ns   797 ns   1998 ns

B.

02 ns   16 ns   135 ns   149 ns   416 ns   500 ns

01 ns   06 ns   48 ns   409 ns   1576 ns   2000 ns

00 ns   01 ns   04 ns   06 ns   16 ns   199 ns

169

C.

01 ns	**68 ns**	**93 ns**	**162 ns**	**294 ns**	**500 ns**
00 ns	**04 ns**	**94 ns**	**178 ns**	**187 ns**	**200 ns**
00 ns	**09 ns**	**38 ns**	**40 ns**	**51 ns**	**200 ns**
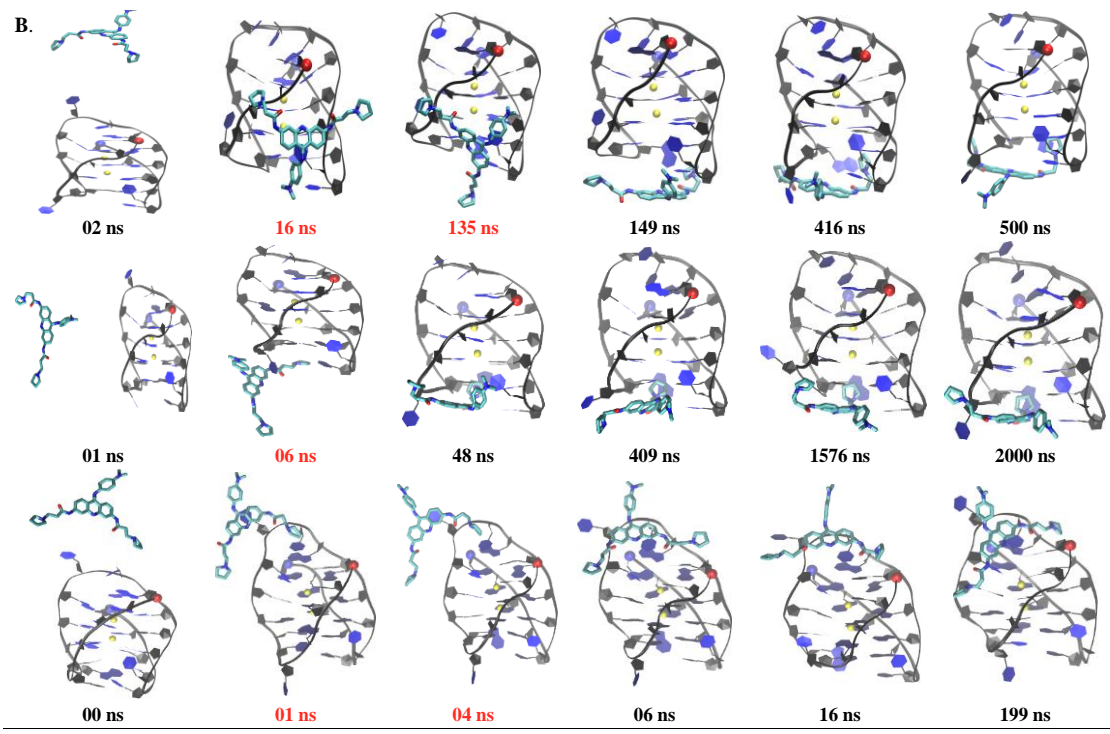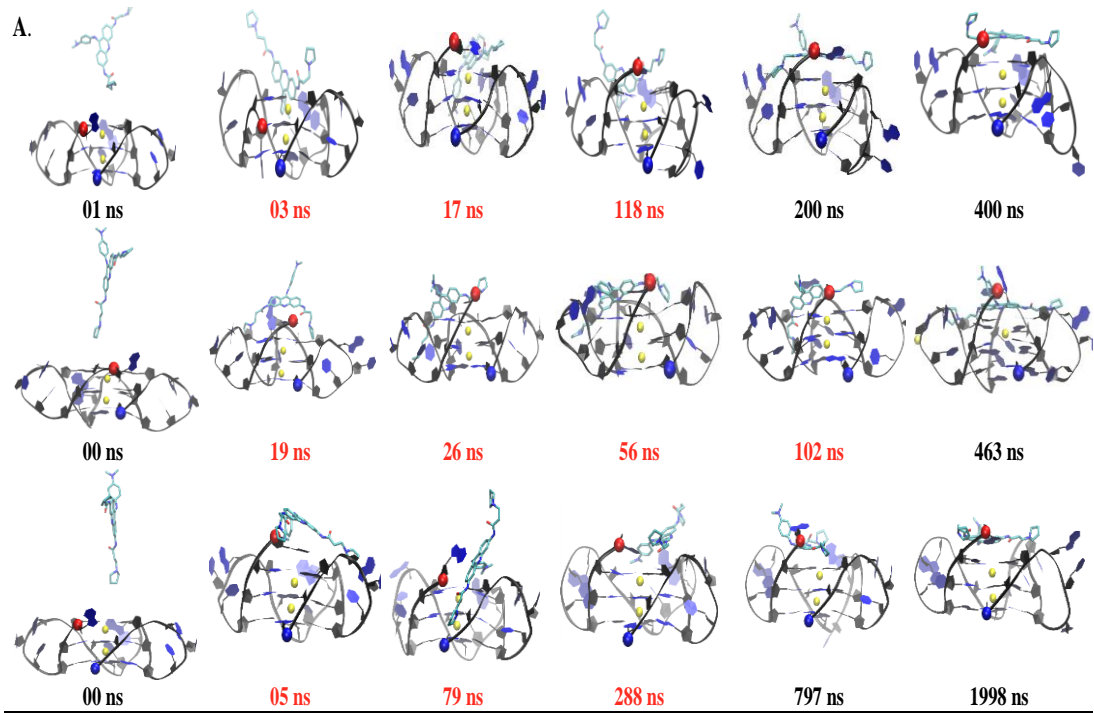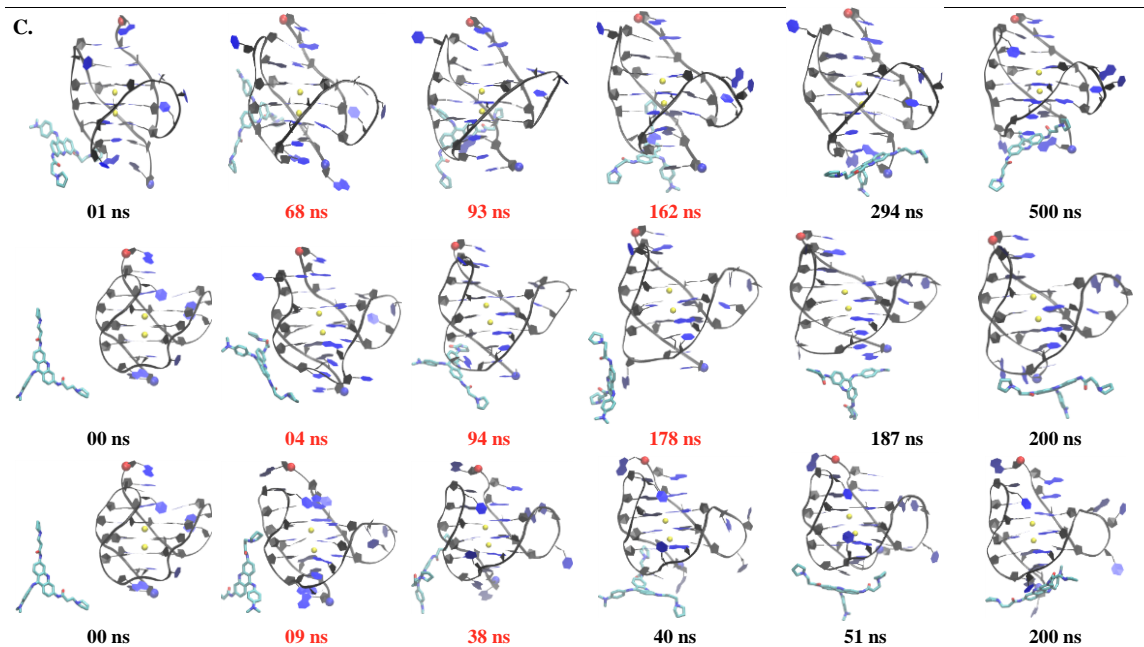
Figure A29. Representative snapshots of BRACO19 using groove binding as an intermediate state in each DNA G-quadruplex system: parallel (A), anti-parallel (B), and hybrid (C). The snapshots that BRACO19 is in a groove binding mode are in red. The 5' residues are represented by a red ball and the 3' residues in blue.

170
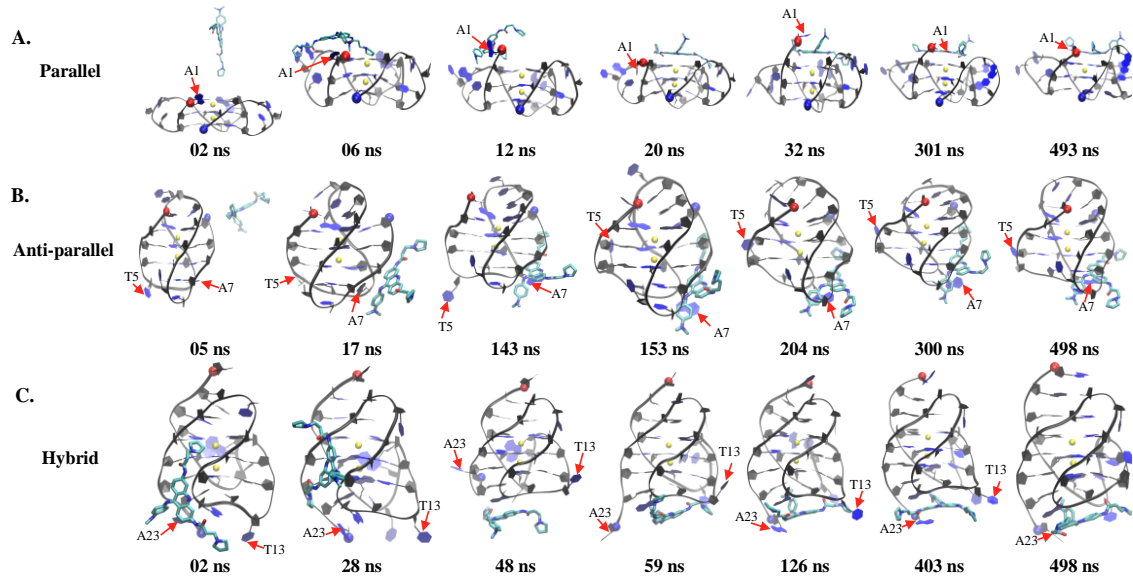
Figure A30. BRACO19 binding to the G-quadruplexes using an induced fit binding mechanism. The 5' residues are represented by a red ball and the 3' residues in blue.

171
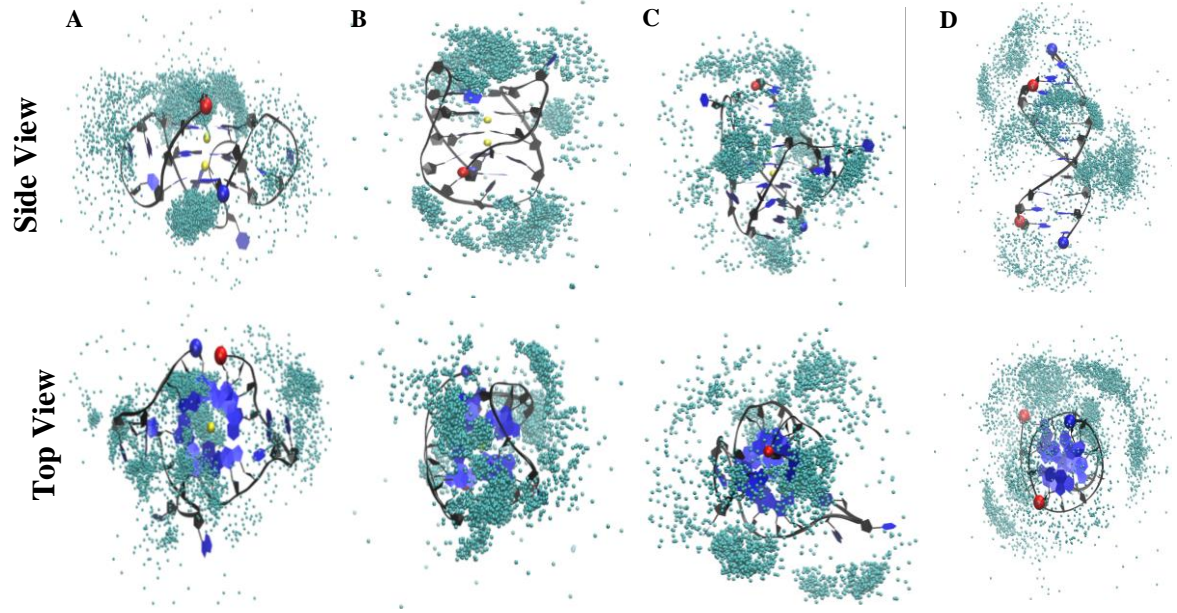
Figure A31. Sampling plot of BRACO19 in complex with the (A) Parallel (B) Anti-Parallel (C) Hybrid and (D) DNA duplex systems.

Appendix B

To Probe the Binding of CX-5461, an Anti-Cancer DNA G-Quadruplex Stabilizer, to Human Telomeric, c-KIT-1, and c-Myc G-Quadruplexes and a DNA Duplex Using Molecular Dynamics Binding Simulations

Supporting Experimental Methods

Stability of Simulation Systems. The root mean square deviation (RMSD) of the DNA backbone for each apo form was calculated against the initial structure (**Figure 2**) and stability of the DNA structures was indicated by the flat and small RMSDs in all runs (**Figure S1**). Last snap shots of the DNA only systems (**Figures S2**), and representative trajectories including order parameter plots (**Figure S3**) are provided in the supporting document.

Convergence of simulations. A plot showing the position of CX-5461, represented by one atom, in each trajectory is presented (**Figure S4**) and indicates good sampling in each system. The root mean square deviation (RMSD) of the DNA backbone was calculated against the initial structure for the complex systems and the stability of the DNA structures was indicated by the flat and small RMSDs in all runs (**Figure S5**). Atom contacts between the DNA structure and the ligand were calculated using a 3.0 Å cutoff (**Figure S6**). The stable contact number indicates the simulation systems reached a steady state in all runs where a stable complex was defined as a complex with a number of atom contacts between the DNA and ligand greater than 30. Last snap shots of the complex systems are provided in the support document (**Figures S7-S10**).

173

G-quadruplex and DNA Duplex Parameters. Following the methods from our previous study[186] the rise, helical rise (H-rise), and helical twist (H-twist) were calculated to gain insight on the inter G4 layer geometry (**Table S2**). Using the Curves+[193] program, we calculated 20 helical parameters to characterize the base pairing of the DNA duplex in regard to the base pair axis (**Table S3**), intra-base pairs (**Table S4**), inter-base pairs (**Table S5**), as well as the puckering conformations of each residue (**Table S6**).

Featurization and Clustering. For each of the four systems, 30 trajectories (501 ns each) were combined into one trajectory. Using VMD, all frames in which there were less than 13 interactions, at a distance less than 3Å, between the G-quadruplex and the ligand were separated as the unbound state [194] for the G-quadruplex systems. 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 [195]. K-means clustering, performed using scikit-learn, was then used to classify the remaining frames into various states [196]. Clustering was performed for K between 2 and 30 inclusively, using the silhouette index as the metric for similarity of clusters [196, 197].

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. Statistics of the clustering analysis are provided in the

174

supporting document (**Table S1**). The number of each unique binding pathway's taken for each system is summarized in **Table S2**. Each structural family identified for the duplex system is reported in **Figure S11**.

Transition Path Theory. Count matrices were then created for lagtimes (τ) of 1, 10, 20, 30 … 500 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 $t \leq t_{max} - \tau$ [198]. The count matrices were symmetrized (sym_{ij}) such that $sym_{ij} = sym_{ji} = \frac{c_{ij}+c_{ji}}{2}$ and then row-normalized ($norm_{ij}$) such that $norm_{ij} = \frac{sym_{ij}}{\sum_{j=1}^{j=n} sym_{ij}}$. 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 S12**). The implied timescale of the first cluster is not included in the plot as the eigenvalue is always 1 and thus contributes no information[199]. The 1KF1 system nodes are labelled 2 for top binding, 3 for side binding, and 4 for bottom binding. The 2MGN system nodes are labelled 2 for top binding, 3 for bottom binding, and 4 for side binding. The 4WO3 system nodes are labelled 2 for side binding, 3 for bottom binding, and 4 for top binding. Further validation that the model had been converged was performed through the Chapman-Kolmogorov test (**Figure S13**)[198]. Network models (**Figure S14**) were then generated based on the count matrices at optimal lag times of 150, 100, and 170 ns for the 1KF1, 2MGN, and 4WO3 systems respectively with the cutoff for a directed edge in the network being set at 50 transitions[200]. Edges in the network diagram are drawn if there were at least 50 transitions (in any single direction) between two nodes. Outgoing

175

transitions are labelled on the edges as the same color as the originating node. The 1KF1 system nodes are labelled 1 for unbound, 2 for top binding, 3 for side binding, and 4 for bottom binding. The 2MGN system nodes are labelled 1 for unbound, 2 for top binding, 3 for bottom binding, and 4 for side binding. The 4WO3 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 times and the standard deviations from the optimal lag times to 300 ns were calculated according to the formula $F_{if} = \tau + \sum_{j \neq f} P_{ij} F_{jf}$, with the boundary condition $F_{ff} = 0$, where τ is the lag time used to construct the transition matrix $P(\tau)$. Summary plots were generated showing the The mean first passage times between the four states (unbound, side transition, top, and bottom) of the CX-5461 and human telomeric (**Figure 4**), c-KIT1 (**Figure 5**), and c-Myc (**Figure 6**) DNA G-quadruplex complex systems.

Order parameters characterize the DNA-drug binding pathway. The DNA-drug binding process was characterized using seven order parameter calculations: hydrogen bond analysis, drug-base dihedral angle, drug center to ligand center distance (R*)*, K^+-K^+ cation distance, DNA and ligand RMSD and MM-GBSA binding energy (ΔE) (**Figures 7-15**;**Figures S15-S25**). The distance cutoff between H-donor and H-acceptor was set to be 3.5Å and the donor-H-acceptor angle cutoff was set to be 120°. The hydrogen bonds were calculated for the top/first, middle/second and bottom/third base layers, and other base pairing when applicable (**Figure 2 E-H**). A visual representation of the hydrogen bond networks are presented in the supporting document (**Figures S29-S31**). The definition for the quadruplexes is the three guanine layers with the 5' side as the first layer. The center-

176

to-center distance (R) was defined in two ways: as the length from the DNA center to the drug molecule center and the length between the two K$^+$ cations within the G-quadruplex structure. The dihedral angle was defined as the dihedral angle between the plane of the stable unbroken base-layer of the DNA that is closest to the drug binding site and the drug's ring plane. MM-PBSA[201] (Molecular Mechanics Poisson Boltzmann Surface Area) module in the AMBER package (PB1 model with mBondi radii set, salt concentration of 0.2 M, and surface tension of 0.0378 kcal/Å^2) was used to analyze the energetics of the bound complexes. The MM-PBSA binding energy for a system was calculated based on three simulations: the ligand only, the DNA only and the DNA-ligand complex using equation 1. The equation is made of four components Eq2: van der Waals interaction energy (VDW), hydrophobic interaction energy (SUR), electrostatic interaction (GBELE) and the change of the conformation energy for DNA and ligand which are calculated using equation 3 and 4. MM-PBSA binding energy is an effective tool for ranking ligand binding affinities proven by up to 1864 crystal complexes tested in systematic benchmarking studies. [202-206]

Eq 1: $\quad \Delta E = E_{complex} - E_{DNA\_free} - E_{lig\_free}$

Eq 2: $\quad \Delta E = \Delta E_{vdw} + \Delta E_{SUR} + \Delta E_{GBELE} + \Delta E_{comformation}$

Eq 3: $\quad \Delta E_x = E_{x_{complex}} - E_{x_{DNA_{complex}}} - E_{x_{lig_{complex}}}$,

$x= vdw, sur$ and $gbele$

Eq 4: $\quad \Delta E_{Comformation} = E_{DNA\_complex} + E_{lig\_complex} - E_{DNA\_free} - E_{lig\_free}$

177

Structural fluctuation of the DNA. For each major binding mode and apo form systems, the root mean square fluctuations (RMSF) of each individual residue of DNA were calculated to characterize the local structural fluctuation (**Figure S35**).

Virtual Screening. The top derivative was chosen from a virtual screening of a CX-5461 derivate library using Maestro 10.3[207]. First, a combinatorial library of 64 ligands prepared using the Interactive Enumeration program. The variants were defined by establishing substitution sites where there were three possible points of substitution to CX-5461. At each possible point of substitution there were 4 functional groups that can be substituted which included hydrogen, fluorine, chlorine, bromine. This suggests that there are 4^3 possible modified versions of CX-5461, and a combinatorial library of these 64 ligands was generated. The charge of each compound at pH=7 was determined by Epik (an empirical pKa prediction program)[208] followed by a geometry optimization that minimized the potential energy using the default parameters. Using the active receptor structure from the most abundant conformation of each system, a grid file was generated using the Receptor Grid Generation program to prepare the complex for the subsequent docking calculation. In each system, ergotamine or dihydroergotamine was selected and a grid box was generated around the ligand with a van der Waals radius scaling factor of 1.0 and a partial cutoff of 0.25. Then, these 64 compounds were docked using Glide with Extra Precision (XP) scoring function, and then filtered using QikProp package[209], to predict the absorption, distribution, metabolism, and excretion (ADME) properties. QikProp ranks the full molecular structure based on pharmaceutically relevant properties by giving each compound a number of stars; compounds with no starts are predicted to be

178

the most drug-like. Finally, three potential compounds for each ligands were manually

chosen based on XP scores (< -9.0 kcal/mol) that were more negative than the docking of

ERG or DHE into the active conformation of $5HT_{1B}$ from the most abundant clusters,

along with the compounds' synthesizability containing fewer substitution groups.

Table B1

Final binding mode statistics of the 30 simulation runs. Reported are the number of
trajectories in each system that intercalates to the top, stacks to the top or bottom, or
binds to a side residue in each of the simulation runs.

System	Top Intercalation	Top Stacking	Bottom Stacking	Side Binding
h-Tel	10	3	7	10
c-KIT1	7	11	6	7
c-Myc	17	2	3	8
Duplex	1	12	10	7

Table B2

DNA G-quadruplex G4 layer geometry parameters for the apo form and two major binding modes of the human telomeric (1KF1), c-KIT1 (4WO3) and c-Myc (2MGN) G-quadruplexes. 1-3: top layer, middle and G4 bottom, respectively.

Parameter	Rise[1]		H-Rise[1]		H-Twist[2]	
Layers[3]	3:2	2:1	3:2	2:1	3:2	2:1
1KF1A	3.56±0.01	3.42±0.01	1.91±0.41	1.86±0.32	12.76±2.55	11.86±1.63
1KF1T	3.55±0.02	3.47±0.02	3.05±0.03	2.94±0.05	22.74±0.15	21.35±1.38
1KF1B	3.51±0.02	3.42±0.03	2.14±0.05	2.07±0.04	13.07±0.56	15.68±3.52
4WO3A	3.49±0.01	3.41±0.01	1.96±0.08	2.05±0.04	17.19±0.01	17.39±0.73
4WO3T	3.49±0.01	3.42±0.01	1.35±0.04	1.48±0.02	16.54±0.17	6.43±0.66
4WO3B	3.50±0.01	3.43±0.00	2.63±0.02	2.69±0.04	23.11 ±0.15	17.18±0.54
2MGNA	3.48±0.01	3.42±0.01	0.32±3.66	0.29±0.08	36.21±3.66	3.90±0.85
2MGNT	3.48±0.02	3.44±0.04	0.87±0.08	0.79±0.05	9.79±1.18	4.67±0.80
2MGNB	3.47±0.00	3.42±0.01	3.17±0.02	3.11±0.02	81.90±0.43	24.29±0.28

[1]Values reported in Å

[2]Values reported in degree

[3]System legend: Apo (A), Top (T), Bottom (B)

Table B3

Base pair axis parameters for the DNA duplex, abnormal values are in red.

Base-pair	A/G[1]	X-displacement[1]	Y-Displacement[2]	Inclination[3]	Tip[3]	Axis-bend[3]
1) G1-C20	A	-0.05±1.44	0.01±0.57	4.66±5.03	2.30±6.77	---
	G	-0.47±1.15	-0.45±1.04	13.21±41.43	-15.35±37.05	---
2) C2-G19	A	-0.47±1.24	-0.13±0.56	4.12±5.38	-0.46±6.05	1.30±0.88
	G	-0.55±0.77	-0.253±0.81	5.58±16.11	-1.65±8.88	3.33±2.79
3) G3-C19	A	-0.27±1.16	-0.04±0.63	7.71±6.59	5.94±4.02	1.49±0.94
	G	-0.57±0.82	0.14±0.78	6.01±7.94	1.28±7.74	2.73±1.89
4) C4-G17	A	-0.15±0.91	0.01±0.64	7.02±4.83	-1.00±5.15	1.52±1.04
	G	-0.75±0.87	0.09±0.67	7.99±6.97	-2.26±7.72	2.32±1.45
5) G5-C16	A	-0.33±0.85	0.31±0.43	5.77±6.68	1.41±2.63	1.72±1.15
	G	-0.60±0.92	0.52±0.35	6.96±3.51	0.53±6.78	2.53±1.07
6) C6-G15	A	-0.12±0.62	-0.14±0.77	4.50±6.60	-0.94±4.34	2.37±1.18
	G	-0.29±1.08	0.16±0.42	6.88±6.62	-3.88±4.44	3.25±0.97
7) G7-C14	A	-0.46±0.65	-0.02±0.79	6.82±4.62	2.42±4.78	2.13±0.95
	G	-0.09±1.12	0.29±0.65	5.05±3.84	1.57±3.79	2.26±1.50
8) C8-G13	A	-0.50±0.87	0.00±0.89	6.54±6.93	-6.18±3.33	1.89±0.90
	G	-0.06±0.85	-0.00±0.56	3.73±9.45	-5.60±2.15	2.19±2.00
9) G9-C12	A	-0.14±0.90	-0.01±0.73	4.54±6.46	-0.56±4.18	1.78±0.85
	G	0.48±1.18	0.01±0.67	0.13±13.15	-1.58±8.15	2.14±2.30
10) C10-G11	A	-0.07±1.33	-0.32±0.78	5.83±7.31	-1.65±7.58	1.87±0.68
	G	0.25±1.56	-0.29±1.15	-6.04±26.73	-9.27±18.46	2.36±2.89

[1]Apo (A) or Groove (G)

[2]Values reported in Å

[3]Values reported in degree

181

Table B4

Base pair step parameters, abnormal values are in red.

Base-pair	A/G[1]	Shear[2]	Stretch[2]	Stagger[2]	Buckle[3]	Propel[3]	Opening[3]
1) G1-C20	A	0.00±0.33	0.06±0.13	0.06±0.45	-0.02±13.44	-0.64±6.01	2.00±4.40
	G	-0.60±1.52	-1.34±3.14	0.06±3.32	7.06±58.97	-35.04±80.92	3.38±17.9
2) C2-G19	A	0.03±0.20	0.04±0.08	0.08±0.39	5.25±10.70	-10.37±10.54	0.08±2.78
	G	-0.05±0.24	0.06±0.10	0.17±0.33	1.72±9.37	-7.10±10.85	1.64±3.24
3) G3-C19	A	-0.12±0.29	0.00±0.13	0.04±0.43	-4.97±7.96	-12.88±8.31	1.74±3.32
	G	-0.15±0.25	-0.01±0.11	0.30±0.26	6.32±9.18	-3.81±8.62	1.49±3.77
4) C4-G17	A	-0.16±0.26	0.04±0.08	0.28±0.25	-2.85±8.39	-10.53±10.67	1.65±3.36
	G	0.04±0.34	0.04±0.11	0.02±0.46	5.97±11.97	-9.31±5.15	1.42±3.41
5) G5-C16	A	-0.09±0.27	0.05±0.10	0.16±0.37	-2.97±10.13	-5.82±10.97	2.33±3.34
	G	-0.03±0.14	0.00±0.07	0.19±0.16	-4.88±6.75	-12.92±9.55	2.28±2.75
6) C6-G15	A	-0.18±0.20	0.00±0.08	0.45±0.44	-4.37±11.10	-12.46±7.58	0.11±2.85
	G	-0.10±0.29	0.08±0.11	0.24±0.39	-2.08±9.95	-12.39±6.90	2.47±2.59
7) G7-C14	A	0.09±0.32	0.03±0.09	0.22±0.32	-1.71±7.21	-8.86±8.23	-0.48±3.69
	G	-0.04±0.19	0.04±0.11	0.28±0.42	-1.52±11.75	-7.63±4.50	-0.32±3.54
8) C8-G13	A	-0.04±0.28	-0.04±0.13	0.00±0.47	-1.74±10.17	-9.98±5.48	0.29±2.29
	G	0.04±0.39	0.04±0.10	0.18±0.46	1.51±9.38	-10.82±5.98	-0.32±3.7
9) G9-C12	A	-0.06±0.37	0.06±0.13	0.23±0.36	2.23±13.20	-6.04±6.27	0.43±4.67
	G	-0.37±0.84	0.09±0.13	0.20±0.57	4.22±14.62	-7.37±15.66	2.52±11.9
10) C10-G11	A	0.19±0.38	-0.03±0.15	0.12±0.51	6.65±12.42	-2.87±8.99	0.42±2.84
	G	-0.34±0.99	-0.82±2.63	1.41±4.93	-5.81±23.54	-11.59±46.49	6.60±17.1

[1]Apo (A) or Groove (G)

[2]Values reported in Å

[3]Values reported in degree

182

Table B5

Paired base- base parameter, abnormal values are in red.

Pair	A/G[1]	Roll[3]	Twist[3]	H-Ris[2]	H-Twi[3]
G1/C2	A	-0.59±4.98	31.51±6.47	3.25±0.25	31.78±6.31
	G	18.97±57.09	22.85±17.69	3.59±2.20	33.99±23.19
C2/G3	A	9.67±6.61	38.35±6.86	3.43±0.35	38.93±6.50
	G	6.45±6.23	33.72±7.08	3.09±0.38	34.15±7.67
G3/C4	A	-1.65±6.55	34.38±4.68	3.22±0.28	34.55±4.46
	G	0.64±4.73	33.65±5.83	3.32±0.28	33.92±5.57
C4/G5	A	6.64±6.60	35.65±8.55	3.37±0.33	36.23±8.01
	G	8.42±4.81	38.71±6.08	3.49±0.16	39.39±5.86
G5/C6	A	1.40±4.55	35.23±3.44	3.29±0.37	35.46±3.31
	G	0.61±6.51	30.48±5.36	3.15±0.22	30.74±5.16
C6/G7	A	6.61±8.28	35.82±8.16	3.21±0.30	36.47±7.60
	G	10.49±5.12	36.51±6.01	3.22±0.39	37.18±6.03
G7/C8	A	-3.90±6.56	36.13±4.92	3.33±0.20	36.24±4.85
	G	-2.96±5.24	32.37±4.88	3.27±0.23	32.42±4.97
C8/G9	A	8.43±4.40	31.86±7.49	3.17±0.33	32.39±7.47
	G	5.59±12.22	36.80±8.78	3.37±0.43	37.80±8.11
G9/C10	A	1.62±7.80	33.71±5.61	3.30±0.29	34.25±5.36
	G	-6.56±13.84	30.36±12.23	3.49±0.41	31.85±8.53

[1]Apo (A) or Groove (G)

[2]Values reported in Å

[3]Values reported in degree

183

Table B5 (continued).

Pair	A/G[1]	Shift[2]	Slide[2]	Rise[2]	Tilt[3]
G1/C2	A	-0.41±0.63	0.15±0.45	3.30±0.31	-0.76±3.71
	G	0.84±2.64	0.30±2.00	2.73±2.00	-1.68±34.15
C2/G3	A	0.26±1.05	0.15±0.84	3.47±0.30	0.88±7.36
	G	0.38±0.86	0.28±0.51	3.15±0.29	0.01±5.48
G3/C4	A	-0.03±0.74	0.24±0.69	3.27±0.29	-2.53±4.23
	G	-0.09±0.58	-0.33±0.25	3.41±0.27	1.87±4.13
C4/G5	A	-0.30±0.99	0.46±0.71	3.39±0.30	-1.49±5.67
	G	-0.14±1.05	0.02±0.51	3.48±0.17	-1.26±6.45
G5/C6	A	0.09±0.42	-0.40±0.39	3.35±0.32	-1.11±3.11
	G	-0.22±0.61	-0.48±0.66	3.26±0.25	-0.27±4.77
C6/G7	A	-0.37±0.95	0.27±0.64	3.22±0.31	1.80±5.10
	G	-0.30±1.23	0.38±0.58	3.22±0.32	-0.97±7.33
G7/C8	A	0.11±0.59	0.14±0.57	3.38±0.24	1.21±3.70
	G	-0.21±0.97	0.12±0.44	3.30±0.23	0.76±7.17
C8/G9	A	0.48±0.89	0.11±0.54	3.17±0.34	0.20±5.38
	G	0.60±1.33	0.45±0.55	3.24±0.38	0.08±6.33
G9/C10	A	0.21±0.73	0.08±0.65	3.33±0.33	1.80±3.64
	G	0.41±0.94	-0.22±0.67	4.18±2.27	-3.35±11.72

[1]Apo (A) or Groove (G)

[2]Values reported in Å

[3]Values reported in degree

184

Table B6

Percentage of each sugar pucker conformation.[1]

Pucker	C1'ex		C2'en		C3'ex		C3'en		C4'ex		O1'en	
Residue	Apo	Groove	Apo	Groove	Apo	Groove	Apo	Groove	Apo	Groove	Apo	Groove
G1	50%	40%	40%	50%	10%	0%	0%	0%	0%	10%	0%	0%
C2	50%	60%	50%	40%	0%	0%	0%	0%	0%	0%	0%	0%
G3	40%	20%	10%	60%	40%	10%	0%	0%	10%	0%	0%	10%
C4	40%	30%	50%	60%	10%	0%	0%	0%	0%	0%	0%	10%
G5	30%	10%	60%	40%	10%	30%	0%	10%	0%	10%	0%	0%
C6	20%	40%	60%	20%	0%	0%	0%	20%	10%	10%	10%	10%
G7	30%	40%	40%	20%	30%	40%	0%	0%	0%	0%	0%	0%
C8	40%	30%	50%	50%	0%	0%	0%	0%	10%	10%	0%	10%
G9	30%	30%	40%	50%	10%	10%	0%	0%	10%	10%	10%	0%
C10	40%	20%	40%	30%	10%	0%	0%	0%	10%	30%	0%	20%
C20	70%	20%	20%	60%	0%	0%	0%	0%	0%	10%	10%	10%
G19	30%	50%	50%	20%	20%	30%	0%	0%	0%	0%	0%	0%
C18	30%	40%	60%	40%	0%	10%	0%	0%	0%	0%	10%	10%
G17	40%	30%	20%	40%	30%	30%	0%	0%	0%	0%	10%	0%
C16	30%	30%	20%	40%	40%	0%	0%	0%	0%	20%	10%	10%
G15	60%	20%	30%	70%	0%	10%	0%	0%	0%	0%	10%	0%
C14	0%	20%	90%	60%	0%	0%	0%	0%	0%	10%	10%	10%
G13	10%	10%	60%	60%	30%	30%	0%	0%	0%	0%	0%	0%
C12	20%	50%	50%	50%	10%	0%	0%	0%	0%	0%	20%	0%
G11	30%	50%	60%	30%	10%	10%	0%	10%	0%	0%	10%	0%

---

[1] Visual representation of the pucker conformations reported in Table B6.

210. Ho, P.; Carter, M., *DNA Structure: Alphabet Soup for the Cellular Soul*. 2011.

185
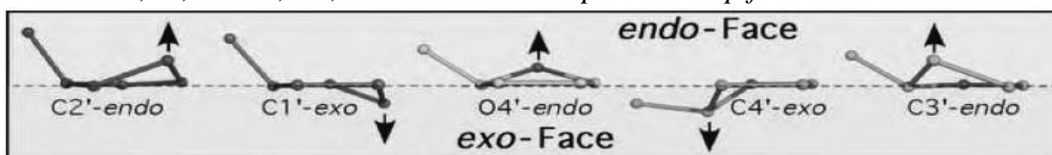
Table B7

Overall binding pathway taken by CX-5461 in each system for each of the 30 runs.

Binding Pathway	h-Tel	c-KIT1	c-Myc	Duplex
Top Intercalation	-	-	-	-
Top Stacking	-	2	1	10
Top Stacking to Top Intercalation	4	-	1	-
5' Terminal to Top Intercalation	-	1	11	-
5' Terminal to Top Stacking	-	1	-	-
Bottom Intercalation	-	1	-	-
Bottom Stacking	6	5	2	8
Bottom Stacking to Bottom Intercalation	-	-	1	-
Side Binding	7	4	5	6
Side Binding to Side Binding (reposition)	3	2	3	1
Side to Top Intercalation	7	3	4	1
Side to Top Stacking	-	7	2	1
Side to Top Stacking to Top Intercalation	-	3	-	-
Side to Top Intercalation to Top Stacking	2	1	-	-
Side Binding to Bottom Stacking	1	-	-	3

186
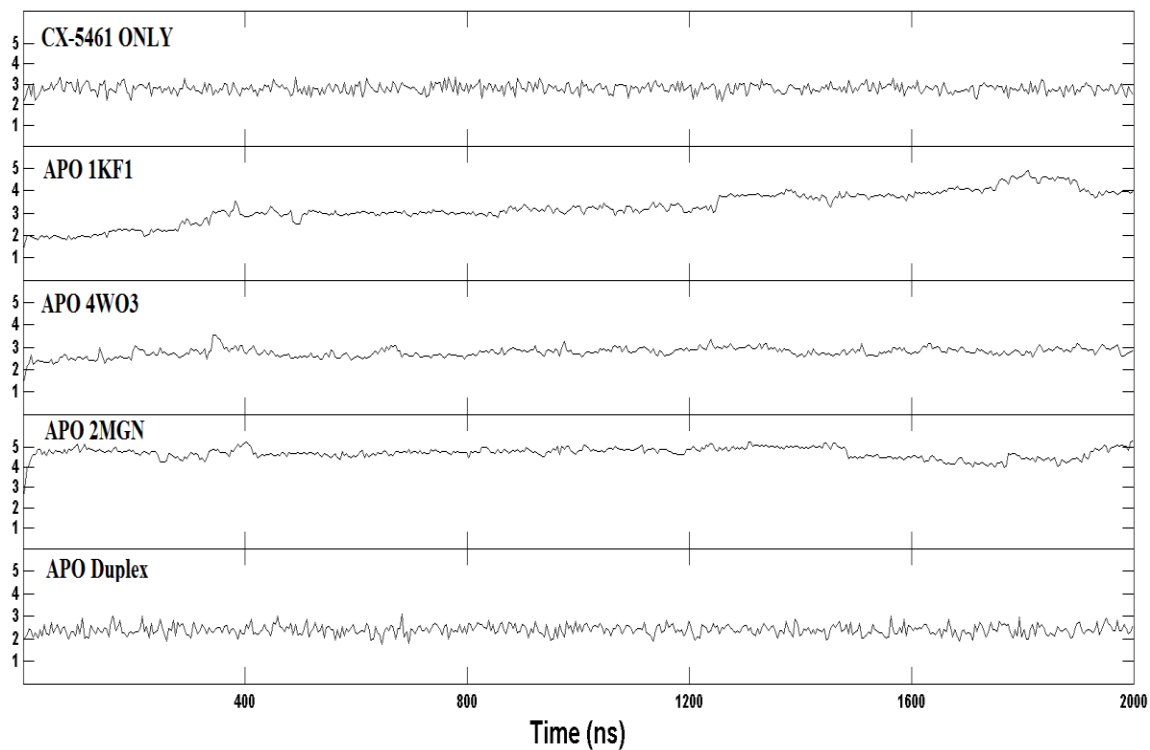
Figure B1. The average RMSD's of the CX-5461 only runs, and the human telomeric G-quadruplex (1KF1), c-KIT1 G-quadruplex (4WO3), c-Myc G-quadruplex (2MGN), and duplex DNA only runs.

187

System	Initial	Run 1	Run 2
h-Tel			
c-KIT1			
c-Myc			
Duplex			
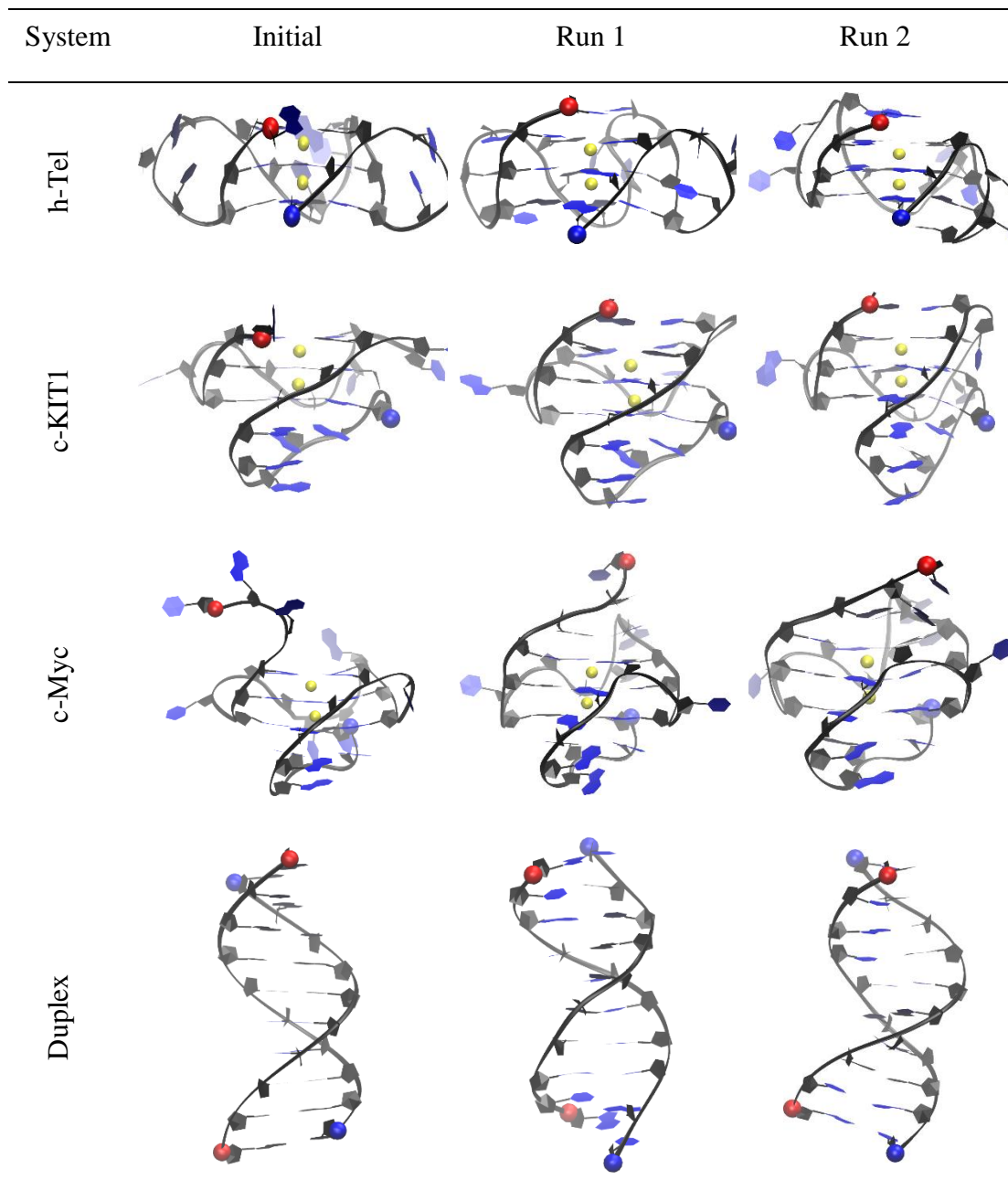
Figure B2. Last snap shots of the receptor only simulation runs.

A

02 ns   713 ns   1573 ns   2000 ns

B

01 ns   774 ns   1353 ns   1996 ns

C

03 ns   862 ns   1415 ns   2000 ns

189
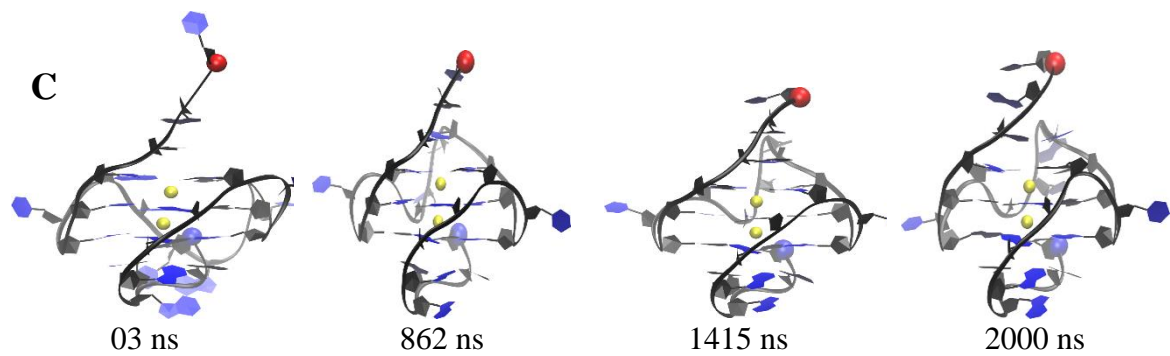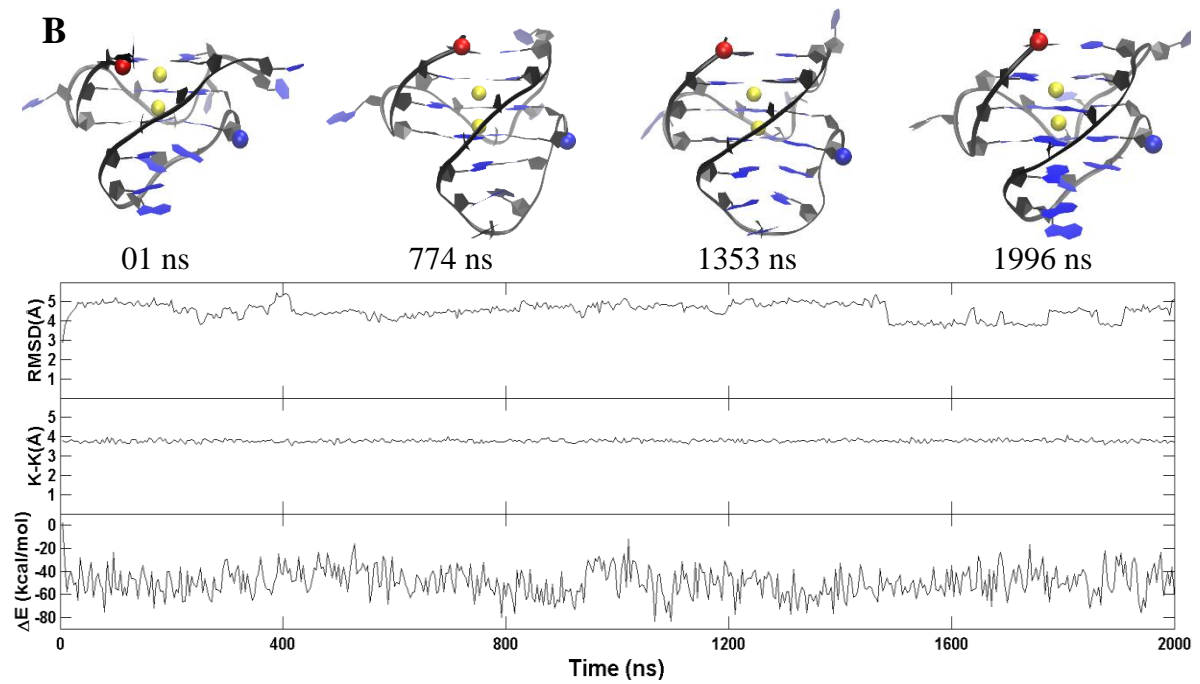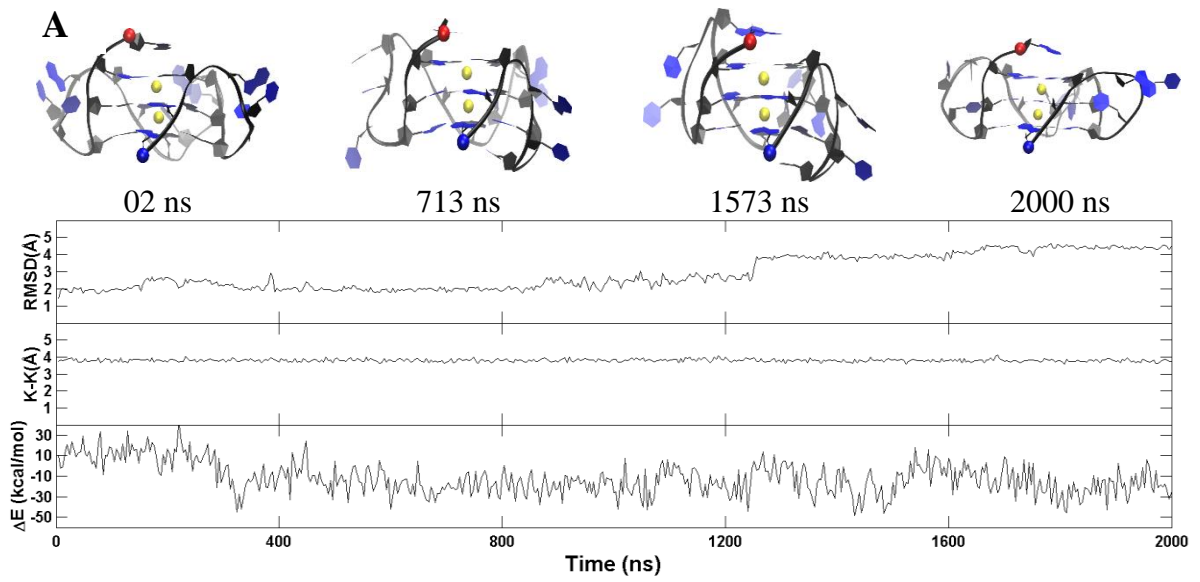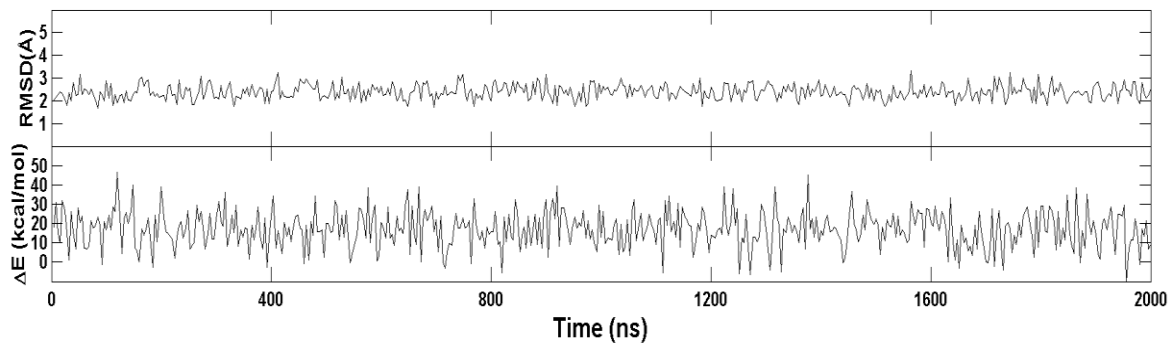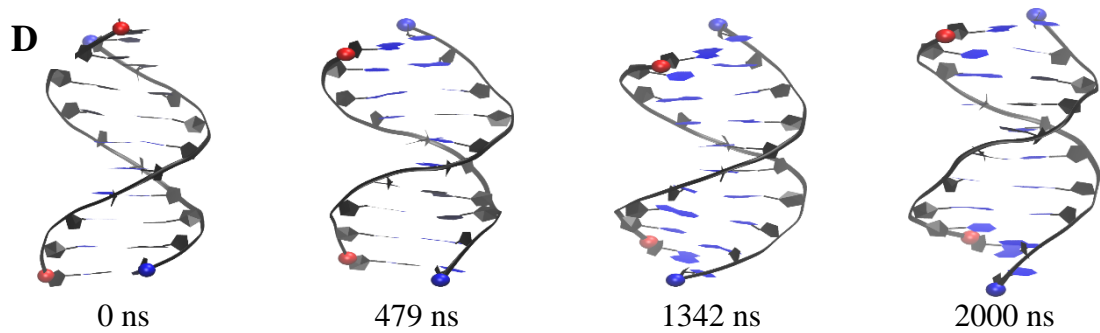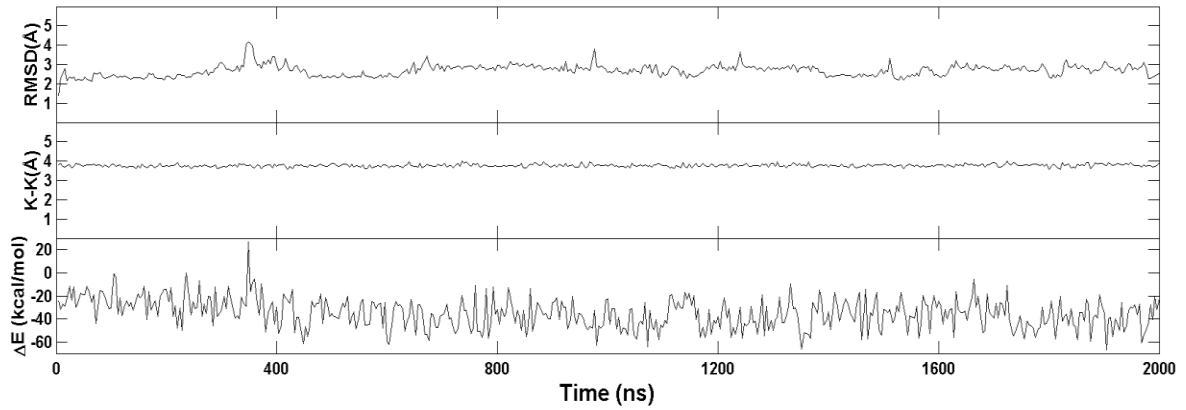
Figure B3. A representative trajectory of the human telomeric(A), c-KIT1(B), c-Myc(C), and duplex(D) DNA only simulations, and the order parameter plot illustrating the RMSD (Å) with reference to the final structure, K+ to K+ distance, and the MM-GBSA binding energy(ΔE in kcal/mol). 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by a yellow ball.
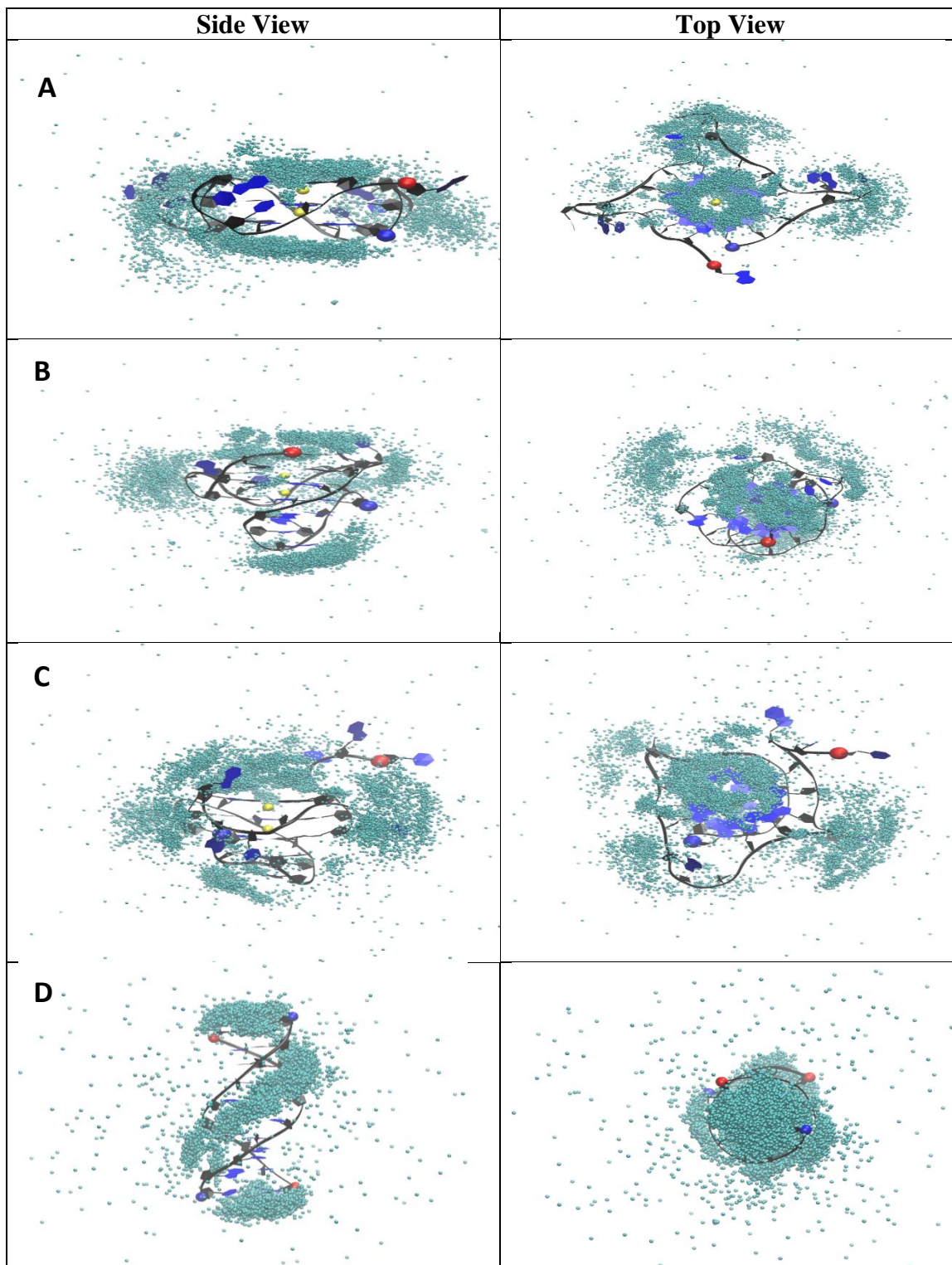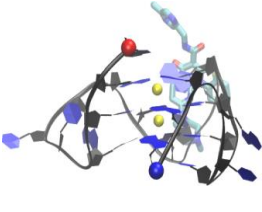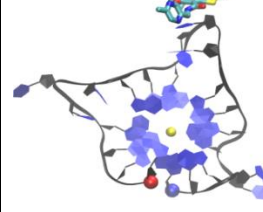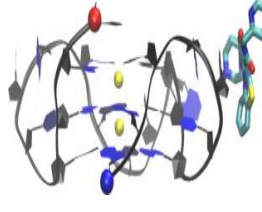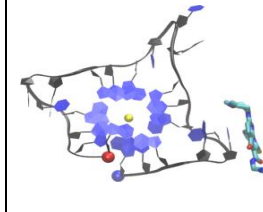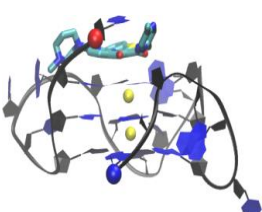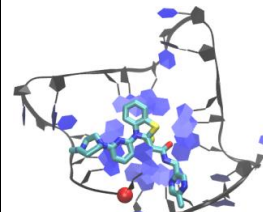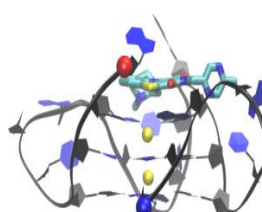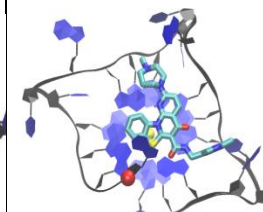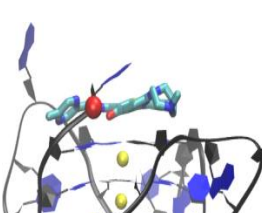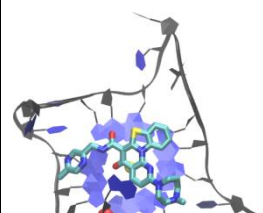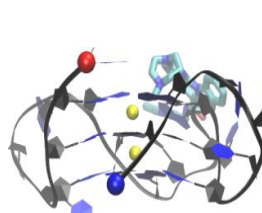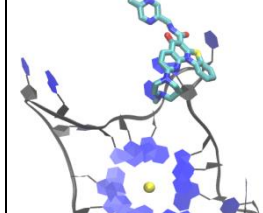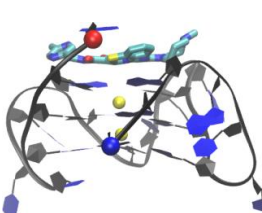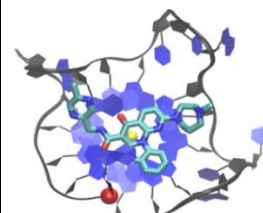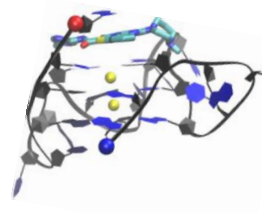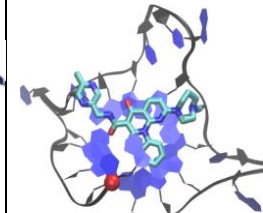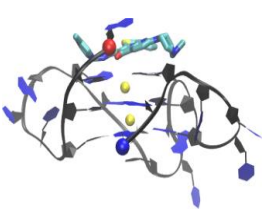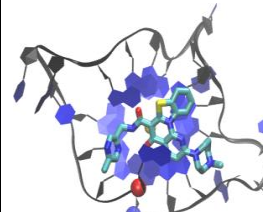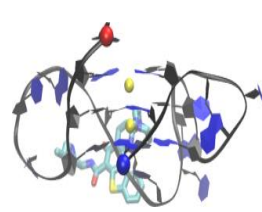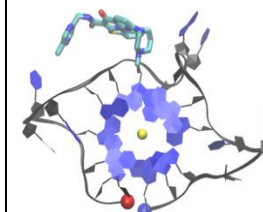
Figure B4. Position plot of CX-5461 in each of the 30 trajectories for the (A) human telomeric complex, (B) c-KIT1 complex, (C) c-Myc complex and (D) DNA duplex systems

191
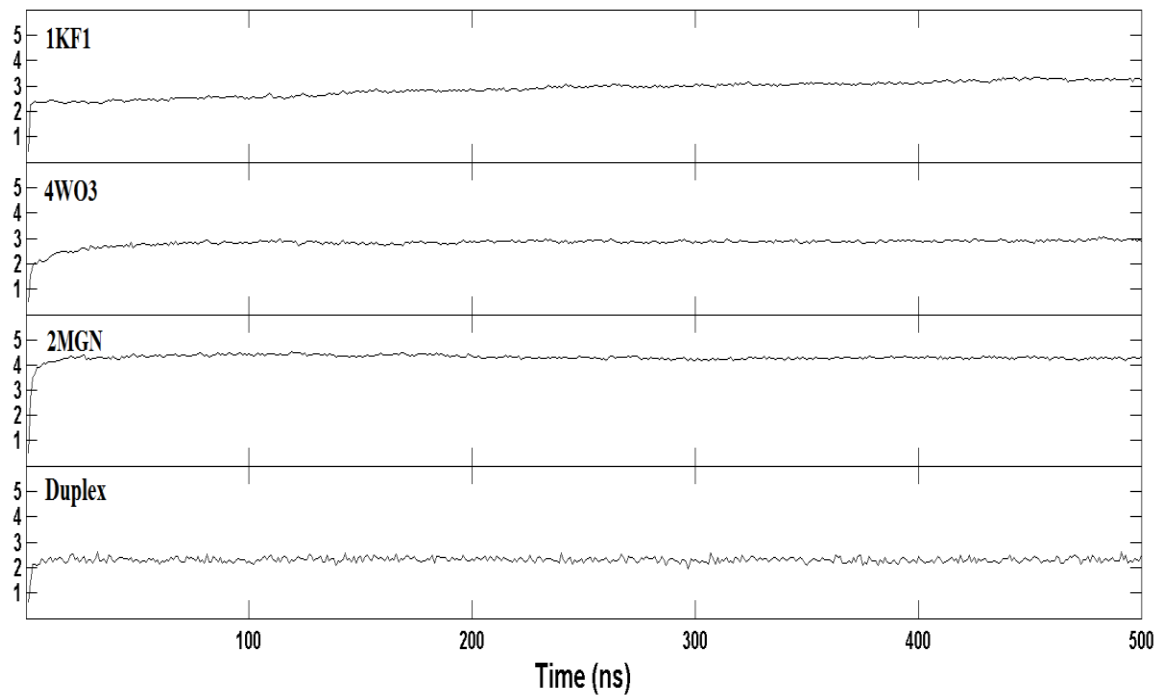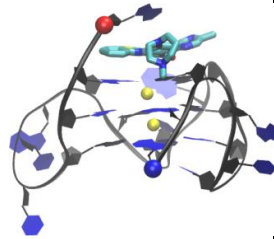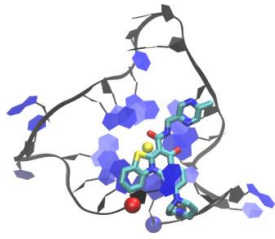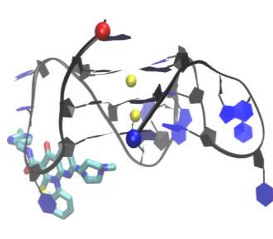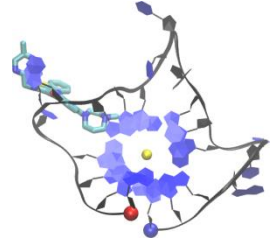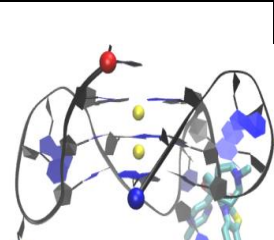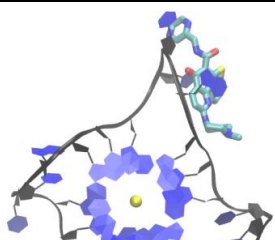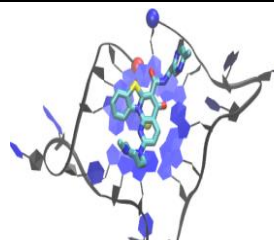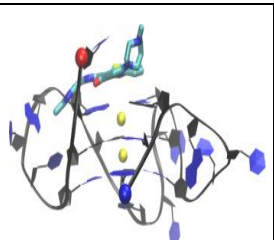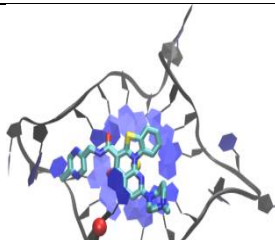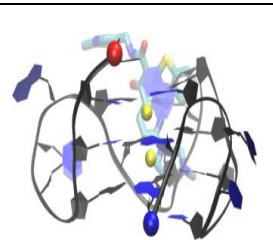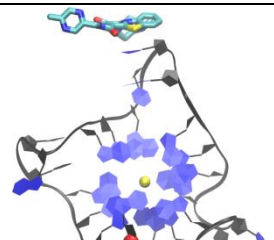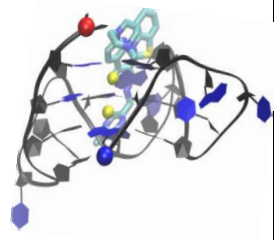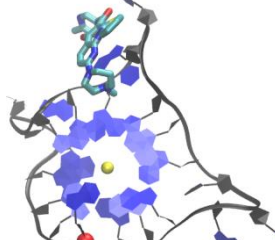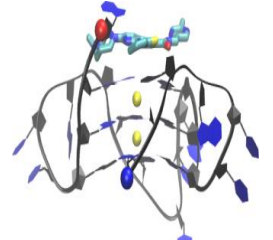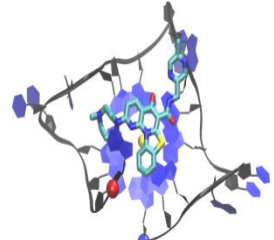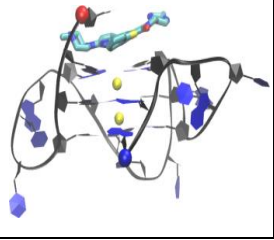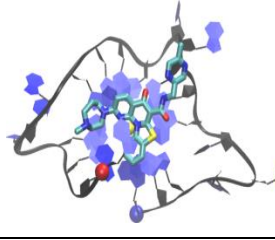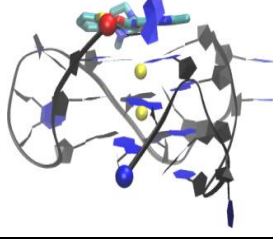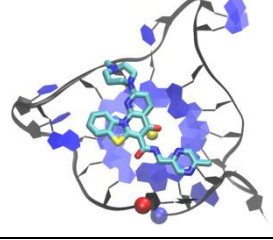
Figure B5. The average RMSD's of the human telomeric/CX-5641, c-KIT1/CX-5641, c-Myc/CX5461, and duplex/CX-5461 complex systems in all runs.
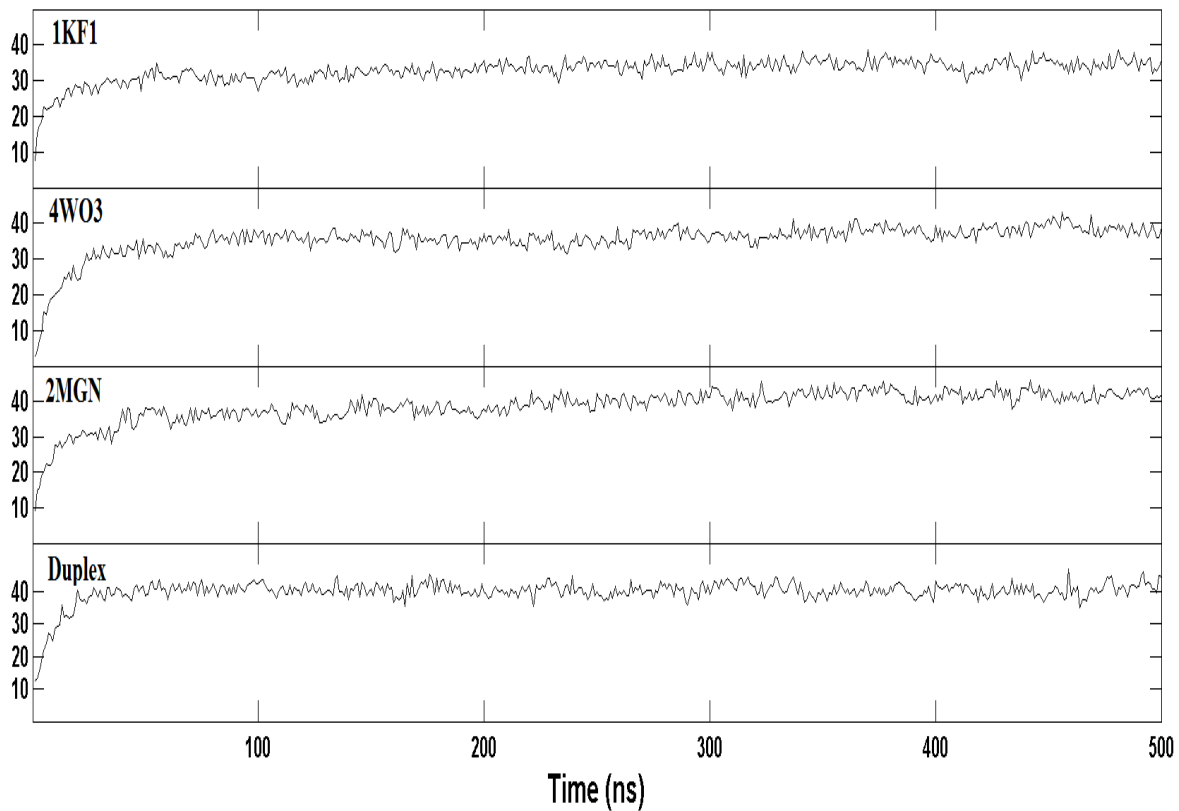
Figure B6. The average contact number between CX-5461 and the human telomeric G-quadruplex, c-KIT1 G-quadruplex, c-Myc G-Quadruplex, and DNA duplex in all runs.

193

Front View	Top/Bottom View	Front View	Top/Bottom View
Run 01		Run 06	
Run 02		Run 07	
Run 03		Run 08	
Run 04		Run 09	
Run 05		Run 10	

194

Front View	Top/Bottom View	Front View	Top/Bottom View
Run 11		Run 16	
Run 12		Run 17	
Run 13		Run 18	
Run 14		Run 19	
Run 15		Run 20	

195

Front View	Top/Bottom View	Front View	Top/Bottom View
Run 21		Run 26	
Run 22		Run 27	
Run 23		Run 28	
Run 24		Run 29	
Run 25		Run 30	
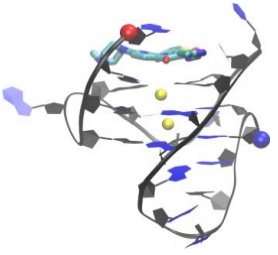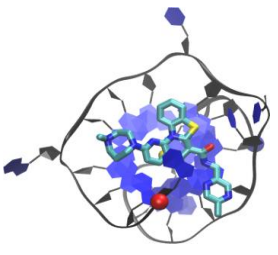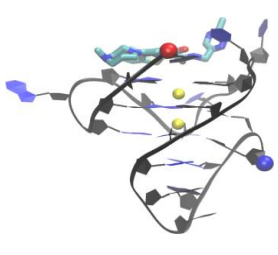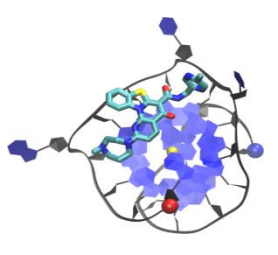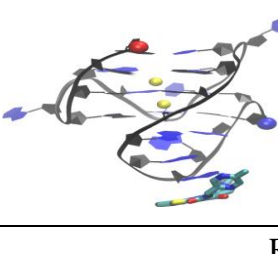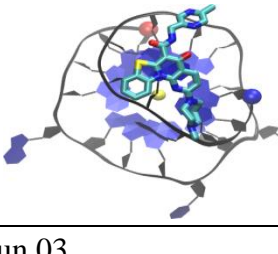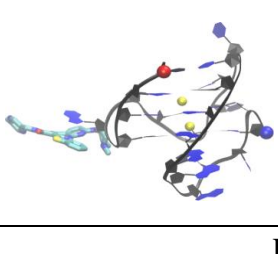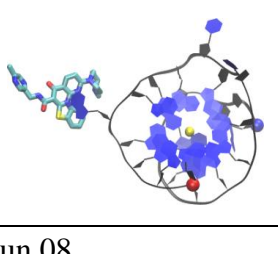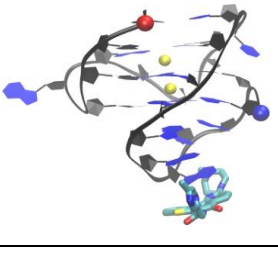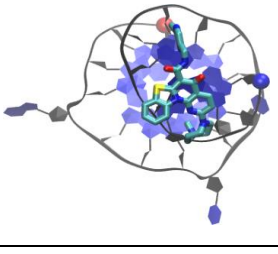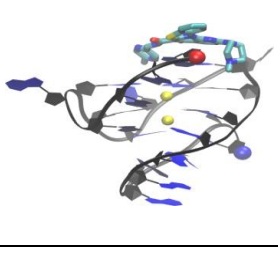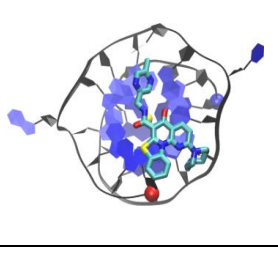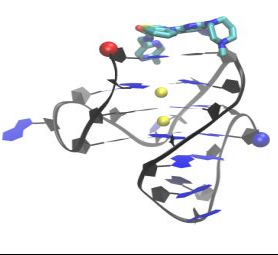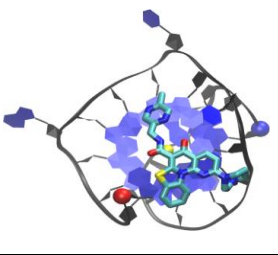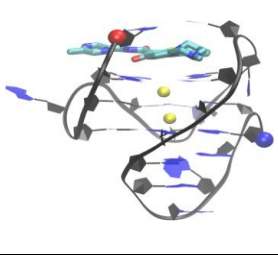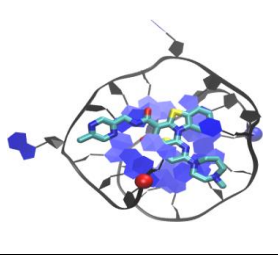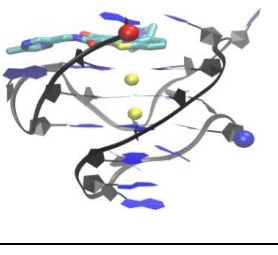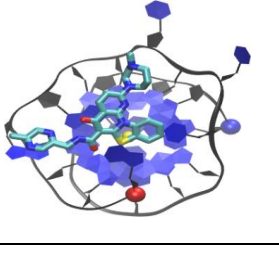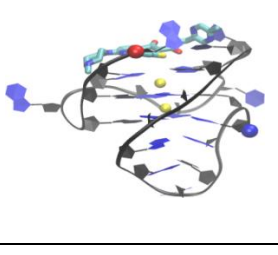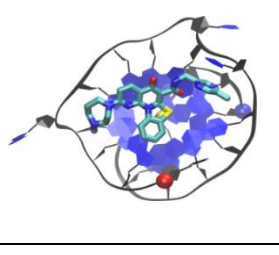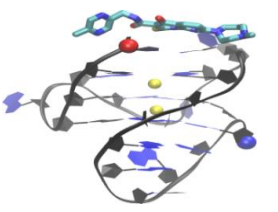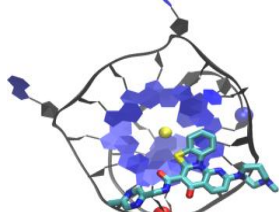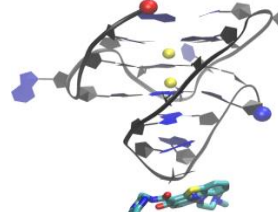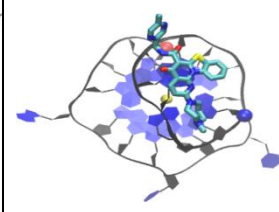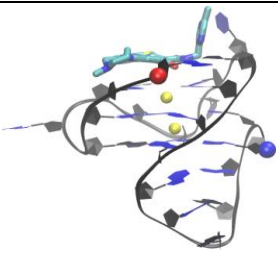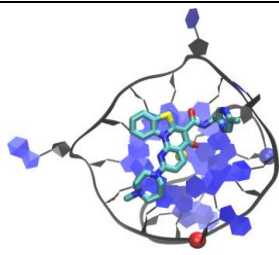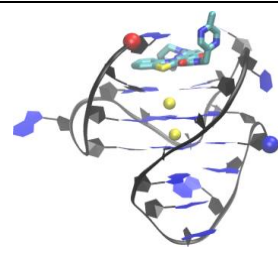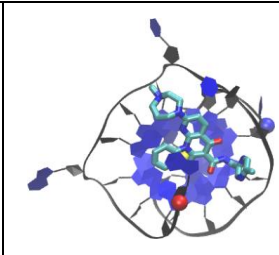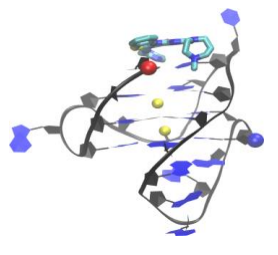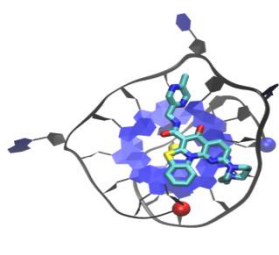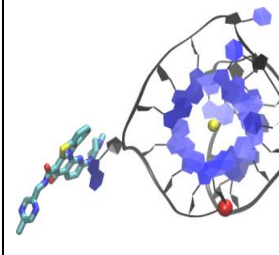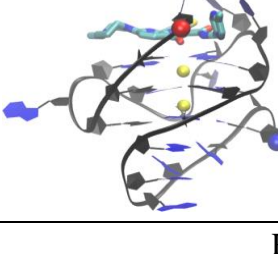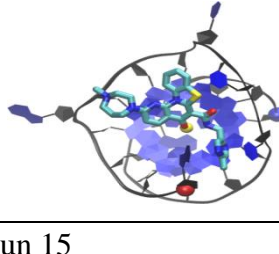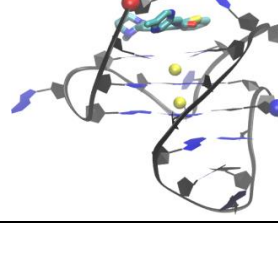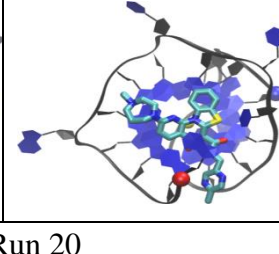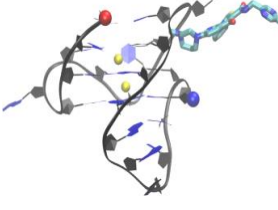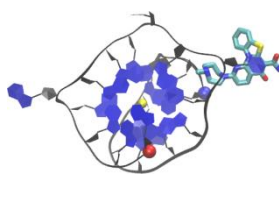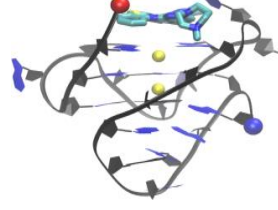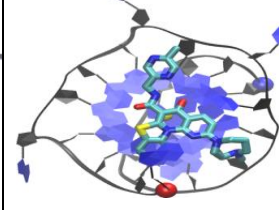
Figure B7. The last snap shots of CX-5461 binding to the human telomeric DNA G-Quadruplex in binding simulations 1-30.

Front View	Top/Bottom View	Front View	Top/Bottom View
Run 01		Run 06	
Run 02		Run 07	
Run 03		Run 08	
Run 04		Run 09	
Run 05		Run 10	

197

Front View	Top/Bottom View	Front View	Top/Bottom View
Run 11		Run 16	
Run 12		Run 17	
Run 13		Run 18	
Run 14		Run 19	
Run 15		Run 20	

198

Front View	Top/Bottom View	Front View	Top/Bottom View
Run 21		Run 26	
Run 22		Run 27	
Run 23		Run 28	
Run 24		Run 29	
Run 25		Run 30	
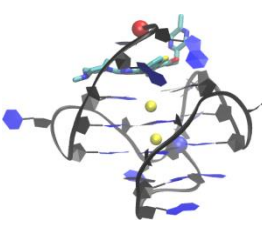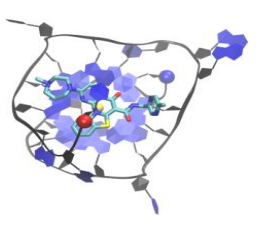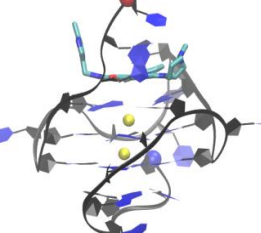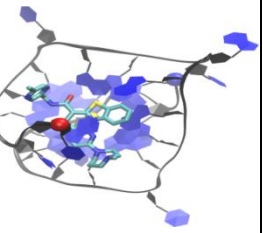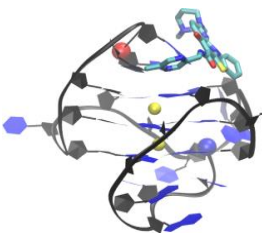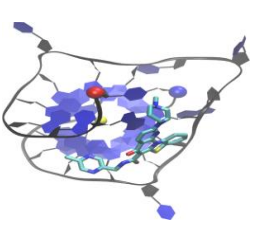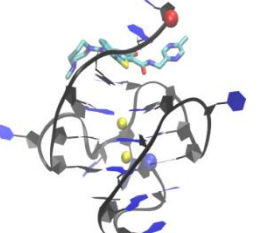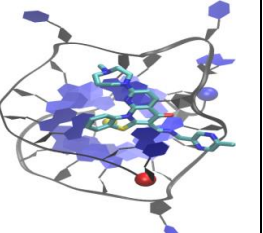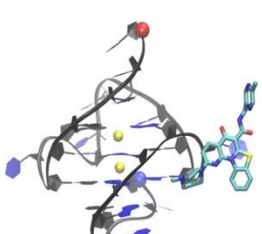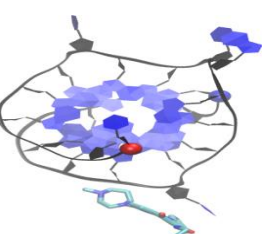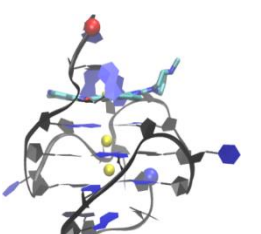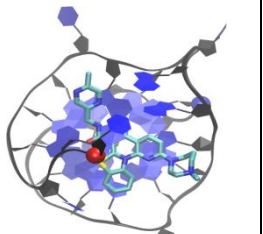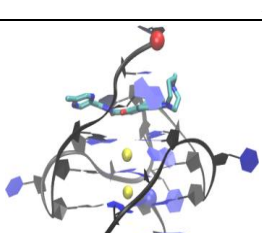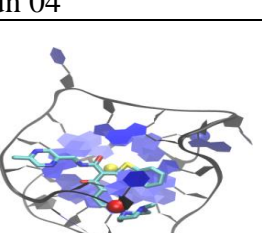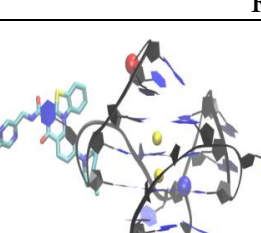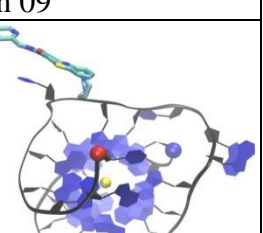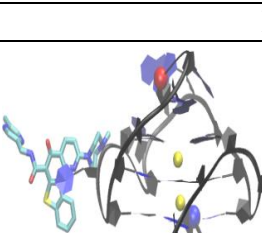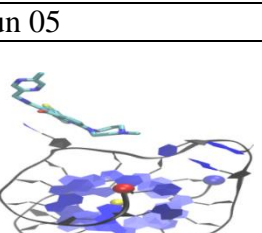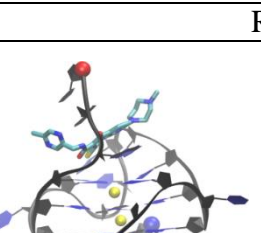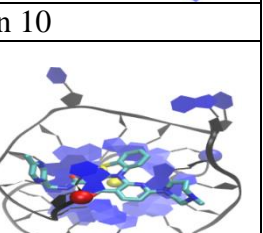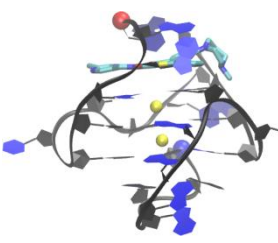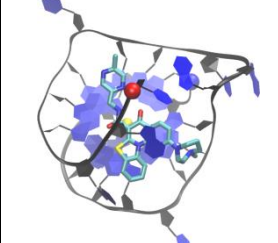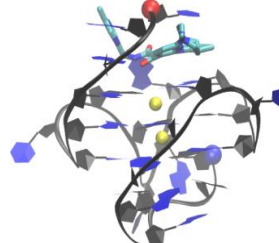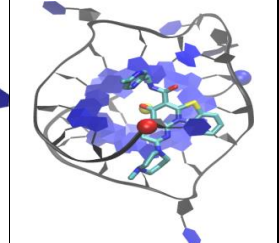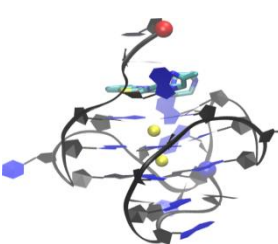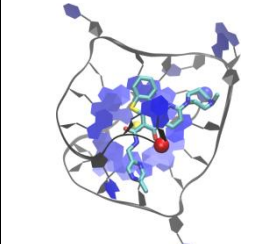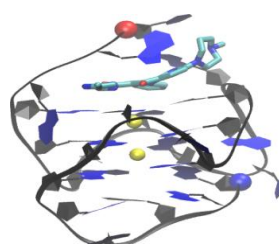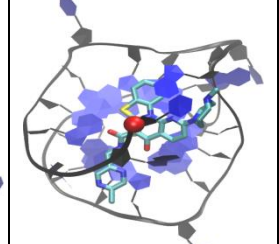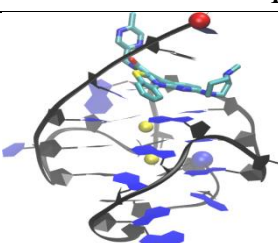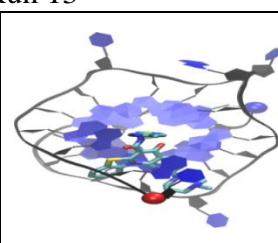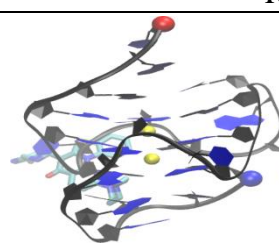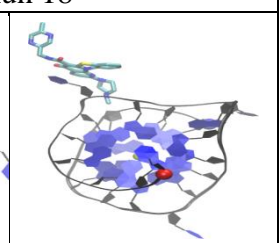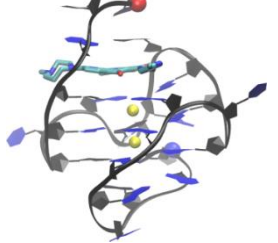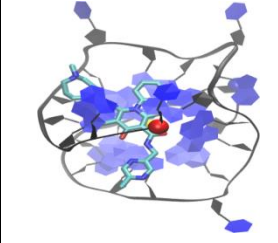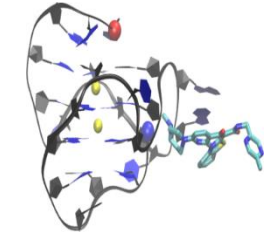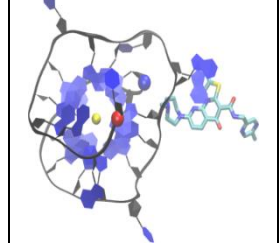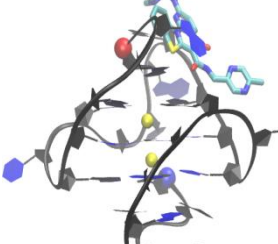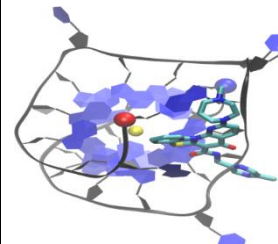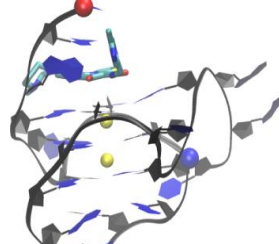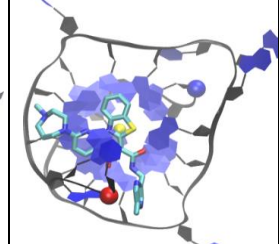
Figure B8. The last snap shots of CX-5461 binding to the c-KIT1 DNA G-Quadruplex in binding simulations 1-30.

Front View	Top/Bottom View	Front View	Top/Bottom View
Run 01		Run 06	
Run 02		Run 07	
Run 03		Run 08	
Run 04		Run 09	
Run 05		Run 10	

200

Front View	Top/Bottom View	Front View	Top/Bottom View
Run 11		Run 16	
Run 12		Run 17	
Run 13		Run 18	
Run 14		Run 19	
Run 15		Run 20	

201

Front View	Top/Bottom View	Front View	Top/Bottom View
Run 21		Run 26	
Run 22		Run 27	
Run 23		Run 28	
Run 24		Run 29	
Run 25		Run 30	
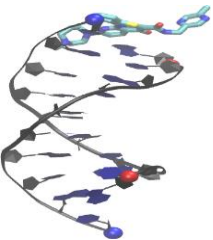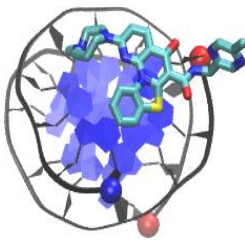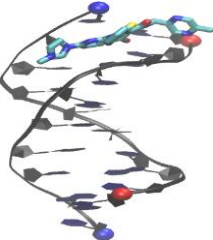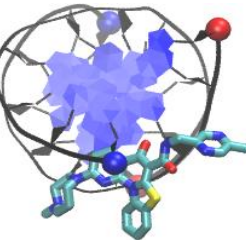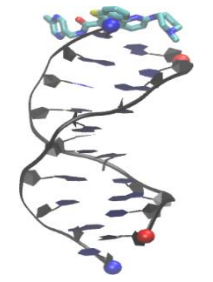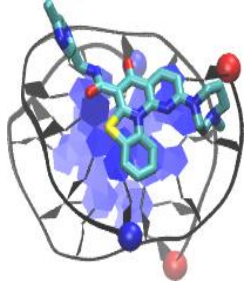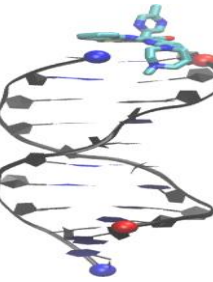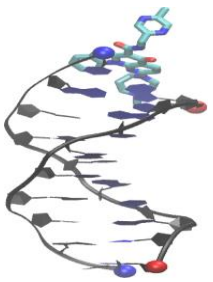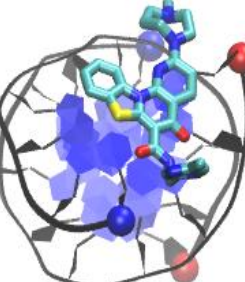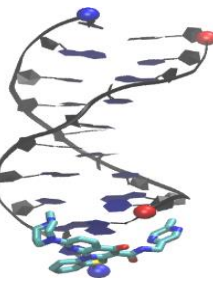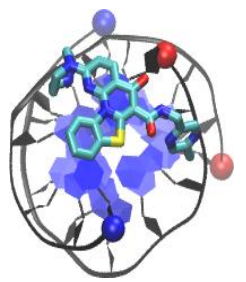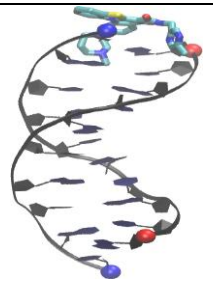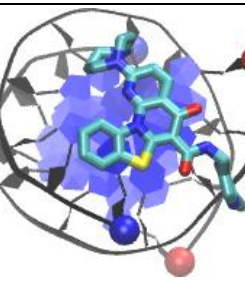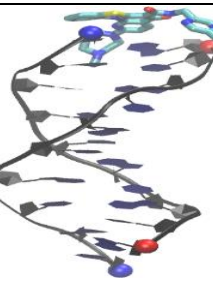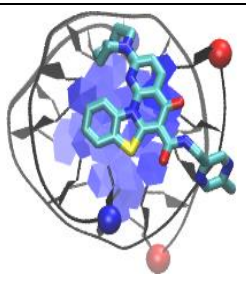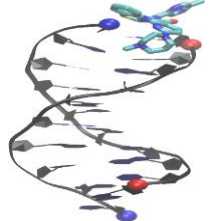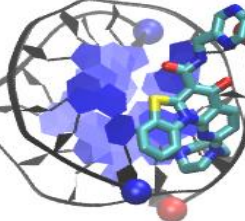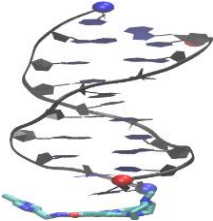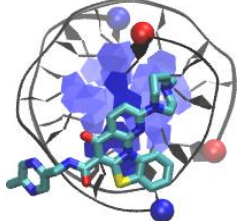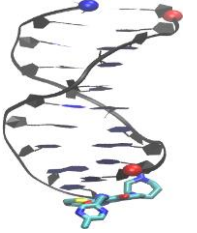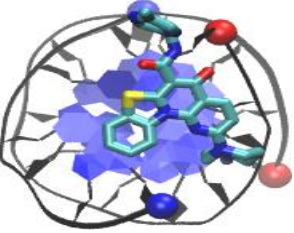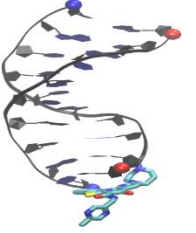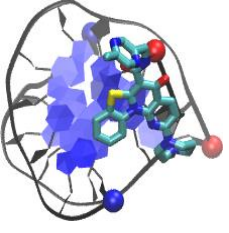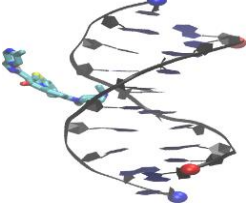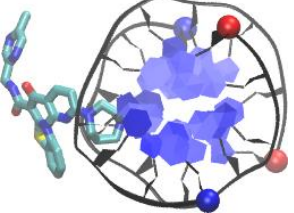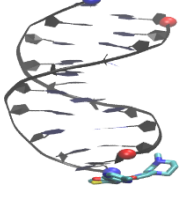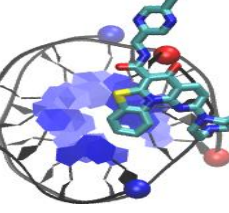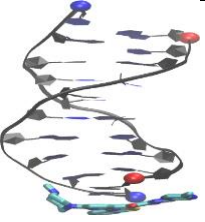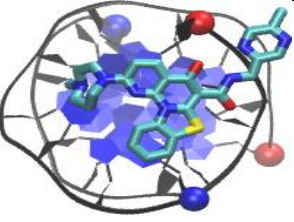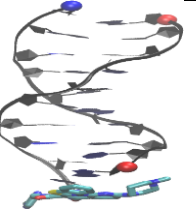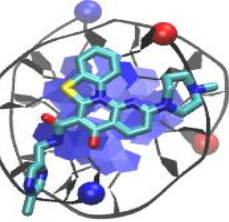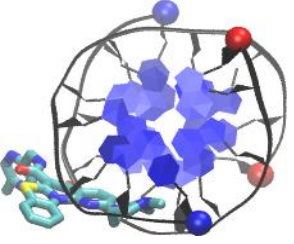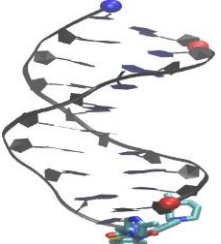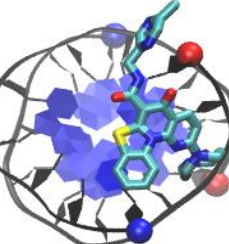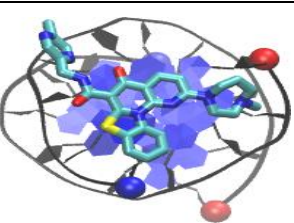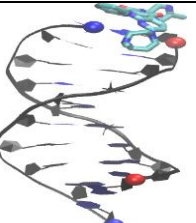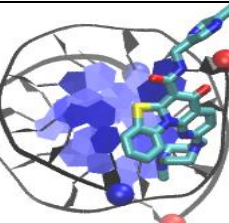
Figure B9. The last snap shots of CX-5461 binding to the c-Myc DNA G-Quadruplex in binding simulations 1-30.

202

Front View	Top/Bottom View	Front View	Top/Bottom View
Run 01		Run 06	
Run 02		Run 07	
Run 03		Run 08	
Run 04		Run 09	
Run 05		Run 10	

203

Front View	Top/Bottom View	Front View	Top/Bottom View
Run 11		Run 16	
Run 12		Run 17	
Run 13		Run 18	
Run 14		Run 19	
Run 15		Run 20	

Front View	Top/Bottom View	Front View	Top/Bottom View
Run 21		Run 26	
Run 22		Run 27	
Run 23		Run 28	
Run 24		Run 29	
Run 25		Run 30	

Figure B10. The last snap shots of CX-5461 binding to the Duplex DNA in binding simulations 1-30.

205

A	2D Interactions	3D Side View	3D Top View
Top Mode			
Bottom Mode			
Groove Mode			

Charged (negative)
Charged (positive)
Glycine
Hydrophobic
Metal

Polar
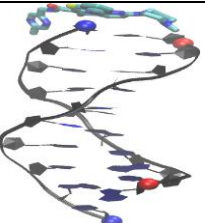Unspecified residue
Water
Hydration site
Hydration site (displaced)

Distance
H-bond
Halogen bond
Metal coordination
Pi-Pi stacking

Pi-cation
Salt bridge
Solvent expos

206

B	2D Interactions	3D Side View	3D Top View
Top Mode			
Bottom Mode			
Groove Mode			

Legend:

- ● Charged (negative)
- Glycine
- Hydrophobic
- Metal
- ● Polar
- Unspecified residue
- Water
- Hydration site
- ✕ Hydration site (displaced)
- ----- Distance
- H-bond
- Halogen bond
- Metal coordination
- ●—● Pi-Pi stacking
- ●— Pi-cation
- —— Salt bridge
- Solvent expos[ure]
- ● Charged (positive)

C	2D Interactions	3D Side View	3D Top View
Top Mode			
Bottom Mode			
Groove Mode			

Legend:
- Charged (negative)
- Charged (positive)
- Glycine
- Hydrophobic
- Metal
- Polar
- Unspecified residue
- Water
- X Hydration site
- X Hydration site (displaced)
- Distance
- H-bond
- Halogen bond
- Metal coordination
- Pi-Pi stacking
- Pi-cation
- Salt bridge
- Solvent exposure

208

D	2D Interactions	3D Side View	3D Top View
Top Mode			
Bottom Mode			
Groove Mode			

Legend:
- Charged (negative)
- Charged (positive)
- Glycine
- Hydrophobic
- Metal
- Polar
- Unspecified residue
- Water
- Hydration site
- Hydration site (displaced)
- Distance
- H-bond
- Halogen bond
- Metal coordination
- Pi-Pi stacking
- Pi-cation
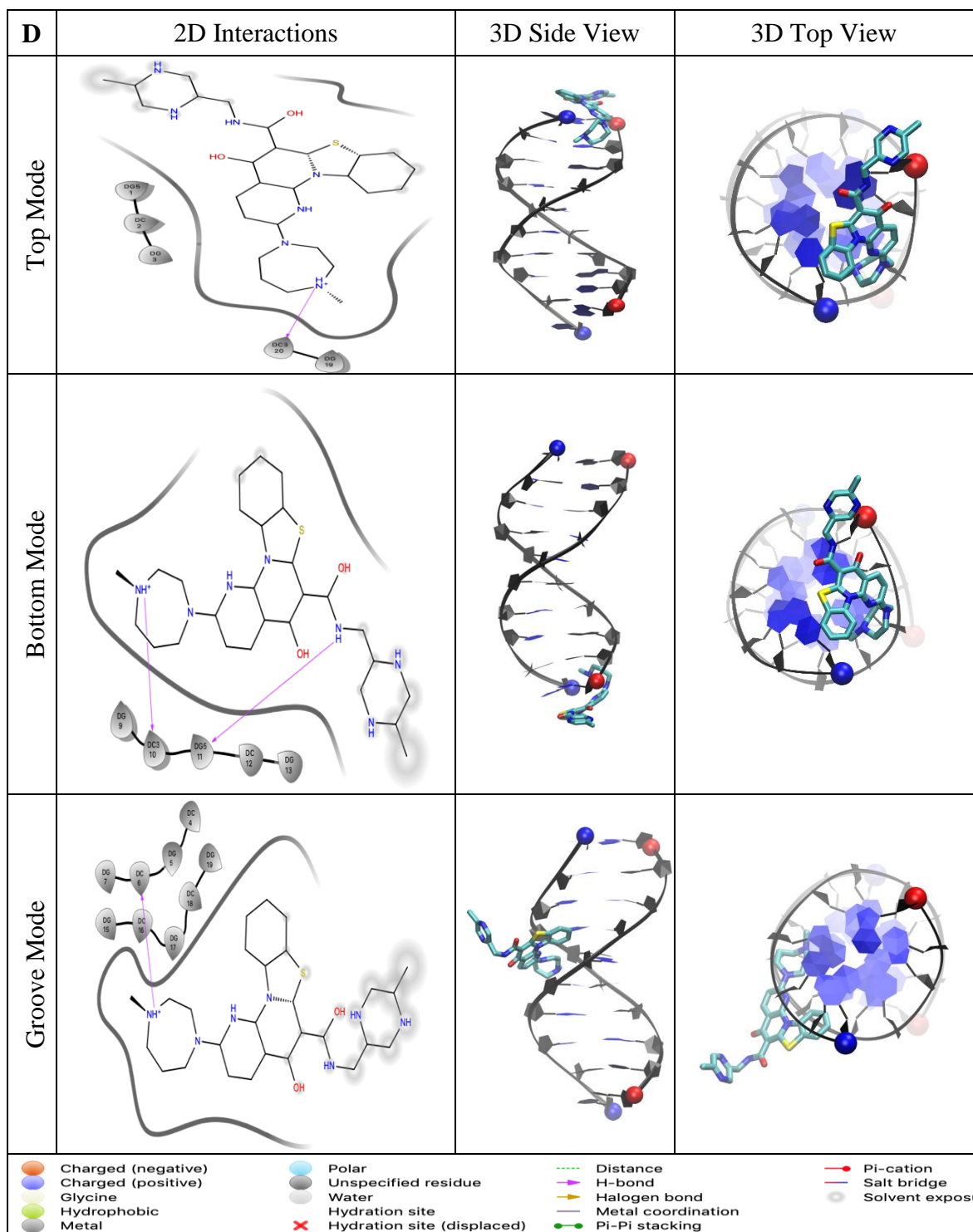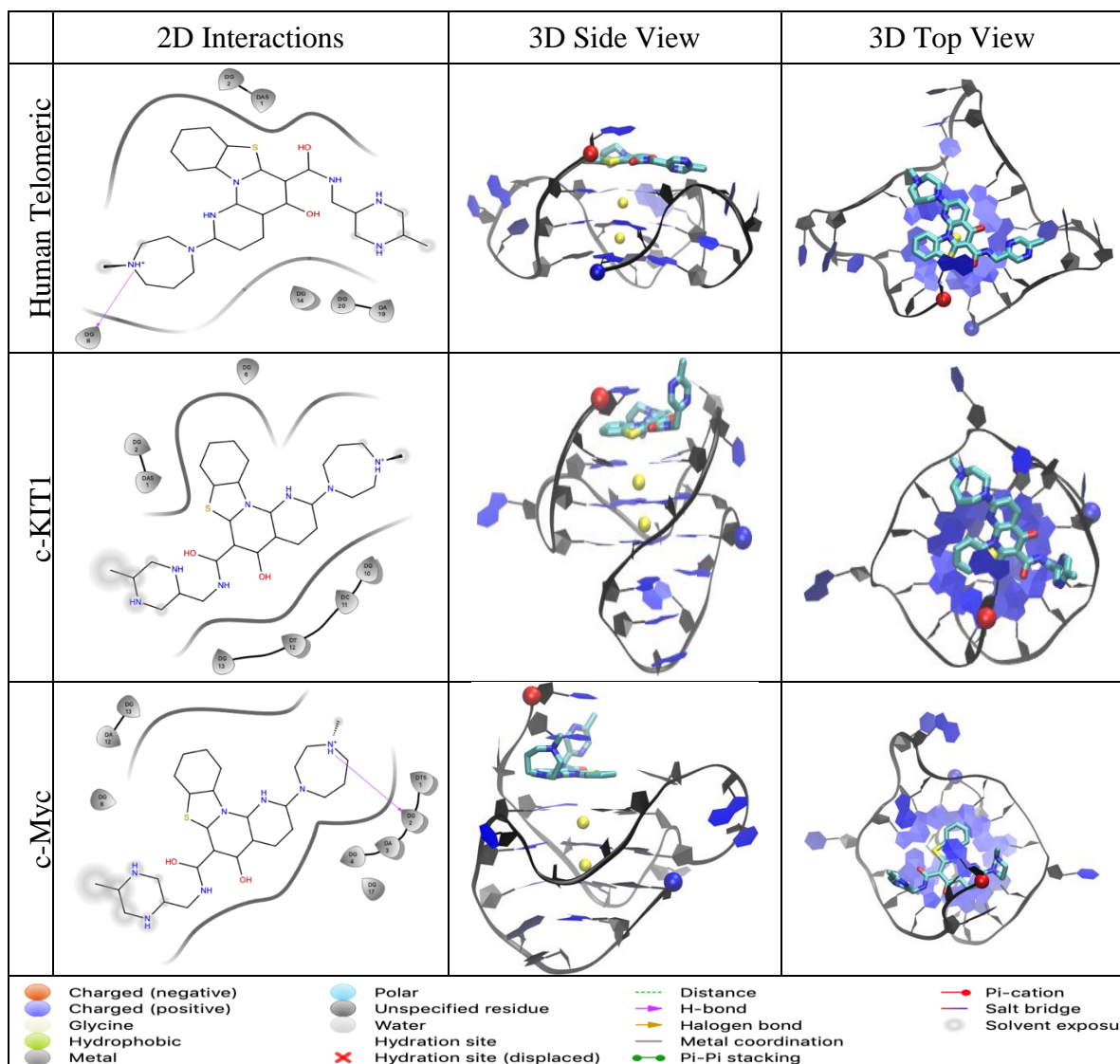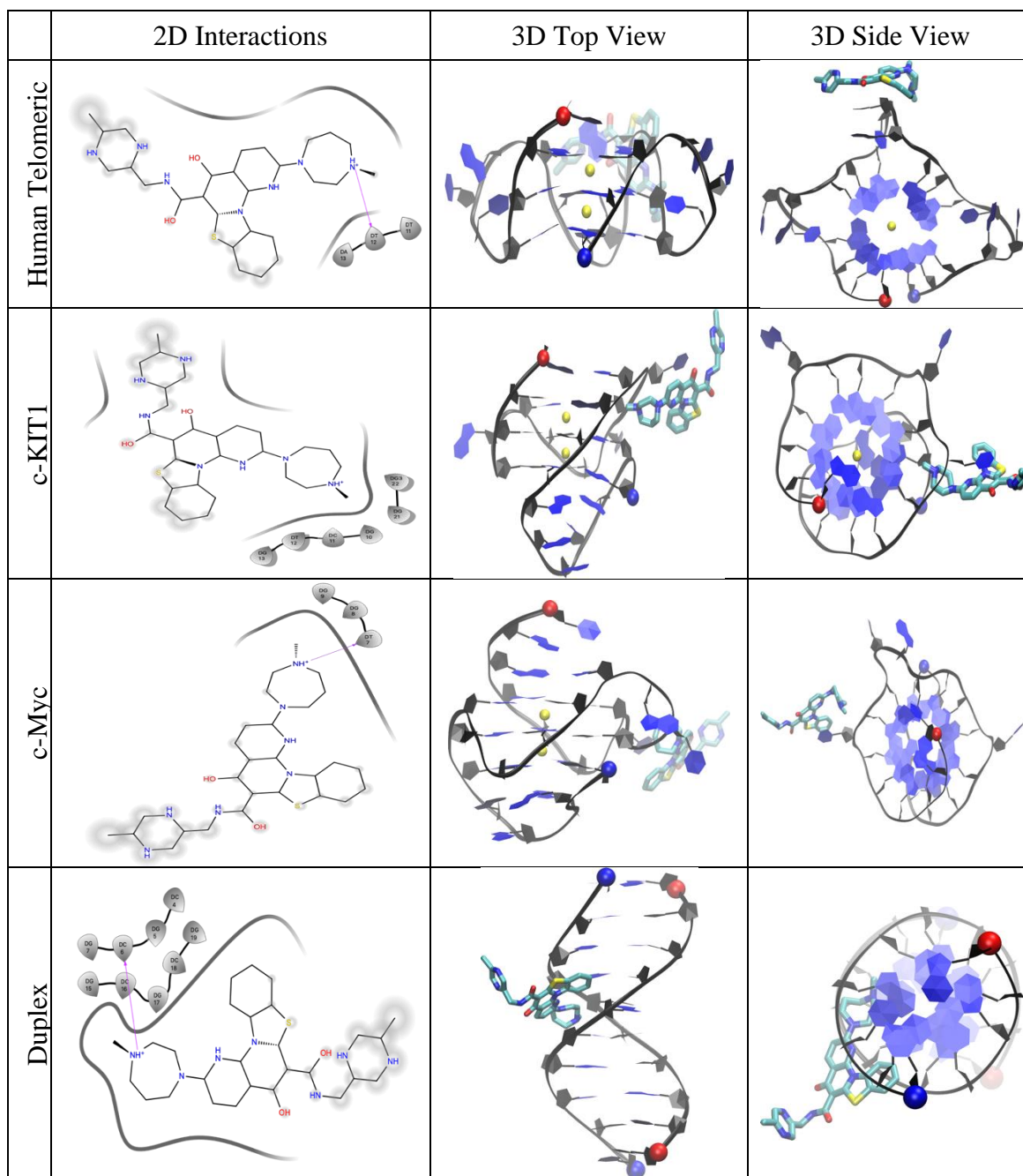- Salt bridge
- Solvent expose

Figure B11. Two dimensional interaction diagrams and three dimensional snapshots from the side and top view showing key interacting residues of the human telomeric (A), c-KIT1 (B), c-Myc (C) and Duplex (D) systems interacting with CX-5461 for the top, bottom and groove binding modes.

	2D Interactions	3D Side View	3D Top View
Human Telomeric			
c-KIT1			
c-Myc			

Legend:
- 🔴 Charged (negative)
- 🔵 Charged (positive)
- Glycine
- Hydrophobic
- Metal
- 🔵 Polar
- Unspecified residue
- Water
- Hydration site
- ✖ Hydration site (displaced)
- ----- Distance
- ➤ H-bond
- ➤ Halogen bond
- Metal coordination
- ●–● Pi–Pi stacking
- ●— Pi–cation
- — Salt bridge
- Solvent exposure

Figure B12. Comparison of the two dimensional interaction diagrams and three dimensional snapshots from the side and top views of the human telomeric, c-KIT1, c-Myc systems interacting with CX-5461 for the top binding mode.

210

Figure B13. Comparison of the two dimensional interaction diagrams and three dimensional snapshots from the side and top views of the human telomeric, c-KIT1, c-Myc, and duplex systems interacting with CX-5461 for the groove binding mode.
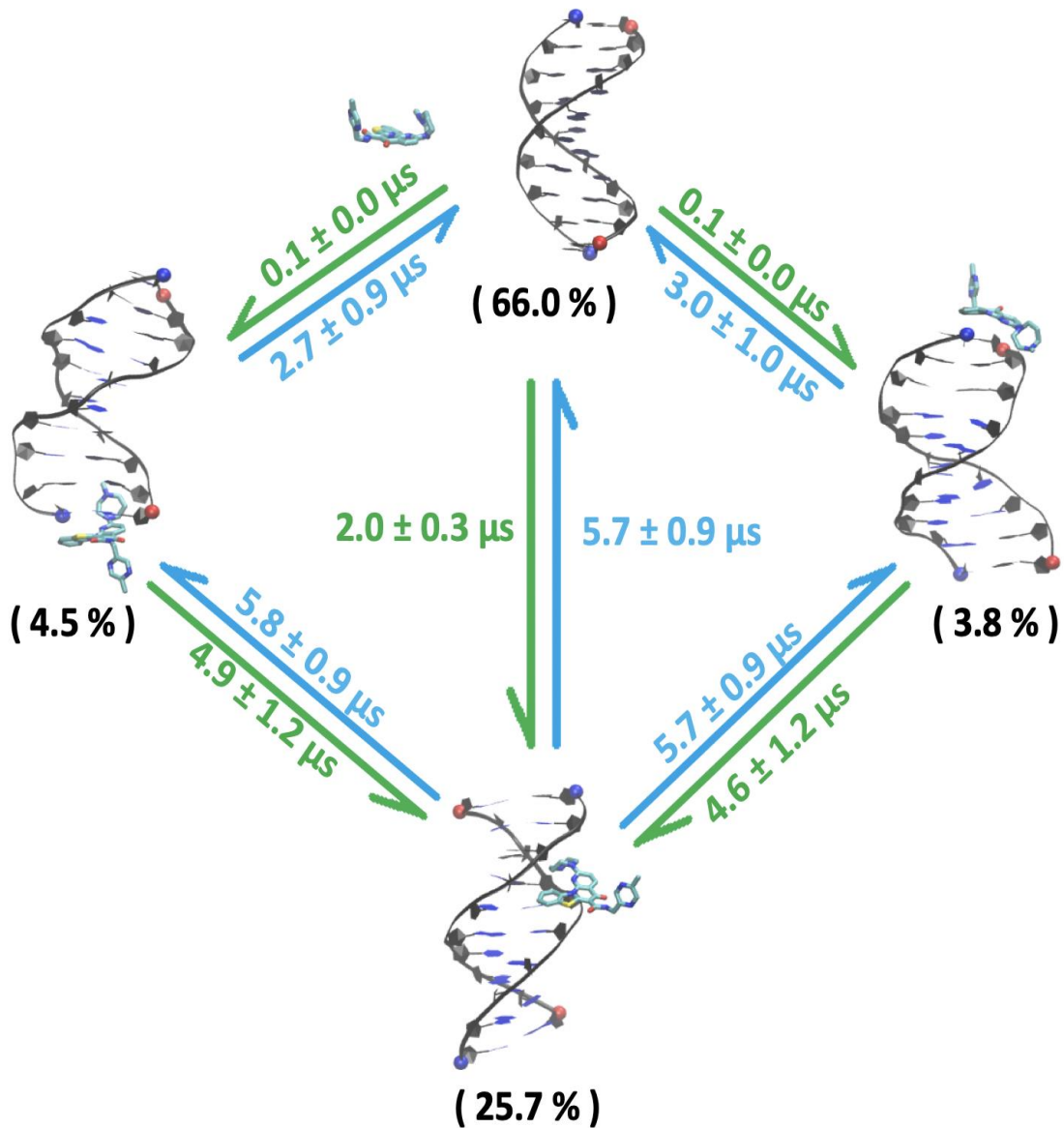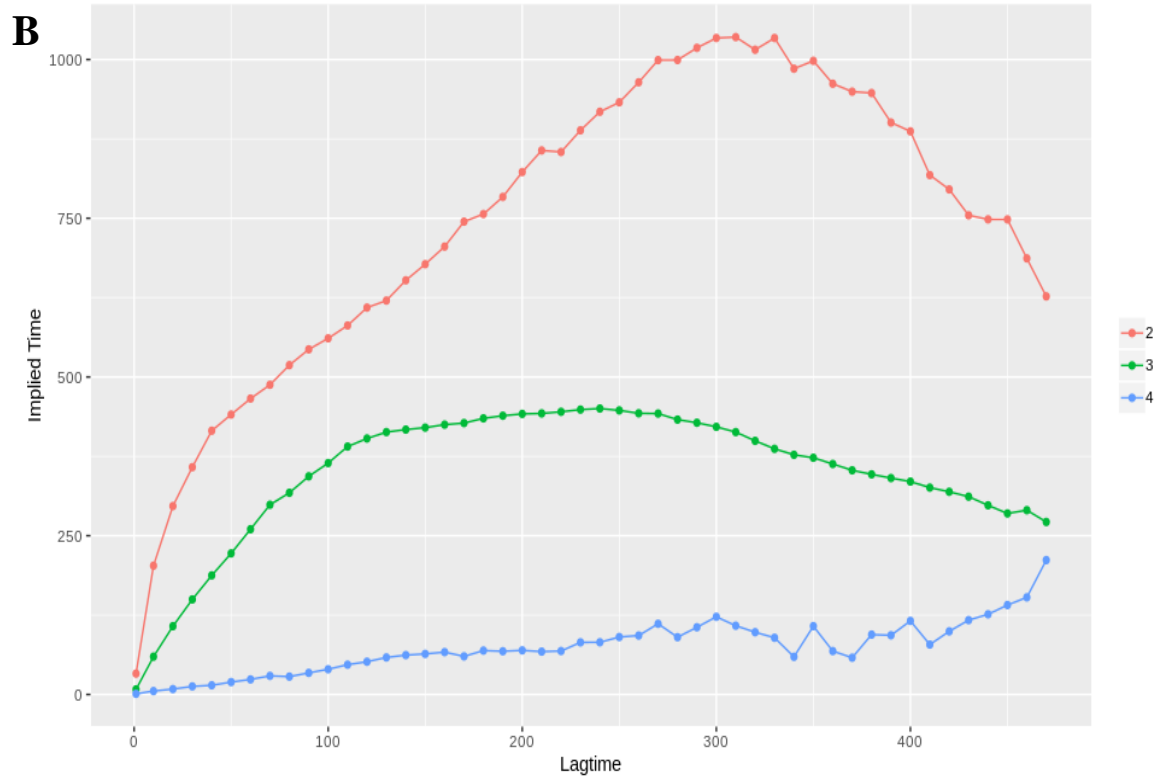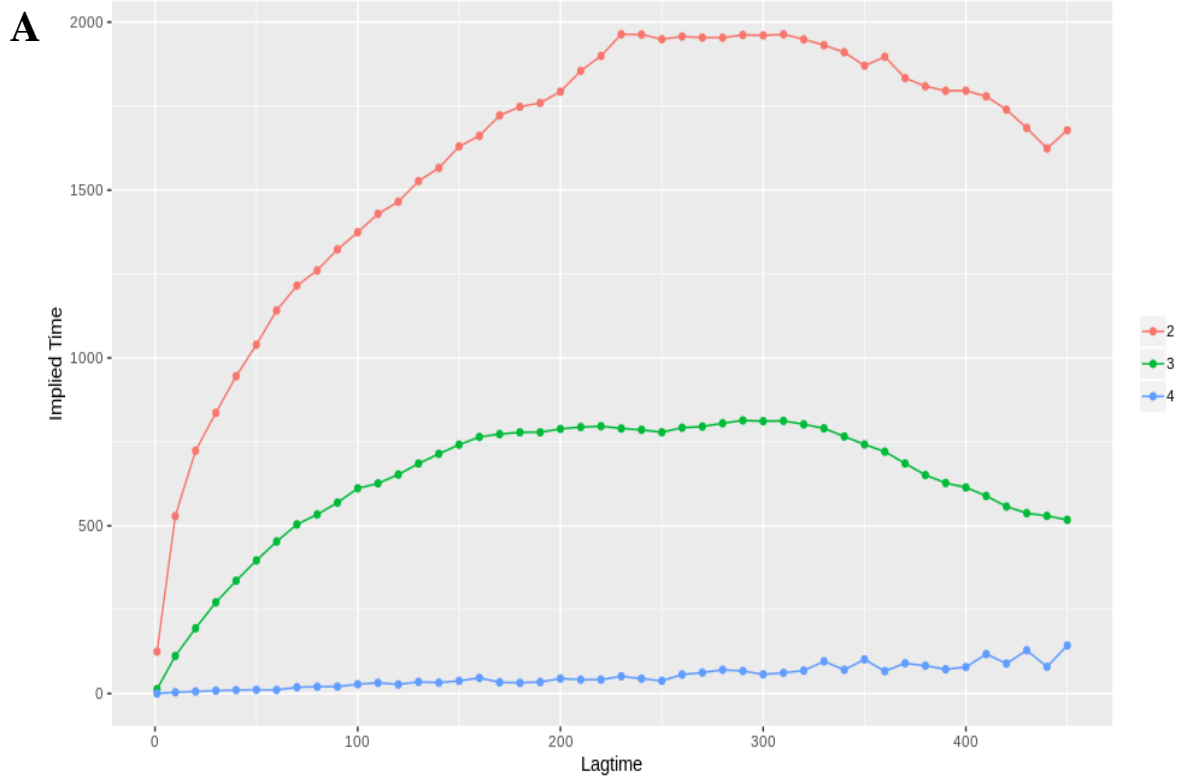
Figure B14. The mean first passage times between the four states (unbound, groove binding, and top and bottom terminal binding) of the DNA duplex and CX-5461 complex system.
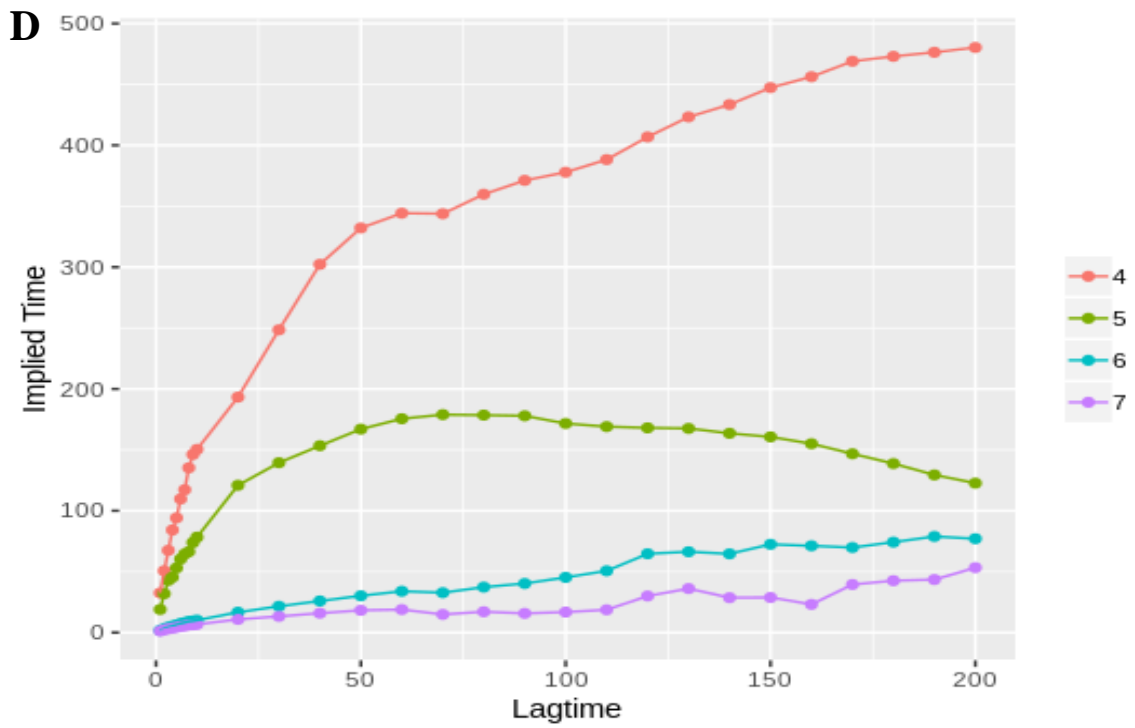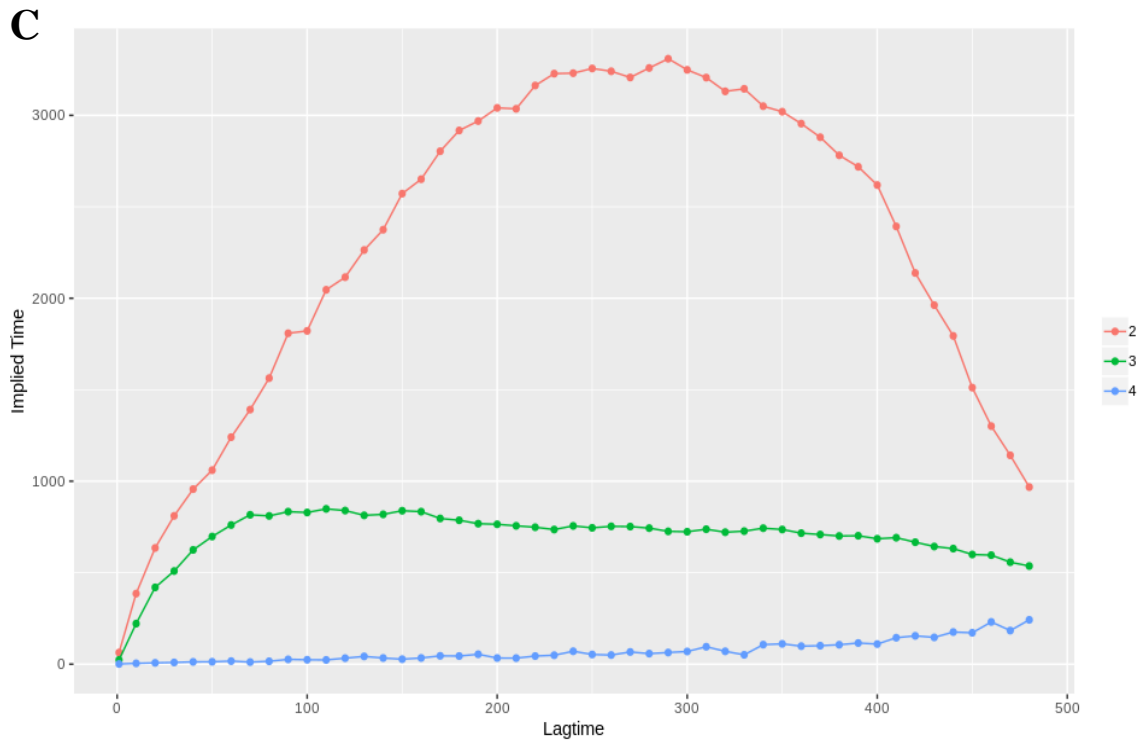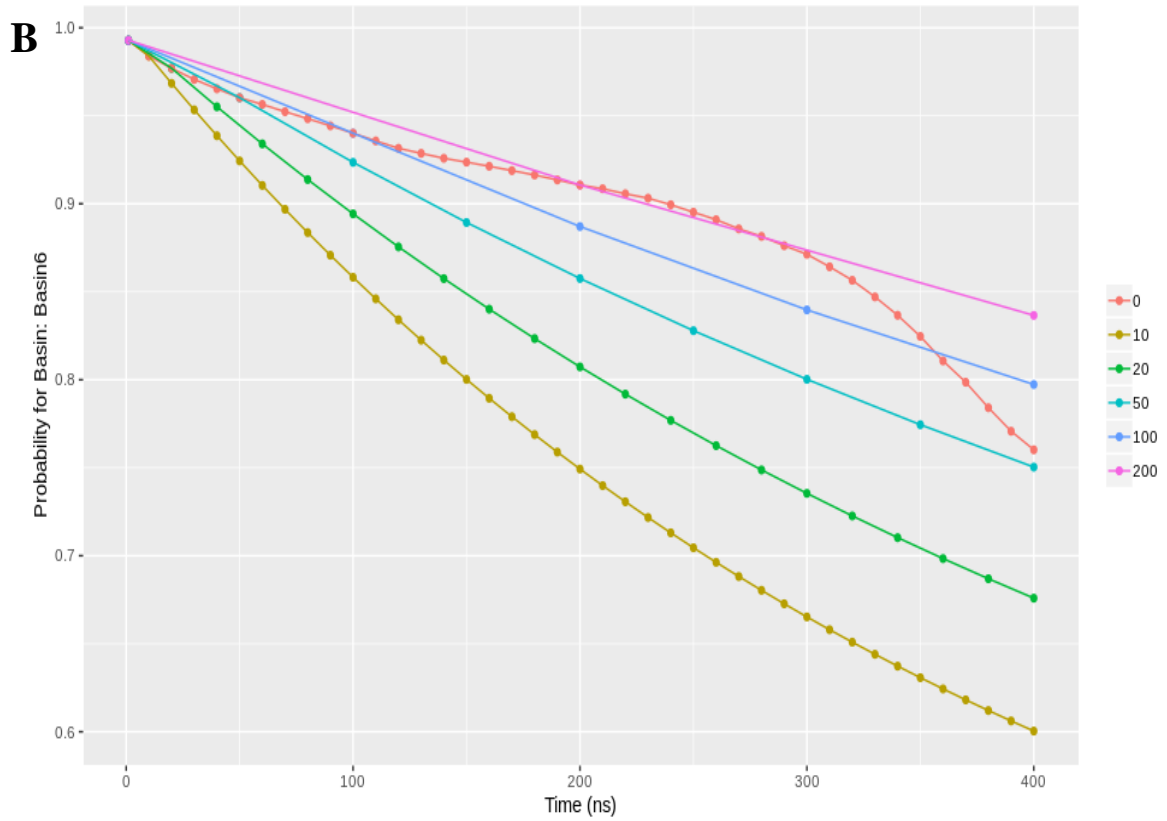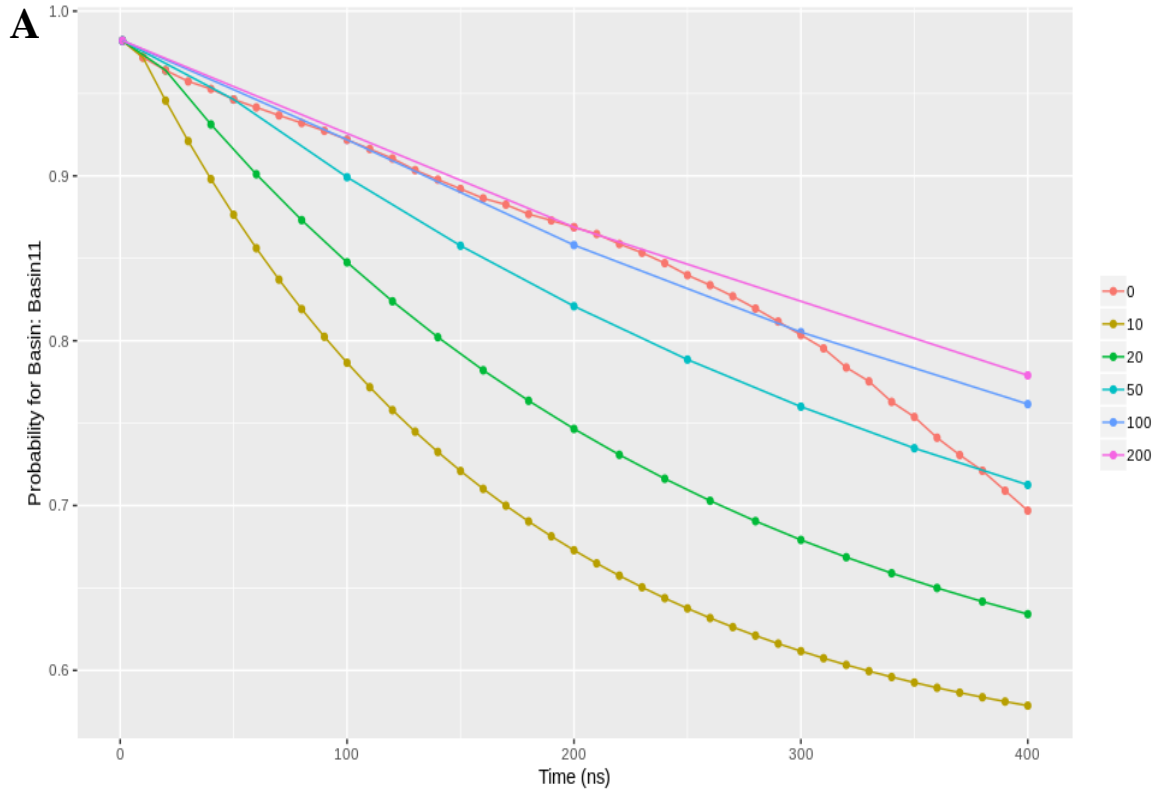
213

Figure B15. The implied timescale of each cluster for all lag times in the human telomeric (A), c-KIT1 (B), c-MYC (C), and duplex (D) complex systems.
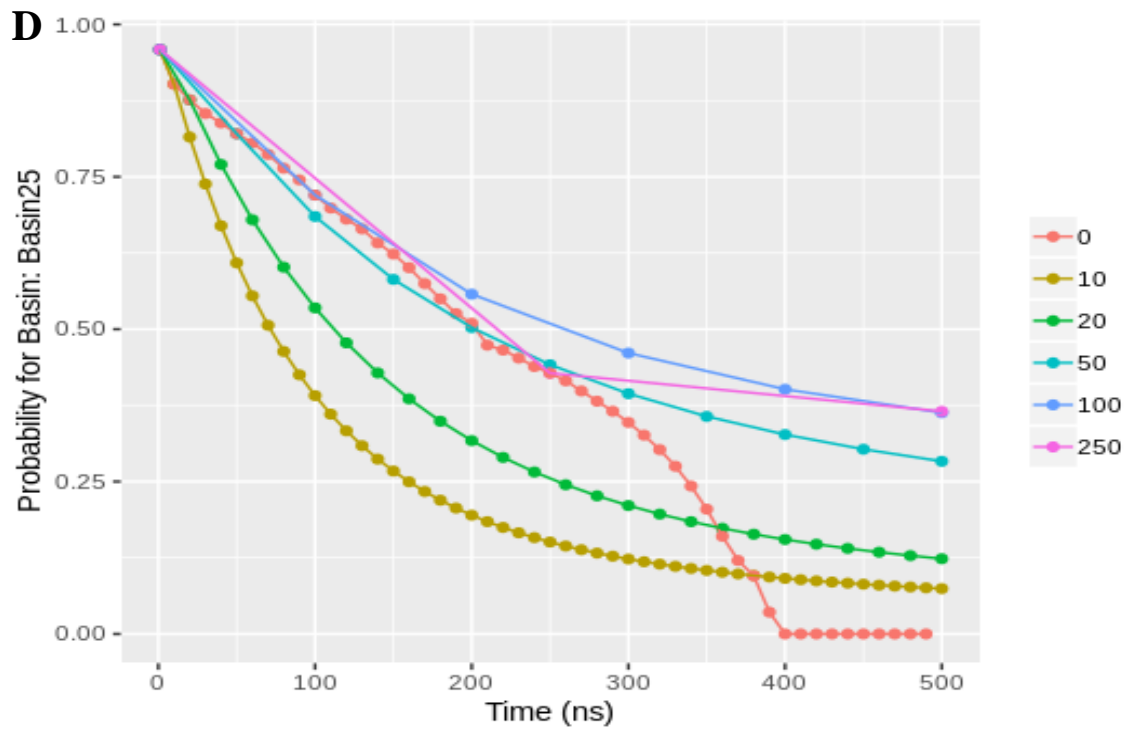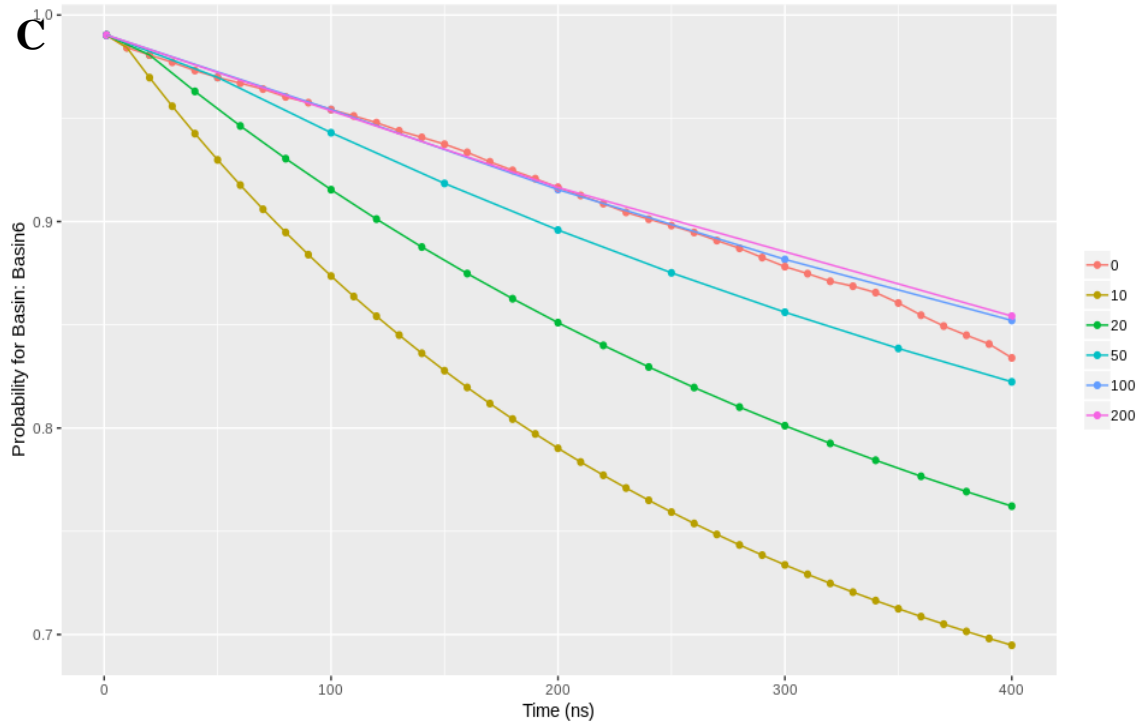
Figure B16. Chapman-Kolmogorov test of the human telomeric (A), c-KIT1 (B), c-MYC (C), and duplex (D) complex systems.
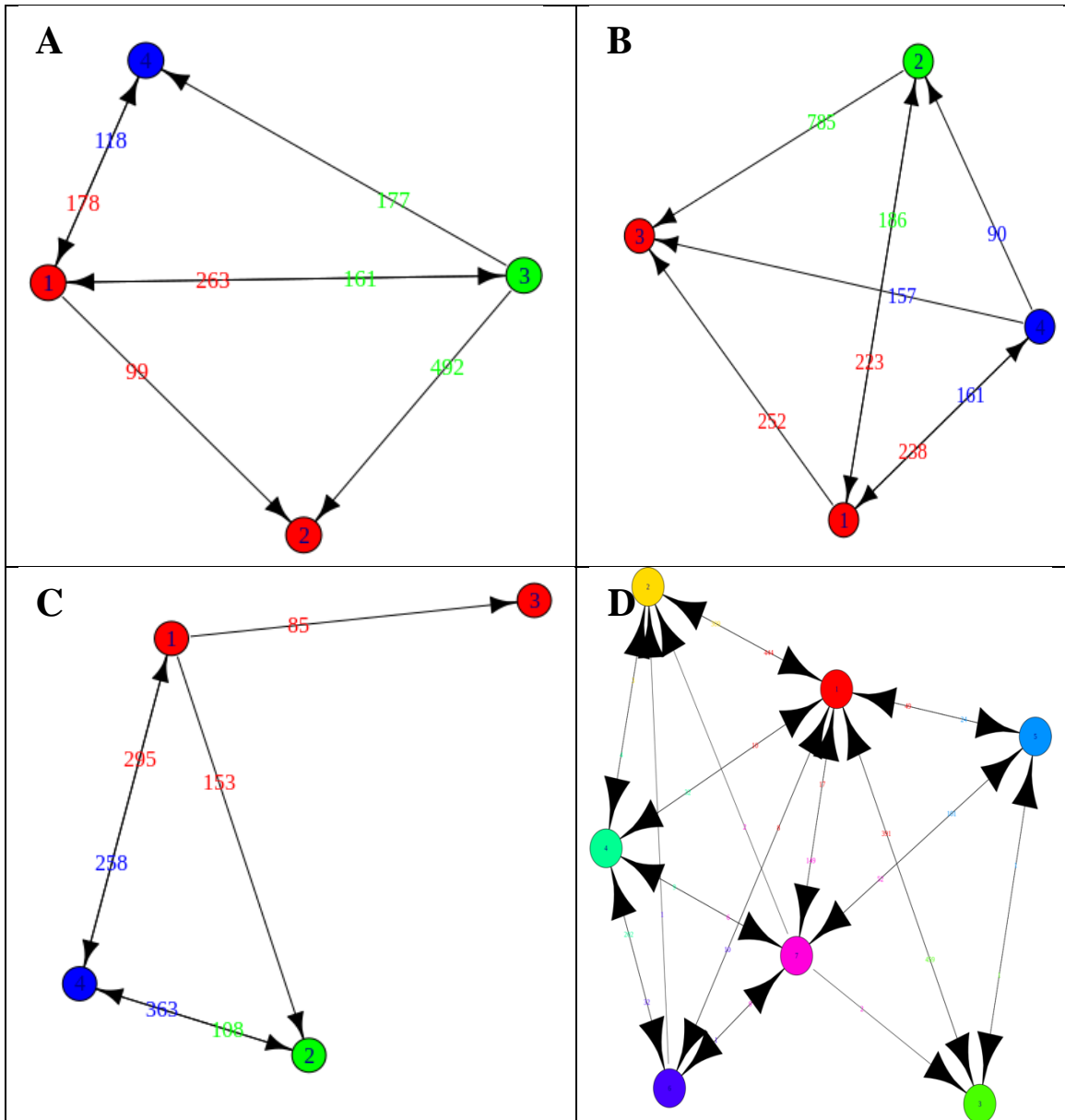
Figure B17. Network Models of the human telomeric (A), c-KIT1 (B), c-MYC (C), and duplex (D) complex systems.
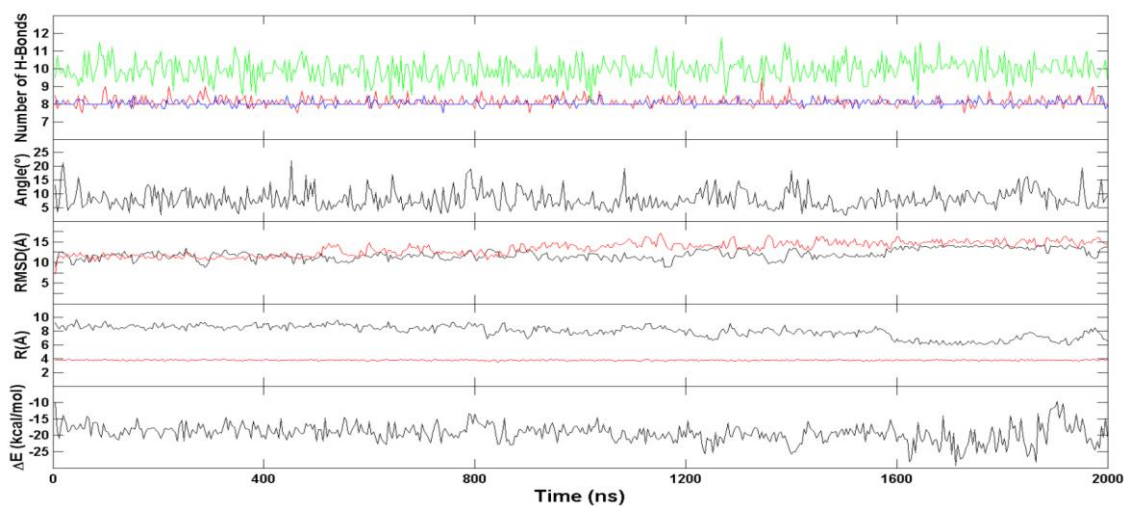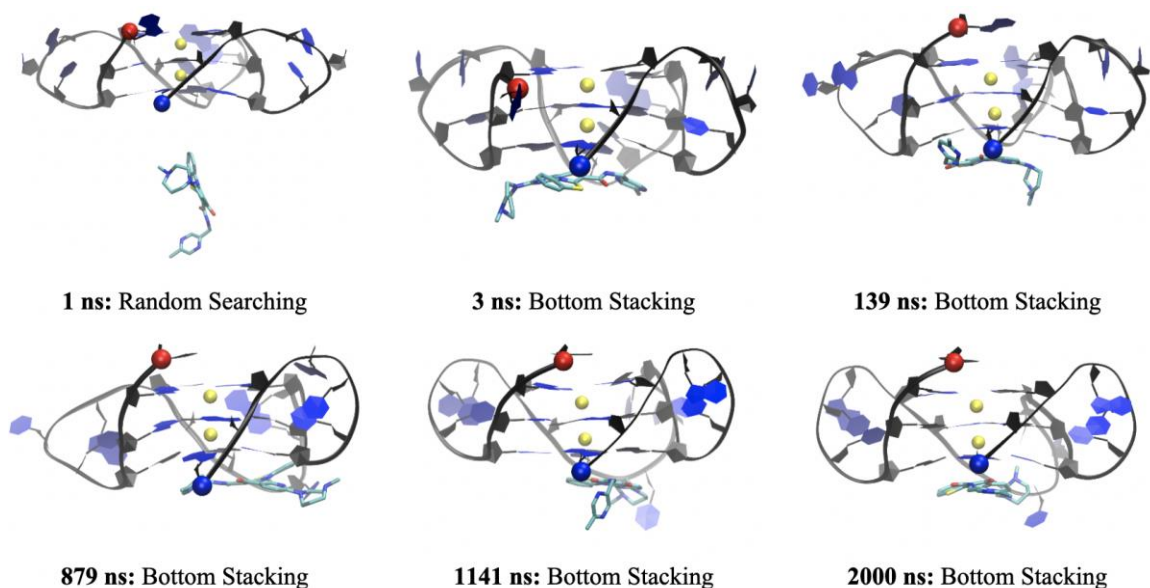
Figure B18. Snapshots from the representative trajectory of the bottom binding mode of the human telomeric G4, run 26, including an order parameter plot. Illustrated in the plot is the breaking and reforming of hydrogen bonds per layer (Top/Red, Middle/Green and Bottom/Blue), drug-base dihedral angle, ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and K+ to K+ distance(red), and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow balls.

218
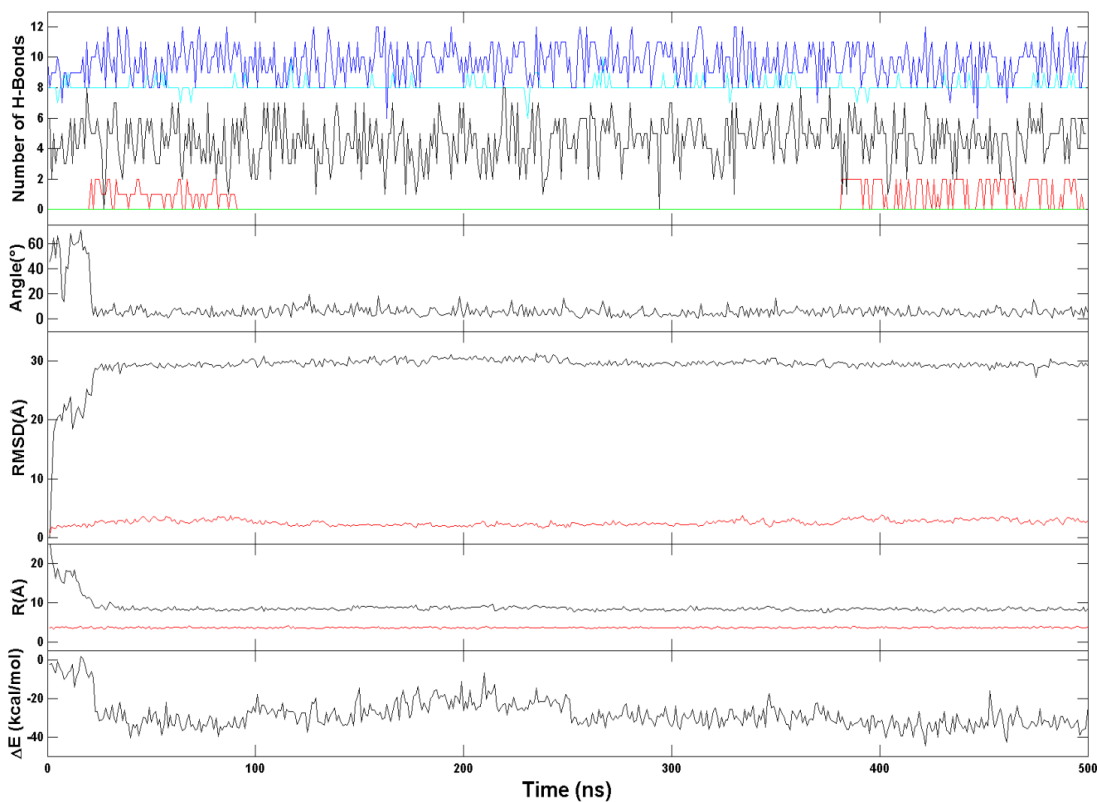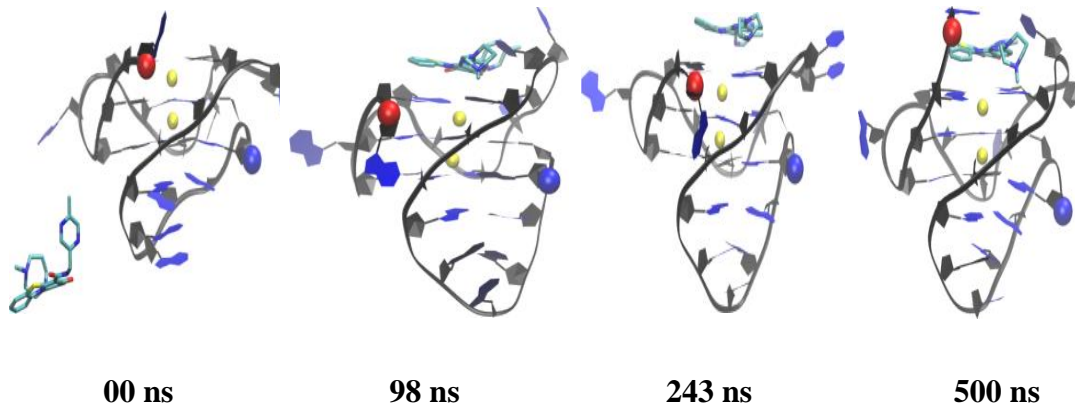
Figure B19. Snapshots from the representative trajectory of the top binding mode of 4WO3, run 20, including an order parameter plot. Illustrated in the plot is the breaking and reforming of hydrogen bonds per layer (5' Face/Red, Top G-Tetrad/Cyan, Middle G-Tetrad /Blue, Bottom G-Tetrad /Black, and 3' Face/Green), drug-base dihedral angle, ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and K+ to K+ distance(red), and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow balls.

219

Figure B20. Snapshots from the representative trajectory of the bottom binding mode to the c-KIT1 G4, run 23, including an order parameter plot. Illustrated in the plot is the breaking and reforming of hydrogen bonds per layer (5' Face/Red, Top G-Tetrad/Cyan, Middle G-Tetrad /Blue, Bottom G-Tetrad /Black, and 3' Face/Green), drug-base dihedral angle, ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and K+ to K+ distance(red), and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow balls.
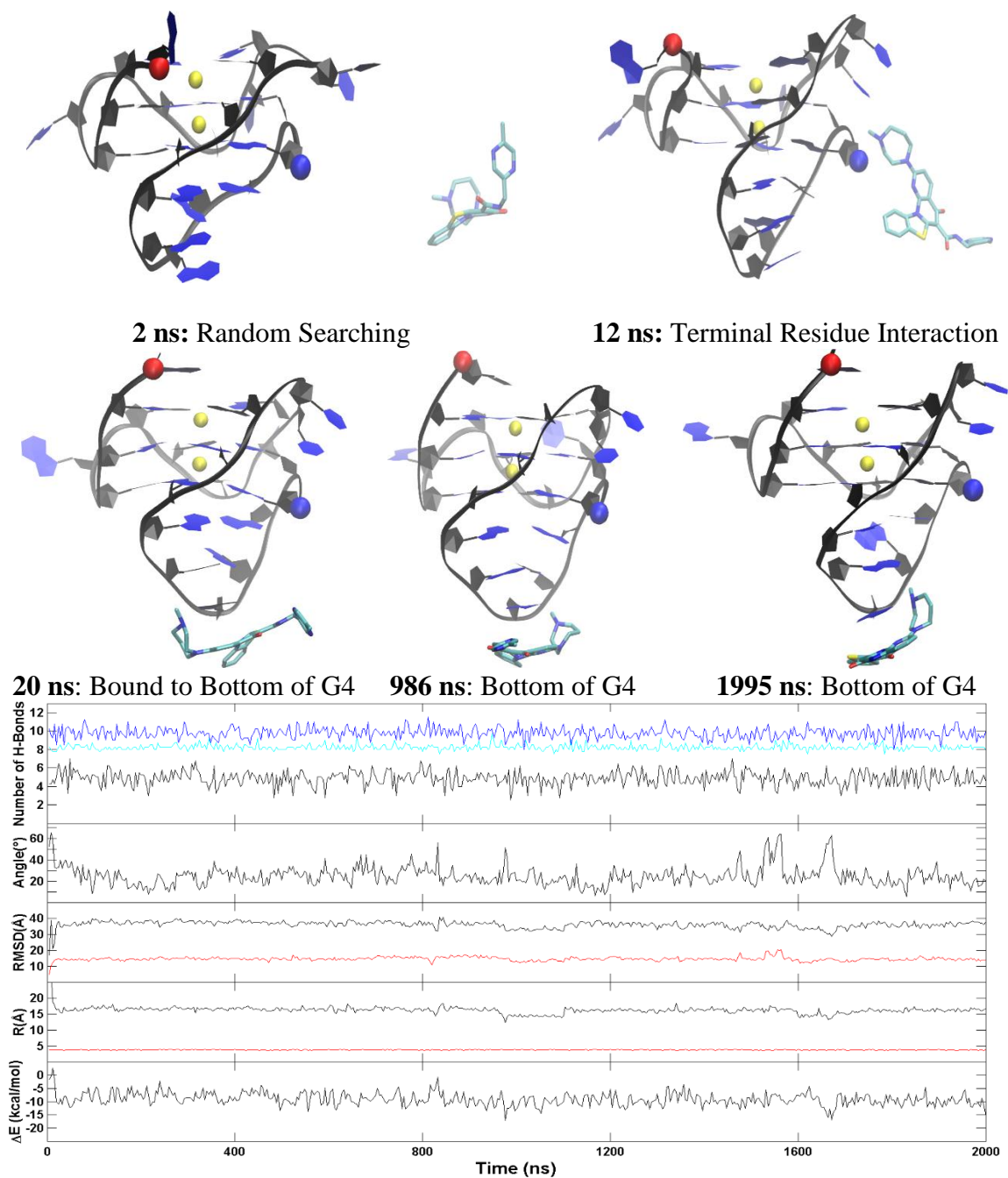
220

05 ns **13 ns** **128 ns** **498 ns**

Figure B21. Snapshots from the representative trajectory of the bottom binding mode of c-KIT1 G4, run 24, including an order parameter plot. Illustrated in the plot is the breaking and reforming of hydrogen bonds per layer (5' Face/Red, Top G-Tetrad/Cyan, Middle G-Tetrad /Blue, Bottom G-Tetrad /Black, and 3' Face/Green), drug-base dihedral angle, ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and K+ to K+ distance(red), and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow balls.
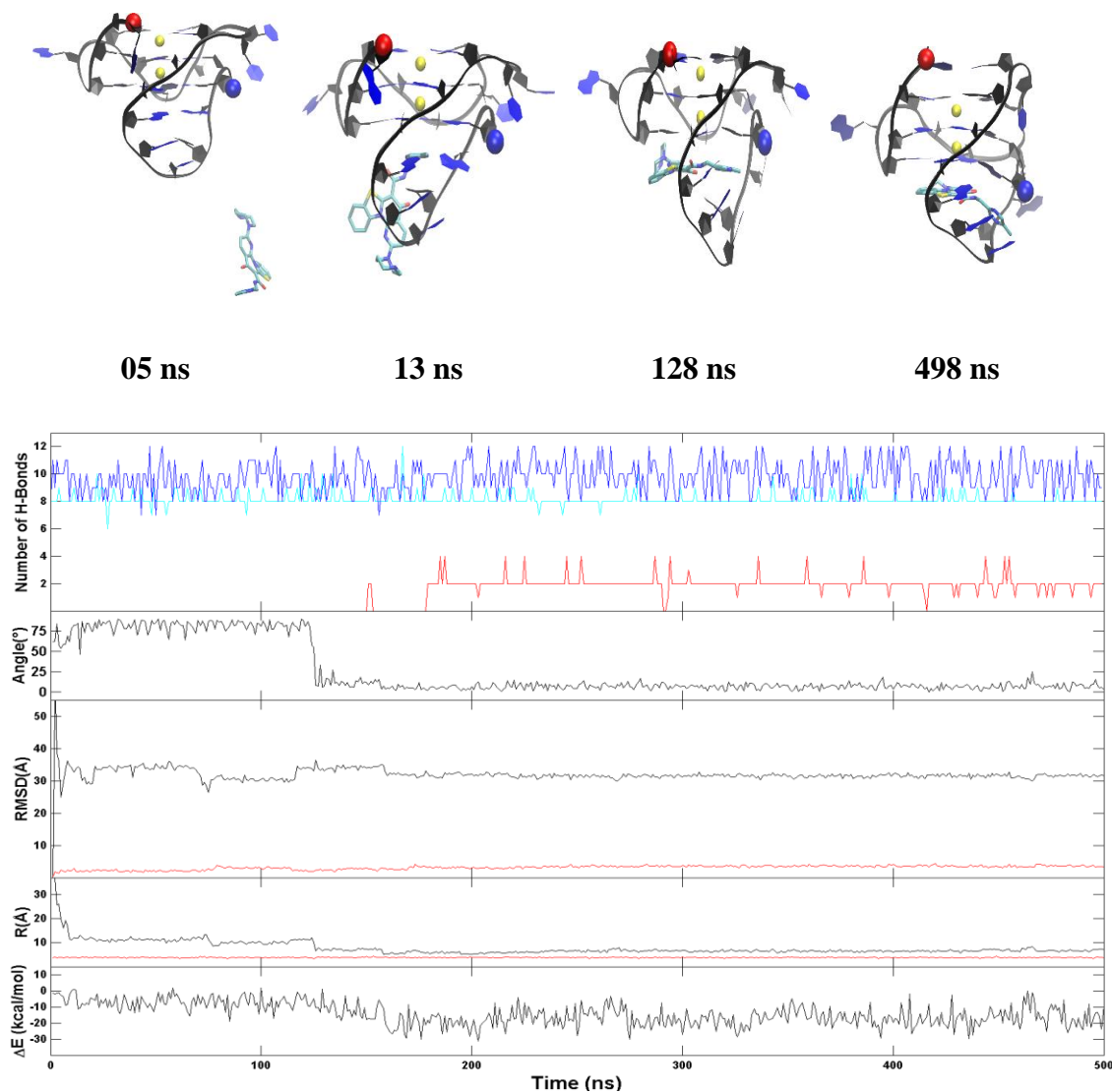
221

Figure B22. Snapshots from the representative trajectory of the side binding intermediate mode of c-KIT1 G4, run 19, including an order parameter plot. Illustrated in the plot is the breaking and reforming of hydrogen bonds per layer (5' Face/Red, Top G-Tetrad/Cyan, Middle G-Tetrad /Blue, Bottom G-Tetrad /Black, and 3' Face/Green), drug-base dihedral angle, ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and K+ to K+ distance(red), and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow balls.

00 ns   **27 ns**   **62 ns**   **146 ns**

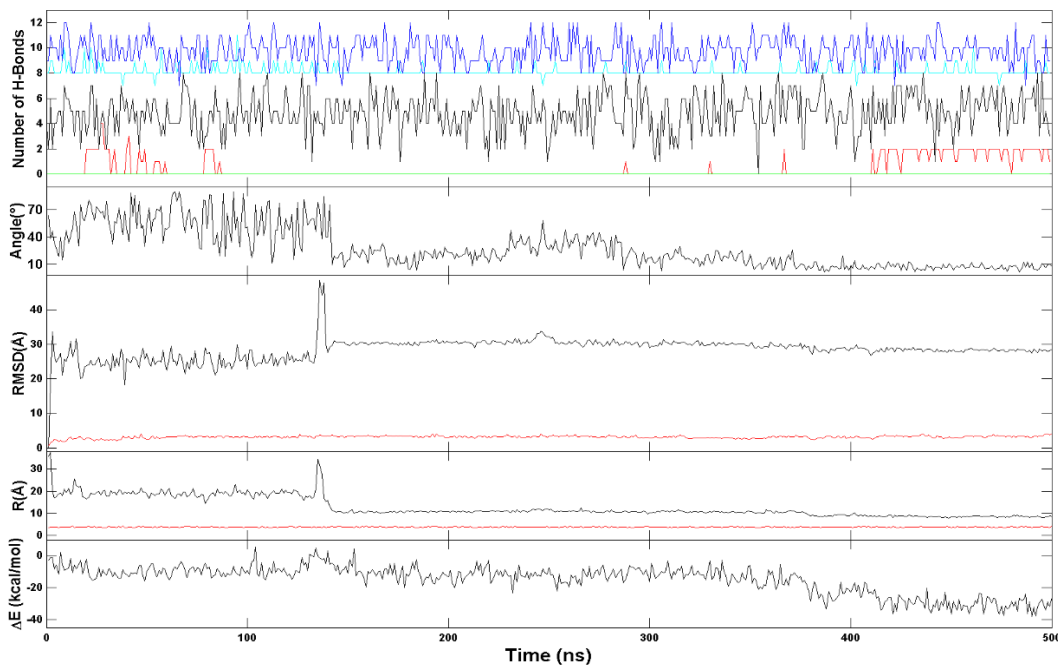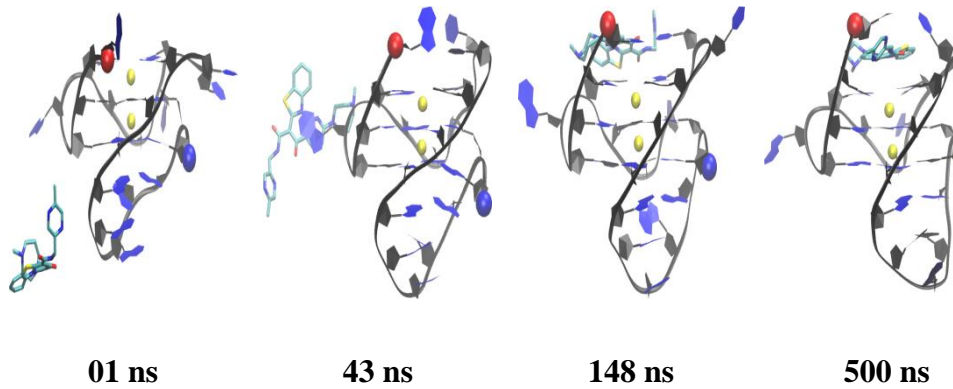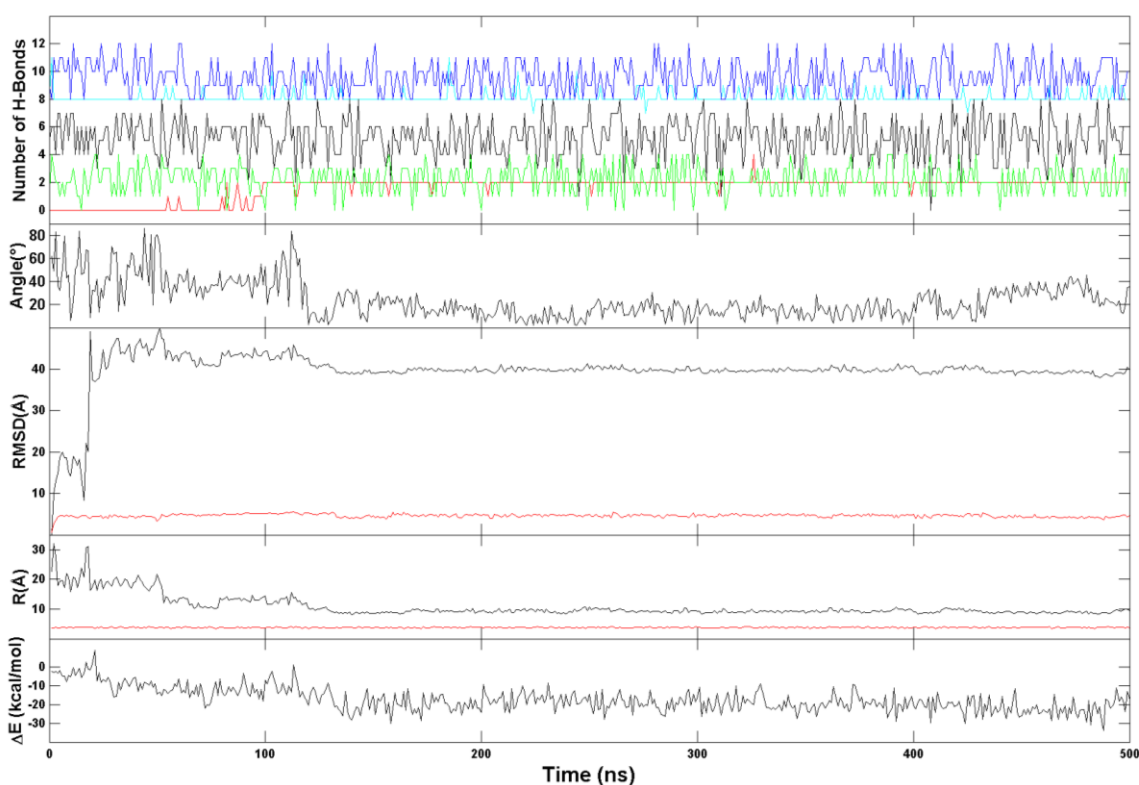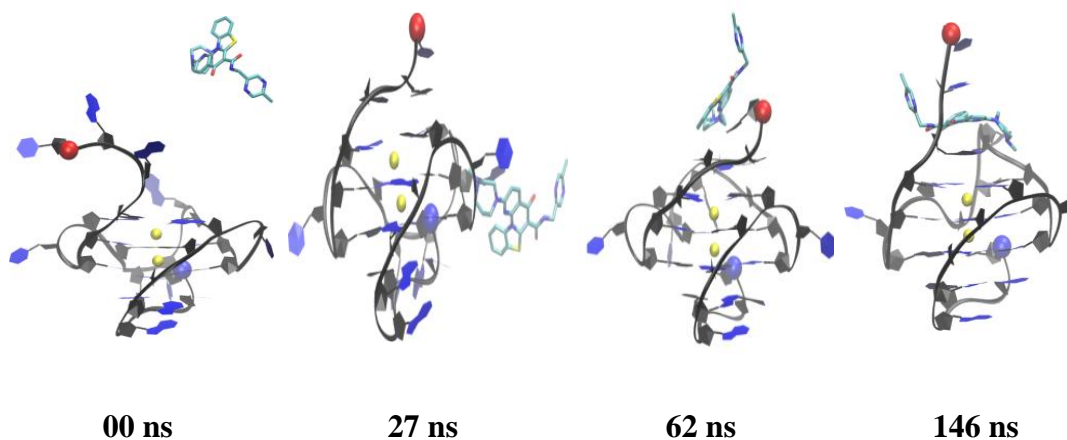*Figure B23.* Snapshots from the representative trajectory of the top binding mode of the c-Myc G4, run 10, including an order parameter plot. Illustrated in the plot is the breaking and reforming of hydrogen bonds per layer (5' Face/Red, Top G-Tetrad/Cyan, Middle G-Tetrad /Blue, Bottom G-Tetrad /Black, and 3' Face/Green), drug-base dihedral angle, ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and K+ to K+ distance(red), and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow balls.
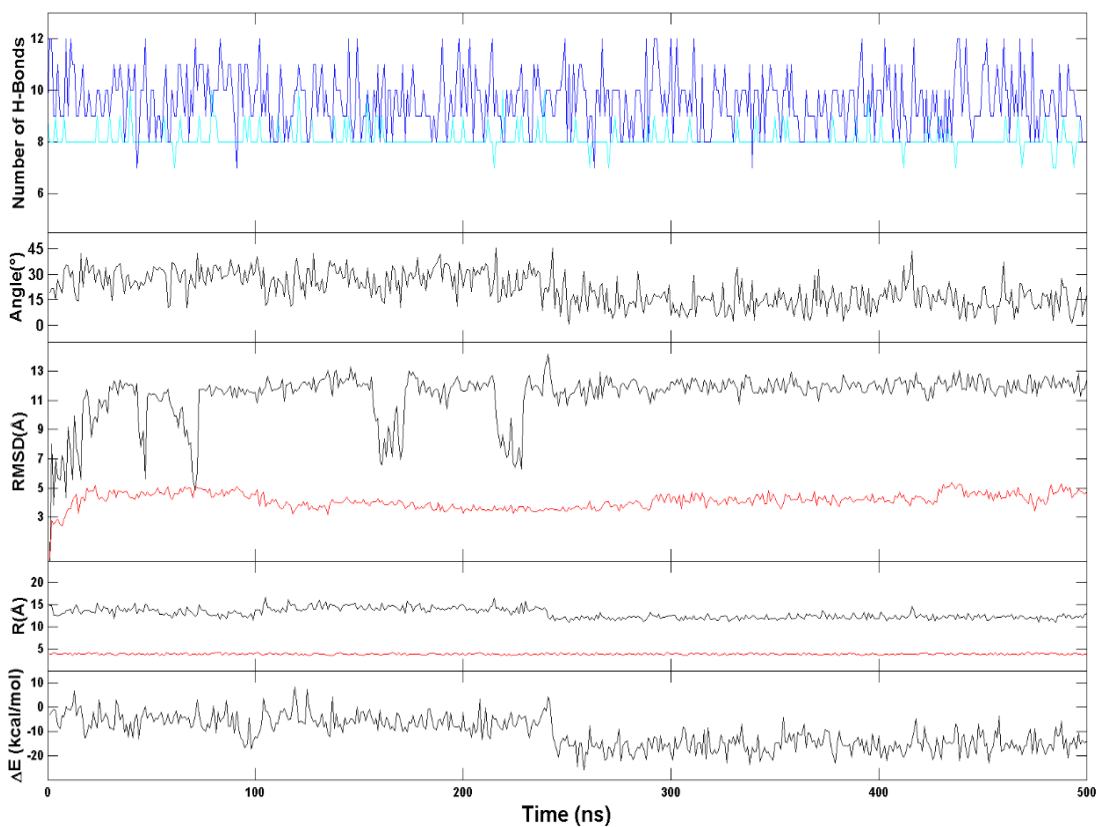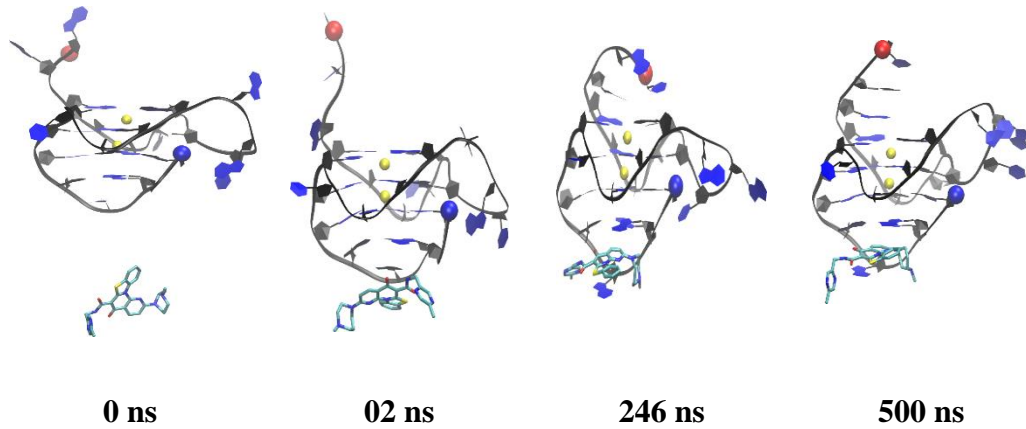
223

0 ns   **02 ns**   **246 ns**   **500 ns**

Figure B24. Snapshots from the representative trajectory of the bottom binding mode of c-Myc G4, run 25, including an order parameter plot. Illustrated in the plot is the breaking and reforming of hydrogen bonds per layer (5' Face/Red, Top G-Tetrad/Cyan, Middle G-Tetrad /Blue, Bottom G-Tetrad /Black, and 3' Face/Green), drug-base dihedral angle, ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and K+ to K+ distance(red), and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow balls.
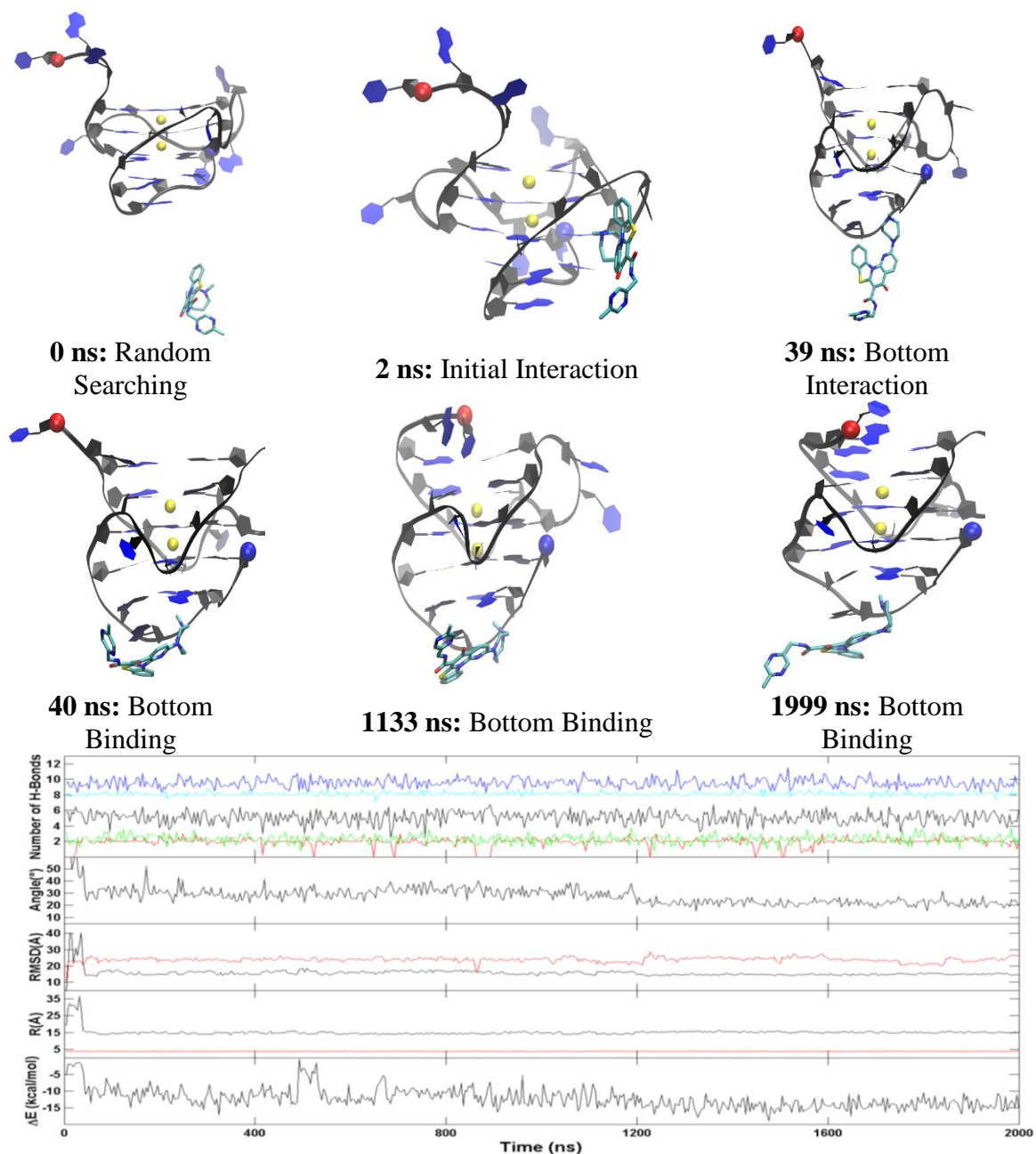
224

0 ns: Random Searching

2 ns: Initial Interaction

39 ns: Bottom Interaction

40 ns: Bottom Binding

1133 ns: Bottom Binding

1999 ns: Bottom Binding

Figure B25. Snapshots from the representative trajectory of the bottom binding mode of c-Myc G4, run 27, including an order parameter plot. Illustrated in the plot is the breaking and reforming of hydrogen bonds per layer (5' Face/Red, Top G-Tetrad/Cyan, Middle G-Tetrad /Blue, Bottom G-Tetrad /Black, and 3' Face/Green), drug-base dihedral angle, ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and K+ to K+ distance(red), and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow balls.
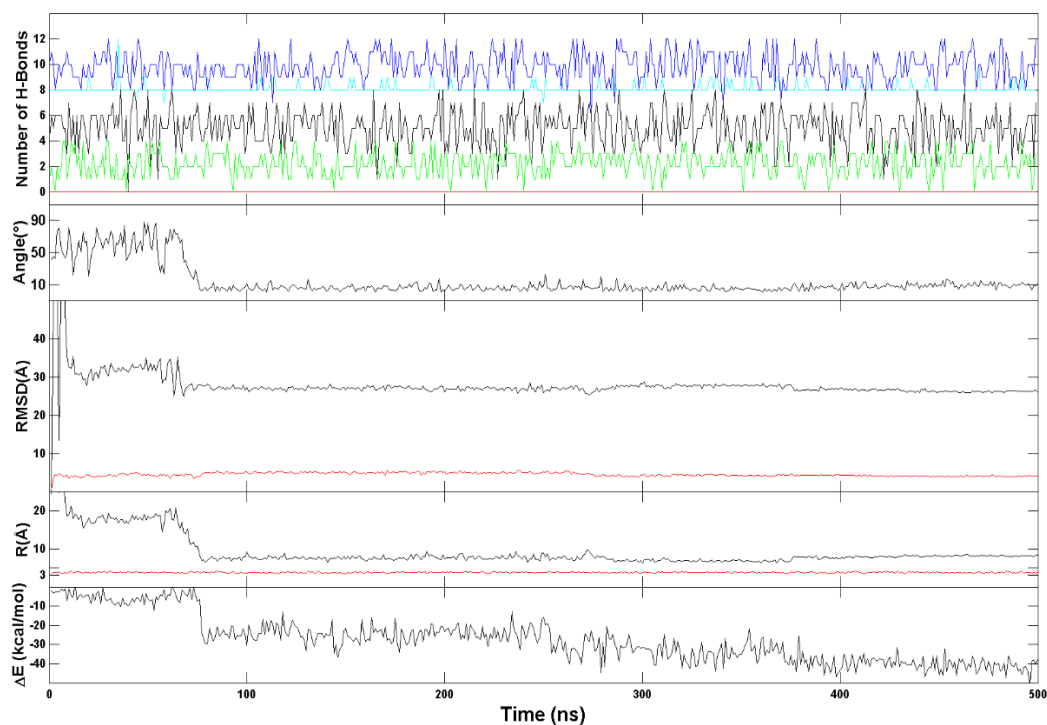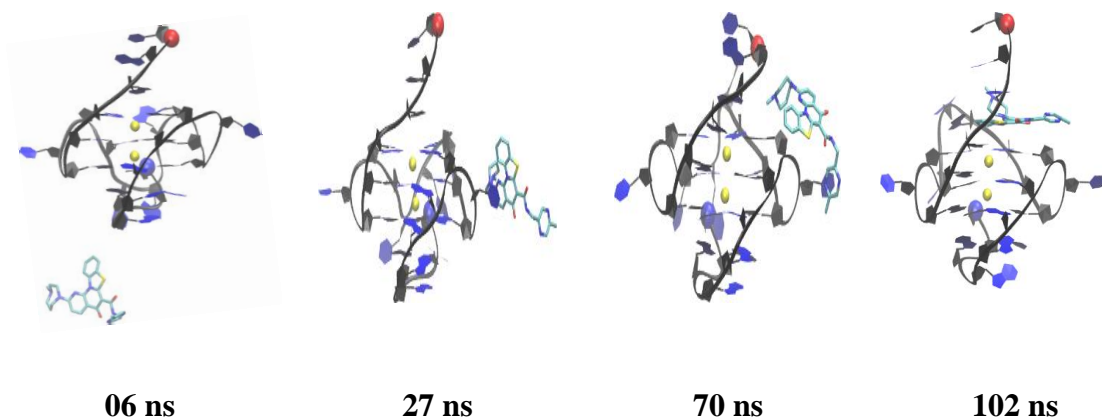
225

06 ns   27 ns   70 ns   102 ns

Figure B26. Snapshots from the representative trajectory of the side binding intermediate mode of the c-Myc G4, run 17, including an order parameter plot. Illustrated in the plot is the breaking and reforming of hydrogen bonds per layer (5' Face/Red, Top G-Tetrad/Cyan, Middle G-Tetrad /Blue, Bottom G-Tetrad /Black, and 3' Face/Green), drug-base dihedral angle, ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and K+ to K+ distance(red), and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow balls.

226
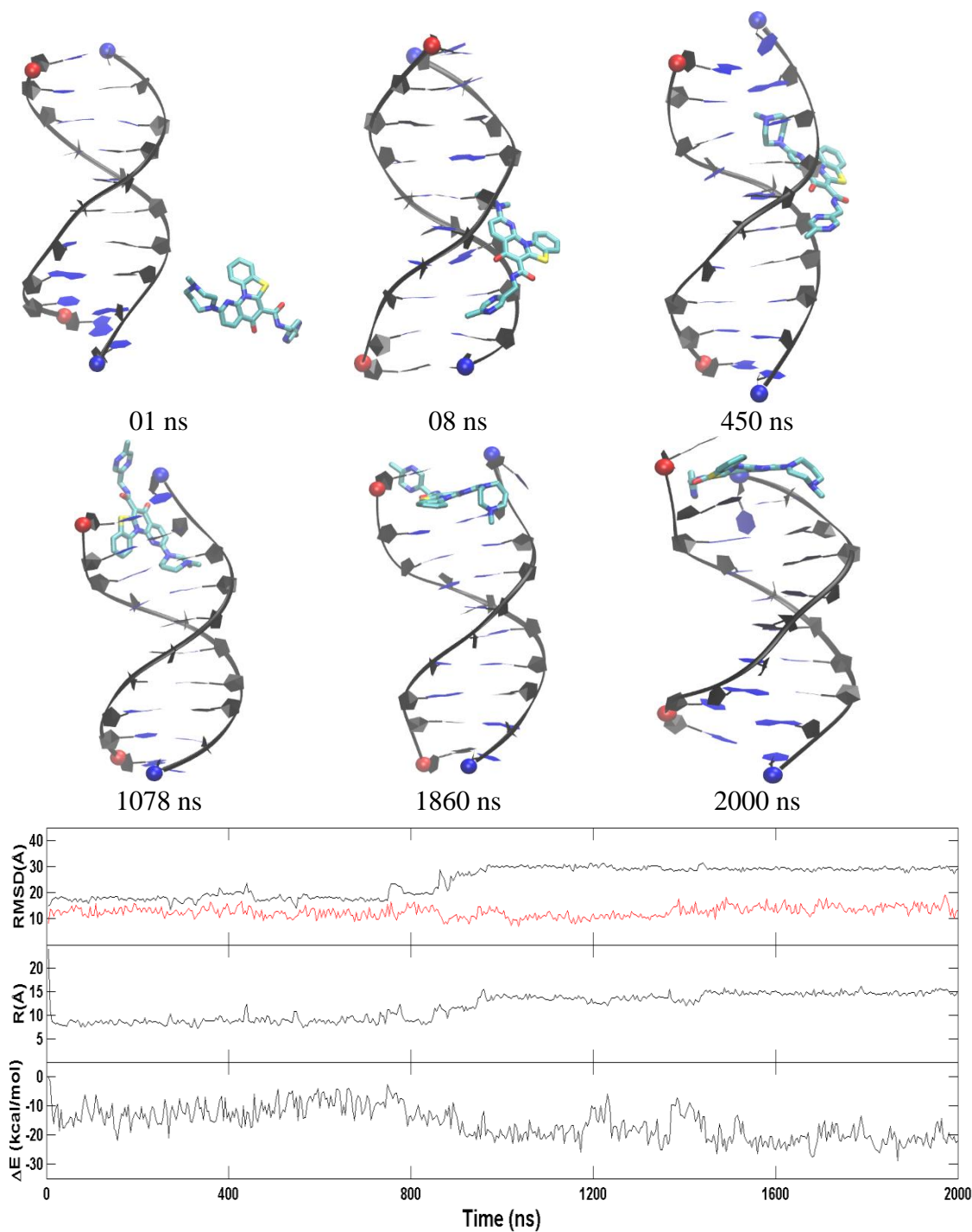
Figure B27. Snapshots from the representative trajectory of the duplex DNA system, run 14, including an order parameter plot. Illustrated in the plot is the ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow balls.

227
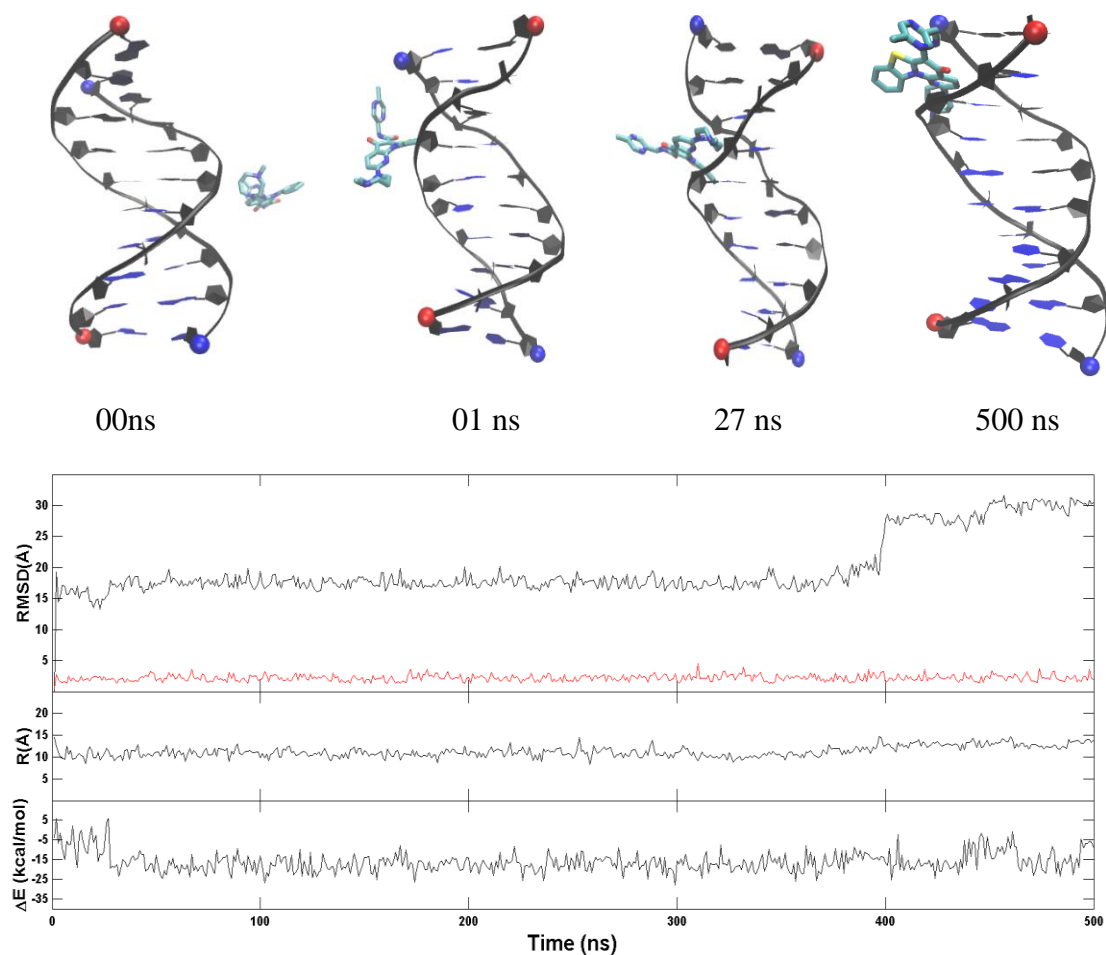
Figure B28. Snapshots from the representative trajectory of the duplex DNA system, run 06, including an order parameter plot. Illustrated in the plot is the ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow balls.

228

01 ns   05 ns   12 ns   339 ns   500 ns
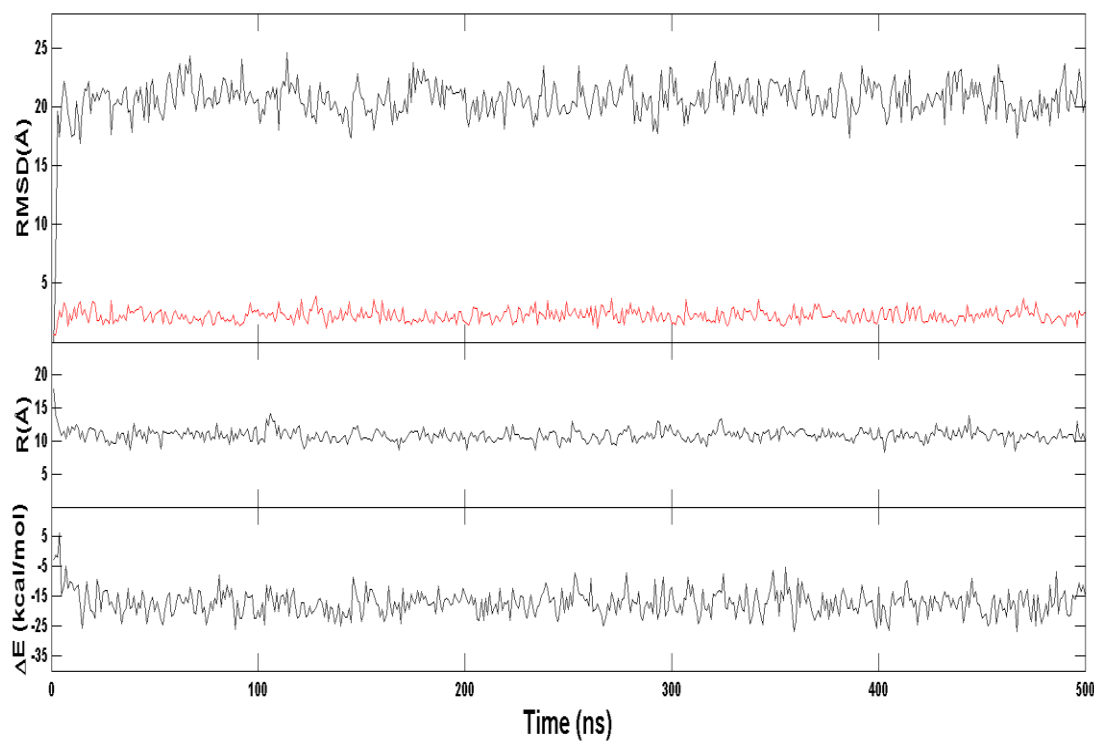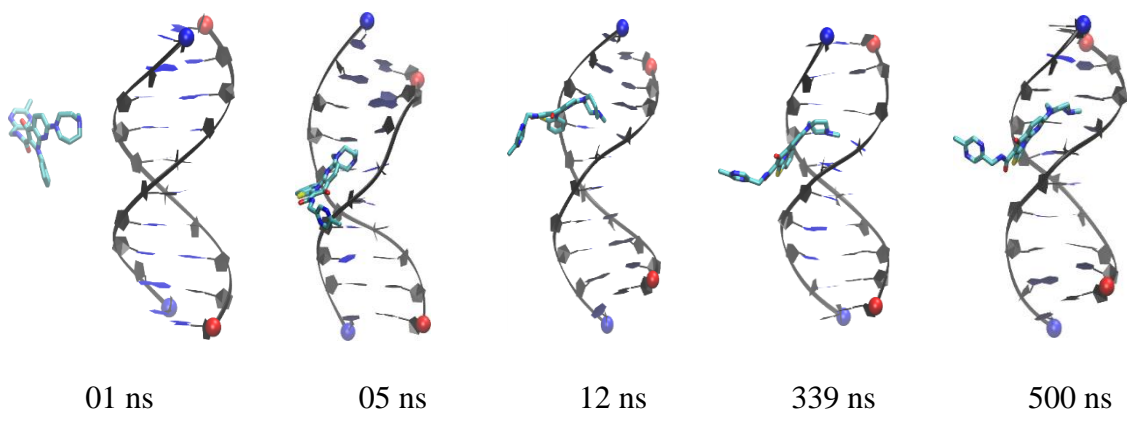
Figure B29. Snapshots from the representative trajectory of the duplex DNA system, run 25, including an order parameter plot. Illustrated in the plot is the ligand(black)/DNA(red) RMSD (Å) with reference to the final structure, ligand center to DNA center distance(black) and the MM-GBSA binding energy (ΔE in kcal/mol) . 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively. K+ ions are indicated by yellow balls.

229

System	Layer 1	Layer 2	Layer 3		
(A) Top					
(B) Bottom					
(C) Side					
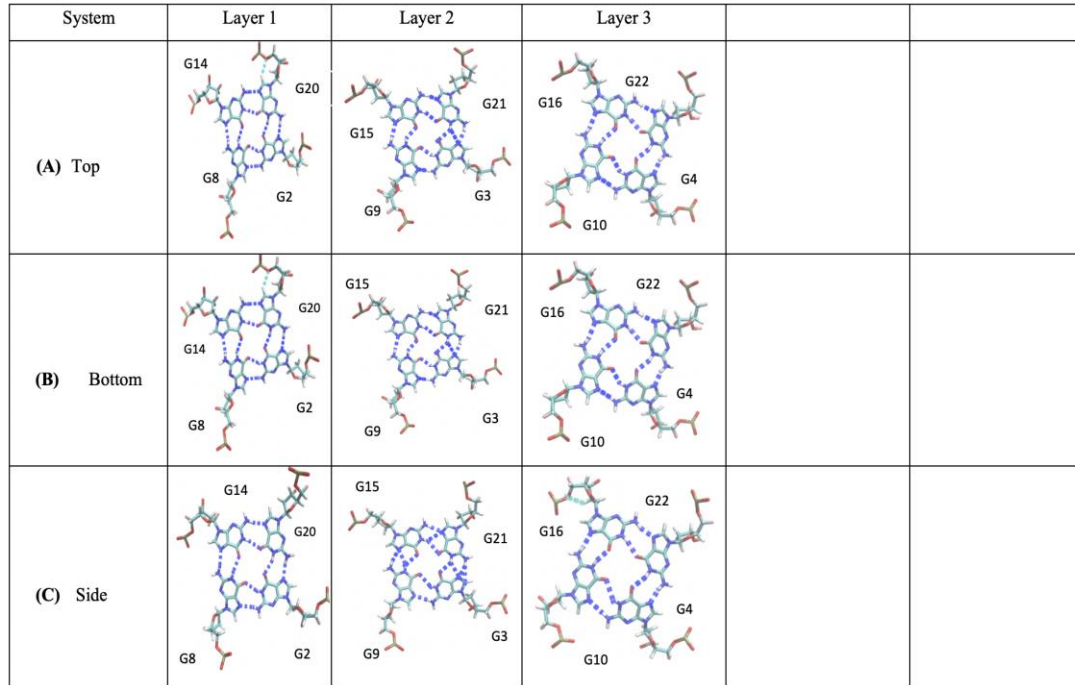
Figure B30. Hydrogen bond network of each layer formed by the human telomeric G-quadruplex in each representative binding modes.
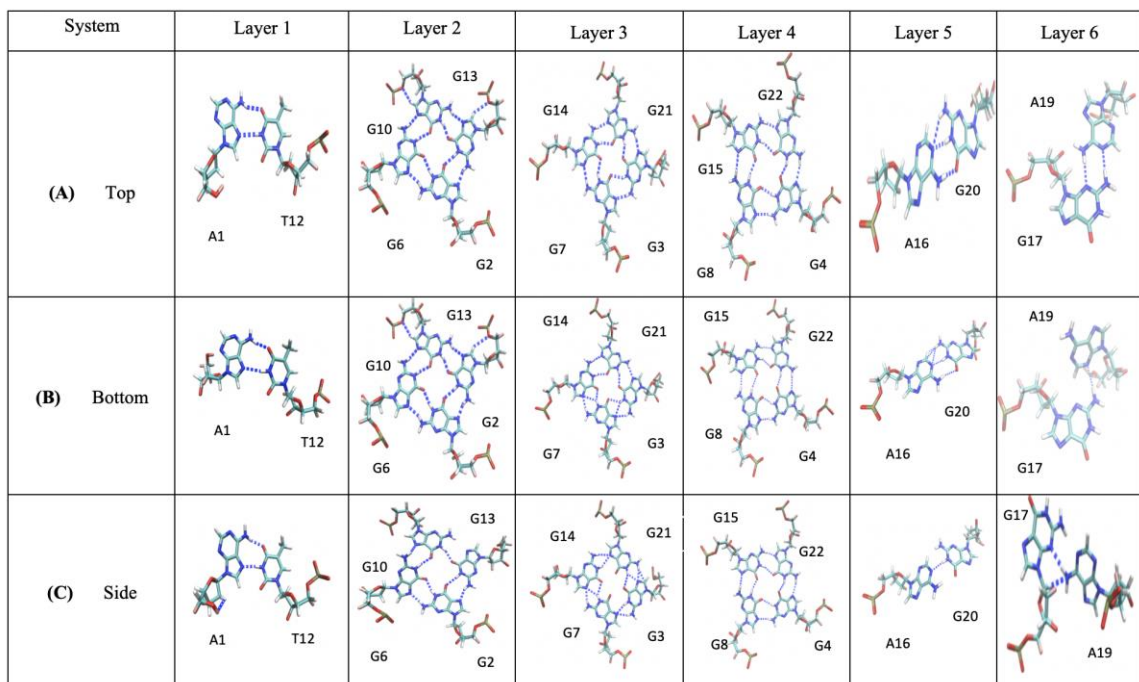
Figure B31. Hydrogen bond network of each layer formed by the c-KIT1 promoter G-quadruplex in each representative binding modes.
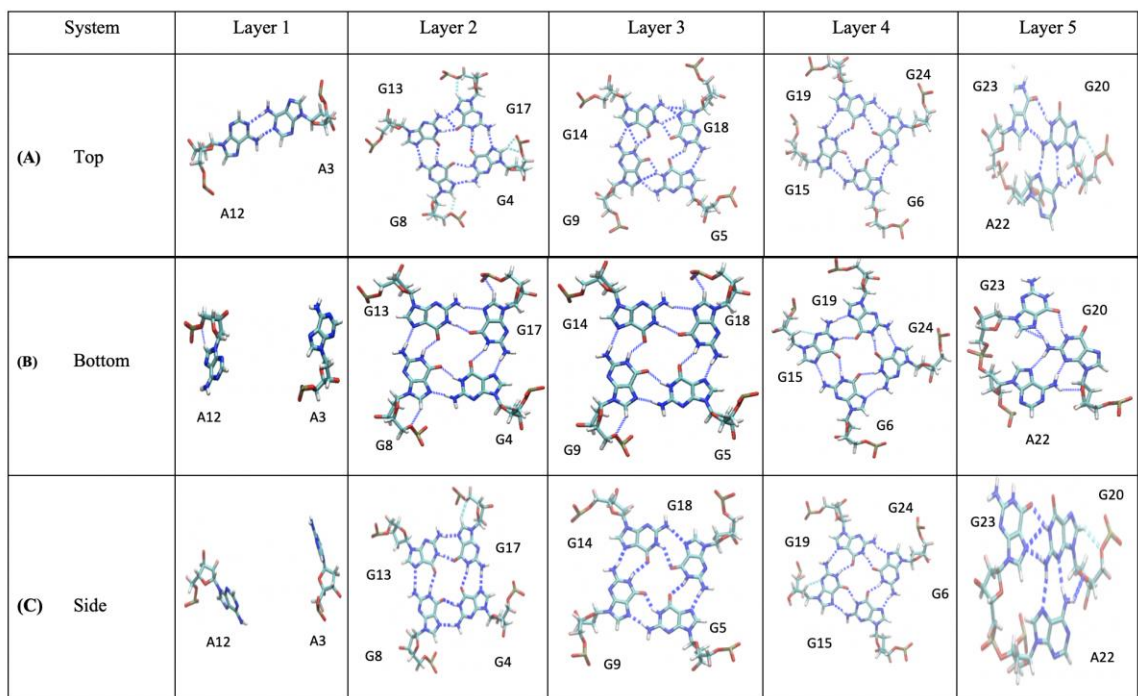
Figure B32. Hydrogen bond network of each layer formed by the c-Myc G-quadruplex in each representative binding modes.
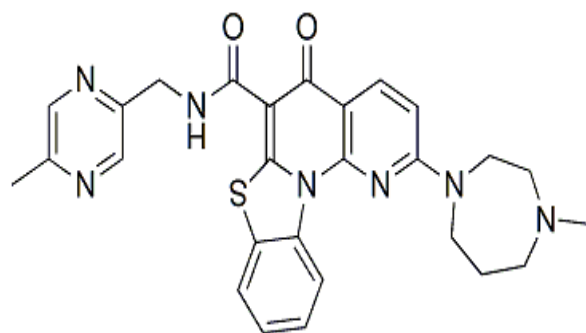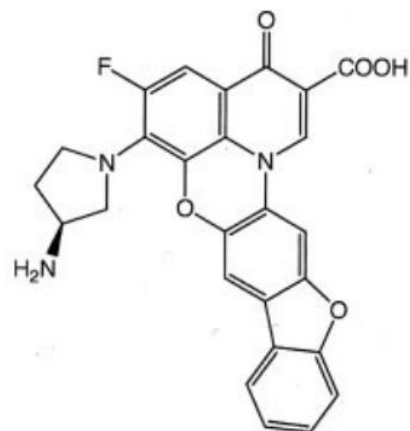
CX-5461   **QQ58**

Figure B33. Comparison of CX-5461 and structurally related ligand QQ58.

Chemical Structure	Duplex	1KF1
RHPS4		
CX-5461		
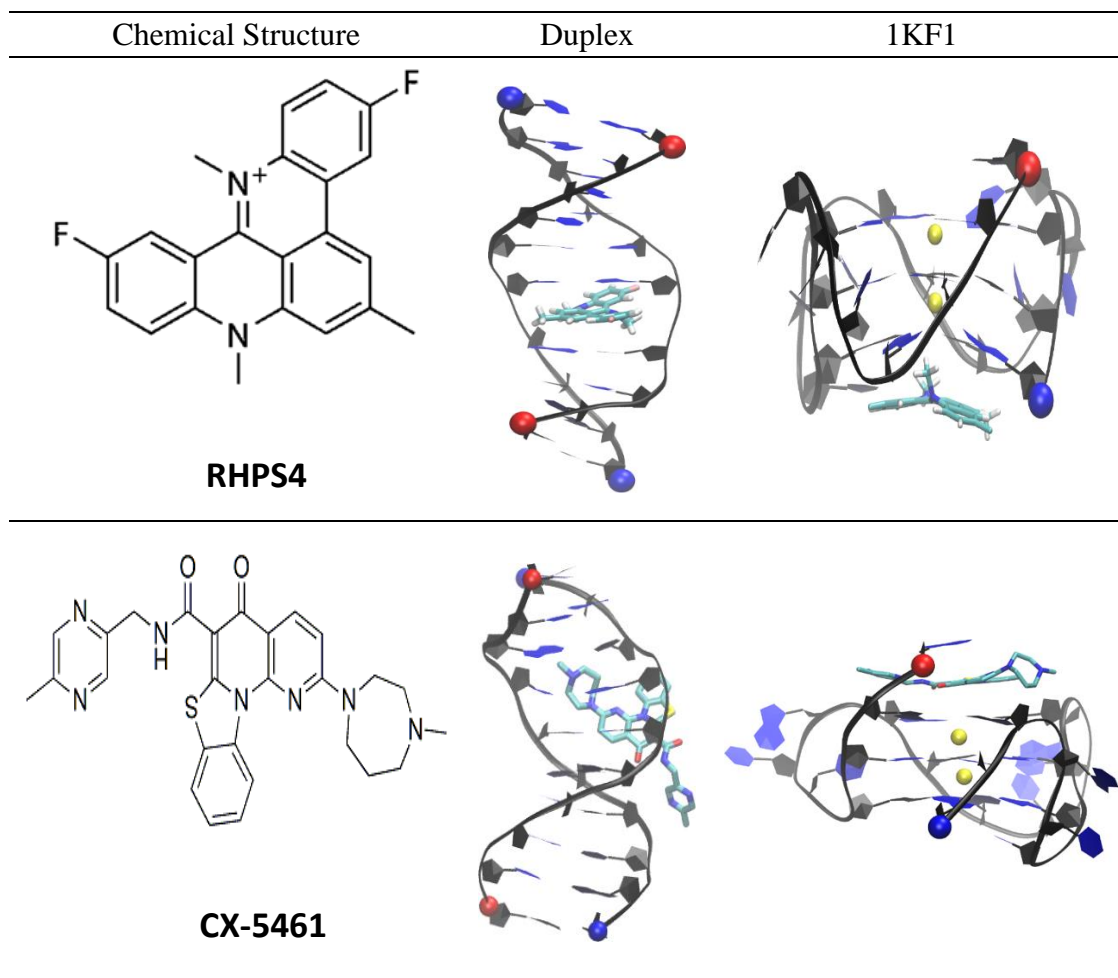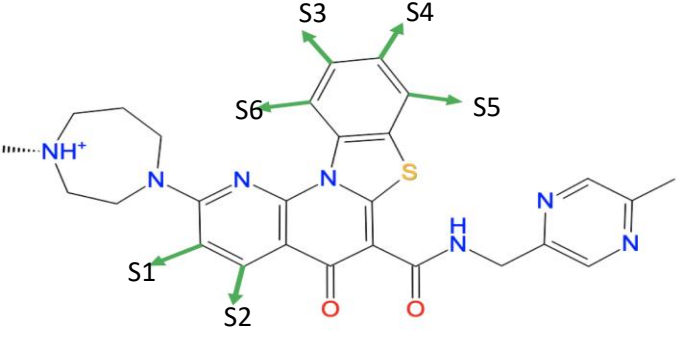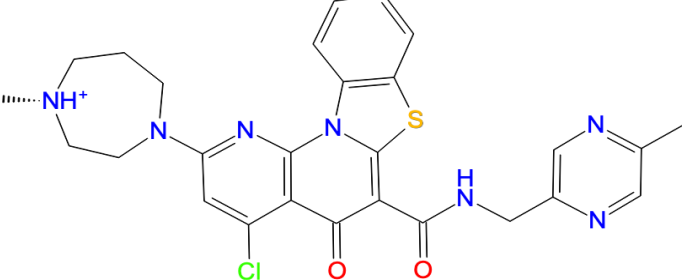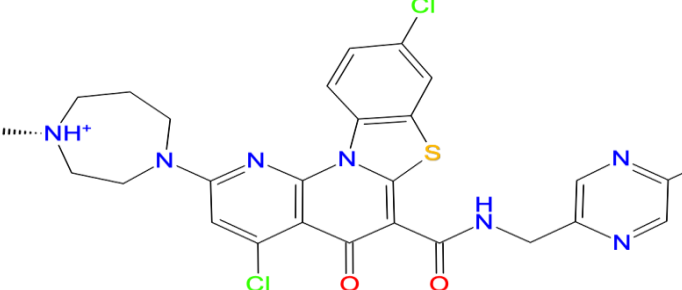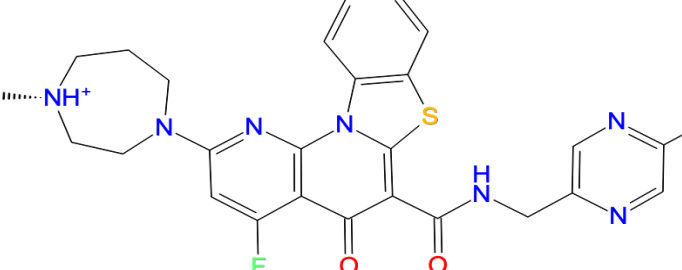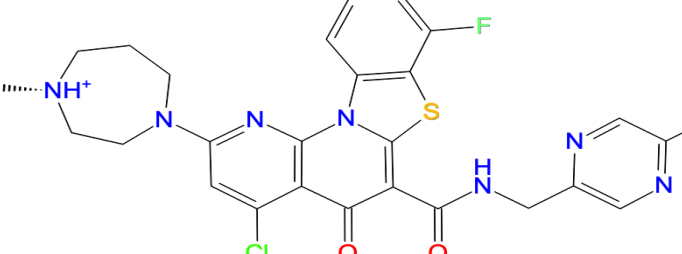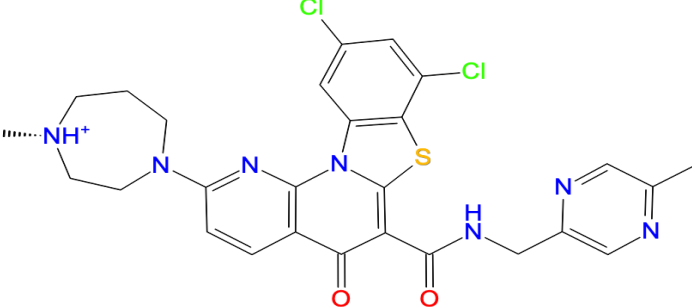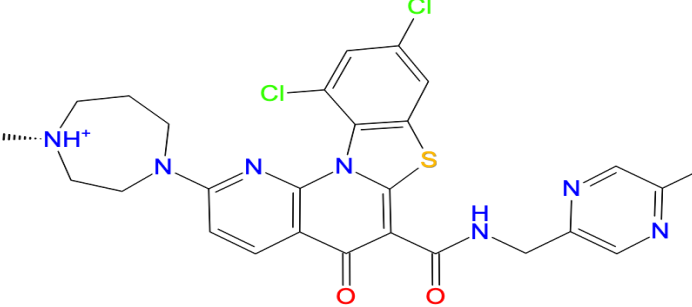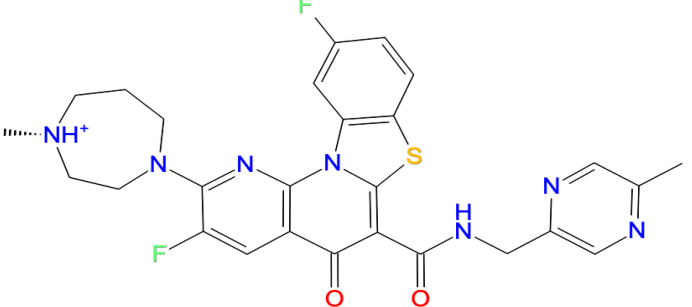
Figure B34. Comparison between RHPS4 and CX-5461.

المنارة للاستشارات

www.manaraa.com

Chemical Structure of CX-5461: Substitution Sites Marked	Docking Score
	h-Tel: -8.386 CKIT1: -8.112 c-Myc: -8.567 Duplex: -5.944

Chemical Structure of New Ligand	Docking Score	△CX-5461
	h-tel: -8.811 c-KIT1: -7.514 c-Myc: -8.847 Duplex: -5.333	h-Tel: -0425 c-KIT1: +0.598 c-Myc: -0.280 Duplex: +0.611
	h-tel: -9.418 c-KIT1: -7.960 c-Myc: -8.073 Duplex: -4.963	h-Tel: -1.032 c-KIT1: +0.152 c-Myc: +0.494 Duplex: +0.981
	h-tel: -8.690 c-KIT1: -7.585 c-Myc: -8.296 Duplex: -5.144	h-Tel: -0.304 c-KIT1: +0.527 c-Myc: +0.271 Duplex: +0.800
	h-tel: -8.898 c-KIT1: -4.641 c-Myc: -8.350 Duplex: -3.856	h-Tel: -0.512 c-KIT1: +3.471 c-Myc: +0.217 Duplex: +2.088

235

	h-tel: -9.207 c-KIT1: -5.637 c-Myc: -7.968 Duplex: -4.932	h-Tel: -0.821 c-KIT1: +2.475 c-Myc: +0.599 Duplex: 1.012
	h-tel: -8.522 c-KIT1: -7.067 c-Myc: -7.166 Duplex: -3.487	h-Tel: -0.136 c-KIT1: +1.045 c-Myc: +1.401 Duplex: +2.457
	h-Tel: -8.487 CKIT1: -6.494 c-Myc: -8.315 Duplex: -3.964	h-Tel: -0.101 c-KIT1: +1.618 c-Myc: +0.252 Duplex: +1.980

Figure B35. Chemical structure of CX-5461 and additional derivates identified through virtual screening, including docking scores. For CX-5461 green arrows indicate substitution sites for the derivates on this table. For the derivates, docking scores are provided as well as the difference between their docking scores and CX-5461.

236

Figure B36. AMBER GAFF2 force field of the ligand CX-5461 (+1) in Mol2 format.

@<TRIPOS>MOLECULE

CX5

  65   70   1   0   0

SMALL

rc

@<TRIPOS>ATOM

1	C1	3.9140	-3.2010	1.5250	c3	1	CX5	-0.003449
2	H1	2.8430	-3.1240	1.3740	hc	1	CX5	0.051021
3	H2	4.0750	-4.1090	2.0970	hc	1	CX5	0.051021
4	C2	4.6010	-3.3270	0.1610	c3	1	CX5	-0.090060
5	H3	5.6690	-3.4690	0.2720	h1	1	CX5	0.117007
6	H4	4.2250	-4.2100	-0.3370	h1	1	CX5	0.117007
7	N1	4.3940	-2.1500	-0.6710	nh	1	CX5	-0.331930
8	C3	5.3770	-1.1020	-0.5600	c3	1	CX5	-0.059912
9	H5	6.3340	-1.4940	-0.8840	h1	1	CX5	0.107942
10	H6	5.1110	-0.3180	-1.2490	h1	1	CX5	0.107942
11	C4	5.5810	-0.5010	0.8510	c3	1	CX5	-0.032701
12	H7	6.4270	-0.9410	1.3590	hx	1	CX5	0.093845
13	H8	5.7610	0.5610	0.7730	hx	1	CX5	0.093845
14	N2	4.3960	-0.6890	1.7550	n4	1	CX5	0.015971
15	H28	3.5750	-0.5830	1.1660	hn	1	CX5	0.134612
16	C5	4.3790	-2.0490	2.4140	c3	1	CX5	0.013568

237

17 H9	3.7170	-1.9720	3.2630	hx	1 CX5	0.091063
18 H10	5.3810	-2.2090	2.7870	hx	1 CX5	0.091063
19 C6	3.0850	-1.8220	-1.0310	ca	1 CX5	0.226522
20 N3	2.6580	-0.6120	-0.7150	nb	1 CX5	-0.023164
21 C7	1.4050	-0.2680	-1.0120	ca	1 CX5	-0.022413
22 C8	0.5010	-1.1540	-1.5710	ca	1 CX5	0.042052
23 C9	0.9720	-2.4230	-1.9110	ca	1 CX5	-0.020359
24 H11	0.2890	-3.1070	-2.3770	ha	1 CX5	0.174135
25 C10	2.2740	-2.7710	-1.6600	ca	1 CX5	-0.294940
26 H12	2.6670	-3.7250	-1.9490	ha	1 CX5	0.191834
27 N4	0.9770	1.0090	-0.6830	na	1 CX5	0.065290
28 C11	-0.3810	1.2860	-0.6570	cc	1 CX5	0.034781
29 C12	-1.3330	0.4440	-1.1520	cd	1 CX5	-0.071965
30 C13	-0.9210	-0.8150	-1.7550	c	1 CX5	0.299734
31 S1	-0.6900	2.8380	0.0980	ss	1 CX5	0.011766
32 C14	1.0350	3.1530	0.1850	ca	1 CX5	0.035038
33 C15	1.7960	2.1070	-0.3030	ca	1 CX5	0.029667
34 C16	-2.7620	0.8690	-1.0420	c	1 CX5	0.396672
35 N5	-3.6880	0.1260	-1.6500	n	1 CX5	-0.246335
36 H13	-3.4080	-0.7230	-2.0910	hn	1 CX5	0.268332
37 C17	-5.1010	0.4490	-1.5570	c3	1 CX5	-0.248195
38 H14	-5.2060	1.5230	-1.5620	h1	1 CX5	0.128459
39 H15	-5.5830	0.0510	-2.4390	h1	1 CX5	0.128459
40 C18	-5.7640	-0.1460	-0.3360	ca	1 CX5	0.351041

238

41 C19	-5.9430	0.5750	0.8380 ca	1 CX5	0.101423
42 H16	-5.6090	1.5930	0.9110 h4	1 CX5	0.092355
43 N6	-6.5140	0.0480	1.9010 nb	1 CX5	-0.504185
44 C20	-6.9220	-1.2080	1.8250 ca	1 CX5	0.465683
45 C21	-6.7340	-1.9280	0.6510 ca	1 CX5	0.012135
46 H17	-7.0600	-2.9500	0.5740 h4	1 CX5	0.106519
47 N7	-6.1630	-1.4020	-0.4130 nb	1 CX5	-0.458448
48 C22	-7.5850	-1.8030	3.0380 c3	1 CX5	-0.523155
49 H18	-6.9280	-1.7350	3.8970 hc	1 CX5	0.145037
50 H19	-7.8440	-2.8430	2.8800 hc	1 CX5	0.145037
51 H20	-8.4910	-1.2550	3.2760 hc	1 CX5	0.145037
52 C23	4.3380	0.3880	2.7880 c3	1 CX5	-0.153992
53 H21	4.2900	1.3450	2.2930 hx	1 CX5	0.106894
54 H22	5.2230	0.3270	3.4050 hx	1 CX5	0.106894
55 H23	3.4560	0.2470	3.3930 hx	1 CX5	0.106894
56 O1	-1.6550	-1.5900	-2.3190 o	1 CX5	-0.488334
57 O2	-3.0290	1.8750	-0.4190 o	1 CX5	-0.464483
58 C24	3.1680	2.2590	-0.4540 ca	1 CX5	-0.114543
59 H24	3.7570	1.4990	-0.9160 ha	1 CX5	0.090937
60 C25	3.7550	3.4440	-0.0390 ca	1 CX5	-0.113961
61 H25	4.8130	3.5790	-0.1730 ha	1 CX5	0.141889
62 C26	2.9950	4.4690	0.5090 ca	1 CX5	-0.181207
63 H26	3.4670	5.3820	0.8220 ha	1 CX5	0.169242
64 C27	1.6230	4.3320	0.6090 ca	1 CX5	-0.178669

239

65 H27   1.0170   5.1350   0.9870 ha   1 CX5   0.220734

@<TRIPOS>BOND

    1   1   2 1

    2   1   3 1

    3   1   4 1

    4   1   16 1

    5   4   5 1

    6   4   6 1

    7   4   7 1

    8   7   8 1

    9   7   19 1

    10   8   9 1

    11   8   10 1

    12   8   11 1

    13   11   12 1

    14   11   13 1

    15   11   14 1

    16   14   15 1

    17   14   16 1

    18   14   52 1

    19   16   17 1

    20   16   18 1

    21   19   20 ar

    22   19   25 ar

240

23   20   21 ar

24   21   22 ar

25   21   27 1

26   22   23 ar

27   22   30 1

28   23   24 1

29   23   25 ar

30   25   26 1

31   27   28 1

32   27   33 1

33   28   29 2

34   28   31 1

35   29   30 1

36   29   34 1

37   30   56 2

38   31   32 1

39   32   33 ar

40   32   64 ar

41   33   58 ar

42   34   35 1

43   34   57 2

44   35   36 1

45   35   37 1

46   37   38 1

241

47	37	39	1
48	37	40	1
49	40	41	ar
50	40	47	ar
51	41	42	1
52	41	43	ar
53	43	44	ar
54	44	45	ar
55	44	48	1
56	45	46	1
57	45	47	ar
58	48	49	1
59	48	50	1
60	48	51	1
61	52	53	1
62	52	54	1
63	52	55	1
64	58	59	1
65	58	60	ar
66	60	61	1
67	60	62	ar
68	62	63	1
69	62	64	ar
70	64	65	1

242

@<TRIPOS>SUBSTRUCTURE

1 CX5   1 TEMP   0 **** ****   0 ROOT