The Drug Binding Pathways and Conformational Deviations of G-quadruplex DNA under Different Chemical and Thermal Conditions: A Computational Study

A Thesis Submitted

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Saikat Pal

to the

Department of Chemistry

Indian Institute of Technology Guwahati, India

2021

DEDICATED TO
MY BELOVED PARENTS

Declaration

I hereby declare that the matter manifested in this thesis entitled "*The Drug Binding Pathways and Conformational Deviations of G-quadruplex DNA under Different Chemical and Thermal Conditions: A Computational Study*" is the result of research carried out by me in the Department of Chemistry, Indian Institute of Technology Guwahati, India under the supervision of Prof. Sandip Paul.

In keeping with the general practice of reporting scientific observations, due acknowledgement has been made wherever the work described is based on the findings of other investigators.

Saikat Pal
IIT Guwahati

Certificate

It is certified that the work contained in this thesis entitled, "*The Drug Binding Pathways and Conformational Deviations of G-quadruplex DNA under Different Chemical and Thermal Conditions: A Computational Study*" has been carried out by Mr. Saikat Pal for the Degree of Doctor of Philosophy under my supervision and the same has not been submitted elsewhere for a degree.

Prof. Sandip Paul
Thesis Supervisor
Department of Chemistry
Indian Institute of Technology Guwahati
Guwahati-781039, India

Preface

On accomplishment of my doctorate studies, I would like to express my sincere gratitude to everyone who directly or indirectly helped me during this wonderful journey of my research career. Firstly, I would like to thank my thesis supervisor, Prof. Sandip Paul for guiding and supporting me over the years. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to my questions and queries so promptly. Prof. Paul introduced me to the world of computer simulations and was always willing to provide insightful comments on my research and encouragement throughout the research process. He gave me the freedom to pursue the research ideas I found most interesting. He also motivated me to pay careful attention to every detail of my research, and worked very hard to provide feedback on each of the articles, reports, and talks I have prepared over the past five years.

Besides my supervisor, I am highly obliged to the doctoral committee members, Prof. Gopal Das, Dr. Lal Mohan Kundu, Dr. Kalyan Raidongia for periodically assessing my work and providing valuable suggestions for its improvement. My sincere thanks go to all other faculty members in the department for their kind help at various stages of my doctoral work. I gratefully acknowledge the Ministry of Human Resource and Development (MHRD), India for financial support and IIT Guwahati for providing the research facilities to carry out my research work. I would like to appreciate the IIT Guwahati super-computing facility PARAM-ISHAN, without which the completion of my dissertation was not possible within this time period.

No words will be sufficient to thank my lab mates Dr. Srijita, Krishna, Rabindranath, Rituparna, Aritra, Madhusmita, and Rimjhim for their unstinting support, helpful discussions, and creating a pleasant working environment in the lab. A heartfelt thank to my seniors Dr. Subrata Paul, Dr. Bhanita Sharma, Dr. Gargi Borgohain, and Dr. Shubhadip Das for sharing their valuable suggestions and experiences during my Ph.D. days.

I take this opportunity to express my sincere thanks to all my teachers in school, college, and university days for helping and encouraging me in various aspects of my life and academics. All the learning from them will be an asset in every walk of my life.

I extend my sincere thanks and bundles of love to my friends Prasenjit, Sounak, Pritam, Avik, Shuvankar, Shubhajit, Sumanta Dr. Shubhrada, Dipanjan, Dr. Keshabda and Dr. Pradipda for listening to me, and being my true friends and mental support in different stages of my life.

Finally, this dream could not have been fulfilled without the endless love, support, and blessings from my family members. I extend my sincere gratitude to my mother for always being my best friend, philosopher, and guide. I am genuinely grateful to my parents for their great sacrifices and understandings, for helping me to be a better human being, and allowing me to fly high. And last but not least, I am blessed to have my brother as one of the strongest pillars of my life. A famous quote reads- Every journey begins with a single step- and today, only because of my family, I move a step uphill towards my zenith.

Saikat Pal

2021

"Natural science, does not simply describe and explain nature; it is part of the interplay between nature and ourselves."

— Werner Karl Heisenberg

Table of Contents

Chapter 1

Introduction

"The nucleic acids, as constituents of living organisms, are comparable in importance to proteins. There is evidence that they are involved in the processes of cell division and growth, that they participate in the transmission of hereditary characters, and that they are important constituents of viruses. An understanding of the molecular structure of the nucleic acids should be of value in the effort to understand the fundamental phenomena of life."

– Linus Pauling

1.1 What is G-quadruplex DNA?

Guanine (G) rich DNA can form different types of diverse structures such as hairpins, cruciforms, triplexes, and G-quadruplexes; on the basis of the conditions, such as, presence of cations, pH, or crowding conditions [1]. Among those different structures, non-canonical guanine-rich DNA (G-DNA) draws a special interest of the scientific world due to its potential therapeutic applications in the field of oncogenic research [2, 3]. G-quadruplex DNA is made of stacked planar arrangements [4–7], formed by four Hoogsteen hydrogen bonds of guanines. These structures are more stable in presence of monovalent (K^+ or

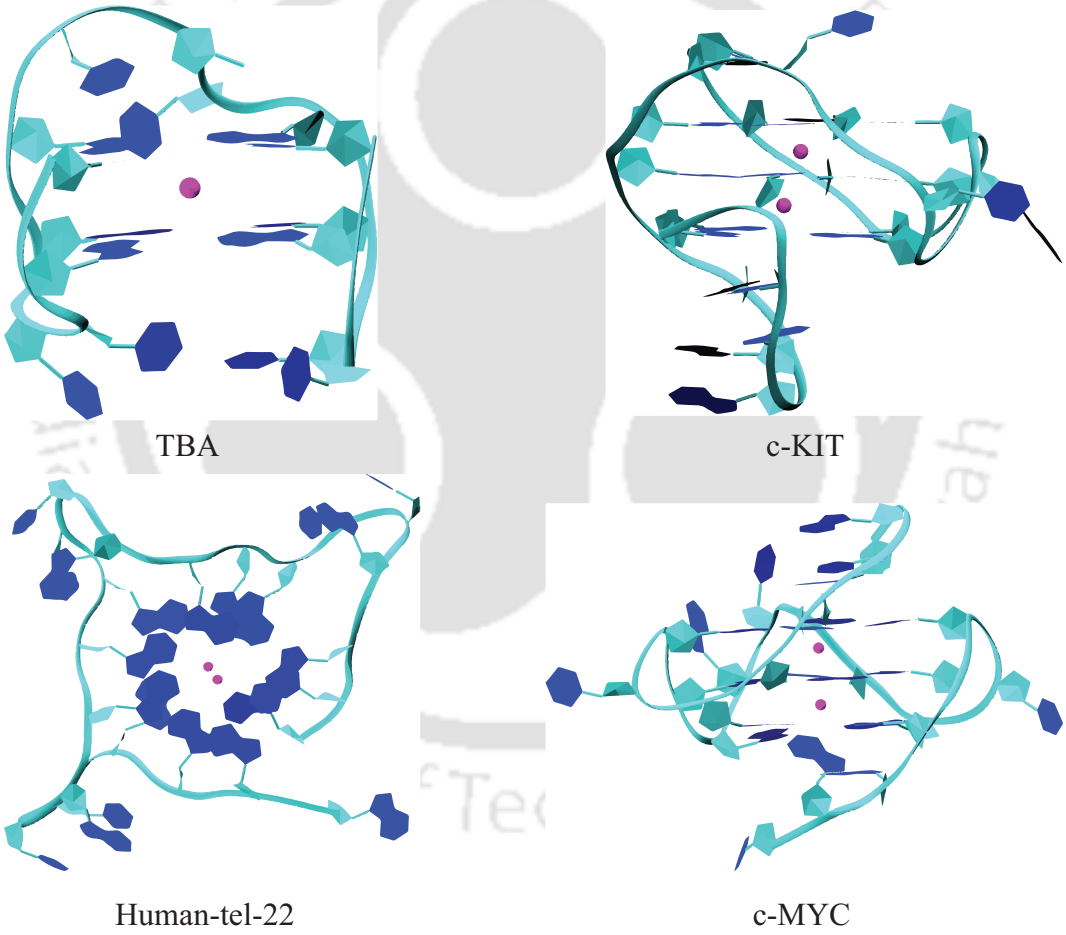


Figure 1.1: *Snapshot representations of different types of G-quadruplex DNAs.*

Na^+) or divalent (Sr^{2+}) cations [8–14]. Even before the human genome was fully mapped in 2004, many putative G-quadruplex forming sequences (PQFS) were found in human genomic DNA. G-quadruplex DNAs in telomere, which is the end of the non-coding portion of the chromosome and oncogene promoter region has been demonstrated in many

recent studies [15–18]. Telomerase which controls the replication of telomere region, was recognized [19, 20] mostly in cancer cells (over 85%). Various numbers of human proto-oncogenes have been identified to control the transcription of the gene regulation, such as c-KIT [15, 21], c-MYC [22, 23], VEGF [24, 25], HIF-1α [26], BCL-2 [27–29], KRAS [30], Rb [31], RET [32], hTERT [33], and PDGFR-β [34, 35]. These G-quadruplexes present in the promoter regions of different oncogene in human cells, attract the special attention in the growing field of therapeutic research [36–38].

In this thesis, we have studied the dynamics of G-quadruplex DNA with the following conditions: (i) different $SrCl_2$ salt concentrations, (ii) different KCl salt concentrations and temperatures under polar and non-polar confinement conditions, (iii) different deep eutectic solvent(DES) concentrations, (iv) different temperatures in DES medium. Moreover, we have also investigated the binding pathways and modes of a drug molecule on G-quadruplex DNA.

1.2 Role of Cations on G-quadruplex DNA

The interaction between cation and G-quadruplex DNA is primarily based on ions placed in the central channel formed by the tetrad. The location of the ions depends on the ion, the configuration of the G-quadruplex DNA and on the tetrad plane or the planes of the tetrad. Ions play a very important role in forming and stabilizing the G-quadruplex structure by screening the charge of O6 atom of tetrad residues and interacting with the grooves and loop residues. Most of the studies on conformational dynamics of G-quadruplex DNA are performed with Na^+ and K^+ ions since these are physiologically relevant monovalent cations. However, apart from these monovalent cations, other monovalent as well as divalent cations affect the stability and conformational changes of the G-quadruplex DNA. In 1992, Hardin et al. reported [39] the ions which stabilize the structure of G-quadruplex DNAs in the following order: $K^+ > Ca^{2+} > Na^+ > Mg^{2+} > Li^+$ and $K^+ > Rb^+ > Cs^+$. Moreover, in 1993, Venczel and Sen et al. suggested [8] the trend of monovalent as well as divalent cations separately as, $K^+ > Rb^+ > Na^+ > Li^+ = Cs^+$ and $Sr^{2+} > Ba^{2+} > Ca^{2+} > Mg^{2+}$. Even though Li^+ ion is commonly believed to destabilize G-quadruplex, its neutral role on the structure of G-quadruplex has been reported, which proves that Li^+ ion does not stabilize or destabilize G-quadruplex [40].

Apart from the physiologically relevant monovalent cations, K^+ and Na^+, there are many other cations which help in stabilizing the conformation of G-quadruplex, including

both monovalent cations, such as Rb^+, Cs^+, NH_4^+, and Tl^+ and divalent cations, like Sr^{2+}, Ba^{2+}, and Pb^{2+} [41–47]. However, the disruption of G-quadruplex is also caused by some divalent cations. Mn^{2+}, Co^{2+} and Ni^{2+}, at low concentrations, can destabilize G-quadruplex, even in the presence of K^+. Further reports of G-quadruplex disruption by other divalent cations such as Ca^{2+}, Co^{2+}, Mn^{2+}, Zn^{2+}, Ni^{2+}, and Mg^{2+}, at either very low concentrations of G-quadruplex and monovalent cations or high temperatures are also ascertained [48, 49]. Miyoshi et al. revealed [50] that the dimeric antiparallel G-quadruplex formed by $d(G4T4G4)_2$ is destabilized, prompting a transition to parallel G-quadruplex structure, by divalent cations, listed in the order, $Zn^{2+} > Co^{2+} > Mn^{2+} > Mg^{2+} > Ca^{2+}$. Regardless, depending upon the structures of G-quadruplex, some divalent cations also have the ability to stabilize G-quadruplexes.

There are numerous reports assessing the properties of divalent cations which are responsible for stabilizing G-quadruplexes. The ionic radii of divalent cations plays a major role in this matter which is proved in a series of biophysical studies on thrombin aptamer in various metal ions, where Sr^{2+} and Ba^{2+} result in thermodynamically more stable G-quadruplex than Mg^{2+} and Ca^{2+} [51]. Moreover, they also identified two dominant contributions of hydration in G-quadruplex folding, which are the dehydration of the guanine O6 and the cations and also the water absorbed when a single strand folds into a G-quadruplex. However, Vairamain et al. conducted an ESI-MS study on the same thrombin-binding aptamer, which did not demonstrate a smooth correlation between the size of the cation and the extent of adduct formation, thus implying that the ionic radii are not the only contributing factor in the stabilization of G-quadruplex [52]. Another study [53] by them on deoxyguanosine formed G-quadruplex in presence of alkali earth metal ions revealed that the bond formed between the metal and oxygen has a significant role in the G-quadruplex stabilization, due to which the following order of stability of G-quadruplex formed with divalent cations is obtained, $Sr^{2+} > Ba^{2+} > Pb^{2+} > Ca^{2+} >> Mg^{2+}$. Moreover, the order of the ability of alkaline earth metal ions to stabilize G-quadruplexes, suggested by a computational study of the interaction between G-tetrads and metal cations, is as follows: $Be^{2+} > Mg^{2+} > Ca^{2+}$. This suggested that the G-quadruplexes are closely bound to the smaller ions and hence the electrostatic interaction plays a dominant role in the cation-tetraplex systems [54]. Further a kinetic study on G-quadruplexes formation with divalent cations revealed that the oligonucleotide sequence, d(GTG3TAG3CG3TTG2), folds into a unimolecular G-quadruplex in presence of Pb^{2+} probably via a three-step pathway, which includes two intermediates [55].

The trivalent metal ions have been hardly used to study its effect on the stabilization of G-quadruplex structures. The trivalent cations have been reported to destabilize G-quadruplexes, but the involvement of metal chelators reduces this effect [56]. On the other hand, a ESI-MS study revealed that the stacking of the quartets is induced by the trivalent lanthanide metal ions (La^{3+}, Eu^{3+}, Tb^{3+}, Dy^{3+}, Tm^{3+}) [57].

1.3 Role of Deep Eutectic Solvents on G-quadruplex DNA

The G-quadruplex DNA is a potential candidate to function as a biosensor, catalyst, nanomachine, and in the construction of functional DNA nanostructure, due to its diverse conformations and robust physiochemical nature, in addition to being physiologically significant [58–63]. However, most of the applications of G-quadruplex DNA are observed to be in aqueous media, although a large number of substantial chemical reactions and devices have been carried out in purely water-free medium [64]. Ito et al. recently reported a covalent attachment between G-quadruplex DNA and poly(ethylene glycol) (PEG) moiety, in order to act as a catalyst in organic solvents [65]. Hence probing for an anhydrous medium which favors G-quadruplex DNA might result in novel approach to DNA-based applications. In last two decades, room-temperature ionic liquids (RTILs), have drawn tremendous attention to the scientific world, due to its a wide range application on chemical reactions [66–70]. These RTILs, which are regarded as green solvents, have been utilized in many reactions, providing a suitable medium [71]. The applications of RTILs have been also recently extended to the biological field, such as, catalytic reactions, biosensors, protein stabilization, and nucleic acids storage [72–74]. The long-term stability of DNA in hydrated ILs is noteworthy, especially the greater stability of A-T base pairs than G-C base pairs for duplex DNA [75, 76]. A new generation of ionic fluids was discovered by Abbott et al. [77], deep eutectic solvents (DESs), which have similar physiochemical properties as that shown by traditional ionic liquids [78]. As compared to the RTILs, DESs are more efficient due to the fact that they are budget friendly, more synthetically accessible, nontoxic, and biodegradable. The properties of various DNA and RNA secondary structures were studied in anhydrous and high-viscous DESs by Hud and co-workers [79]. They also investigated the folding of G-quadruplex in both anhydrous DES and mixed DES-water solvents, which concluded that the diffusion-limited kinetics is consistent with Kramers rate theory [80]. This implies that introduction of G-quadruplex DNAs in water-free ILs or

DESs might alter or improve the properties of DNAs and thus, can be eventually applied in anhydrous environments. Zhao et al. [81, 82] observed that various G-quadruplexes can form intra molecular, intermolecular, and even higher-order structures in anhydrous DES, even though parallel structure is the most favored conformation. Moreover, DES provided ultra-stability to G-quadruplex, compared to water medium.

1.4 Effect of Drugs on G-quadruplex DNA

The first G-quadruplex having a therapeutic significance was spotted in the human telomeric DNA at its 3′-end overhang [1], which consists of repetition of d-(TTAGGG) sequences and various related proteins [83–86]. DNA replication truncates the length of the telomere sequence, limiting cell growth and proliferation. However, telomerase, which is a reverse transcriptase, has the potential to add nucleotides to the telomeric end of the cancer cells, due to which the cells are not exhausted [87]. Thus, the stabilization of G-quadruplex by ligands, subsequently obstructing the telomerase activity, as a novel anti-cancer therapeutic strategy is currently under research [88, 89].

In addition, the G-quadruplex sequences in the promoter of cancer-related genes have also been the center of attraction for potential antitumor therapeutic agents [2]. One such ligand which binds with G-quadruplex, quarfloxin, has passed the Phase II trials in treating various tumors, such as, carcinoid tumors, neuroendocrine tumors and lymphoma [90]. The G-quadruplex-nucleolin complexes of DNA ribosomes in the nucleolus is destabilized by quarfloxin, causing redistribution of nucleolin into the nucleoplasm. Quarfloxin then interacts with G-quadruplex in the promoter of c-MYC proto-oncogene, prohibiting its gene expression [2]. However, due to high binding of albumin, the Phase III trials for quarfloxin is at a halt [91]. The stabilization of G-quadruplex formation in other tumor related genes, such as, hTERT [33], c-KIT [15], KRAS [92, 93], BCL2 [28], and VEGF [24], in transcriptional regulation, by small molecules impairs promoter activity and, thus, destroys tumors.

The structure at the ends of a chromosome is known as the telomere, and it consists of repeated microsatellite sequences and is capped by Shelterin complexes (POT1 and TRF2), which are required for maintaining and protecting telomeres from DNA repair mechanisms and regulating the activity of telomerase in mammals [86]. A single microsatellite repeat sequence of $(GGGTTA)_n$ is present in the human telomeric region at its 3′-end overhang, containing about 200 ± 75 nucleotides. About 50-200 base pairs

are curtailed off from the telomeric region after every cell replication, which eventually results in cell senescence and apoptosis as the telomere is exhausted and the Hayflick limit is reached [94, 95]. The single strand structure in this region favors G-quadruplex formation, which initially has been of interest to the scientific community. Interestingly, G-quadruplex-binding proteins, such as, ATRX, POT1, TPP1 act in the telomere processing-like recruitment of a specific protein, while the telomere G-quadruplex structures are unwounded by various G-quadruplex helicases, like Pif1 and BLM to stabilize the shelterin complex [96, 97]. Moreover, over expression of a protein complex responsible for adding telomere sequences, telomerase, in various tumor cells elongates the telomeric end of the cancer cells and thus immortalize them. Now the folding of the 3′ overhang into a G-quadruplex unable the telomerase to hybridize the telomere, which inhibits telomerase activity and consequently, cell apoptosis [3]. Moreover, the guanidine-rich 3′ overhang formed G-quadruplex telomere intercepts the binding of telomere protection proteins, leading to chromosomal fusions and stimulation of cell apoptosis [1, 98]. Hence small molecules binding with telomere G-quadruplex has been the therapeutic target to prevent telomerase activity in tumor cells [88, 89, 99–101]. Telomerase inhibitors, such as Telomestatin and 2,6-diamidoanthraquinone derivatives binds to G-quadruplex telomers at an early stage [88, 89, 99], while RHPS4 interferes with the shelterin complexes, promoting telomere malfunction in mammalian cancer cells [102]. The canonical telomere G-quadruplex also give rise to numerous higher-order structures at the 3′-end overhang [103–106].

c-MYC encodes a multi functional transcription factor, which function not only as a transcription activator of cell proliferation genes, but also as a transcription repressor of some other genes accountable for arresting growth [107, 108]. The crucial cellular functions, such as, metastatic capacity, metabolic transformation and cell proliferation involves a variety of c-MYC-responsive genes [109]. The c-MYC protein function is mainly activated via upstream oncogenic pathways and is over expressed in tumor cells (almost 80% of solid tumors). Hence maintaining the regulation of the gene is a plausible therapeutic approach against cancer [107, 108]. Nevertheless, the short half-life, large dimension, and unstructured nature of c-MYC makes it a difficult target for binding with small drug molecules [2]. The preferential stabilization of c-MYC G-quadruplex over G-quadruplexes in other promoters (c-KIT1, c-KIT2, and BCL2) and also in duplex DNA by a crescent-shaped thiazole peptide, TH3, has been reported by Dash and colleagues [110]. The K_D values in fluorimetric titration studies further validate that the binding affinity of TH3 to c-MYC G-quadruplex increases by 4-5-fold ($K_D \sim 0.3\ \mu M$). The binding preference

has also been approved biologically. A new molecule, IZCZ-3, which is a conjugate of diaryl-substituted imidazole and a carbazole moiety, has been developed by Tan and his group, who reported the preferential binding and stabilization of c-MYC G-quadruplex by this four-leaf-clover-like molecule [111]. Earlier, the triaryl-substituted imidazole scaffold was found to be selective to G-quadruplex, and its derivatives were used to distinguish different G-quadruplexes. The carbazole moiety and its derivatives, on the other hand, exhibited crucial binding preference toward c-MYC G-quadruplex [111–114]. Hence, linking the triaryl-substituted imidazole unit with a carbazole moiety resulted in high affinity and selectivity of c-MYC G-quadruplex ligands, as expected by the researchers.

The c-KIT oncogene is involved in encoding a receptor tyrosine kinase, which is bridged and activated by the binding of dimerized stem cell factors (SCF), and it further leads to proliferation, differentiation, and survival in hemopoietic precursor cells [115–117]. Overexpression or mutations causes c-KIT to dysfunction, which is related to various diseases, such as mastocytosis, acute myelogenous leukemia (AML) and gastrointestinal stromal tumors (GIST) [118–120]. Long term exposure of a drug for GIST, the kinase inhibitor Imatinib (Glivec), which even though is approved by FDA, results in secondary mutations at exons 13, 14, or 17, which encodes tyrosine kinase domains [121]. The mutations at exon 17 causes resistance to the drug and weakens its therapeutic effect [122]. Hence focus on the procedure for the suppression of c-KIT expression is of utmost importance. Balasubramanian and his colleagues reported two ligands, which consists of three N,N-dimethyl amine-substituted and one N,N-dimethyl amine/2 fluorine-substituted tails, respectively, that preferably bind with the c-KIT G-quadruplexes rather than the telomere G-quadruplex in an SPR-binding assay. Moreover, the ligands also inhibit the gene expression in two different cancer cell lines, as demonstated by RT-qPCR assays [123].

In this thesis, we have used the following G-quadruplex DNAs: (i) Thrombin binding aptamer (TBA), (ii) human telomeric G-quadruplex DNA, (iii) c-KIT G-quadruplex DNA.

Thrombin binding aptamer (TBA) with a sequence of 15-mer d(GGTTGGTGTGGTTGG) is the simplest G-quadruplex DNA and it binds with thrombin protein (**Fig. 1.2**). Thus, the important role of TBA in cell biology attracts the scientific world in the field of molecular biology. The structure of TBA G-quadruplex DNA is already well established by NMR and X-ray crystallography studies [124–128]. These studies reveal that TBA contain two guanine tetrads (namely tetrad-1 and tetrad-2), two TT loops (i.e loop-1 and loop-2) and one TGT loop (loop-3) and it can fold into an

anti-parallel chair like structure.

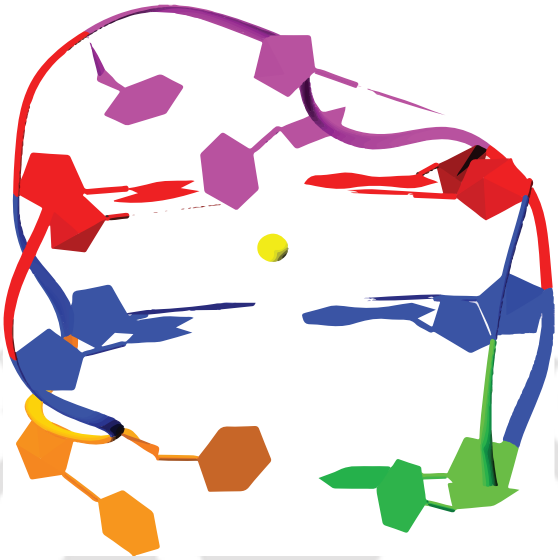


Figure 1.2: *Snapshot representation of TBA G-quadruplex DNA.*

A parallel topology of 22 mer human telomeric quadruplex DNA (d[AGGG(TTAGGG)$_3$]) consists of three tetrads (namely tetrad-1, tetrad-2, and tetrad-3), three propeller loops (thymine-thymine-adenine) and an adenine-1 flanking base (**Fig. 1.3**).

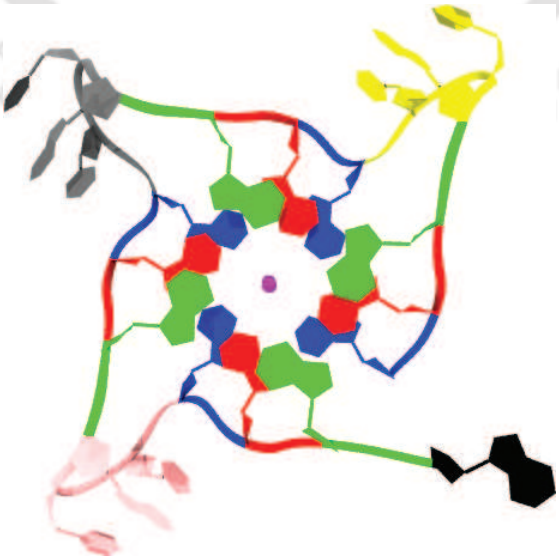


Figure 1.3: *Snapshot representation of human telomeric G-quadruplex DNA.*

The proto-oncogene human c-KIT G-quadruplex DNA consists of 22 mer d(AGGGAGGGCGCTGGGAGGGAGGG) sequence (**Fig. 1.4**). c-KIT DNA contains three G-tetrads (namely tetrad-1, tetrad-2 and tetrad-3) along with four loop regions (namely loop-1, loop-2, loop-3, loop-4). Moreover, loop-1 and loop-2 contain adenosine-5 (A5) and cytosine-9 (C9) nucleoside bases respectively, and loop-3 has cytosine-11 (C11) and thymine-12 (T12) nucleoside bases. Furthermore, c-KIT G-quadruplex DNA has one long loop (namely loop-4) which consists of A16-G17-G18-A19-G20 nucleoside bases.

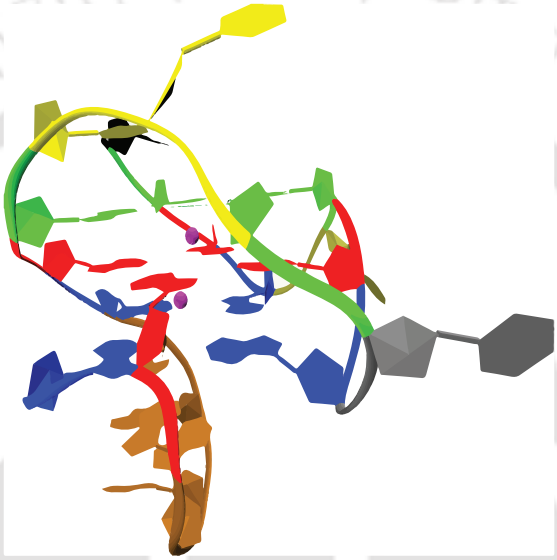


Figure 1.4: *Snapshot representation of c-KIT G-quadruplex DNA.*

1.5 Methodology

In this thesis, we have used classical MD simulation technique. It has been widely used to investigate the structure and dynamics of biomolecular systems, such as proteins, nucleic acids, and small molecules like amino acids, sugars and drugs. In MD simulation, the potential energy function (U) is described by all interactions between the atoms that are covalently bonded as well as non-bonded interactions between atoms and molecules in the condensed phase. The interactions between particles are governed by the so-called force field parameterization [129].

The potential energy function is written as a sum of bonded and non-bonded interaction terms

$$U = U_{bond} + U_{angle} + U_{dihedral} + U_{vdw} + U_{Coulomb} \tag{1.1}$$

The first three terms (U_{bond}, U_{angle}, $U_{dihedral}$) are the bonded terms, which describe the bond stretching, angle bending, and torsion rotation, and the last two terms are for the non-bonded potential. In bonded terms, the bond and angle contributions are described by harmonic potentials and all of the interactions between directly bonded atoms (1-2 interactions), angles (1-3 interactions, where two atoms bonded to a common atom), and torsion (interactions between pairs of 1-4 atoms) are defined as:

$$U_{bond} = \sum_{bonds} K_b(b_{ac} - b_{eq})^2 \tag{1.2}$$

$$U_{angle} = \sum_{angles} K_\theta(\theta_{ac} - \theta_{eq})^2 \tag{1.3}$$

$$U_{dihedral} = \sum_{dihedrals} \frac{V_n}{2}(1 + cos(n\phi - \delta)) \tag{1.4}$$

The letters b, θ, ϕ, and δ represent the bond length, bond angle, dihedral angle, and phase angle, respectively. The subscripts ac stands for actual and eq stands for equilibrium. The parameters K_b, K_θ, and V_n are the force constants for bond, bond angle, and dihedral angle, respectively.

The non-bonded potentials are calculated using two terms, the first one is the Lennard-Jones term (U_{vdw}) [130] describing the van der Waals interaction [131], and the second one is the Coulomb term ($U_{coulomb}$) [132] that deals with the electrostatic interactions between particles having partial charges on them. The non-bonding interaction terms are defined as:

$$U_{vdw} = \sum_i \sum_{i<j} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{1.5}$$

$$U_{coulomb} = \sum_i \sum_{i<j} [\frac{q_i q_j}{4\pi\epsilon_o r_{ij}}] \tag{1.6}$$

where the overall sum is over all the atom pairs i and j. Lennard-Jones parameters σ and ϵ are the diameter of atomic sites and well depth energy, respectively. r_{ij} is the inter-atomic distance. q_i and q_j are the partial charges on interaction sites i and j and ϵ_o is the electrical permittivity.

The aim of the MD simulation is to observe the evolution of atomic coordinates in time. We consider an N-particle system characterized by the following Hamiltonian

$$H = \sum_{i=1}^{N} \frac{p_i^2}{2m} + U(\mathbf{r}^N) \tag{1.7}$$

where m is the mass of each particle, p_i is the momentum of the i-th particle and $U(\mathbf{r}^N)$ is the total potential energy of the system which includes the all particle-particle interactions. The coordinates of the particles are denoted by $\mathbf{r}^N = \{\mathbf{r}_1, \cdots, \mathbf{r}_N\}$. The position and velocity of i-th particle is represented by $\mathbf{r}_i$ and $\mathbf{v}_i$, respectively. The method of molecular dynamics consists of solving the equation

$$a_i = \frac{\mathbf{F_i}}{m_i} \tag{1.8}$$

where $i = 1, 2, ...N$, m_i is the mass of i-th particle and $\mathbf{F}_i$ is the force acting on particle i. This equation is obtained easily from the Lagrangian

$$L = \frac{1}{2}\sum_{i=1}^{N} m_i \mathbf{v_i}.\mathbf{v_i} - \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} u(r_{ij}) \tag{1.9}$$

where the potential U has been assumed to be the sum of pair potentials u_{ij}. The Lagrangian equation of motion is

$$\frac{d}{dt}(\frac{\partial L}{\partial \dot{q}_i}) - \frac{\partial L}{\partial \dot{q}_i} = 0 \tag{1.10}$$

It is clear from Eq. 1.10 that the dynamics of particles is described by 3N number of second order differential equations.

It is also possible to write down the Hamiltonian (H) for the system and solve the the Hamiltonian equations of motion

$$\dot{\mathbf{q}}_k = \frac{\partial H}{\partial p_k} \tag{1.11}$$

$$\dot{\mathbf{p}}_k = -\frac{\partial H}{\partial q_k} \tag{1.12}$$

where $\mathbf{q}_k$ and $\mathbf{p}_k$ represent generalized coordinates and momenta. For a system with pairwise interaction potential, the Hamiltonian is

$$H = \frac{1}{2}\sum_{i=1}^{N} m_i \mathbf{v_i}.\mathbf{v_i} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} u(r_{ij}) \tag{1.13}$$

and Eqs. 1.11 and 1.12 yield

$$\frac{dr_i}{dt} = \frac{\mathbf{p_i}}{m_i} \tag{1.14}$$

$$-\dot{\mathbf{p}}_i = -\nabla \mathbf{u} = \mathbf{F_i} \tag{1.15}$$

where i=1,2,....N. There are now $6N$ first order differential equations to be solved.

The equation of motion is solved numerically to yield particle velocities and positions as a function of time. It is usually integrated by using finite difference approach. The

Verlet algorithm is one of the most commonly used algorithm for this purpose [133]. The advantage of the use of Verlet algorithm is that its implementation is straightforward and storage requirements are modest. Although, it has the disadvantage of moderate precision during the calculation and velocity does not appear explicitly in the Verlet integration. As an improvement to the Verlet algorithm, the leap-frog algorithm [134] has been developed. But, it has a disadvantage that the positions and velocities are not synchronized. As an alternative of Verlet or the leapfrog algorithm, Velocity Verlet algorithm has been developed and the following relations are used to calculate new position and velocity at the same time:

$$r(t+dt) = r(t) + v(t)dt + \frac{1}{2}a(t)dt^2 \tag{1.16}$$

$$v(t+dt) = v(t) + \frac{1}{2}[a(t) + a(t+dt)]dt \tag{1.17}$$

To calculate the velocities at time t+dt, this method requires acceleration at time t and t+dt. In the present work, we have employed Velocity Verlet algorithm.

Next, we have calculated ΔH_{bind} using MM-GBSA [135] method. ΔH_{bind} is calculated with the help of series of following equations:

$$\Delta H_{bind} = \Delta E_{vac} + \Delta G_{solv} \tag{1.18}$$

where ΔE_{vac} is the energy in vacuum (gas phase) and ΔG_{solv} represents the solvation free energy [136, 137]. They can be calculated as:

$$\Delta E_{vac} = \Delta E_{ele} + \Delta E_{vdw} \tag{1.19}$$

$$\Delta G_{solv} = \Delta G_{GB} + \Delta G_{NP} \tag{1.20}$$

where ΔE_{ele} and ΔE_{vdw} are the electrostatic and van der Waals components of energies. The polar component, ΔG_{GB}, is estimated using generalized-Born (GB) approach [138] and the nonpolar contribution, ΔG_{NP}, of the solvation free energy ΔG_{solv} is calculated as follows [136, 138]:

$$\Delta G_{NP} = \gamma(SASA) + \beta \tag{1.21}$$

where γ=0.005 kcal/Å^2 and β=0.0, and SASA represents the solvent accessible surface area.

1.6 Present Work

Nucleic acid sequences containing series of guanine can form G-quadruplex structures, which are of special interest because of their high thermal stability, resistance to denaturing conditions, and stiffness. Moreover, due to the diverse and polymeric nature, it draws the attention of the scientific world in significant and diverse fields.

Chapter 1 introduces the formation and stability of the G-quadruplex DNA in different chemical and thermal conditions along with its applications through literature survey. In **Chapter 2**, we have reported about the conformational changes of the thrombin-binding G-quadruplex aptamer (TBA) under dissimilar $SrCl_2$ salt concentrations. **Chapter 3** deals with the conformational changes of the human telomeric G-quadruplex DNA at varying KCl salt concentrations and temperatures under non-polar and polar confinements. In **Chapter 4**, we have discussed about the effect of hydrated and non-hydrated choline chloride-urea deep eutectic solvent (namely reline) on TBA G-quadruplex DNA. **Chapter 5** describes the role of reline, a natural DES, on temperature induced conformational changes of c-KIT G-quadruplex DNA. In **Chapter 6**, we have investigated the binding modes and pathway of APTO-253 on c-KIT G-quadruplex DNA. In the last chapter i.e., **Chapter 7** we have summarized the overall findings to reflect upon the interpretation about the conformational deviation of the different types of G-quadruplex DNAs under different chemical and thermal conditions and its drug binding pathway.

References

[1] O. Doluca, J. M. Withers and V. V. Filichev, *Chem. Rev.*, 2013, **113**, 3044–3083.

[2] S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat. Rev. Drug Discov.*, 2011, **10**, 261–275.

[3] A. M. Zahler, J. R. Williamson, T. R. Cech and D. M. Prescott, *Nature*, 1991, **350**, 718–720.

[4] P. Stadlbauer, M. Krepl, I. Cheatham, Thomas E., J. Koča and J. Šponer, *Nucleic Acids Res.*, 2013, **41**, 7128–7143.

[5] P. Hazel, J. Huppert, S. Balasubramanian and S. Neidle, *J. Am. Chem. Soc.*, 2004, **126**, 16405–16415.

[6] N. H. Campbell and S. Neidle, in *G-Quadruplexes and Metal Ions*, ed. A. Sigel, H. Sigel and R. K. O. Sigel, Springer Netherlands, Dordrecht, 2012, pp. 119–134.

[7] S. Haider and S. Neidle, in *Molecular Modeling and Simulation of G-Quadruplexes and Quadruplex-Ligand Complexes*, ed. P. Baumann, Humana Press, Totowa, NJ, 2010, pp. 17–37.

[8] E. A. Venczel and D. Sen, *Biochemistry*, 1993, **32**, 6220–6228.

[9] S. Chowdhury and M. Bansal, *J. Phys. Chem. B*, 2001, **105**, 7572–7578.

[10] N. Špačková, I. Berger and J. Šponer, *J. Am. Chem. Soc.*, 1999, **121**, 5519–5534.

[11] B. Islam, P. Stadlbauer, M. Krepl, J. Koca, S. Neidle, S. Haider and J. Sponer, *Nucleic Acids Res.*, 2015, **43**, 8673–8693.

[12] T. Cragnolini, D. Chakraborty, J. Sponer, P. Derreumaux, S. Pasquali and D. J. Wales, *J. Chem. Phys.*, 2017, **147**, 152715.

[13] S. Pal and S. Paul, *Int. J. Biol. Macromol.*, 2019, **121**, 350 – 363.

[14] J. V. Burda, J. Šponer and P. Hobza, *J. Phys. Chem.*, 1996, **100**, 7250–7255.

[15] S. Rankin, A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian and S. Neidle, *J. Am. Chem. Soc.*, 2005, **127**, 10584–10589.

[16] T. Simonsson, P. Pecinka and M. Kubista, *Nucleic Acids Res.*, 1998, **26**, 1167–1172.

[17] H. Sun, J. K. Karow, I. D. Hickson and N. Maizels, *J. Biol. Chem.*, 1998, **273**, 27587–27592.

[18] H. Sun, A. Yabuki and N. Maizels, *Proc. Nat. Acad. Sc.*, 2001, **98**, 12444–12449.

[19] N. Kim, M. Piatyszek, K. Prowse, C. Harley, M. West, P. Ho, G. Coviello, W. Wright, S. Weinrich and J. Shay, *Science*, 1994, **266**, 2011–2015.

[20] G. W. Collie and G. N. Parkinson, *Chem. Soc. Rev.*, 2011, **40**, 5867–5892.

[21] H. Fernando, A. P. Reszka, J. Huppert, S. Ladame, S. Rankin, A. R. Venkitaraman, S. Neidle and S. Balasubramanian, *Biochemistry*, 2006, **45**, 7854–7860.

[22] T. Simonsson, M. Kubista and P. Pecinka, *Nucleic Acids Res.*, 1998, **26**, 1167–1172.

[23] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 11593–11598.

[24] D. Sun, K. Guo, J. J. Rusche and L. H. Hurley, *Nucleic Acids Res.*, 2005, **33**, 6070–6080.

[25] P. Agrawal, E. Hatzakis, K. Guo, M. Carver and D. Yang, *Nucleic Acids Res.*, 2013, **41**, 10584–10592.

[26] R. De Armond, S. Wood, D. Sun, L. H. Hurley and S. W. Ebbinghaus, *Biochemistry*, 2005, **44**, 16341–16350.

[27] J. Dai, T. S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R. A. Jones and D. Yang, *J. Am. Chem. Soc.*, 2006, **128**, 1096–1098.

[28] T. S. Dexheimer, D. Sun and L. H. Hurley, *J. Am. Chem. Soc.*, 2006, **128**, 5404–5415.

[29] P. Agrawal, C. Lin, R. I. Mathad, M. Carver and D. Yang, *J. Am. Chem. Soc.*, 2014, **136**, 1750–1753.

[30] S. Cogoi and L. E. Xodo, *Nucleic Acids Res.*, 2006, **34**, 2536–2549.

[31] Y. Xu and H. Sugiyama, *Nucleic Acids Res.*, 2006, **34**, 949–954.

[32] K. Guo, A. Pourpak, K. Beetz-Rogers, V. Gokhale, D. Sun and L. H. Hurley, *J. Am. Chem. Soc.*, 2007, **129**, 10220–10228.

[33] S. L. Palumbo, S. W. Ebbinghaus and L. H. Hurley, *J. Am. Chem. Soc.*, 2009, **131**, 10878–10891.

[34] R. V. Brown, T. Wang, V. R. Chappeta, G. Wu, B. Onel, R. Chawla, H. Quijada, S. M. Camp, E. T. Chiang, Q. R. Lassiter, C. Lee, S. Phanse, M. A. Turnidge, P. Zhao, J. G. N. Garcia, V. Gokhale, D. Yang and L. H. Hurley, *J. Am. Chem. Soc.*, 2017, **139**, 7456–7475.

[35] Y. Qin, J. S. Fortin, D. Tye, M. Gleason-Guzman, T. A. Brooks and L. H. Hurley, *Biochemistry*, 2010, **49**, 4208–4219.

[36] G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nat. Chem.*, 2013, **5**, 182–186.

[37] G. Biffi, M. Di Antonio, D. Tannahill and S. Balasubramanian, *Nat. Chem.*, 2014, **6**, 75–80.

[38] P. Murat and S. Balasubramanian, *Curr. Opin. Genet. Dev.*, 2014, **25**, 22 – 29.

[39] C. C. Hardin, T. Watson, M. Corregan and C. Bailey, *Biochemistry*, 1992, **31**, 833–841.

[40] W. A. Harrell Jr, *Quadruplex Nucleic Acids*, The Royal Society of Chemistry, 2006, pp. i–302.

[41] J. S. Lee, *Nucleic Acids Res.*, 1990, **18**, 6057–6060.

[42] F. M. Chen, *Biochemistry*, 1992, **31**, 3769–3776.

[43] E. A. Venczel and D. Sen, *Biochemistry*, 1993, **32**, 6220–6228.

[44] N. Nagesh and D. Chatterji, *Journal of Biochemical and Biophysical Methods*, 1995, **30**, 1–8.

[45] S. Basu, A. A. Szewczak, M. Cocco and S. A. Strobel, *J. Am. Chem. Soc.*, 2000, **122**, 3240–3241.

[46] I. V. Smirnov, F. W. Kotch, I. J. Pickering, J. T. Davis and R. H. Shafer, *Biochemistry*, 2002, **41**, 12133–12139.

[47] M. Cai, X. Shi, V. Sidorov, D. Fabris, Y. fai Lam and J. T. Davis, *Tetrahedron*, 2002, **58**, 661–671.

[48] S. Blume, *Nucleic Acids Res.*, 1997, **25**, 617–625.

[49] C. C. Hardin, A. G. Perry and K. White, *Biopolymers*, 2000, **56**, 147–194.

[50] D. Miyoshi, A. Nakao, T. Toda and N. Sugimoto, *FEBS Lett.*, 2001, **496**, 128–133.

[51] B. I. Kankia and L. A. Marky, *J. Am. Chem. Soc.*, 2001, **123**, 10799–10804.

[52] M. Vairamani and M. L. Gross, *J. Am. Chem. Soc.*, 2003, **125**, 42–43.

[53] M. Sravani, V. Nagaveni, S. Prabhakar and M. Vairamani, *Rapid Commun. Mass Spectrom.*, 2011, **25**, 2095–2098.

[54] P. Deepa, P. Kolandaivel and K. Senthilkumar, *Comput. Theor. Chem.*, 2011, **974**, 57–65.

[55] W. Liu, H. Zhu, B. Zheng, S. Cheng, Y. Fu, W. Li, T.-C. Lau and H. Liang, *Nucleic Acids Res.*, 2012, **40**, 4229–4236.

[56] H. Lu, S. Li, J. Chen, J. Xia, J. Zhang, Y. Huang, X. Liu, H. chen Wu, Y. Zhao, Z. Chai and Y. Hu, *Metallomics*, 2015, **7**, 1508–1514.

[57] I. C. M. Kwan, Y.-M. She and G. Wu, *Chem. Commun.*, 2007, 4286.

[58] S. Roe, D. J. Ritson, T. Garner, M. Searle and J. E. Moses, *Chem. Commun.*, 2010, **46**, 4309–4311.

[59] P. Alberti and J.-L. Mergny, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 1569–1573.

[60] J. J. Li and W. Tan, *Nano Lett.*, 2002, **2**, 315–318.

[61] Y. Xiao, V. Pavlov, T. Niazov, A. Dishon, M. Kotler and I. Willner, *J. Am. Chem. Soc.*, 2004, **126**, 7430–7431.

[62] C. Zhao, L. Wu, J. Ren and X. Qu, *Chem. Commun.*, 2011, **47**, 5461–5463.

[63] K. Dutta, T. Fujimoto, M. Inoue, D. Miyoshi and N. Sugimoto, *Chem. Commun.*, 2010, **46**, 7772–7774.

[64] Z. Wang, L. Zhu, F. Yin, Z. Su, Z. Li and C. Li, *J. Am. Chem. Soc.*, 2012, **134**, 4258–4263.

[65] H. Abe, N. Abe, A. Shibata, K. Ito, Y. Tanaka, M. Ito, H. Saneyoshi, S. Shuto and Y. Ito, *Angew. Chem. Int. Ed.*, 2012, **51**, 6475–6479.

[66] T. Welton, *Chem. Rev.*, 1999, **99**, 2071–2084.

[67] J. G. Huddleston, A. E. Visser, W. M. Reichert, H. D. Willauer, G. A. Broker and R. D. Rogers, *Green Chem.*, 2001, **3**, 156–164.

[68] R. Sheldon, *Chem. Commun.*, 2001, 2399–2407.

[69] R. D. Rogers and K. R. Seddon, *Science*, 2003, **302**, 792–793.

[70] T. Welton, *Coord. Chem. Rev.*, 2004, **248**, 2459 – 2477.

[71] Z. Yang and W. Pan, *Enzyme Microb. Technol.*, 2005, **37**, 19 – 28.

[72] F. van Rantwijk and R. A. Sheldon, *Chem. Rev.*, 2007, **107**, 2757–2785.

[73] K. Fujita and H. Ohno, *Chem. Commun.*, 2012, **48**, 5751–5753.

[74] K. Fujita, D. R. MacFarlane and M. Forsyth, *Chem. Commun.*, 2005, 4804–4806.

[75] R. Vijayaraghavan, A. Izgorodin, V. Ganesh, M. Surianarayanan and D. MacFarlane, *Angew. Chem. Int. Ed.*, 2010, **49**, 1631–1633.

[76] H. Tateishi-Karimata and N. Sugimoto, *Angew. Chem. Int. Ed.*, 2012, **51**, 1416–1419.

[77] A. P. Abbott, G. Capper, D. L. Davies, R. K. Rasheed and V. Tambyrajah, *Chem. Commun.*, 2003, 70–71.

[78] A. P. Abbott, D. Boothby, G. Capper, D. L. Davies and R. K. Rasheed, *J. Am. Chem. Soc.*, 2004, **126**, 9142–9147.

[79] I. Mamajanov, A. Engelhart, H. Bean and N. Hud, *Angew. Chem. Int. Ed.*, 2010, **49**, 6310–6314.

[80] F. M. Lannan, I. Mamajanov and N. V. Hud, *J. Am. Chem. Soc.*, 2012, **134**, 15324–15330.

[81] C. Zhao, J. Ren and X. Qu, *Langmuir*, 2013, **29**, 1183–1191.

[82] C. Zhao and X. Qu, *Methods*, 2013, **64**, 52 – 58.

[83] W. J. Chung, B. Heddi, M. Tera, K. Iida, K. Nagasawa and A. T. Phan, *J. Am. Chem. Soc.*, 2013, **135**, 13495–13501.

[84] R. K. Moyzis, J. M. Buckingham, L. S. Cram, M. Dani, L. L. Deaven, M. D. Jones, J. Meyne, R. L. Ratliff and J. R. Wu, *Proc. Natl. Acad. Sci. U.S.A.*, 1988, **85**, 6622–6626.

[85] W. E. Wright, V. M. Tesmer, K. E. Huffman, S. D. Levene and J. W. Shay, *Genes Dev.*, 1997, **11**, 2801–2809.

[86] T. De Lange, *Genes Dev.*, 2005, **19**, 2100–2110.

[87] A. D. Moorhouse, A. M. Santos, M. Gunaratnam, M. Moore, S. Neidle and J. E. Moses, *J. Am. Chem. Soc.*, 2006, **128**, 15972–15973.

[88] K. Shin-ya, K. Wierzba, K.-i. Matsuo, T. Ohtani, Y. Yamada, K. Furihata, Y. Hayakawa and H. Seto, *J. Am. Chem. Soc.*, 2001, **123**, 1262–1263.

[89] M.-Y. Kim, H. Vankayalapati, K. Shin-ya, K. Wierzba and L. H. Hurley, *J. Am. Chem. Soc.*, 2002, **124**, 2098–2099.

[90] D. Drygin, A. Siddiqui-Jain, S. O'Brien, M. Schwaebe, A. Lin, J. Bliesath, C. B. Ho, C. Proffitt, K. Trent, J. P. Whitten, J. K. C. Lim, D. Von Hoff, K. Anderes and W. G. Rice, *Cancer Res.*, 2009, **69**, 7653–7661.

[91] R. K. Morgan and T. A. Brooks, in *Small-molecule Transcription Factor Inhibitors in Oncology*, The Royal Society of Chemistry, 2019, pp. 169–193.

[92] S. Cogoi, F. Quadrifoglio and L. E. Xodo, *Biochemistry*, 2004, **43**, 2512–2523.

[93] R. K. Morgan, H. Batra, V. C. Gaerig, J. Hockings and T. A. Brooks, *Biochim. Biophys. Acta Gene Regul. Mech.*, 2016, **1859**, 235 – 245.

[94] C. B. Harley, A. B. Futcher and C. W. Greider, *Nature*, 1990, **345**, 458–460.

[95] V. A. Zakian, *Science*, 1995, **270**, 1601–1607.

[96] T. Tian, Y.-Q. Chen, S.-R. Wang and X. Zhou, *Chem*, 2018, **4**, 1314 – 1344.

[97] J. Amorim, G. Santos, J. Vinagre and P. Soares, *Genes*, 2016, **7**, 66.

[98] E. L. Denchi and T. de Lange, *Nature*, 2007, **448**, 1068–1071.

[99] D. Sun, B. Thompson, B. E. Cathers, M. Salazar, S. M. Kerwin, J. O. Trent, T. C. Jenkins, S. Neidle and L. H. Hurley, *J. Med. Chem.*, 1997, **40**, 2113–2116.

[100] R. T. Wheelhouse, D. Sun, H. Han, F. X. Han and L. H. Hurley, *J. Am. Chem. Soc.*, 1998, **120**, 3261–3262.

[101] S. Neidle, *FEBS J.*, 2010, **277**, 1118–1125.

[102] H. J. Lipps and D. Rhodes, *Trends Cell Biol.*, 2009, **19**, 414 – 422.

[103] D. J. E. Yue, K. W. Lim and A. T. Phan, *J. Am. Chem. Soc.*, 2011, **133**, 11462–11465.

[104] V. T. Mukundan and A. T. Phan, *J. Am. Chem. Soc.*, 2013, **135**, 5017–5028.

[105] J. Abraham Punnoose, Y. Cui, D. Koirala, P. M. Yangyuoru, C. Ghimire, P. Shrestha and H. Mao, *J. Am. Chem. Soc.*, 2014, **136**, 18062–18069.

[106] I. Russo Krauss, S. Ramaswamy, S. Neidle, S. Haider and G. N. Parkinson, *J. Am. Chem. Soc.*, 2016, **138**, 1226–1233.

[107] S. Adhikary and M. Eilers, *Nature Reviews Molecular Cell Biology*, 2005, **6**, 635–645.

[108] C. Dang, *Cell*, 2012, **149**, 22 – 35.

[109] D. M. Miller, S. D. Thomas, A. Islam, D. Muench and K. Sedoris, *Clin. Cancer Res.*, 2012, **18**, 5546–5553.

[110] D. Dutta, M. Debnath, D. Mller, R. Paul, T. Das, I. Bessi, H. Schwalbe and J. Dash, *Nucleic Acids Res.*, 2018, **46**, 5355–5365.

[111] M.-H. Hu, Y.-Q. Wang, Z.-Y. Yu, L.-N. Hu, T.-M. Ou, S.-B. Chen, Z.-S. Huang and J.-H. Tan, *J. Med. Chem.*, 2018, **61**, 2447–2459.

[112] S.-B. Chen, J.-H. Tan, T.-M. Ou, S.-L. Huang, L.-K. An, H.-B. Luo, D. Li, L.-Q. Gu and Z.-S. Huang, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 1004 – 1009.

[113] M.-H. Hu, S.-B. Chen, Y.-Q. Wang, Y.-M. Zeng, T.-M. Ou, D. Li, L.-Q. Gu, Z.-S. Huang and J.-H. Tan, *Biosens. Bioelectron.*, 2016, **83**, 77 – 84.

[114] S.-B. Chen, W.-B. Wu, M.-H. Hu, T.-M. Ou, L.-Q. Gu, J.-H. Tan and Z.-S. Huang, *Chem. Commun.*, 2014, **50**, 12173–12176.

[115] T. J. Shaw, E. J. Keszthelyi, A. M. Tonary, M. Cada and B. C. Vanderhyden, *Exp. Cell Res.*, 2002, **273**, 95 – 106.

[116] C. E. Edling and B. Hallberg, *Int. J. Biochem. Cell Biol.*, 2007, **39**, 1995 – 1998.

[117] L. K. Ashman, *Int. J. Biochem. Cell Biol.*, 1999, **31**, 1037 – 1051.

[118] S. C. Robertson, J. A. Tynan and D. J. Donoghue, *Trends Genet.*, 2000, **16**, 265 – 271.

[119] P. Blume-Jensen and T. Hunter, *Nature*, 2001, **411**, 355–365.

[120] C. L. Corless, J. A. Fletcher and M. C. Heinrich, *J. Clin. Oncol.*, 2004, **22**, 3813–3825.

[121] E. Wardelmann, S. Merkelbach-Bruse, K. Pauls, N. Thomas, H.-U. Schildhaus, T. Heinicke, N. Speidel, T. Pietsch, R. Buettner, D. Pink, P. Reichardt and P. Hohenberger, *Clin. Cancer Res.*, 2006, **12**, 1743–1749.

[122] G. Spitaleri, R. Biffi, M. Barberis, C. Fumagalli, F. Toffalorio, C. Catania, C. Noberasco, C. Lazzari, F. de Marinis and T. D. Pas, *Onco. Targets Ther.*, 2015, 1997.

[123] M. Bejugam, S. Sewitz, P. S. Shirude, R. Rodriguez, R. Shahid and S. Balasubramanian, *J. Am. Chem. Soc.*, 2007, **129**, 12926–12927.

[124] I. Russo Krauss, A. Merlino, A. Randazzo, E. Novellino, L. Mazzarella and F. Sica, *Nucleic Acids Res.*, 2012, **40**, 8119–8128.

[125] H. Ranpura and P. H. Bolton, *Biophys. J.*, 2014, **106**, 65a.

[126] P. Schultze, R. F. Macaya and J. Feigon, *J. Mol. Biol.*, 1994, **235**, 1532 – 1547.

[127] R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe and J. Feigon, *Proc. Natl. Acad. Sc.*, 1993, **90**, 3745–3749.

[128] V. M. Marathias and P. H. Bolton, *Nucleic Acids Res*, 2000, **28**, 1969–1977.

[129] J. W. Ponder and D. A. Case, *Adv. Prot. Chem.*, 2003, **66**, 27.

[130] J. E. Lennard-Jones, *Proc. R. Soc. London, Ser. A*, 1924, **106**, 463.

[131] J. D. van der Waals, *Verhandelingen der Koninklijke Akademie der Wetenschappen*, 1893,**1**, 1.

[132] C. A. Coulomb, *Collection de ḿemoires relatifs à la physique*, Gauthier-Villars, 1884, 569-638.

[133] L. Verlet, *Phys. Rev.*, 1967, **159**, 98.

[134] R. W. Hockney, *Meth. Comp. Phys.*, 1970 **9**, 136.

[135] J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, and D. A. Case, *J. Am. Chem. Soc.*, 1998, **120**, 9401.

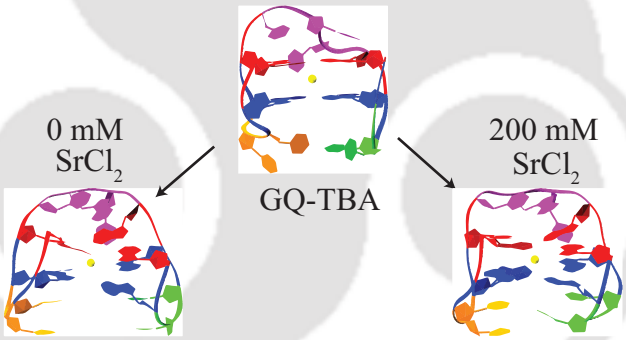
[136] Kasimova, O. A.; Pavan. M. G.; Danani. A.; Modan. K.; Cristiani, A.; Scapozza, L.; Gurny, R. and Moller, M. *J. Phys. Chem. B* **116** , 4338 (2012)..

[137] B. Jayaram, D. Sprous, and D. L. Beveridge, *J. Phys. Chem. B*, 1998, **102**, 9571.

[138] A. Onufriev, D. Bashford, and D. A. Case, *Proteins. Struct. Funct. Bioinf.*, 2004, **55**, 383.

Chapter 2

Conformational Deviation of Thrombin Binding G-quadruplex Aptamer (TBA) in Presence of Divalent Cation Sr^{2+}



Overview

Thrombin binding TBA-G-quadruplex aptamer (TBA) plays a major role in blood coagulation cascade. The 15-mer TBA sequence tends to form a four-stranded TBA-G-quadruplex structure. In this research work, a series of explicit solvent classical MD simulations of the TBA is carried out using different salt ($SrCl_2$) concentrations (0, 50, 100 and 200 mM). Here we have also testified the effect of salt concentration of divalent cation Sr^{2+} on the conformational change of quadruplex DNA. The structural deviations, fluctuations, torsional angles and the affinity of the ion are explored at different salt concentrations. It is found that the conformation of TBA-G-quadruplex at 0 mM and 50 mM salt concentrations, are very much different than the other salt concentrations (100 mM and 200 mM). Also observed are: (i) no exchange of Sr^{2+} ion between inside and outside of the channel, (ii) an enhancement in the Sr^{2+} ion density around the phosphate region of the loop residues as salt concentration increases and (iii) the stacking of T3 and T4 residues of loop-1 that appears up to 50 mM concentration, vanishes as the salt concentration is increased further.

2.1 Introduction

DNA, a polymeric unit of nucleotides in living cells, is involved in transmitting, encoding and expressing genetic information. It can adopt different confirmations based on its sequence and physiologically relevant conditions [1]. DNA is most commonly present as double helical B-DNA, but it can also form various hairpins, triplexes, motifs, quadruplex and hybrid structures. Among these structures, non-canonical quadruplex nucleic acids [2–9] have the potential to form into a four-stranded helical structure using four guanines hoogsteen-paired hydrogen bonded to each other and π-π stacking of quartet of guanines.

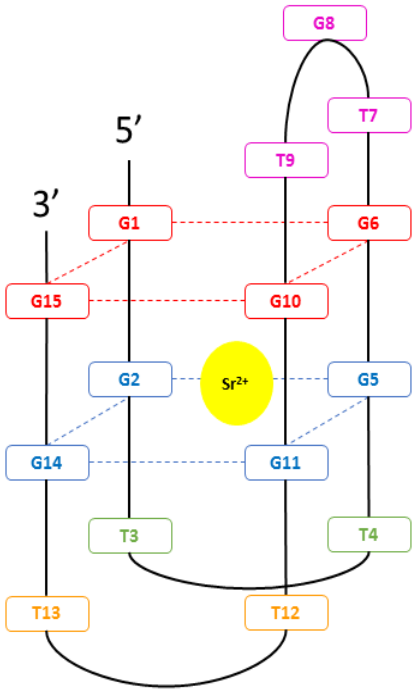


Figure 2.1: *Schematic representation of the TBA-G-quadruplex aptamer and* $Sr^{2+}{}_{in}$ *ion. Tetrad-1, tetrad-2, loop-1, loop-2 and loop-3 are shown in red, blue, green, magenta and orange respectively.*

G-quadruplex molecules have been also found to communicate with different proteins like α-thrombin [10–13] , nucleolin [14–16], STAT3 protein [17–19]. Malfunctioning of α-thrombin causes hemorrhage and thrombosis, which further leads to coronary heart disease and thrombotic disorders. Therefore, to affect the action of the α-thrombin, there

is a need for a specific inhibitor. In 1992, Bock et al. found that the 15-mer quadruplex DNA acts as an inhibitor of thrombin, called a Thrombin Binding Aptamer (TBA) [20–37] ,which particularly binds to the serine protease thrombin. Therefore, the structural change and features of TBA are important for both in modern biology and pharmacological industry, which further proves that quadruplex DNA plays a major role in drug design and is also a therapeutic target [38–41]. In **Fig. 2.1**, we have represented the schematic diagram of the 15-mer TBA.

The formation and stability of G-quadruplex depend on several factors, such as stacking interactions and hoogsteen hydrogen bonds between guanine bases, electrostatic interactions and hydration shell. Along with these, there is another factor that plays a significant role in the stabilization of quadruplexes. This important stabilizing factor arises due to the coordination of carbonyl (O6) with the metal cations inside the quadruplex DNA tetrad residues [42–48].

Previous, experimental and simulation studies proposed that the absence of coordinated metal cation at the center of the tetrad destabilizes the quadruplex structure as it is electrostatically not favorable [49, 50]. Most studies have examined the role of monovalent cations mainly Na^+ and K^+ and their influences on the quadruplex structure and stability [51]. Though different monovalent and divalent cations have been tested to affect the structure, stability and folding efficiency of tetrad and loop residues of quadruplexes, a complete understanding of the driving force of the folding of quadruplex is yet to be achieved [52–64]. The position of the inner cation(s) in the quadruplex structure depends not only on the size of the cation but also on its charge. In case of Na^+ ion, the position was found either being sandwiched by the tetrads or in the plane of a tetrad [65]. On the other hand, K^+ and $NH_4{}^+$, being larger in size, are fitted in between two consecutive tetrads [66–68].

Most of the quadruplex structures are acquired with the NMR experiments, that have restricted capability to identify the binding site of the metal ion. In general, from the available experimental methods, further information about the mechanism of the ion binding of the dynamics and the quadruplex structure is difficult to obtain. Experimentally it has been observed that the divalent ions stabilize G-quadruplex DNA in the following order Sr^{2+}>Ba^{2+}>Ca^{2+}>Mg^{2+} [69, 70]. Although the effects of monovalent cations such as Na^+ and K^+ and group IIA cations such as Sr^{2+}, Ba^{2+} and Pb^{2+} have been studied extensively primarily by different experimental techniques [71–76], very few molecular dynamics simulations have been devoted to examine the effect of divalent cations on guanine-rich

quadruplex DNA. Yang et al calculated the potential mean force (PMF) of the 15-TBA unfolding [77] in the presence of central metal ion Sr^{2+} (using Aqvist parameter [78]). Reshetnikov et al showed the spontaneous exchange of the ion between the quadruplex molecule and solvent in bulk at the atomistic level [79].

In this study, by using classical molecular simulation, we study the effect of divalent cation (Sr^{2+}) on TBA-quadruplex-DNA at different salt concentrations. Our study provides information about the behavior of loop and tetrad residues at different ionic concentrations.

2.2 Models and Simulation Method

Molecular dynamics simulations of thrombin binding TBA-G-quadruplex aptamer in aqueous solutions of different salt concentrations were performed by the using AMBER14 [80] simulation package. Out of the 11 models of NMR based structure of the thrombin binding DNA aptamer d(GGTTGGTGTGGTTGG) (15-TBA), from Protein Data Bank (PDB) entry 1RDE model-1 was used [81]. In this PDB, Sr^{2+} ions were not included. Thus, one Sr^{2+} ion was placed in the center of the TBA-G-quadruplex core. The standard DNA simulation force field ff99-bsc0 [82] was used for DNA in this study. This version of the Cornell et al. AMBER force-field provides a decisive stabilization of DNA simulations as reported earlier [83–86]. Following a previous study for Sr^{2+} ion, the Lennard-Jones parameters, σ =1.777 Å and ε =0.19470705 kcal/mol, were adopted [87]. In aqueous solution, these parameters for Sr^{2+} ion reproduce the experimental hydration free energies (HFE) as well as ion-oxygen distances (IOD) in the first solvation shell and coordination numbers (CN) accurately. Initially, the quadruplex-DNA was solvated in a truncated octahedral box of TIP3P [88] water molecules with a 20 Å buffer. Since, the DNA aptamer considered in this study carries the electronic charge of -14e, six more Sr^{2+} ions were added (without any counterion) to neutralize the system (containing one more Sr^{2+} ion in the core of the DNA as mentioned earlier). This procedure was adapted for constituting system without any added salt. In order to prepare the systems in which DNA was immersed in an aqueous salt solution with varying salt concentrations, an adequate number of Sr^{2+} and Cl^{-} ions were added by replacing water molecules. TIP3P specific Amber-adapted Joung and Cheatham parameters [89] were used for Cl^{-}. The different systems considered in this study are presented in **Table 2.1**. The periodic boundary conditions were applied in all three directions. At first, each of the systems was energy minimized by a two step process.

Table 2.1: *N_{water}, $N_{Sr^{2+}}$, and N_{Cl^-} respectively, are the number of water molecules, Sr^{2+} and Cl^- ions. C_{SrCl_2} is the concentration of $SrCl_2$ in molarity (M).*

Systems	N_{water}	$N_{Sr^{2+}}$	N_{Cl^-}	C_{SrCl_2}	Box length(Å)
0 mM	6669	7	0	0	64.53
50 mM	6648	14	14	0.056	64.43
100 mM	6627	21	28	0.113	64.41
200 mM	6585	35	56	0.227	64.24

Specifically, for minimization, each of the systems was subjected to 500 steepest descent cycles followed by an equal number of conjugate gradient cycles with a restraint of 100 kcal mol^{-1} $Å^{-2}$ on the G-quadruplex DNA atoms including the channel Sr^{2+} ion in canonical (NVT) ensemble. After minimization, all the systems were heated slowly from 0 to 300 K over 150ps in NVT ensemble with 100 kcal mol^{-1} $Å^{-2}$ positional restraints (using the collision frequency of 0.2 ps^{-1}) on the DNA atomic sites including the Sr^{2+} ion placed in the core of the quadruplex. Then a short equilibration run of 250 ps in NVT ensemble was carried out. The positional restraints were slowly taken off by performing an analogous series of alternating minimizations and followed by short equilibrations (of 250 ps in NVT ensemble, like before) using the restraint force constants of 25, 20, 15, 10, 5 and finally, 0 kcal mol^{-1} $Å^{-2}$, sequentially. For each of these simulations, to calculate the long-ranged nonbonding electrostatic interactions, particle mesh Ewald (PME) method [90] with a charge grid spacing of $\backsim$1Å and a DSUM_TOL value of 10^{-5} Å were used. A cut-off radius of 12 Å was considered for all nonbonding short-ranged interactions. SHAKE algorithm [91] with a tolerance of 10^{-5} Å was used to constrain the covalent bonds involving hydrogen atoms and the integration time step was set to 2 fs. Langevin dynamics [92] with the collision frequency of 1 ps^{-1} was applied for controlling the temperature of the systems. Finally, for each of the systems without restraints, MD simulation was performed for 1000 ns at 300K temperature and 1 atm pressure condition in isothermal-isobaric (NPT) ensemble. The physical pressure was maintained by the application of Berendsen barostat with the collision frequency of 1.0 ps^{-1} [93, 94] . For all systems, to analyze the trajectories obtained from 1000 ns production run, Cpptraj program [95] of the AMBER14 toolkit as well as Visual Molecular Dynamics (VMD) [96] were utilized as and when required. All torsional angle plot was calculated using python code, inspired by Do_X3DNA package [97]. The spatial distribution functions (SDFs) for all salt concentrations were calculated using TRAVIS [98, 99].

2.3 Results and Discussions

2.3.1 Root Mean Square Deviations (RMSDs)

The all-atom RMSD values (as a function of simulation time) over 1000 ns MD trajectories for the TBA-G-quadruplex at different salt concentrations are shown in **Fig. 2.2**. In the

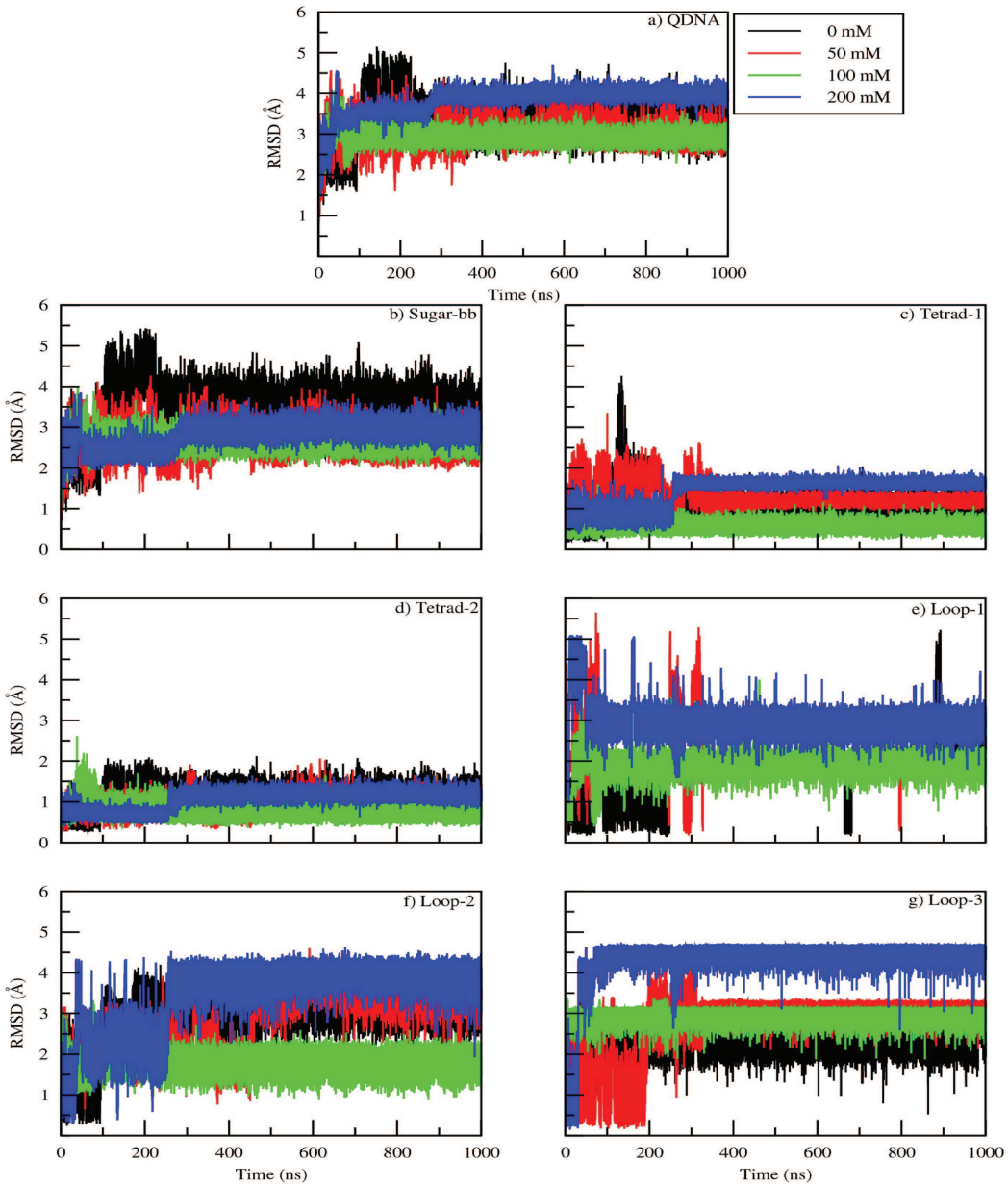


Figure 2.2: *The changes in the root mean square deviation (RMSD) values of (a) DNA, (b) Sugar phosphate backbone (bb), (c) tetrad-1, (d) tetrad-2, (e) loop-1, (f) loop-2 and (g) loop-3 residues of TBA with respect to the NMR structure.*

same figure, we also show the change in the RMSD values of the DNA as well as its

sugar phosphate backbone, tetrad-1, tetrad-2, loop-1, loop-2 and loop-3 as the simulation progresses. For different systems, in order to calculate the RMSD values of DNA, the NMR structure (model-1) is taken as the reference structure. Concentrating on the effect of salt concentration on the RMSD values of the DNA, first **Fig. 2.2 (a)**, we find that the addition of salt causes a reduction in the structural displacement of the DNA from that of its reference structure and this effect is pronounced up to 100 mM of salt concentration. Interestingly, when 200 mM salt concentration is used a much larger structural displacement of the DNA is observed. These observations apparently, imply that the addition of small amount of salt (up to 100 mM salt concentration) induces the deviation of DNA molecule (relative to that of the system with no salt) and a further addition of a salt causes structural change. To find the region where RMSD values deviate more, we calculate the change in the RMSD values separately for three different loops and as well as for two tetrads of the DNA at different salt concentrations. A close examination in the changes in the RMSD values of the DNA sugar phosphate backbone, tetrads and loops at different concentrations (**Fig. 2.2 (b)-(g)**) suggest that for 200 mM salt concentration, the changes in the RMSD values of tetrad (above 250 ns) and loop contribute to the effective enhancement of the RMSD value of the DNA. Moreover, as expected the contribution of RMSD value of the loops is much higher than that of tetrads in the effective change in DNA RMSD. In case of loop-2 and loop-3, at 200 mM salt concentration, the change in the RMSD values are higher than that of loop-1. However, after 250 ns simulation there is no significant change in the RMSD values for all salt concentrations.

2.3.2 Root Mean Square Fluctuations (RMSFs)

In **Fig. 2.3**, we present the RMSFs of all atomic positions of the DNA at different salt concentrations. In the same figure, the atom positions of three loops (namely, loop-1, loop-2 and loop-3) and two tetrads (i.e., tetrad-1 and tetrad-2) of the DNA are indicated in order to understand the effect of salt concentrations on different loops and tetrads. As earlier, the NMR structure (model-1) is again taken as the reference structure to calculate the RMSF values of DNA. The RMSFs displayed in **Fig. 2.3**, reveal the effect of the presence of salt on the fluctuations of the loop and tetrad atoms, where a non-monotonous change without any proper trend in the RMSF values (as the salt concentration changed) is observed. It is found that the tetrad residues are less fluctuating than the loop residues. In fact, the RMSF values of tetrad residues are very much close to that of the reference structure. Furthermore, the comparison of the fluctuations of three different loop indicates

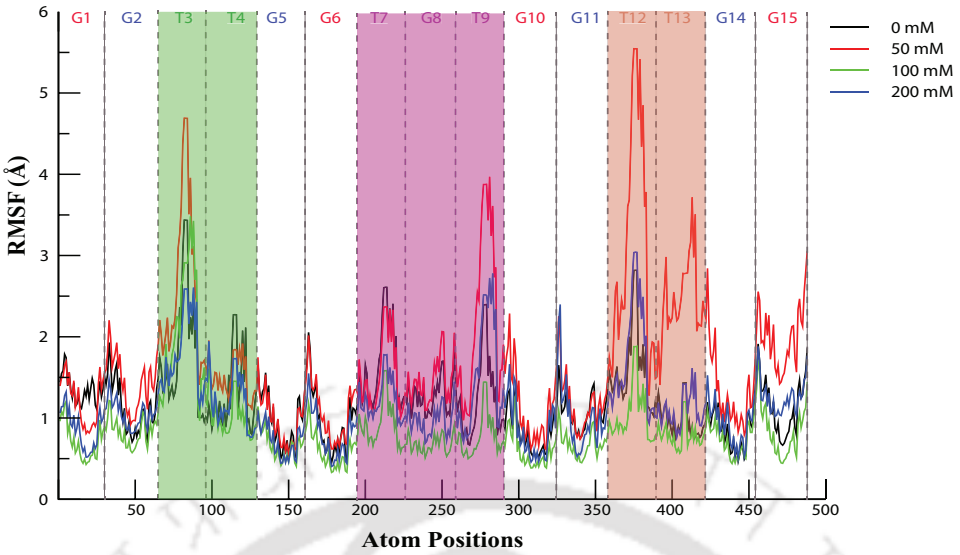


Figure 2.3: *The changes in the root mean square fluctuation (RMSF) values of different atomic positions with respect to the NMR structure. Loop-1, Loop-2 and Loop-3 regions are shown in green, magenta and range respectively.*

a much higher variation for loop-1 and loop-3 than for loop-2. Moreover, for loop-1 the fluctuation is maximum for the system without any salt, (which is closely followed by 100 mM salt concentration) and for loop-3, it is at 50 mM concentration where the fluctuation reaches the maximum value, but, for systems with no added salts and those at 200 mM salt concentration, the fluctuations are observed for all atom positions. Thus, overall it can be stated that increasing the salt concentration up to 200 mM results in a conformation which is different from that of intermediate concentrations.

We have achieved important information about structural changes from RMSD and RMSF values. However more precise analyses are required to determine the effect of different parameters on the structural conformation. Therefore, we have carried out additional analysis such as RDF, SDFs of the ions, torsional angles, hydrogen bonding throughout the simulation trajectories.

2.3.3 Torsional Angels

To exhibit the flexibility of TBA-G-quadruplex DNA structure we need to calculate four major angle parameters: (i) backbone torsional angles (α/γ), (ii) backbone torsional angles ($\zeta - \epsilon$), (iii) backbone torsional angles (β-δ), (iv) glycosidic torsional angles (χ). Due to conformational restrictions of force field [55, 100, 101] enforced by the topology of the system in this structure, the torsional angles of TBA-G-quadruplex DNA are quite different in contrast to those of canonical duplex DNA. All torsional angles of tetrads and as well

as that of loop residues are shown in a wheel representation in **Figures 2.4 and 2.5** respectively. We have calculated the time distribution of all torsional angles in each tetrad and loop residues throughout the 1000 ns trajectory. Using 15° resolution, the acquired distributions have been clustered and then plotted with respect to the probabilities.

In the case of tetrad residues, it is observed that only G-5, G-14 and in case of loop residues G-8, T-9 and T-13 are in the g^-g^+ ($\alpha \sim 280°$, $\gamma \sim 60°$) region. For all different salt concentrations, in most of the residues, the α torsion angle deviates from the NMR structure except for G-6, T-7 and G-8 residues. Further, for all systems, the α torsion angles of the residues G-2, G-5, G-6, T-9, G-11, T-13 and G-15 are very similar, even though they deviate from the NMR structure. Moreover, the α torsional angles of the residues G-5, G-8, T-9, T-12, T-13 and G-14 are in g^- region and T-3, T-4 and T-12 region are in trans (t) region ($g^- \sim 280°$, $t \sim 180°$). At 100 mM salt concentration, for all the residues, α torsional angles are either same or very close to that observed in the NMR structure. The γ torsional angles of the residues G-2, G-6, G-8 and T-9 at all concentrations do not deviate much from that of the NMR structure. Similarly the γ torsional angles of the residues G-1, T-4, G-5, T-7, G-11, G-14 and G-15 are very similar for all concentrations but different from that of the NMR structure. Thus, the γ torsional angles of most of the residues are in g^+ region except for T-3, T-4, G-10 and T-12 residues ($g^+ \sim 60°$). It is also observed, that at 100 mM salt concentration, for all the residues, γ torsion angles are very close to that of the NMR structure. Thus it can be concluded that the conformation of the TBA-G-quadruplex DNA at 100 mM salt concentration is very similar to the NMR structure (model-1). Moreover, the detailed analysis concludes that for all simulated systems α/γ torsion angles of the TBA-G-quadruplex DNA is not in the g^-g^+ region like canonical duplex DNA.

The β torsional angles of all residues are in 't' region except for T-7 residue ($t \sim 180°$). Only at 100 mM salt concentration, the β torsional angle of T-7 falls in 't' region. Moreover, the β torsional angles of all residues (except for G-2, G-6 and G-15 residues) fall in the different region than that of NMR structure. Interestingly the β torsional angles of all the residues at different salt concentrations are equivalent. For all systems the δ torsional angles that fall in the a^+ ($\sim 130°$) region of all residues appear in the similar region to that of the NMR structure (excluding for the G-5, G-10 and G-14 residues).

In case of BI duplex DNA ϵ-ζ torsional angles are in g^- region ($g^- \sim 270°$) for all residues. The ϵ torsional angles of the residues T-4, G-6, T-9 and G-11 are in g^- region ($g^- \sim 270°$) while that of others have deviated from it. The ζ torsional angles of almost

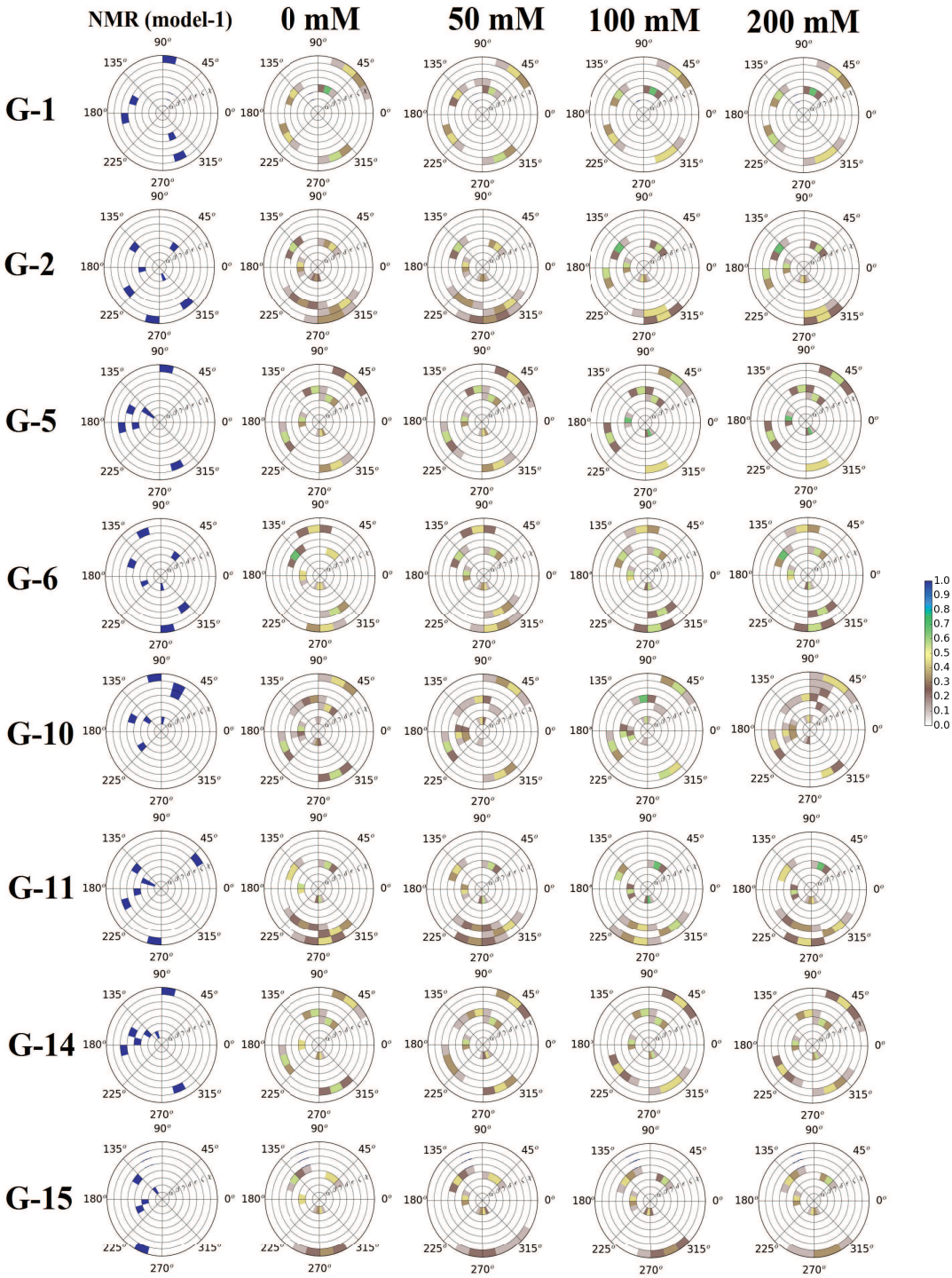


Figure 2.4: *All torsion angles of tetrad residues in a wheel representation at all salt concentrations and NMR structure. The color scale between 0 to 1 represents the probability of the respected angle.*

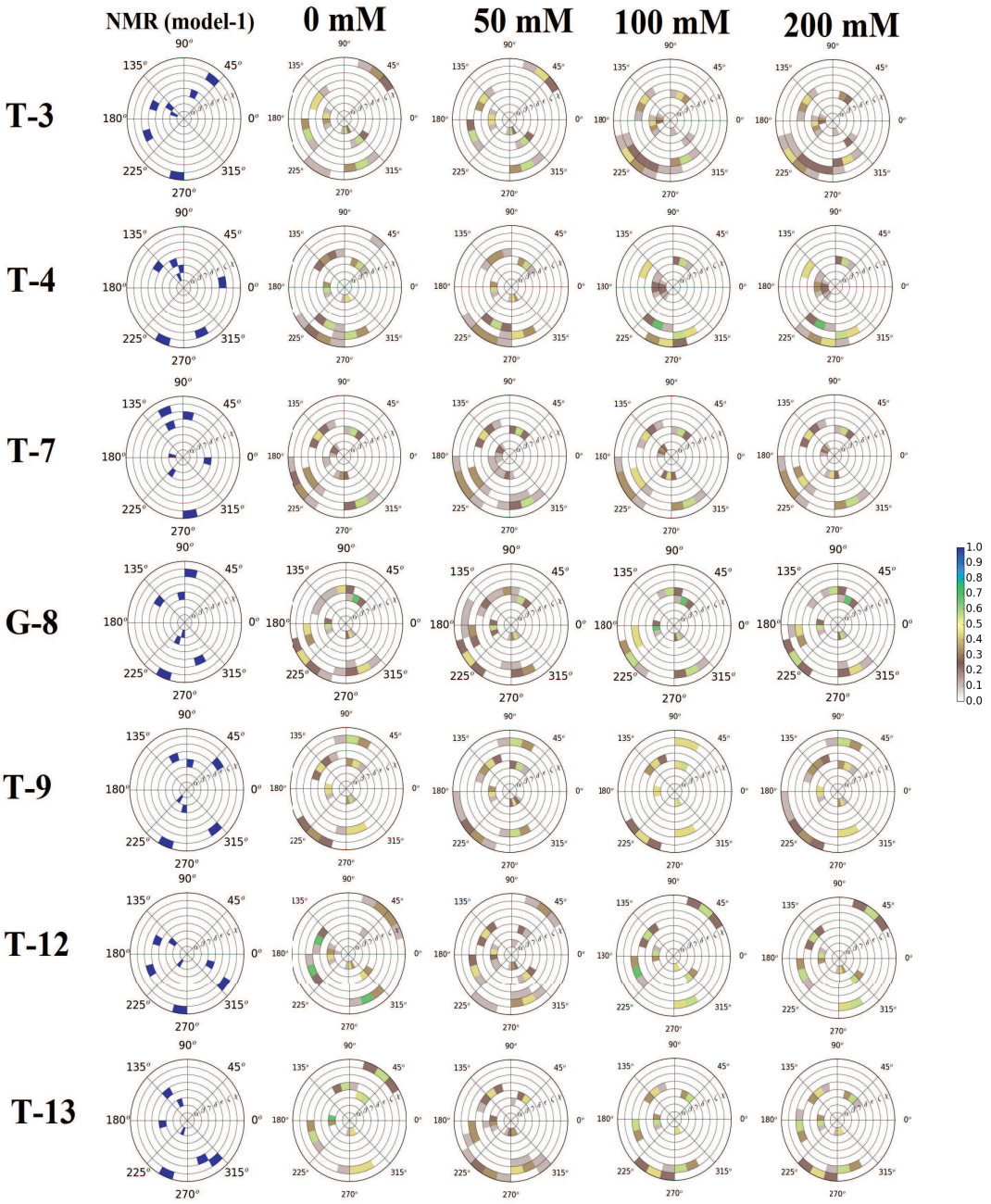


Figure 2.5: *All torsion angles of loop residues in a wheel representation at all salt concentrations and NMR structure. The color scale between 0 to 1 represents the probability of the respected angle.*

all residues are in g^- region except G-6 residue. For T-3 residue, at both 100 mM and 200 mM salt concentrations, the ζ torsional angles are analogous and they deviate from the corresponding NMR structure value at the same time.

The structure and conformation of quadruplex DNA mainly depend on the glycosidic χ torsional angle. For all salt concentrations, the χ torsional angles of all the residues are similar to the values of the NMR structure except for the T-3, T-7 and T-12 residues. Moreover, for the system with no salt, χ torsional angle varies for T-13 residue significantly from that of the NMR structure. It indicates that quadruplex DNA loop residues are more fluctuating than tetrad residues throughout the simulation and these findings are in good accordance with the observed RMSF and RMSD values for the loop residues discussed above. The reason behind the higher stability of the tetrad residues when compared to that of the loop residues lies in the strong hydrogen bonding interactions of the former.

2.3.4 Distance

In **Table 2.2**, we have shown the average distance between the oxygen atom of guanine (O6) and Sr^{2+} ($Sr^{2+}{}_{in}$) ion of the channel of G-quadruplex DNA for all systems. In every

Table 2.2: *The average distance between the oxygen atom of guanine (O6) in tetrad residues and* $Sr^{2+}{}_{in}$.

Residues	0 mM (Å)	50 mM (Å)	100 mM (Å)	200 mM (Å)
G-1-O6	2.75 ± 0.20	2.70 ± 0.30	2.69 ± 0.23	2.70 ± 0.54
G-2-O6	2.59 ± 0.35	2.59 ± 0.17	2.60 ± 0.11	2.60 ± 0.14
G-5-O6	2.65 ± 0.17	2.70 ± 0.64	2.62 ± 0.20	2.63 ± 0.09
G-6-O6	2.61 ± 0.21	2.63 ± 0.24	2.66 ± 0.20	2.65 ± 0.28
G-10-O6	2.70 ± 0.27	2.69 ± 0.09	2.67 ± 0.24	2.67 ± 0.14
G-11-O6	2.64 ± 0.14	2.58 ± 0.22	2.58 ± 0.12	2.58 ± 0.19
G-14-O6	2.63 ± 0.11	2.63 ± 0.51	2.63 ± 0.16	2.64 ± 0.17
G-15-O6	2.69 ± 0.02	2.64 ± 0.01	2.64 ± 0.22	2.65 ± 0.09

salt concentration the average distance between the oxygen atom of guanine (O6) and $Sr^{2+}{}_{in}$ falls in the range of 2.5 - 2.7 Å. The $Sr^{2+}{}_{in}$ remains almost fixed in its position in the middle of the quartet tetrad. It suggests that $Sr^{2+}{}_{in}$ ion of the channel of G-quadruplex DNA does not exchange throughout the simulation. At 0 mM and 50 mM salt concentrations, the distances are higher than that for the other (100 mM and 200 mM) salt concentrations. Further, the systems with 100 mM and 200 mM salt concentrations have

similar distances. The average distance values of $Sr^{2+}{}_{in}$ and the oxygen atom of guanine (O6) are well comparable to experimental and as well as theoretical studies [102–106] and the results validate the force field parameters of Sr^{2+} ion considered in this study [87].

2.3.5 Radial Distribution Functions (RDFs)

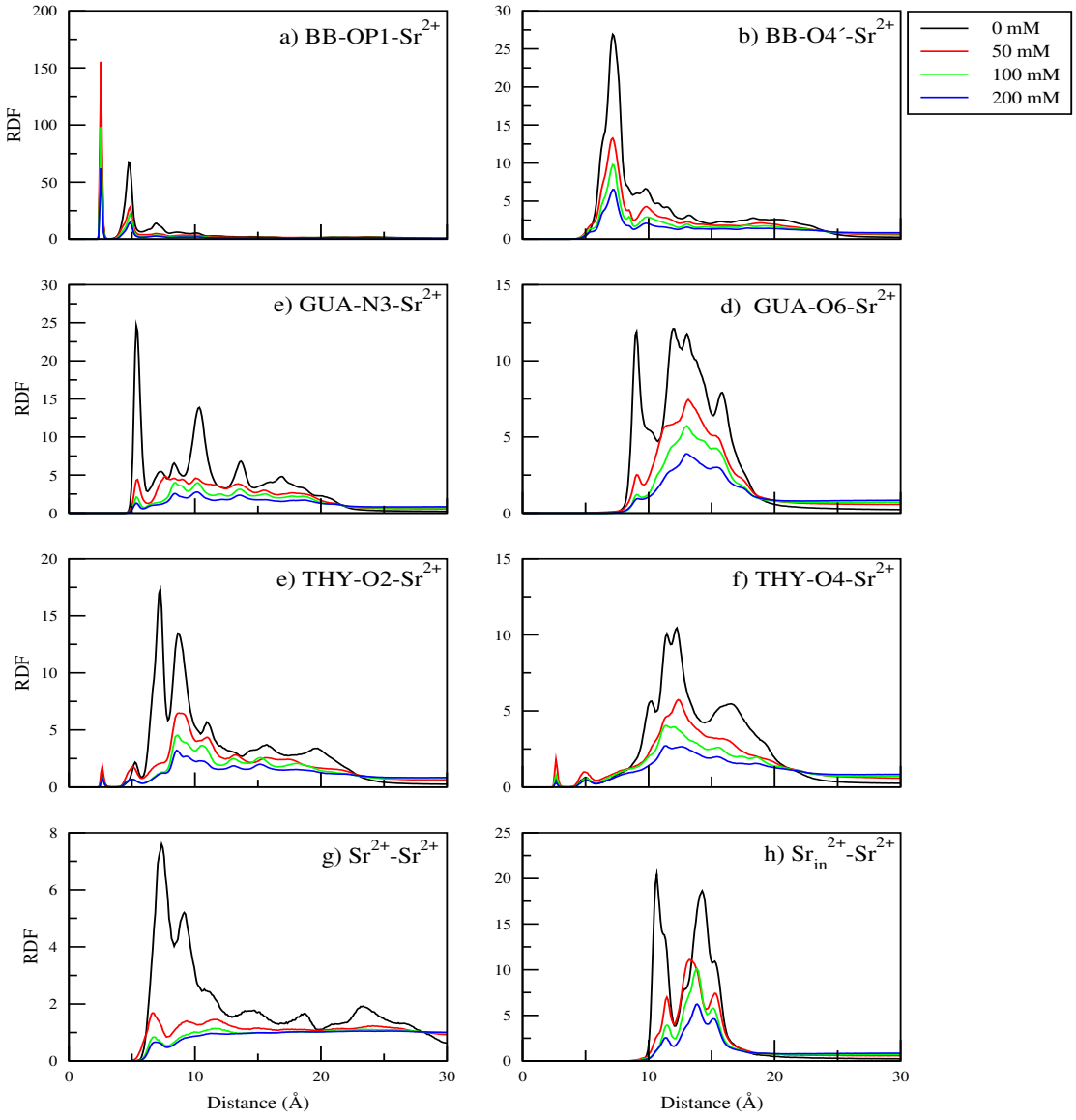


Figure 2.6: *Atom-atom radial distribution functions (RDF) for (a) OP1 atom of the backbone (BB), (b) O4′ atom of the backbone (BB), (c) N3 atom of the Guanine, (d) O6 atom of the Guanine, (e) O2 atom of the Thymine, (f) O4 atom of the Thymine, (g) Sr^{2+} and (h) $Sr^{2+}{}_{in}$ atom with Sr^{2+} outside channel.*

The nitrogenous bases are hydrophobic in nature and on the other hand the sugar-phosphate backbone is negatively charged and hydrophilic. In **Fig. 2.6**, for different systems, we have shown the selected site-site radial distribution functions (rdfs) involving different atomic sites of the DNA such as oxygen atom (OP1) of sugar-phosphate containing backbone (BB), oxygen (O6) and nitrogen (N3) atoms of guanine (G), oxygen atoms (O4 and O2) of thymine (T), and the Sr^{2+} counterions outside the channel. For all the systems, in the same figure, we have also included the pair-correlation functions involving the counterion inside the channel ($Sr^{2+}{}_{in}$) and counterions outside the channel (Sr^{2+}) and between the Sr^{2+}-Sr^{2+}. Note that, different site-site distribution functions provide the information about the interactions between corresponding species, albeit indirectly, without providing much knowledge quantitatively. For example, the height of the first peak in a particular rdf, g(r), gives the estimation of the number of species in the first solvation shell of a reference molecule only after carrying out the integration of it. Since, we mainly focus on the qualitative change in the rdfs involving the counterions Sr^{2+} and different atomic sites of DNA quadruplex and the ion present in the channel ($Sr^{2+}{}_{in}$) as outside salt is added, the quantitative estimation of the number of counterions in the first solvation shell in the respective atomic sites are not considered in this study. Nevertheless, we mainly find: (i) The addition of salt decreases the first peak heights of OP1-Sr^{2+} at 2.65 Å and Sr^{2+}-Sr^{2+} distribution functions profoundly; (ii) For the distribution functions involving guanine O6-Sr^{2+} and guanine N3-Sr^{2+} a sharp decrease is observed as one moves from 0 mM to 50 mM salt concentration, which is followed by essentially no change in the peak height at 9.05 and 5.35 Å as salt concentration increases further; (iii) For the distribution functions involving thymine O2 and O4 and Sr^{2+}, the heights of the first peaks are at 2.65 Å for both the atoms are not a strong function of added salt; (iv) A dramatic change in the Sr^{2+}-$Sr^{2+}{}_{in}$ rdf is observed, in which a maximum of first peak height is noticed for 0 mM salt concentration followed by 200 mM salt concentration and the first peak heights for systems with 50 and 100 mM salt concentrations are almost similar. These suggest a more favorable interaction between the counterion and the ion present in the channel for systems having of 0 and 200 mM (salt concentrations) than that for 50 and 100 mM concentrations.

In **Fig. 2.7**, for different systems, we have shown the selected site-site rdfs involving different atomic sites of the DNA such as oxygen atoms (O4′, OP1 and OP2) of sugar-phosphate containing backbone (BB), oxygen (O6) and nitrogen (N3 and N2) atoms of guanine (G), nitrogen (N3) and oxygen atoms (O4 and O2) of thymine (T), the counterion inside the channel ($Sr^{2+}{}_{in}$) and the oxygen atom (O_w) of water. The rdf values of

$Sr^{2+}{}_{in}$ and the oxygen atom of water (O_w) at 2.65 Å are well comparable to experimental and as well as theoretical study [102–106]. Since, the nitrogenous bases are hydrophobic and the sugar-phosphate backbone is hydrophilic, the peak heights of backbone atoms are much higher than that of nitrogenous base atoms. In the case of $Sr^{2+}{}_{in}$ among all the salt concentrations, the first peak height is highest at 200 mM and the lowest peak height is observed for 100 mM salt concentration.

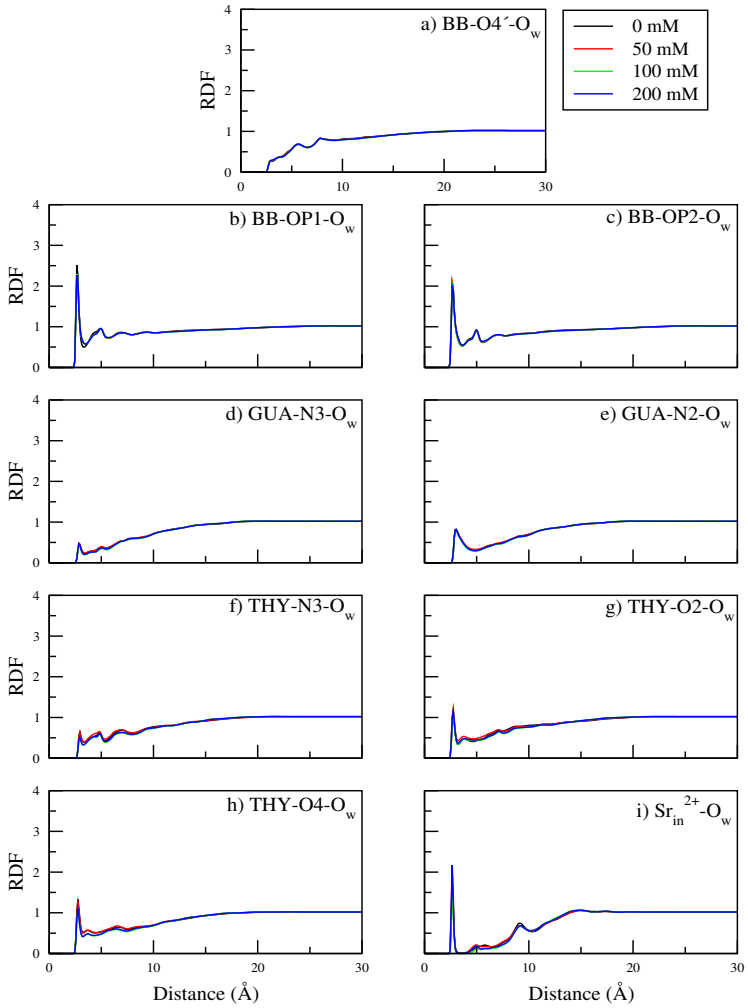


Figure 2.7: *Site-site radial distribution functions (RDF) of (a) O4′ atom of the backbone(BB), (b) OP1 atom of the backbone(BB), (c) OP2 atom of the backbone(BB), (d) N3 atom of the Guanine, (e) N2 atom of the Guanine, (f) N3 atom of the Thymine, (g) O2 atom of the Thymine, (h) O4 atom of the Thymine and (i) $Sr^{2+}{}_{in}$ atom residues with oxygen(O_W) in water.*

2.3.6 Hydrogen Bonding

The stability of G-quadruplex DNA is strongly influenced by hydrogen bonding interaction between the tetrads, loop residues and the water present in the solution. We have calculated the average hydrogen bonds for tetrads and loop residues with water for all systems (**Table 2.3**). A hydrogen bond is considered if the distance between the donor (D) and acceptor

Table 2.3: *The average number of hydrogen bonds between water and tetrad (and loop) residues.*

Residues	0 mM	50 mM	100 mM	200 mM
Tetrad-1	27.68	29.07	28.36	28.09
Tetrad-2	38.82	35.47	35.06	36.45
Loop-1	21.76	21.3	18.5	19.25
Loop-2	29.44	27.55	28.12	27.67
Loop-3	20.91	21.76	19.72	19.36

(A) atoms is less than or equal to 3.5 Å and simultaneously the angle D-A-H is less than or equal to 45° [107–109]. The number of hydrogen bonds for loop residues for all the systems are comparable to each other. Due to the presence of excess number of guanine residues in loop-2, the average number of hydrogen bonds is more than that in other loops. In case of tetrad residues, although both have the same number of guanine residues, the number of hydrogen bonds is different except at 50 mM salt concentration. The number of hydrogen bonds at all salt concentration is similar for each tetrad, with the exception occurring at 50 mM for tetrad-1 and at the system without any salt for tetrad-2.

2.3.7 Spatial Distribution Functions (SDFs)

The spatial distribution functions (SDFs) can be utilized to determine the three dimensional distribution of molecules in solution around a reference molecule. Moreover, using SDFs, the plots of the RDFs can be easily interpreted. In **Fig. 2.8**, we have shown the density of Sr^{2+} ions around the DNA by using the last 4 ns of average trajectory PDB of DNA. An isovalue of 2.5 was used for all different salt concentrations. It is apparent that with increasing salt concentration, the density of Sr^{2+} ions is increasing up to 100 mM salt concentrations. The Sr^{2+} ion density at salt concentration of 200 mM is almost same as that at 100 mM. The Sr^{2+} ions are particularly localized near the phosphate backbone for all the concentrations mainly in case of loop residues. This is in accordance with the calculated

RDFs discussed above. On increasing salt concentration, the ion density increases near the O2 and O4 atoms of thymine of loop residues. For the O6 atom of the guanine of the tetrad residues, a strong interaction with $Sr^{2+}{}_{in}$ ion inside the channel is expected. On the other hand, N3 atoms of guanine of tetrad residues interact strongly with $Sr^{2+}{}_{in}$ ions inside the ion channel. This interaction increases as the salt concentration is increased.

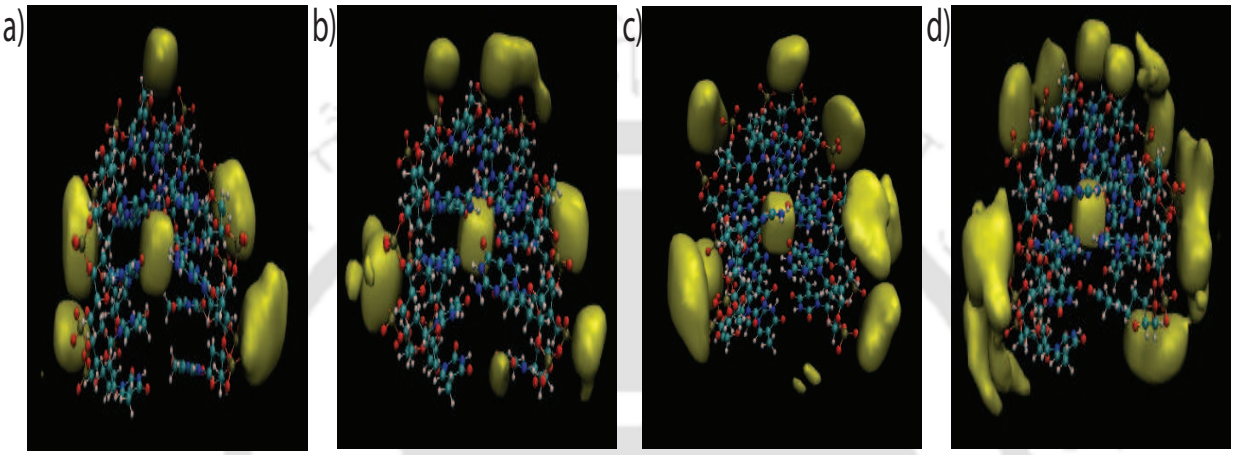


Figure 2.8: *The spatial distribution function (SDF) of Sr^{2+} ion around the loop residue at (a) 0 mM, (b) 50 mM, (c) 100 mM and (d) 200 mM different salt concentrations.*

2.3.8 Loop Dynamics

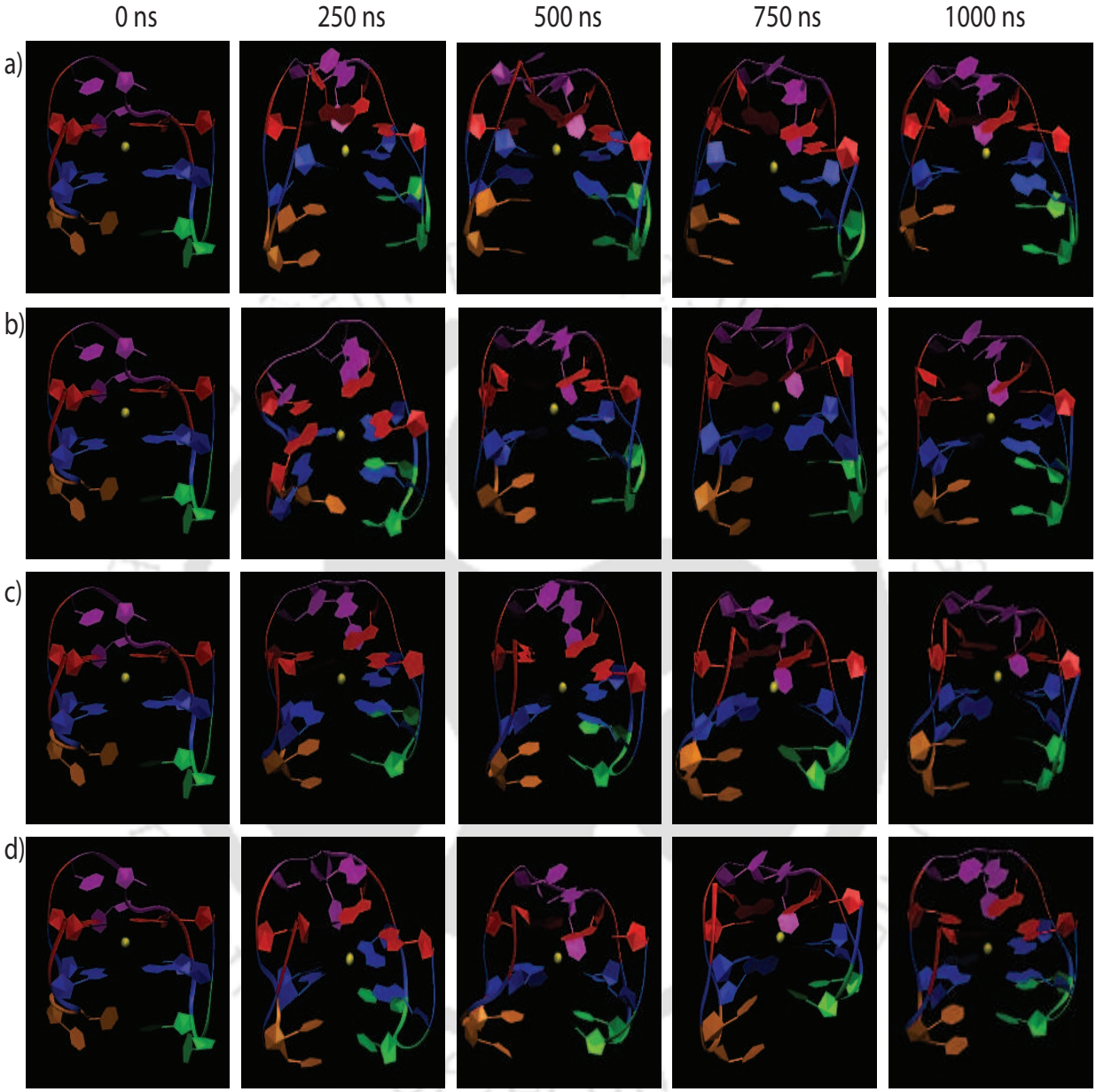


Figure 2.9: *Snapshots of the TBA for different systems (from top to bottom) (a) 0 mM, (b) 50 mM, (c) 100 mM, (d) 200 mM . From left to right refer to the snapshots of 0, 250, 500, 750 and 1000 ns. The* $Sr^{2+}{}_{in}$ *ion is shown in yellow; tetrad-1, tetrad-2, loop-1, loop-2 and loop-3 are shown in red, blue, green, magenta and orange respectively.*

In **Fig. 2.9**, we have shown the progression of the TBA structure at different time intervals (0, 250, 500, 750 and 1000 ns). It is clear that the final structure is very different from that of the initial (NMR) structure at all salt concentrations. In case of NMR structure, there is no π-π stacking between loop residues but for all the systems considered here, the

presence of π-π interactions is observed. A π-π stacking interaction is considered if the distance between the center of mass of two base residues is less than 5.0 Å and the angle between the corresponding plane of the two base residues is less than 20°. From **Fig. 2.10** and **Table 2.4** it is clear that at 0 mM and 50 mM salt concentrations T3-T4 (loop-1), G8-T9 (loop-2), T12-T13 (loop-3) of all the residues have π-π stacking interactions. In case of 100 mM and 200 mM salt concentrations, all loop residues, other than those in loop-1, have π-π stacking interaction.

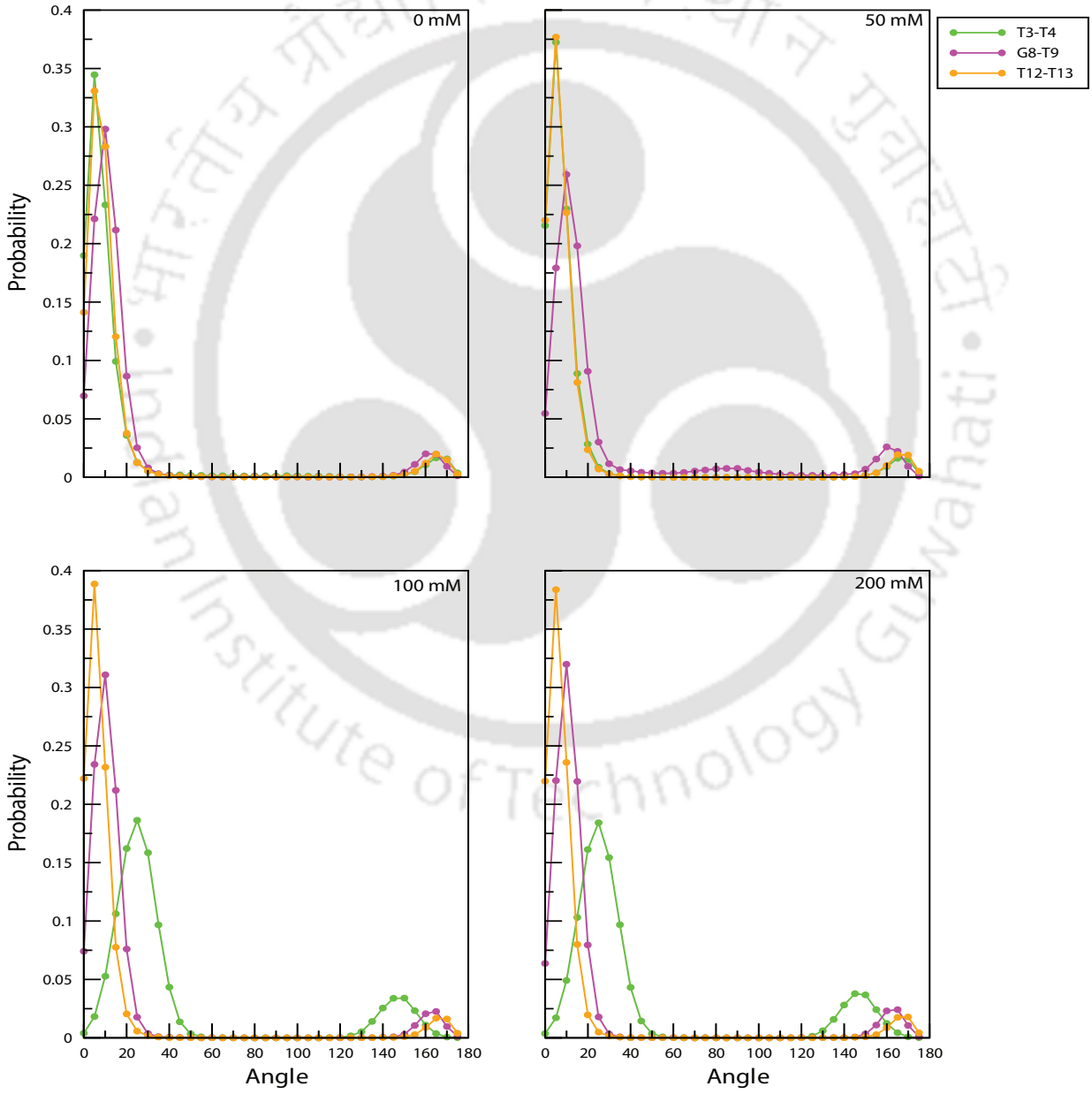


Figure 2.10: *The stacking probability of different loop residues at different salt concentrations.*

Table 2.4: *The average distance between the residues inside the loop.*

Residues	0 mM (Å)	50 mM (Å)	100 mM (Å)	200 mM (Å)
T3-T4	4.39 ± 0.36	4.33±0.43	8.57±0.31	8.60±0.32
G8-T9	4.58 ± 0.33	4.62±0.35	4.76±0.42	4.63±0.51
T12-T13	4.01 ± 0.34	4.19±0.49	3.97±0.35	3.94±0.39

2.3.9 Limitations

In this study, the latest force field parameters for different atomic sites have been used for G-quadruplex DNA and ions. In fact, the current AMBER force-field parameters are not assured in a priori to be entirely compatible with the reline system and force field. This issue is very difficult to solve because of the lack of experimental and simulation results. We can therefore assume that the conformational sampling is a far less serious problem for the current system.

2.4 Summary and Conclusions

Conformational dynamics of guanine-rich quadruplex DNA molecules are very much dependent on the presence of divalent cation ($Sr^{2+}{}_{in}$) inside its channel. Despite few limitations, all-atom molecular dynamics simulation techniques can shed some light on binding and the effect of salt concentrations on the stability and conformational change of guanine-rich quadruplex DNA. Here, we put forward a study of thrombin binding TBA-G-quadruplex aptamer in presence of divalent cation (Sr^{2+}) at different salt concentrations. The dynamical properties of loop residues, as well as the conformational changes of the DNA with the change in different salt concentrations, are explored by means of classical MD simulation.

The determination of the values of root mean square deviation (RMSD) of DNA TBA-G-quadruplex for different salt concentrations apparently indicates that the stability of DNA is attained at all different salt concentrations, although at 200 mM salt concentration, a slightly higher deviation of RMSD value is observed. However, the calculations of root mean square fluctuations (RMSFs) at different concentrations show a different trend from that of RMSD analysis. In brief, the decomposition of RMSFs for tetrad and loop residues show that the former is very much stable at all salt concentrations whereas the latter is very flexible throughout the simulation. It is further observed that the fluctuations are more for residues of loop-1 and loop-3 than that for loop-2 residues. Interestingly, the analysis of backbone torsional angles suggests that the change in salt concentration does

not have much influence on the values of different torsional angles. The estimation of the average distance between the oxygen atom (O6) of the guanine and $Sr^{2+}{}_{in}$ excludes the possibility of Sr^{2+} ion exchange between outside and inside the channel. It further indicates that the $Sr^{2+}{}_{in}$ ion plays an important role in maintaining the overall stability of the G-tetrad residues. The effect of salt concentration is clearly seen in spatial distribution function (SDF) analysis. The determination of SDFs at different concentrations, conclude that the Sr^{2+} ion density is increasing at the loop residues, mainly in the phosphate region of the loop residues as the salt concentration is increased. The loop dynamics display different behaviors of loop residues at different salt concentrations. For the first time it is shown that in presence of minimal or low salt concentrations (0 mM and 50 mM) T3-T4 residues are stacked to each other which is not observed at considerably high salt concentrations (100 mM and 200 mM).

References

[1] D. Sen and W. Gilbert, *Nature*, 1988, **334**, 364.

[2] H. J. Lipps and D. Rhodes, *Trends Cell Biol.*, 2009, **19**, 414–422.

[3] W. E. Wright, V. M. Tesmer, K. E. Huffman, S. D. Levene and J. W. Shay, *Genes Dev*, 1997, **11**, 2801–2809.

[4] H. Han and L. H. Hurley, *Trends Pharmacol. Sci.*, 2000, **21**, 136–142.

[5] G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nat. Chem.*, 2013, **5**, 182.

[6] N. Maizels and L. T. Gray, *PLOS Genet.*, 2013, **9**, 1–6.

[7] N. Maizels, *EMBO Rep.*, 2015, **16**, 910–922.

[8] R. F. Hoffmann, Y. M. Moshkin, S. Mouton, N. A. Grzeschik, R. D. Kalicharan, J. Kuipers, A. H. Wolters, K. Nishida, A. V. Romashchenko, J. Postberg *et al.*, *Nucleic Acids Res.*, 2015, **44**, 152–163.

[9] M. L. Bochman, K. Paeschke and V. A. Zakian, *Nat. Rev. Genet.*, 2012, **13**, 770.

[10] L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas and J. J. Toole, *Nature*, 1992, **355**, 564.

[11] W. Bode, I. Mayr, U. Baumann, R. Huber, S. R. Stone and J. Hofsteenge, *EMBO J.*, 1989, **8**, 3467–3475.

[12] T. J. Rydel, K. Ravichandran, A. Tulinsky, W. Bode, R. Huber, C. Roitsch and J. W. Fenton, *Science*, 1990, **249**, 277–280.

[13] J. M. Harlan, P. J. Thompson, R. R. Ross and D. F. Bowen-Pope, *J. Cell Biol.*, 1986, **103**, 1129–1133.

[14] V. Đapić, V. Abdomerović, R. Marrington, J. Peberdy, A. Rodger, J. O. Trent and P. J. Bates, *Nucleic Acids Res.*, 2003, **31**, 2097–2107.

[15] H. Ginisty, H. Sicard, B. Roger and P. Bouvet, *J. Cell Sci.*, 1999, **112**, 761–772.

[16] M. Derenzini, V. Sirri, D. Trere and R. L. Ochs, *Lab. Invest.*, 1995, **73**, 497–502.

[17] N. Jing, Y. Li, W. Xiong, W. Sha, L. Jing and D. J. Tweardy, *Cancer Res.*, 2004, **64**, 6603–6609.

[18] A. Kotha, M. Sekharam, L. Cilenti, K. Siddiquee, A. Khaled, A. S. Zervos, B. Carter, J. Turkson and R. Jove, *Mol. Cancer Ther.*, 2006, **5**, 621–629.

[19] S. Lassmann, I. Schuster, A. Walch, H. Göbel, U. Jütting, F. Makowiec, U. Hopt and M. Werner, *J. Clin. Pathol.*, 2007, **60**, 173–179.

[20] L. C. Griffin, G. F. Tidmarsh, L. C. Bock, J. J. Toole and L. Leung, *Blood*, 1993, **81**, 3271–3276.

[21] W. Li, A. Kaplan, G. Grant, J. Toole and L. Leung, *Blood*, 1994, **83**, 677–682.

[22] R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe and J. Feigon, *Proc. Natl. Acad. Sci.*, 1993, **90**, 3745–3749.

[23] Y. Wang, F. Yang and X. Yang, *Biosens Bioelectron*, 2010, **25**, 1994–1998.

[24] A. Pasternak, F. J. Hernandez, L. M. Rasmussen, B. Vester and J. Wengel, *Nucleic Acids Res.*, 2010, **39**, 1155–1164.

[25] P. Schultze, R. F. Macaya and J. Feigon, *J. Mol. Biol.*, 1994, **235**, 1532–1547.

[26] L. J. Aaldering, V. Poongavanam, N. Langkjær, N. A. Murugan, P. T. Jørgensen, J. Wengel and R. N. Veedu, *ChemBioChem*, 2017, **18**, 755–763.

[27] W. Kotkowiak, J. Lisowiec-Wachnicka, J. Grynda, R. Kierzek, J. Wengel and A. Pasternak, *Mol. Ther. Nucleic Acids*, 2018, **10**, 304 – 316.

[28] Q. W. Hughes, B. T. Le, G. Gilmore, R. I. Baker and R. N. Veedu, *Molecules*, 2017, **22**, 1770.

[29] C. Yang, M. Kulkarni, M. Lim and Y. Pak, *Nucleic Acids Res.*, 2017, **45**, 12648–12656.

[30] C. Riccardi, I. Russo Krauss, D. Musumeci, F. Morvan, A. Meyer, J.-J. Vasseur, L. Paduano and D. Montesarchio, *ACS Appl Mater Interfaces*, 2017, **9**, 35574–35587.

[31] A. Roloff, A. S. Carlini, C. E. Callmann and N. C. Gianneschi, *J. Am. Chem. Soc.*, 2017, **139**, 16442–16445.

[32] J. Lietard, H. Abou Assi, I. Gomez-Pinto, C. González, M. M. Somoza and M. J. Damha, *Nucleic Acids Res.*, 2017, **45**, 1619–1632.

[33] S. Y. Tan, C. Acquah, S. Y. Tan, C. M. Ongkudon and M. K. Danquah, *Process Biochem.*, 2017, **60**, 42–51.

[34] J. Zhang, R. R. O. Loo and J. A. Loo, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 1815–1822.

[35] I. R. Krauss, V. Napolitano, L. Petraccone, R. Troisi, V. Spiridonova, C. A. Mattia and F. Sica, *Int. J. Biol. Macromol.*, 2018, **107**, 1697–1705.

[36] W. Jianhao, G. Yaqin, L. Li, W. Cheli, W. Jianpeng, D. Shumin, L. Jinping, Q. Lin and J. Pengju, *J. Sep. Sci.*, **40**, 3161–3167.

[37] J. Ye, Y. Zhang, H. Huang, J. Wang, Z. Jiang and M. Bai, *Sens Actuators B: Chemical*, 2017, **243**, 1255–1260.

[38] S. M. Kerwin, *Curr. Pharm. Des*, 2000, **6**, 441–471.

[39] J.-L. Mergny and C. Hélène, *Nat. Med.*, 1998, **4**, 1366.

[40] S. Neidle and G. N. Parkinson, *Biochimie*, 2008, **90**, 1184 – 1196.

[41] H. Han and L. H. Hurley, *Trends Pharmacol. Sci.*, 2000, **21**, 136 – 142.

[42] A. N. Lane, J. B. Chaires, R. D. Gray and J. O. Trent, *Nucleic Acids Res.*, 2008, **36**, 5482–5515.

[43] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res.*, 2006, **34**, 5402–5415.

[44] B. A. Tucker, J. S. Hudson, L. Ding, E. Lewis, R. D. Sheardy, E. Kharlampieva and D. Graves, *ACS Omega*, 2018, **3**, 844–855.

[45] J. Šponer, G. Bussi, P. Stadlbauer, P. Kührová, P. Banáš, B. Islam, S. Haider, S. Neidle and M. Otyepka, *Biochim. Biophys. Acta*, 2017, **1861**, 1246–1263.

[46] E. D. M., N. Julia and C. G. H., *Angew. Chem. Int. Ed.*, 2017, **56**, 11640–11644.

[47] T. Fujii, P. Podbevšek, J. Plavec and N. Sugimoto, *J. Inorg. Biochem.*, 2017, **166**, 190–198.

[48] R. Sha, L. Xiang, C. Liu, A. Balaeff, Y. Zhang, P. Zhang, Y. Li, D. N. Beratan, N. Tao and N. C. Seeman, *Nature Nanotechnol.*, 2018, 1.

[49] N. Špacková, I. Berger and J. Šponer, *J. Am. Chem. Soc.*, 1999, **121**, 5519–5534.

[50] S. Chowdhury and M. Bansal, *J. Phys. Chem. B*, 2001, **105**, 7572–7578.

[51] M. Havrila, P. Stadlbauer, B. Islam, M. Otyepka and J. Sponer, *J. Chem. Theory Comput.*, 2017, **13**, 3911–3926.

[52] D. Miyoshi, A. Nakao and N. Sugimoto, *Nucleic Acids Res.*, 2003, **31**, 1156–1163.

[53] P. Stadlbauer, L. Mazzanti, T. Cragnolini, D. J. Wales, P. Derreumaux, S. Pasquali and J. Šponer, *J. Chem. Theory Comput*, 2016, **12**, 6077–6097.

[54] B. Islam, P. Stadlbauer, M. Krepl, J. Koca, S. Neidle, S. Haider and J. Sponer, *Nucleic Acids Res.*, 2015, **43**, 8673–8693.

[55] M.-H. Li, Q. Luo, X.-G. Xue and Z.-S. Li, *J. Mol. Struct.: THEOCHEM*, 2010, **952**, 96–102.

[56] E. Zavyalova, G. Tagiltsev, R. Reshetnikov, A. Arutyunyan and A. Kopylov, *Nucleic Acid Ther.*, 2016, **26**, 299–308.

[57] I. Smirnov and R. H. Shafer, *Biochemistry*, 2000, **39**, 1462–1468.

[58] P. Hazel, J. Huppert, S. Balasubramanian and S. Neidle, *J. Am. Chem. Soc.*, 2004, **126**, 16405–16415.

[59] N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E. S. Baker, M. T. Bowers, E. De Pauw and V. Gabelica, *J. Am. Chem. Soc.*, 2008, **130**, 10208–10216.

[60] A. Guedin, J. Gros, P. Alberti and J.-L. Mergny, *Nucleic Acids Res.*, 2010, **38**, 7858–7868.

[61] A. Risitano and K. R. Fox, *Nucleic Acids Res.*, 2004, **32**, 2598–2606.

[62] Y. Bian, W. Ren, F. Song, J. Yu and J. Wang, *J. Chem. Phys*, 2018, **148**, 204107.

[63] T. Cragnolini, D. Chakraborty, J. poner, P. Derreumaux, S. Pasquali and D. J. Wales, *J. Chem. Phys*, 2017, **147**, 152715.

[64] A. E. Bergues-Pupo, I. Gutiérrez, J. R. Arias-Gonzalez, F. Falo and A. Fiasconaro, *Sci. Rep.*, 2017, **7**, 11756.

[65] K. Phillips, Z. Dauter, A. I. Murchie, D. M. Lilley and B. Luisi, *J. Mol. Biol.*, 1997, **273**, 171–182.

[66] R. Ida and G. Wu, *J. Am. Chem. Soc.*, 2008, **130**, 3590–3602.

[67] P. Sket and J. Plavec, *J. Am. Chem. Soc.*, 2010, **132**, 12724–12732.

[68] N. V. Hud, P. Schultze, V. Sklenář and J. Feigon, *J. Mol. Biol.*, 1999, **285**, 233–243.

[69] C. C. Hardin, T. Watson, M. Corregan and C. Bailey, *Biochemistry*, 1992, **31**, 833–841.

[70] E. A. Venczel and D. Sen, *Biochemistry*, 1993, **32**, 6220–6228.

[71] J. S. Lee, *Nucleic Acids Res.*, 1990, **18**, 6057–6060.

[72] F. M. Chen, *Biochemistry*, 1992, **31**, 3769–3776.

[73] N. Nagesh and D. Chatterji, *J. Biochem. Biophys. Methods*, 1995, **30**, 1–8.

[74] S. Basu, A. A. Szewczak, M. Cocco and S. A. Strobel, *J. Am. Chem. Soc.*, 2000, **122**, 3240–3241.

[75] D. Miyoshi, A. Nakao, T. Toda and N. Sugimoto, *FEBS Lett.*, 2001, **496**, 128–133.

[76] M. Cai, X. Shi, V. Sidorov, D. Fabris, Y.-f. Lam and J. T. Davis, *Tetrahedron*, 2002, **58**, 661–671.

[77] C. Yang, S. Jang and Y. Pak, *J. Chem. Phys.*, 2011, **135**, 225104.

[78] J. Aqvist, *J. Phys. Chem.*, 1990, **94**, 8021–8024.

[79] R. V. Reshetnikov, J. Sponer, O. I. Rassokhina, A. M. Kopylov, P. O. Tsvetkov, A. A. Makarov and A. V. Golovin, *Nucleic Acids Res.*, 2011, **39**, 9789–9802.

[80] D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg,

C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, Walker, J. Wang, R. M. Wolf, X. Wu and P. A. Kollman, *Amber 14*, University of California, San Francisco, 2014.

[81] X.-a. Mao, L. A. Marky and W. H. Gmeiner, *J. Biomol. Struct. Dyn.*, 2004, **22**, 25–33.

[82] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham III, C. A. Laughton and M. Orozco, *Biophys. J.*, 2007, **92**, 3817–3829.

[83] J. Šponer, X. Cang and T. E. Cheatham III, *Methods*, 2012, **57**, 25–39.

[84] E. Fadrna, N. Spackova, J. Sarzyñska, J. Koca, M. Orozco, T. E. Cheatham III, T. Kulinski and J. Sponer, *J. Chem. Theory Comput.*, 2009, **5**, 2514–2530.

[85] *Innovations in Biomolecular Modeling and Simulations*, ed. T. Schlick, The Royal Society of Chemistry, 2012, vol. 2, pp. P001–330.

[86] J. L. Gelpí, S. G. Kalko, X. Barril, J. Cirera, X. de la Cruz, F. J. Luque and M. Orozco, *Proteins: Struct., Funct., Bioinf.*, 2001, **45**, 428–437.

[87] P. Li and K. M. Merz, *J. Chem. Theory Comput.*, 2014, **10**, 289–297.

[88] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

[89] I. S. Joung and T. E. Cheatham III, *J. Phys. Chem. B*, 2008, **112**, 9020–9041.

[90] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

[91] J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

[92] R. W. Pastor, B. R. Brooks and A. Szabo, *Mol. Phys.*, 1988, **65**, 1409–1419.

[93] B. P. Uberuaga, M. Anghel and A. F. Voter, *J. Chem. Phys.*, 2004, **120**, 6363–6374.

[94] D. J. Sindhikara, S. Kim, A. F. Voter and A. E. Roitberg, *J. Chem. Theory Comput.*, 2009, **5**, 1624–1631.

[95] D. R. Roe and T. E. Cheatham III, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.

[96] W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.

[97] R. Kumar and H. Grubmüller, *Bioinformatics*, 2015, **31**, 2583–2585.

[98] M. Brehm and B. Kirchner, *TRAVIS-a free analyzer and visualizer for Monte Carlo and molecular dynamics trajectories*, 2011.

[99] T.-A. F. Analyzer, *J. Chem. Inf. Model.*, 2011, **51**, 2007–2023.

[100] M. Rebic, A. Laaksonen, J. Sponer, J. Ulicny and F. Mocci, *J. Phys. Chem. B*, 2016, **120**, 7380–7391.

[101] X. Cang, J. Šponer and T. E. Cheatham III, *Nucleic Acids Res.*, 2011, **39**, 4499–4512.

[102] I. Persson, M. Sandström and H. Yokoyama, *Z. Naturforsc.*, 1995, **50**, 21–37.

[103] R. Caminiti, A. Musinu, G. Paschina and G. Pinna, *J. Appl. Crystallogr.*, 1982, **15**, 482–487.

[104] D. M. Pfund, J. G. Darab, J. L. Fulton and Y. Ma, *J. Phys. Chem.*, 1994, **98**, 13102–13107.

[105] T. Seward, C. Henderson, J. Charnock and T. Driesner, *Geochim. Cosmochim. Acta*, 1999, **63**, 2409–2418.

[106] D. Harris, J. Brodholt and D. Sherman, *J. Phys. Chem. B*, 2003, **107**, 9056–9058.

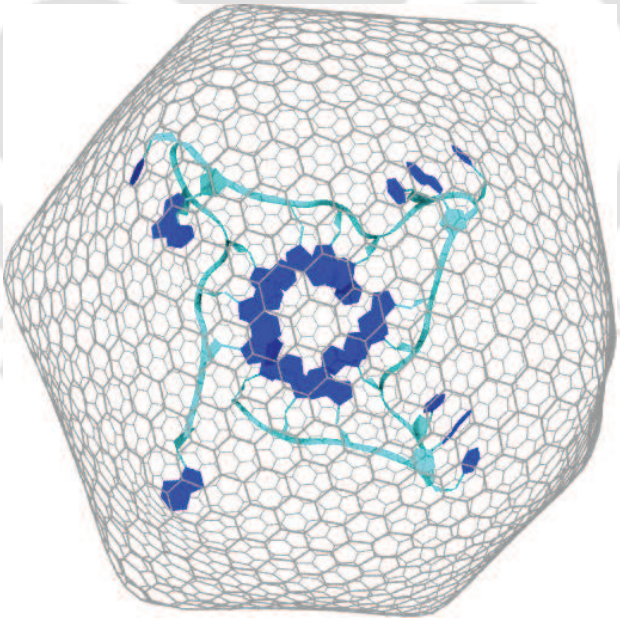
[107] S. Das and S. Paul, *J. Phys. Chem. B*, 2017, **121**, 8774–8785.

[108] B. Sharma and S. Paul, *J. Phys. Chem. B*, 2016, **120**, 9019–9033.

[109] G. Borgohain and S. Paul, *Comput. Theor. Chem.*, 2018, **1131**, 78 – 89.

Chapter 3

Conformational Deviation of the Human Parallel Telomeric G-quadruplex DNA in Presence of Different Salt Concentrations and Temperatures under Confinement



Overview

Recently, the stability of G-quadruplex DNA inside the close confinement such as α-hemolysin, nanocavity water pool and different metal-organic-frameworks (MOFs), draws attention of the scientific world. To understand the conformational change of G-quadruplex DNA at atomistic level, we have carried out a total of 40 μs simulation run under both non-polar and polar confinement conditions. To investigate the dynamics, we have considered two different KCl salt concentrations i.e., 0 M (minimal salt concentration) and higher than 2 M salt concentration (higher salt concentration) at two different temperatures, 300 K and 350 K. Here, we have observed that the human telomeric G-quadruplex DNA deviates more from its crystal structure at minimal salt concentration in both non-polar and polar confinement conditions. Moreover, loop regions deviate and fluctuate more compared to the other regions, i.e., sugar-phosphate backbone and tetrad regions. The K^+ ions are found to be primarily responsible for this phenomenon. From the spatial density function (SDF) plots , higher density of K^+ ions are observed in the backbone region. Moreover, from the residue-wise first solvation shell estimation, we have noticed that the K^+ ions mainly accumulate in the tetrad region under both the non-polar and polar confinement conditions due to which the tetrad regions are more rigid than the loop regions. Moreover, higher salt concentration results in increased rigidity of the G-quadruplex DNA. Our study enlightens valuable insight about the conformational deviation of the G-quadruplex DNA in the field of nanotechnology.

3.1 Introduction

Various studies [1–4] reported that G-quadruplex DNA is commonly found in oncogene promoter and telomere (end of the noncoding part of the chromosome) regions. Telomere replication is mainly governed by the telomerase enzyme. The length of the human telomere shortens during each cell division cycle in regular somatic cells, which leads to its apoptosis [5]. This ultimate fate is prevented by cancer cells and is immortalised by activating the telomerase enzyme, which stabilizes the length of telomere. Thus, it is important to form G-quadruplex structure, which can prevent the activation of the telomerase enzyme. Therefore, for genomic stability and cell ageing, telomere structure and stability play a very crucial role [6–8] in the medicinal field.

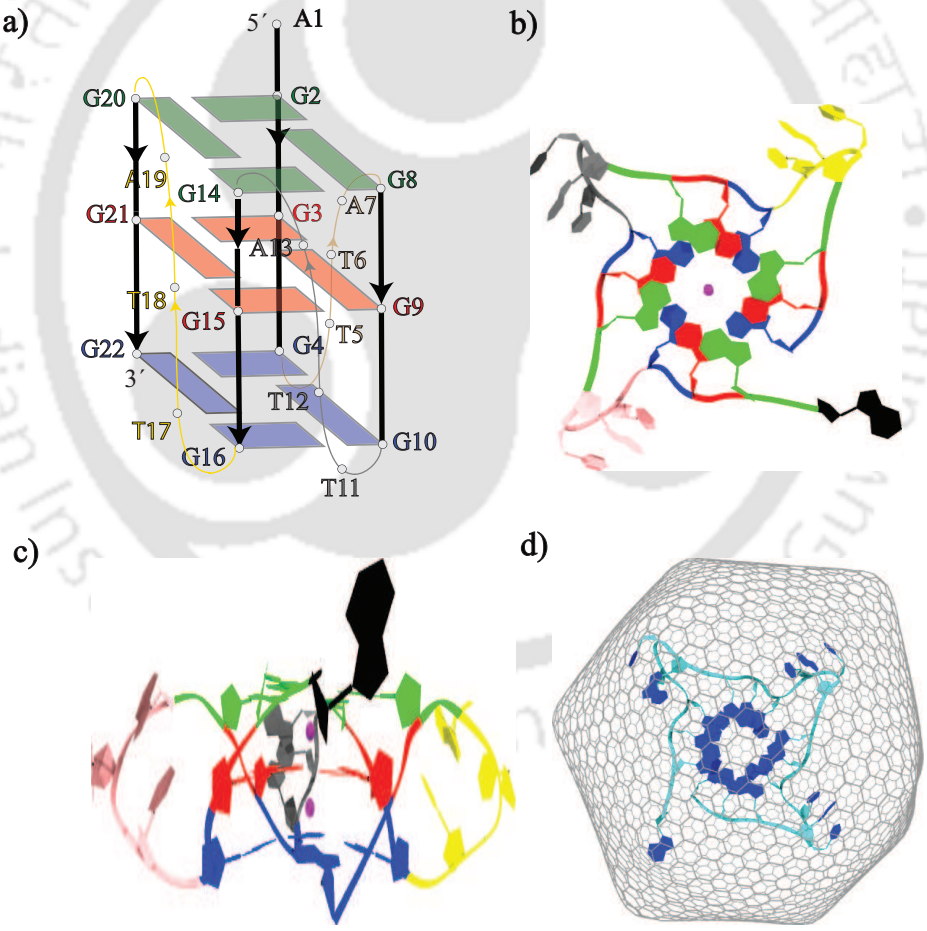


Figure 3.1: *Schematic depiction of (a) human telomeric G-quadruplex DNA, (b) top view, (c) side view picture in new ribbon representations of human telomeric G-quadruplex DNA, and (d) human telomeric G-quadruplex DNA inside the confinement. Two central K^+ cations are shown in purple color. Tetrad-1, tetrad-2, loop-1, loop-2, and loop-3 are shown in green, red, blue, pink, gray, and yellow, respectively. Guanine and thymine, adenine are represented as G, T, and A respectively.*

A parallel topology of 22 mer human telomeric quadruplex DNA (d[AGGG(TTAGGG)$_3$]) consists of three tetrads (namely tetrad-1, tetrad-2, and tetrad-3), three propeller loops (thymine-thymine-adenine) and an adenine-1 flanking base (**Fig. 3.1**).

Recently, the stability of G-quadruplex DNA in nanopore protein cavity, has drawn attention in the scientific world. In 2008, Shim et al., reported [9] the encapsulating of thrombin binding aptamer (TBA) in α-hemolysin (αHL) protein nanocavity and also the folding and unfolding behavior of the TBA in another work [10]. Moreover, they have also shown the entry of TBA in β-barrel like nanocavity. They have observed that, among the various cations, in presence of K^+ cations, TBA forms G-quadruplex structure, having the most stable conformation.

In 2013, another group of scientists, Burrows and her coworkers, investigated the structural dependency of human telomere sequence in α-hemolysin nanopore [11]. Moreover, they studied the interaction between the human telomere DNA and α-hemolysin nanopore. Furthermore, they observed, the existence of two interchangeable conformations (namely hybrid-1 and hybrid-2) of human telomere G-quadruplex DNA, in presence of K^+ cations. In another pioneer work [12], they also examined the entry and stabilization of four types (namely hybrid-1, hybrid-2, basket and propeller) of conformational structures of human telomere G-quadruplex DNA in α-hemolysin nanopore cavity. They found that hybrid and basket like human telomere G-quadruplex DNA structures are capable of trans-location to the nanopore cavity in presence of optimal current. However, propeller fold human telomere G-quadruplex DNA cannot pass the entry gate or latch zone of α-hemolysin nanopore due to its disk-like shape. Thus, it unfolds outside the α-hemolysin nanopore much faster than the other hybrid and basket structure.

In 2019, Ma et al., investigated [13] the femtosecond time-resolved fluorescence coupled with transient absorption on human telomeric G-quadruplex DNA. They found that basket type topology of human telomeric G-quadruplex DNA converts to hybrid type topologies in presence of K^+ ions, inside the confinement by nanocavity water pool. They also examined the effect of metal cations like Na^+ and K^+ on folding topology of human telomeric G-quadruplex DNA.

Very recently Peng and his coworker reported [14] the separation of different types of DNA and RNA nucleic acid by using metal-organic framework. Moreover, they identified three critical factors for encapsulation of different nucleic acid, which are (i) steric hindrance of size, (ii) stability of conformational transition energy, and (iii) different molecular weight

of nucleic acids. Using these above-mentioned criteria, they have extracted the flexible and unstable structures of nucleic acids from different mixture.

However, all these experimental studies failed to provide the conformational changes of nucleic acids at atomic level. Thus, molecular dynamics studies are utilized to acquire useful information of the conformational changes in nanocavity or close confinement. Till now, the conformational deviation and behavior of G-quadruplex DNA inside these kind of nanocavity or confinement, have not been studied via theoretical approach. Thus, to understand the dynamics of G-quadruplex DNA, we have carried out total of 40 μs all atom molecular dynamics simulation. In this study, we have studied the conformational deviation of human parallel telomeric G-quadruplex DNA inside the non-polar as well as polar confinement. Here, we have performed molecular dynamics under two different temperatures i.e., 300 K and 350 K, and at two different KCl salt concentrations i.e., 0 M (minimal salt concentration) and higher than 2 M salt concentration.

3.2 Models and Simulation Method

Following our previous works [15, 16] a nearly spherical fullerene like ball was taken as a confinement, which consists of a total of, 2940 carbon atoms. There is no charge for non-polar confinement, but for polar confinement, half of the carbon atoms contain -0.15e charge and the rest of the carbon atoms contain +0.15e to obtain an overall neutral charge for the confinement. To construct the confinement, CaGe software [17] was used. All atom molecular dynamics simulation of the aqueous solution of the parallel telomeric DNA were performed inside a non-polar as well as a polar nearly spherical confinement using AMBER18 package [18]. The crystal structure of the 22-mer parallel telomeric quadruplex DNA (d[AGGG(TTAGGG)$_3$]) was obtained from the RCSB Protein Data Bank (RCSB PDB) with PDB ID 1KF1 [19]. Initially, this crystal structure contains three K^+ ions which were placed in the central channel core of the quadruplex DNA. Thus, 18 K^+ more ions were added to neutralize the aqueous quadruplex DNA system. 2500 TIP3P [20] water molecules were used to solvate the parallel telomeric quadruplex DNA. Moreover, to maintain the KCl salt concentrations, required amount of the K^+ and Cl^- ions were added to the system. The details of the system were shown in **Table 3.1**. Initial configuration

Table 3.1: *N_{Q-DNA}, N_{water}, N_{K^+}, and N_{Cl^-}, denote the number of the human parallel telomeric G-quadruplex DNA, water, K^+, Cl^- molecules respectively, used for different systems. C_{KCl} represent the concentration in molarity.*

System	N_{Q-DNA}	N_{water}	N_{K^+}	N_{Cl^-}	C_{KCl}(M)	Confinement	Time (μs)
N0_300	1	2500	21	0	0	Non-polar	5
N0_350	1	2500	21	0	0	Non-polar	5
N2_300	1	2500	282	261	2.77	Non-polar	5
N2_350	1	2500	282	261	2.77	Non-polar	5
P0_300	1	2500	21	0	0	Polar	5
P0_350	1	2500	21	0	0	Polar	5
P2_300	1	2500	282	261	2.77	Polar	5
P2_350	1	2500	282	261	2.77	Polar	5

for every system was generated using PACKMOL [21] software. The widely used [22–30] latest AMBER DNA force field leaprc.DNA.OL15 [31, 32] , which includes parmbsc0 [33] and OL15 [31] corrections to the ff10 forcefield, was employed for the parallel telomeric quadruplex DNA. TIP3P water specific AMBER-adapted Joung and Cheatham force-field [34] parameters were used for K^+ and Cl^- ions. Following our previous studies [15, 16], for fullerene like ball, General Amber Force Field (GAFF) [35] was used. Considering Waghe et al., work [36], the Lennard Jones parameter "ϵ" of carbon atom of the confinement was reduced to half of its original value.

Considering our previous works [37–39], to perform the simulations, following steps were employed for all systems: (i) At first, to remove the bad contacts between the parallel telomeric quadruplex DNA, confinement, waters and ions, energy minimization was done for 5000 steps by steepest descent method, followed by the equal number of the steps by conjugate gradient method. (ii) Then, each system was heated gradually from 0 K to the desired temperature at an interim of 50 K for energy 25 ps in canonical ensemble (NVT) using a collision frequency of 1.0 ps^{-1}. (iii) After that, every system was equilibrated in canonical (NVT) ensemble for 250 ps at a temperature of 300 K. (iv) All atoms of the telomeric quadruplex DNA, including the central core two K^+ ions and the confinement, were fixed with a harmonic restraint of force constant 100 kcal mol^{-1} $Å^{-2}$ for all the before mentioned steps. (v) After that, the harmonic restraint of force constant was gradually decreased to 25, 20, 15, 10, 5 and finally 0 kcal mol^{-1} $Å^{-2}$ with a series of minimization and equilibration. (vi) Next, a 50 ns equilibration was performed in NPT ensemble to maintain the density of the system. (vii) Finally, a 5 μs production run was performed for each

system at 1 atm physical pressure and the considered temperatures with a 5 ps^{-1} collision frequency.

To maintain the temperature and pressure, Langevin dynamics [40] and Berendsen barostat [41–43] with a pressure relaxation time of 2 ps, were used. SHAKE algorithm [44] was implemented to restrain the covalent bonds involving hydrogen atoms with a tolerance of 10^{-5} Å. A 2 fs time integration step was employed. A cut off radius of 12 Å was set. Particle mesh Ewald method (PME) [45] was implemented to estimate all non bonded short-ranged interactions and the long-ranged electrostatic interactions. CPPTRAJ [46] of Amber-Tools19 [18] and Visual Molecular Dynamics (VMD) [47] were used to analyze the trajectories and also as a visualization tool respectively. To estimate spatial distribution functions (SDFs), TRAVIS [48] was used.

3.3 Results and Discussions

3.3.1 Structural Changes of Parallel Telomeric Quadruplex DNA

3.3.1.1 Root Mean Square Deviations (RMSDs)

Considering previous studies [49, 50] root-mean-square deviation (RMSD) is used to measure the deviations of a DNA from its reference state. In **Fig. 3.2**, we have shown the rms fit RMSD values of all the heavy atoms of parallel telomeric quadruplex DNA for all systems. Moreover, the crystal structure is considered as a reference structure. Here, the RMSD values (**Fig. 3.2 (a) and (e)**) below 2 Å indicate that the parallel telomeric quadruplex DNA slightly deviates from its reference structure for all the considered conditions. From overall RMSD values, it is very difficult to recognize the regions which deviate more from its crystal structure. Thus, we have calculated the RMSDs (**Fig. 3.2 (b)-(d) and (f)-(h)**) of separated regions of the parallel telomeric quadruplex DNA i.e., (i) sugar-phosphate backbone, (ii) tetrad, and (iii) loop regions. Here, we notice that the RMSD values of loop regions deviate more from its crystal structure, followed by the sugar-phosphate backbone and the tetrad regions. Moreover, at higher salt concentration, it deviates much less than that at minimal salt concentration. Furthermore, the difference between the RMSD values at 300 and 350 K temperatures for all considered regions of parallel telomeric quadruplex DNA, are very similar. For further study of the structural

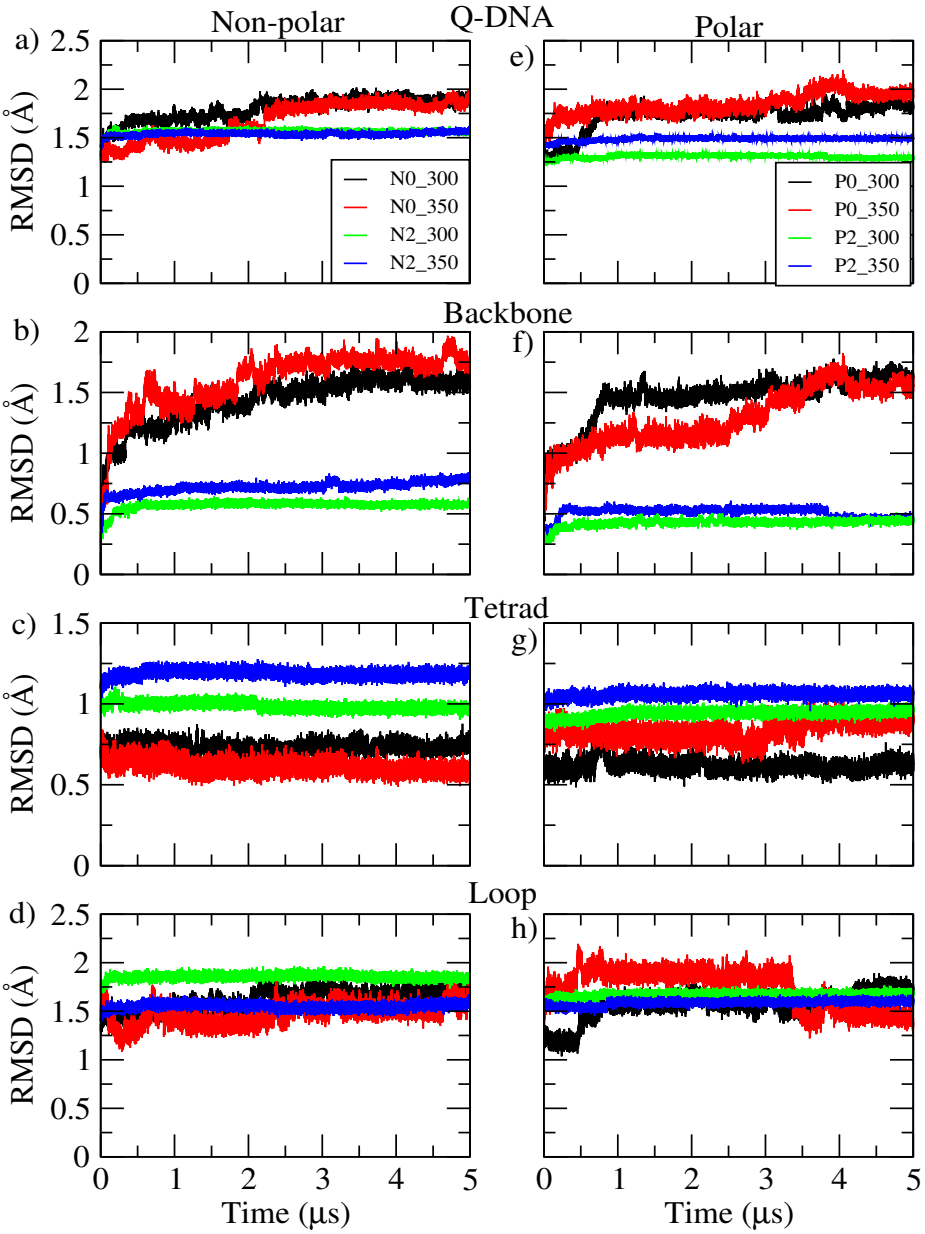


Figure 3.2: *Time progression of the root-mean-square deviations (RMSDs) of all heavy atoms of human telomeric G-quadruplex DNA for different systems.*

change of the parallel telomeric quadruplex DNA, we have performed pair-wise RMSD for all systems (**Fig. 3.3**). From **Fig. 3.3**, it is clear that the conformational deviations of the parallel telomeric quadruplex DNA are more for polar confinement than the non-polar confinement. However, at higher salt concentration, these deviations are very much lower as compared to the minimal salt concentration for both the considered temperatures.

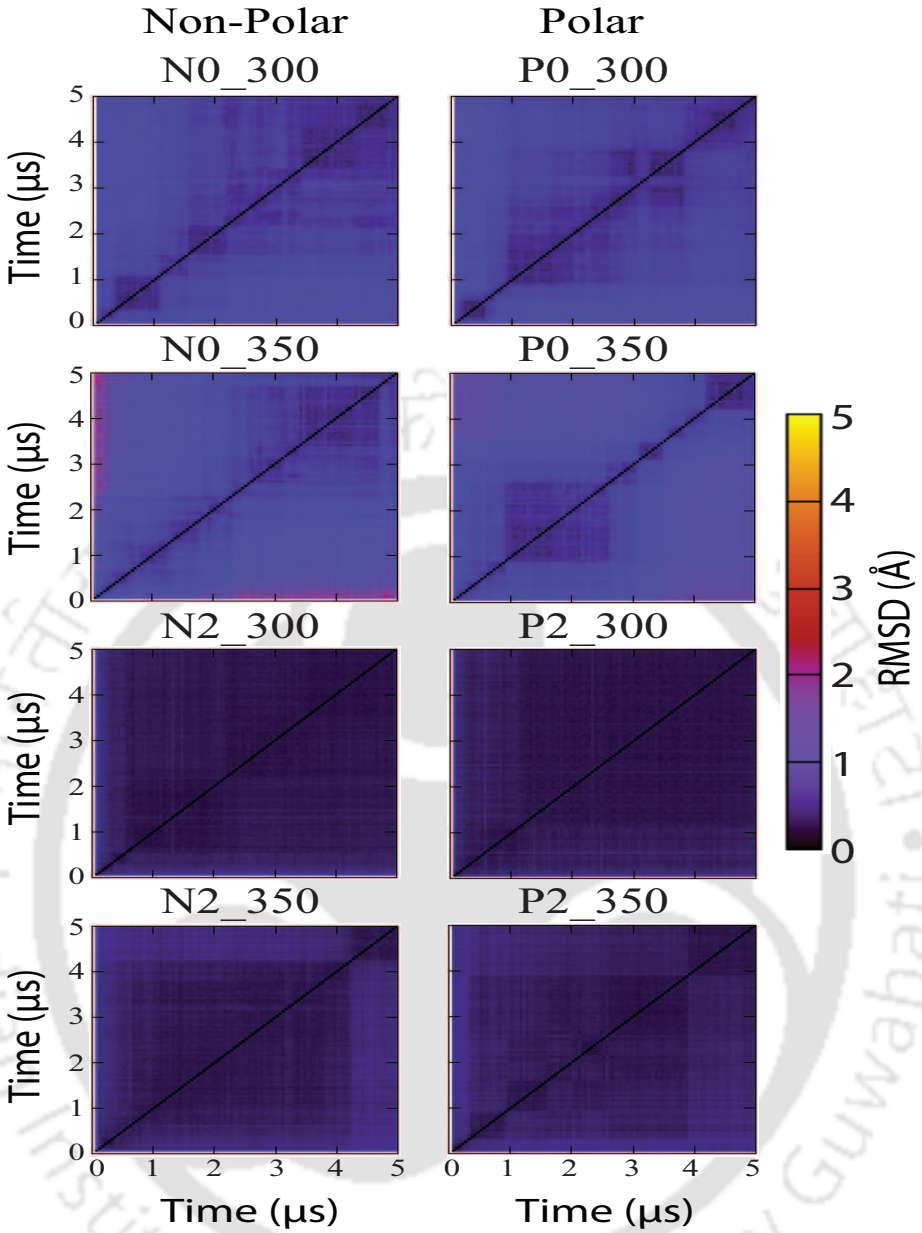


Figure 3.3: *Pair-wise root-mean-square deviations (2D-RMSDs) of all heavy atoms of human telomeric G-quadruplex DNA for different systems.*

3.3.1.2 Root Mean Square Fluctuations (RMSFs)

Next, we evaluate the root-mean-square-fluctuation (RMSF) for the examining conformational change of a particular residue of the parallel telomeric quadruplex DNA. In **Fig. 3.4**, we have plotted the RMSF values of the parallel telomeric quadruplex DNA considering all

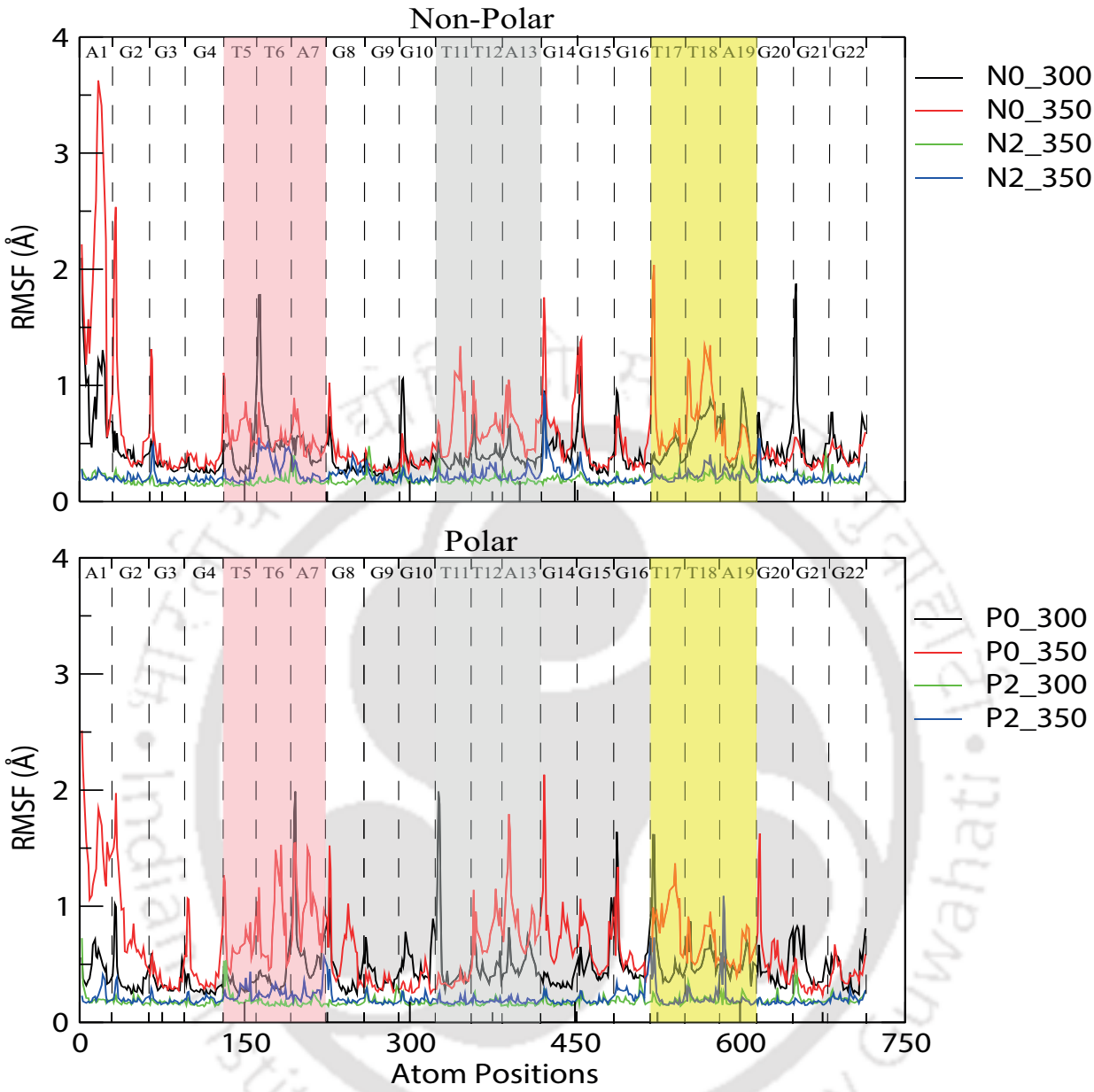


Figure 3.4: *Root-mean-square fluctuations (RMSFs) of all heavy atoms of human telomeric G-quadruplex DNA for different systems. Loop-1, loop-2, and loop-3 regions are shown in pink, gray, and yellow, respectively.*

the heavy atoms. For these estimations, we have considered the crystal structure as the reference structure. Here, it is noticed that, mainly loop residues fluctuate more in minimal salt conditions than the higher salt concentration. Moreover, the same scenario is observed for both the non-polar and the polar confinement. In more details, highest RMSF values are observed for N0_350 (upto 3.6 Å) and P0_350 (upto 2.5 Å) systems. Among, the non-polar and polar confinement conditions, maximum number of atoms of the parallel telomeric quadruplex DNA fluctuate more in polar confinement conditions except for the

flanking base A1 atoms. When the temperature changes from 300 to 350 K, at higher salt conditions, the RMSF values do not differ very much in either non-polar or polar confinement conditions. However, this scenario is completely different at minimal salt concentration. Here, higher RMSF values of the atoms of the parallel telomeric quadruplex DNA, is observed for higher temperature.

From RMSD, 2D-RMSD, and RMSF analyses we have observed that the parallel telomeric quadruplex DNA deviates and fluctuates more from its crystal structure at minimal salt concentration for both the polar and non-polar confinement conditions. On the other way, we can say that the higher salt concentration make the parallel telomeric quadruplex DNA more rigid than the lower salt concentration.

3.3.2 Tetrad Dynamics of Parallel Telomeric Quadruplex DNA

Now, it is more important to check the dynamics of the tetrads of the parallel telomeric quadruplex DNA. For this investigation, we have checked two parameters of the tetrads. These are hydrogen bond estimation of the tetrad residues and the distance of the central cations which are placed in the central core of the G-quadruplex ion channels.

3.3.2.1 Hydrogen Bond Estimation of Tetrad Residues

To estimate the hydrogen bonds of the tetrads residues, considering earlier studies [51–56], following two criteria are set: (i) cutoff distance between donor (D) and acceptor is 3.5 Å and simultaneously (ii) the D-A-H angle cutoff is 45°. In **Fig. 3.5**, we have shown the number of hydrogen bonds of the three tetrads (namely tetrad-1, tetrad-2, and tetrad-3) for all the systems. Here, at minimal salt concentration, the number of hydrogen bonds observed is 8 for each of the tetrads in both the non-polar and polar confinement conditions. However, this number is reduced to 5 for tetrad-1 of N0_300 and P0_350 systems. On the other hand, for higher salt concentration systems, the number of hydrogen bonds slightly decrease from 8 to 5 for both the non-polar and polar confinement conditions. This is the reason behind the slightly higher RMSD values of the tetrad residues (**Fig. 3.2 (c) and (g)**) that are noticed for the above mentioned systems.

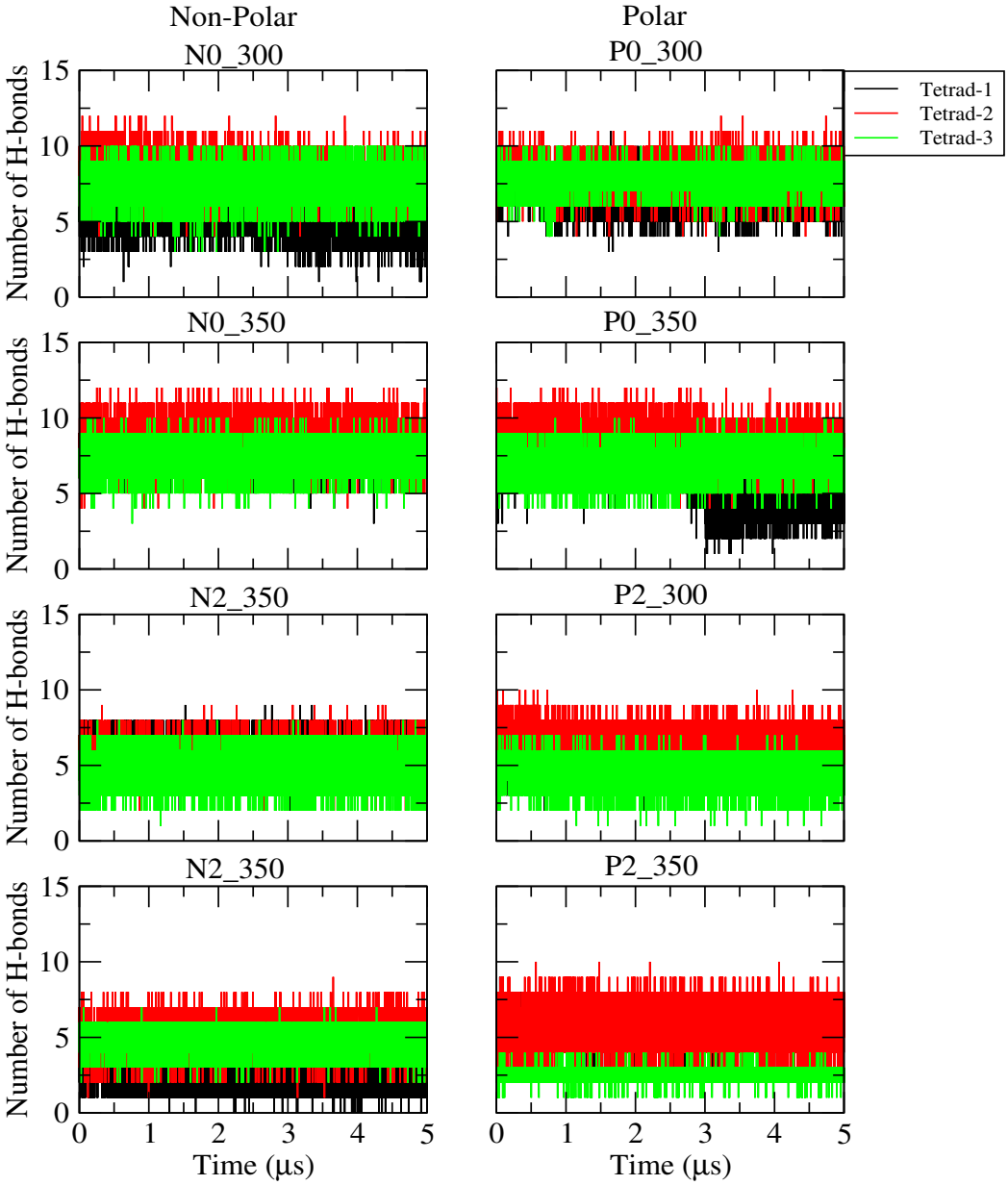


Figure 3.5: *Number of hydrogen bonds for tetrad-1, tetrad-2, and tetrad-3 of human telomeric G-quadruplex DNA for different systems.*

3.3.2.2 Distance Between the Channel Core Central Cations

Central K^+ cations which are placed at the channel core of the g-quadruplex DNA, play a very important role in the formation and stabilization of the G-quadruplex DNA. Thus, it is very important to check the occurrence of any ion-exchange occurred for all the systems. Previously, Rebič et al. [57] reported that one central cation exits from the channel rim of the parallel telomeric quadruplex DNA using TIP3P water [20] model and TIP3P water specific AMBER-adapted Joung and Cheatham force-field [34] parameters for K^+ and Cl^-

ions. Moreover, during this ion-exchange simulation time, larger deviations from the crystal structure of the considered G-quadruplex are noticed, which is reflected in the RMSD values

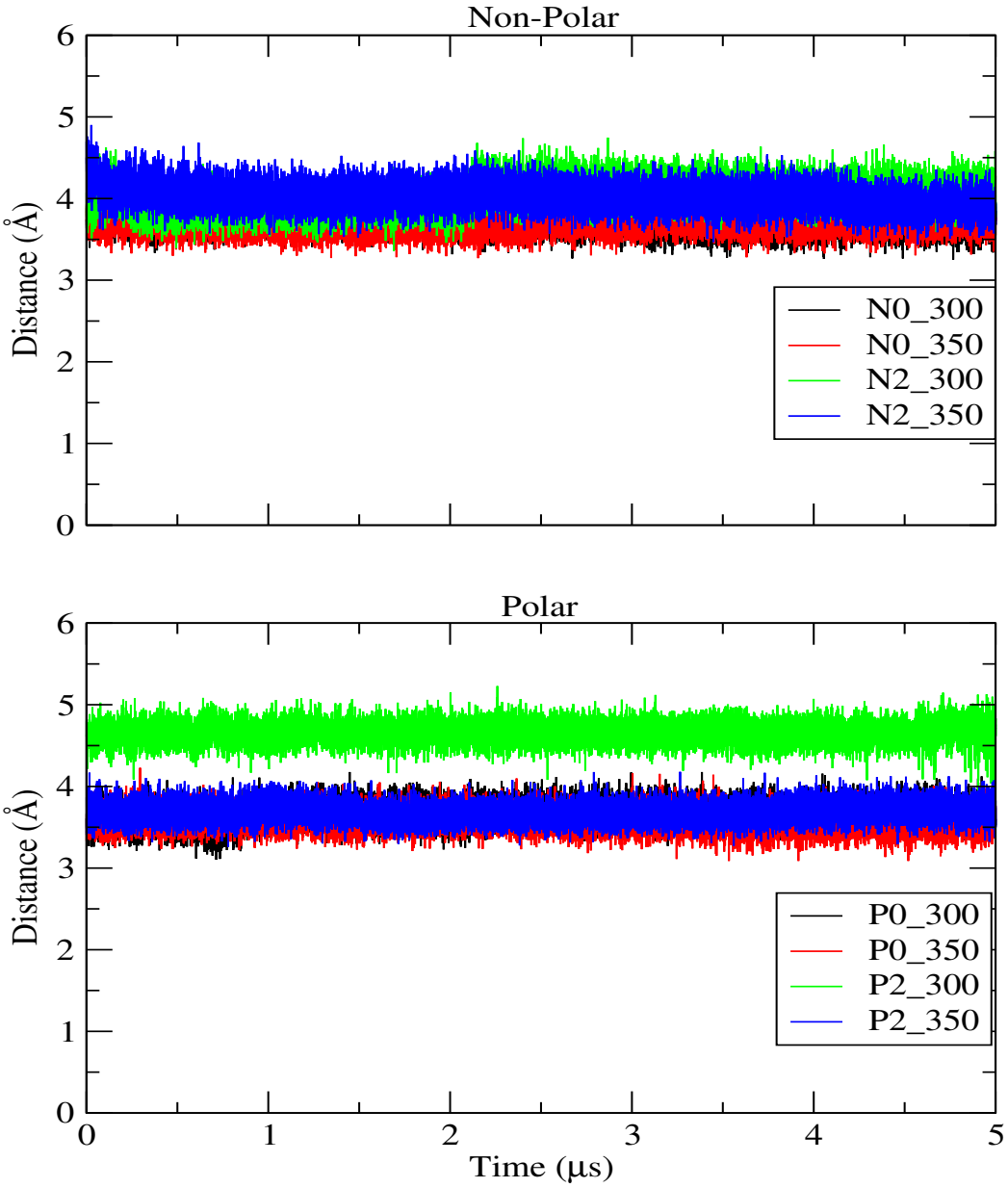


Figure 3.6: *Time progression of the distance between two central cations of channel core of human telomeric G-quadruplex DNA for different systems.*

of the G-quadruplex DNA. In **Fig. 3.6**, we have shown the distance between the two central cations of the parallel telomeric quadruplex DNA through out the 5 μs simulation run for all systems. Here, we have observed that for non-polar confinement the distance

between the two central cations lies between 3.7 and 4.2 Å. On the other hand, this distance varies from 3.6 to 3.7 Å for all systems under polar confinement conditions, except for P2_300 system where slightly higher distance (4.6 Å) is noticed. Thus, it can be concluded that the central cations of the channel core of the parallel telomeric quadruplex DNA are stable in the channel core through out the simulations under both the polar and non-polar confinement conditions.

3.3.3 Loop Dynamics of Parallel Telomeric Quadruplex DNA

The parallel telomeric quadruplex DNA contains three thymine-thymine-adenine (TTA) propeller loops on the side of the G-quadruplex ion channel core. Earlier pioneer works by Islam et al. [22, 26] reported that the propeller loops of the human telomeric quadruplex DNA play an important role in conformational deviations of the G-quadruplex DNAs. From our previous RMSD and RMSF analyses, we have noticed that the major deviation from the crystal structure of the parallel telomeric quadruplex DNA is observed for the loop regions in compared to that of with namely backbone and tetrad regions. Considering the above information, it is very crucial to investigate the loop dynamics of the parallel telomeric quadruplex DNA under non-polar as well as polar confinement conditions with different salt concentrations and temperatures. Now, to check the loop dynamics following parameters are used : (i) Principal Component Analysis, and (ii) π-π stacking interactions.

3.3.3.1 Principal Component Analysis for Different Propeller Loop Regions

Data from an MD simulation containing plenty of degrees of freedom (dof) cannot always be interpreted. Thus, to monitor these large and huge complicated data, principal component analysis (PCA) is used due to its simplicity. In the anharmonic motion of an atomic molecule, PCA eliminates the dominant modes from the trajectory. Moreover, PCA produces the configurational space that consists of a few degrees of freedom of molecule by reducing the dimensionality of the data. Hence, PCA is very important tool to check the structural deviations of the biological molecules [49, 58–61]. So, we have performed PCA analysis considering the backbone atoms of the individual loop (namely loop-1, loop-2, and loop-3) residues. In **Fig. 3.7**, we have presented the cumulative fraction of eigenvalues of the first 50 eigenvector for all systems. Here, we have observed that the majority (more

than 50%) of the structural changes are covered by the first three eigenvectors. **Fig. 3.7**

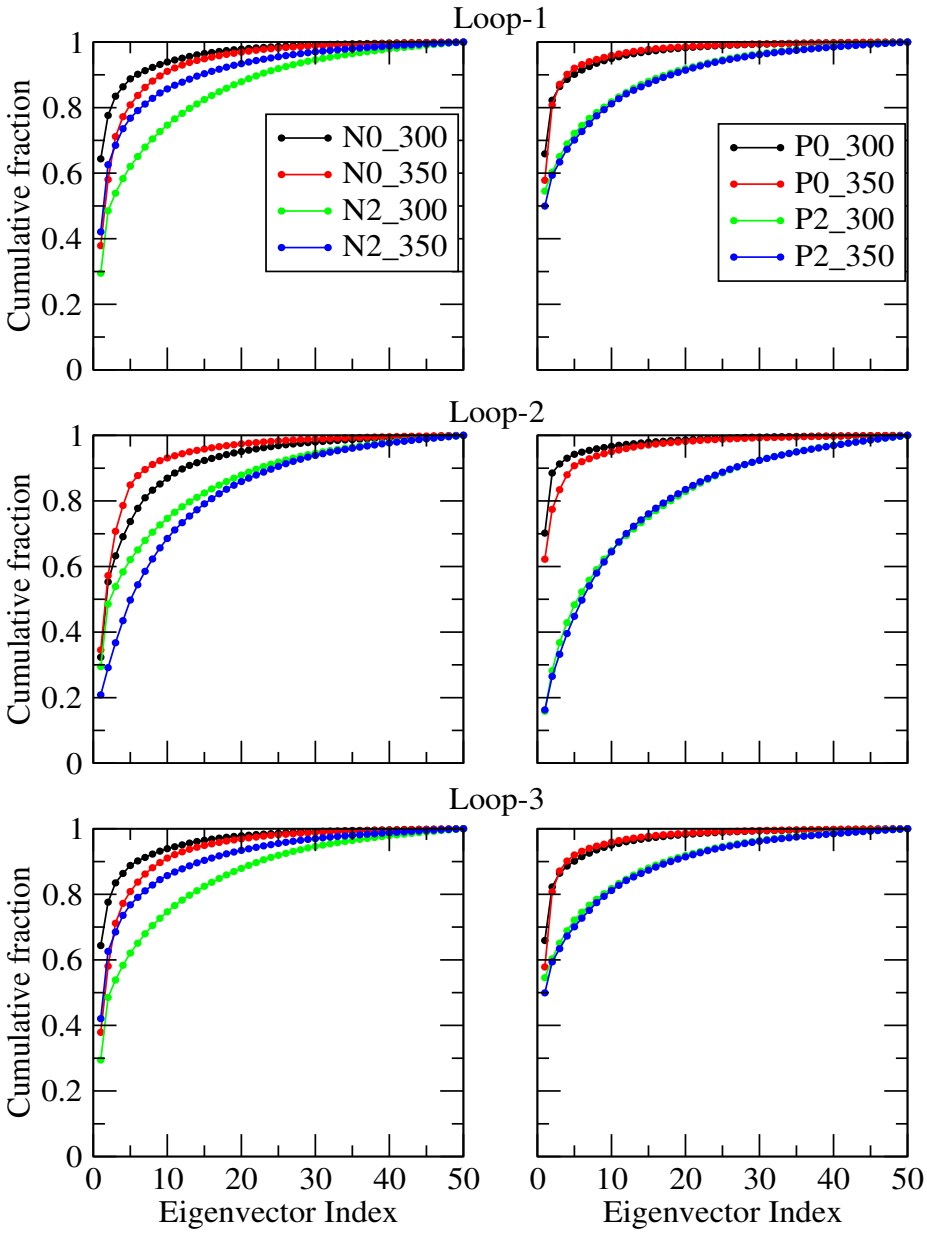


Figure 3.7: *Cumulative fraction of the eigenvalues taken from principal component analysis (PCA) of loop-1, loop-2, and loop-3 of human telomeric G-quadruplex DNA for different systems.*

also indicates that the overall motions (≈100%) are covered by the first 50 eigenvalues. In more details, the most of the dominant modes are noticed by the first principal component (PC1) of the three loop regions, under the non-polar and polar confinement conditions. In **Figures 3.8 and 3.9**, we have shown the dominant modes of the backbone atoms of

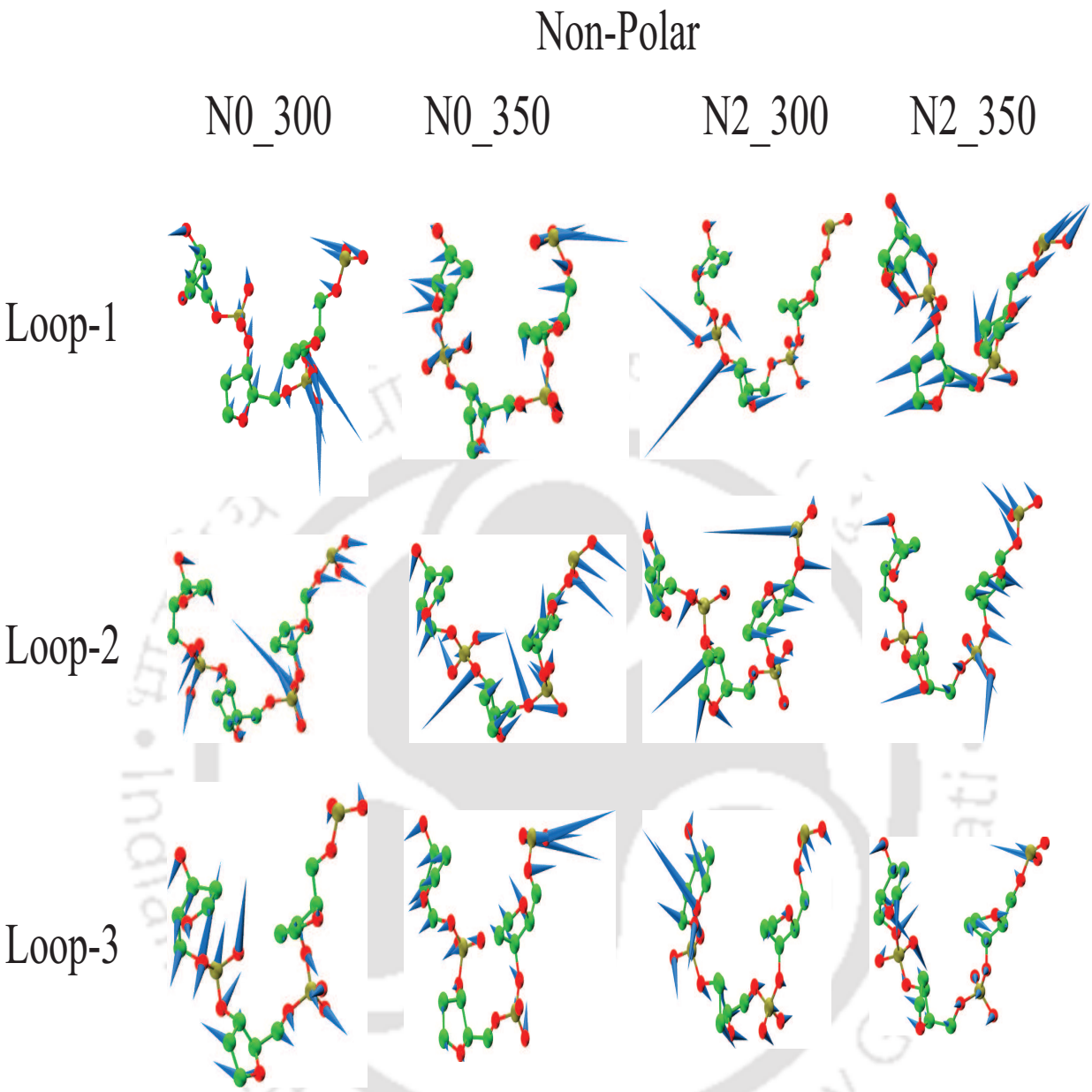


Figure 3.8: *Representing the dominant motions of human telomeric G-quadruplex DNA along with the first eigenvector in porcupine plot for different systems under non-polar confinement. The models are shown as a backbone trace. The arrows attached to each backbone atom indicate the direction of the eigenvector and the size of each arrow shows the magnitude of the corresponding eigenvalue.*

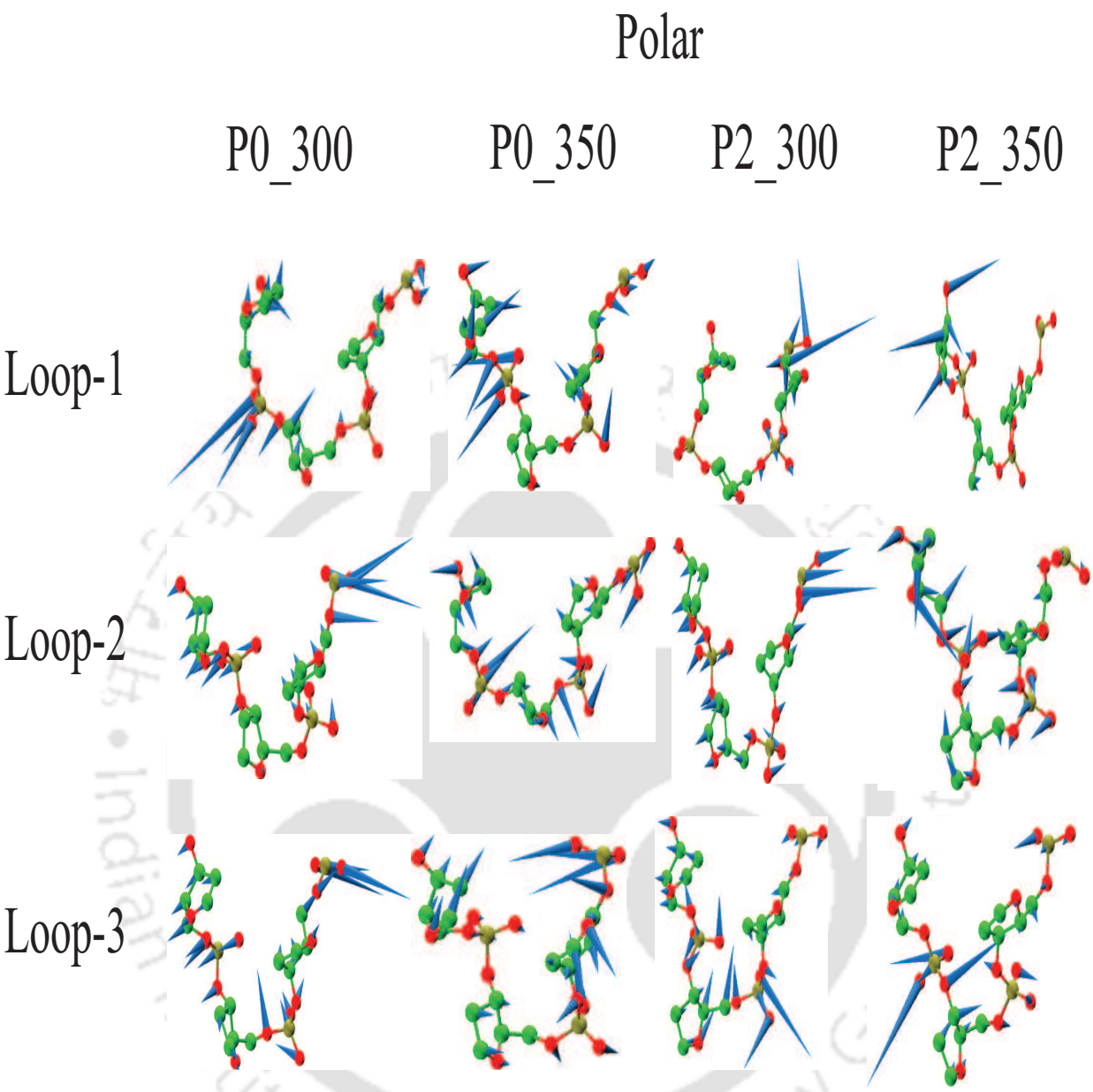


Figure 3.9: *Representing the dominant motions of human telomeric G-quadruplex DNA along with the first eigenvector in porcupine plot for different systems under polar confinement. The models are shown as a backbone trace. The arrows attached to each backbone atom indicate the direction of the eigenvector and the size of each arrow shows the magnitude of the corresponding eigenvalue.*

individual loop regions along with the first eigenvector at different salt concentrations and temperatures. In case of non-polar confinement, the major dominant mode of the backbone atoms of all the three loop regions are noticed at minimal salt concentration as compared to higher salt concentration. Moreover, at minimal salt concentration, for N0_300 and N0_350 systems, mainly major dominant mode for all three loop regions are noticed for

N0_350 systems. Among all the three propeller loop regions, loop-1 and loop-3 show major dominant motions. Similar scenario is also noticed under polar confinement conditions. On further examination, we have plotted the two principal components, which are PC1 and PC2 with regard to free energy change (ΔG_i) following the equation

$$\Delta G_i = -k_B T \left[\ln P_i - \ln P_{\max}\right] \tag{3.1}$$

where k_B represents the Boltzmann constant, T is the temperature of the system, P_i denotes the probability distribution of the principal component which is subtracted from its

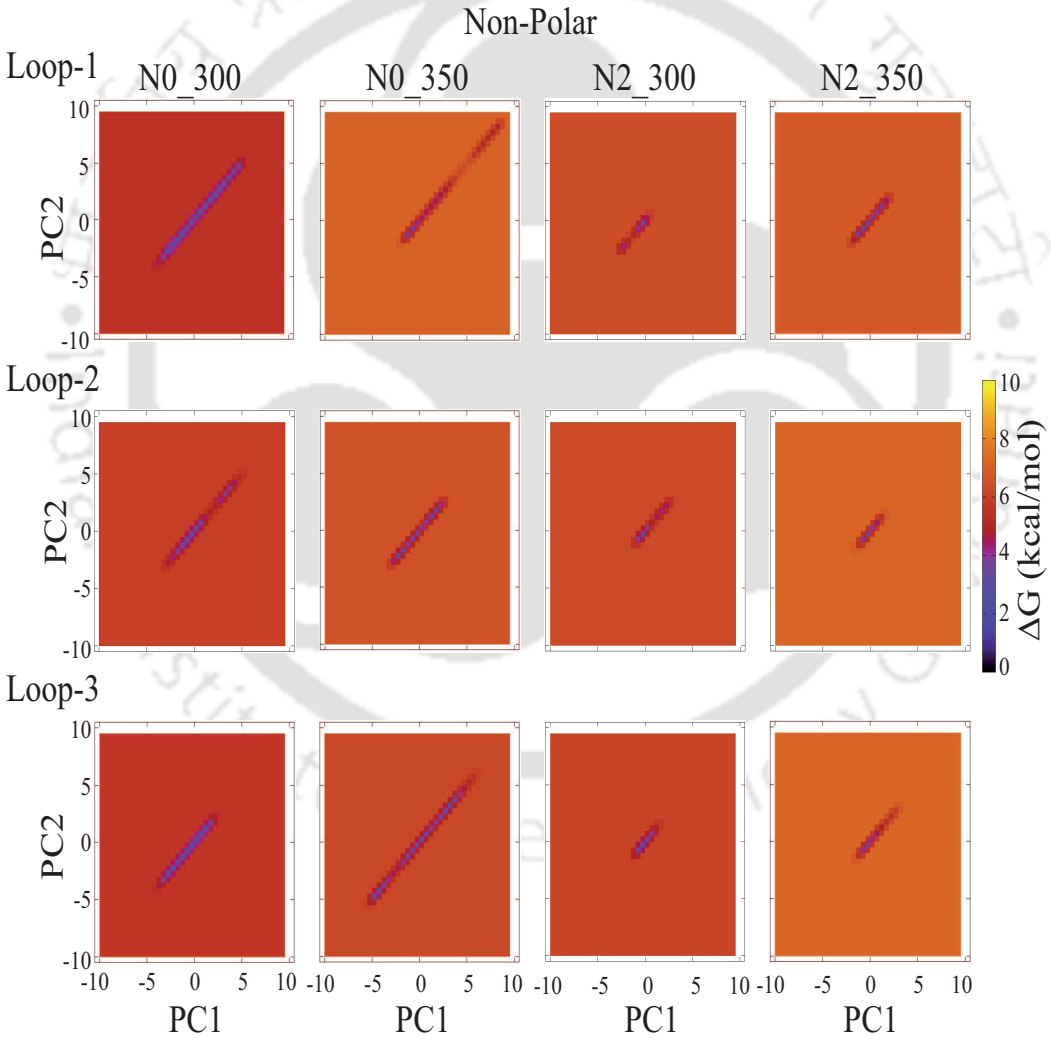


Figure 3.10: *Free energy landscapes (FELs) plotted as a function of principal components of PC1 and PC2 for the backbone atoms of loop-1, loop-2, and loop-3 of human telomeric G-quadruplex DNA for different systems under non-polar confinement.*

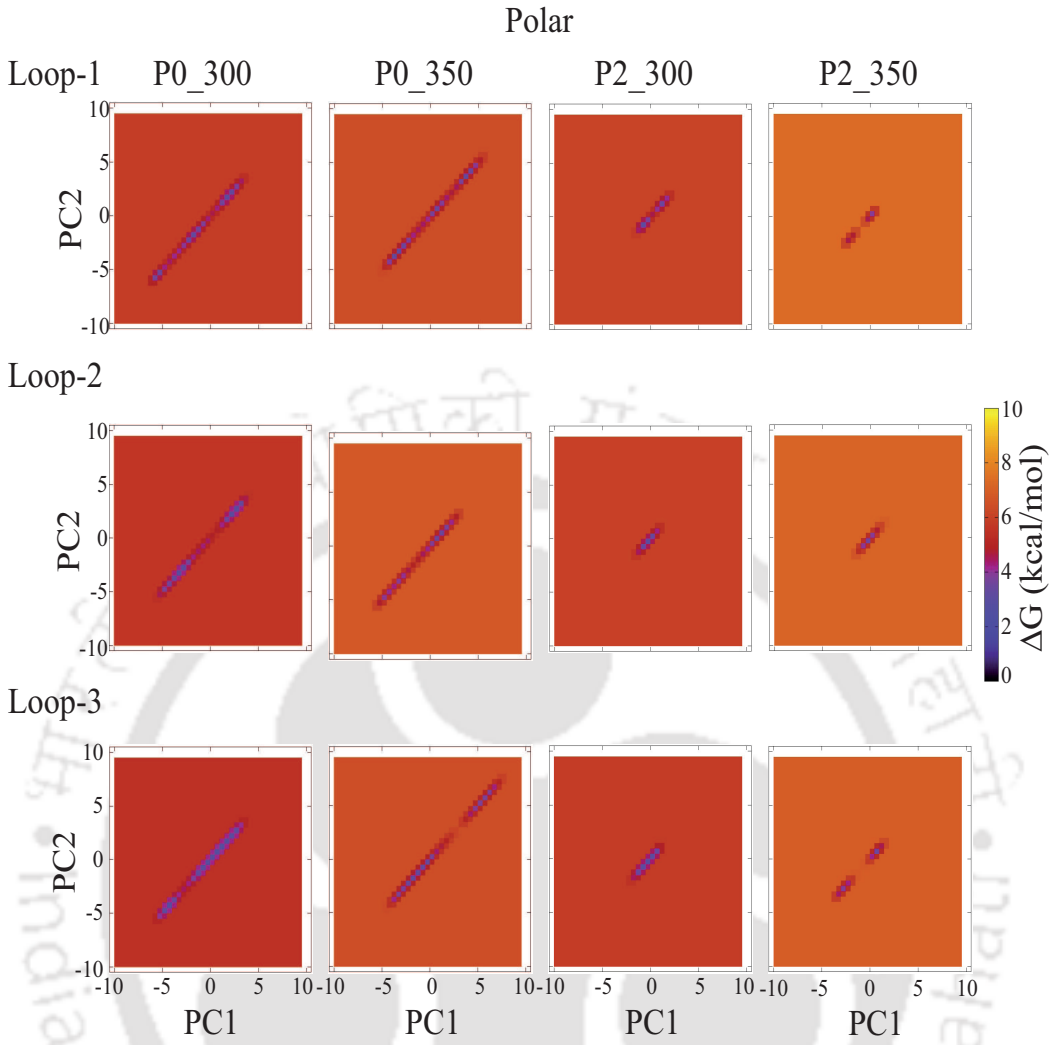


Figure 3.11: *Free energy landscapes (FELs) plotted as a function of principal components of PC1 and PC2 for the backbone atoms of loop-1, loop-2, and loop-3 of human telomeric G-quadruplex DNA for different systems under polar confinement.*

maximum value i.e. P_{max}. In **Figures 3.10 and 3.11**, we have presented the cartesian projection of the PC1 and PC2 in terms of free energy landscapes for all systems. From the **Figures 3.10 and 3.11**, it is clear that, at higher salt concentration the free energy basin/well emerges where both PC1 and PC2 of the individual loop regions are in the range of 2.5 to -2.5. On the other hand, at minimal salt concentration, the free energy basin/well is observed at higher PC1 and PC2 values. Similar results are also noticed for both polar and non-polar confinement conditions. Moreover, at 350 K temperature, the area of free energy basin/well is increased for loop-1 and loop-3 for N0_350 and P0_350 systems. This indicates that the above mentioned loop regions deviate far more from the crystal structure. However, at higher salt concentration all the three loops maintain its native state.

3.3.3.2 π-π Stacking of Different Propeller Loop Regions

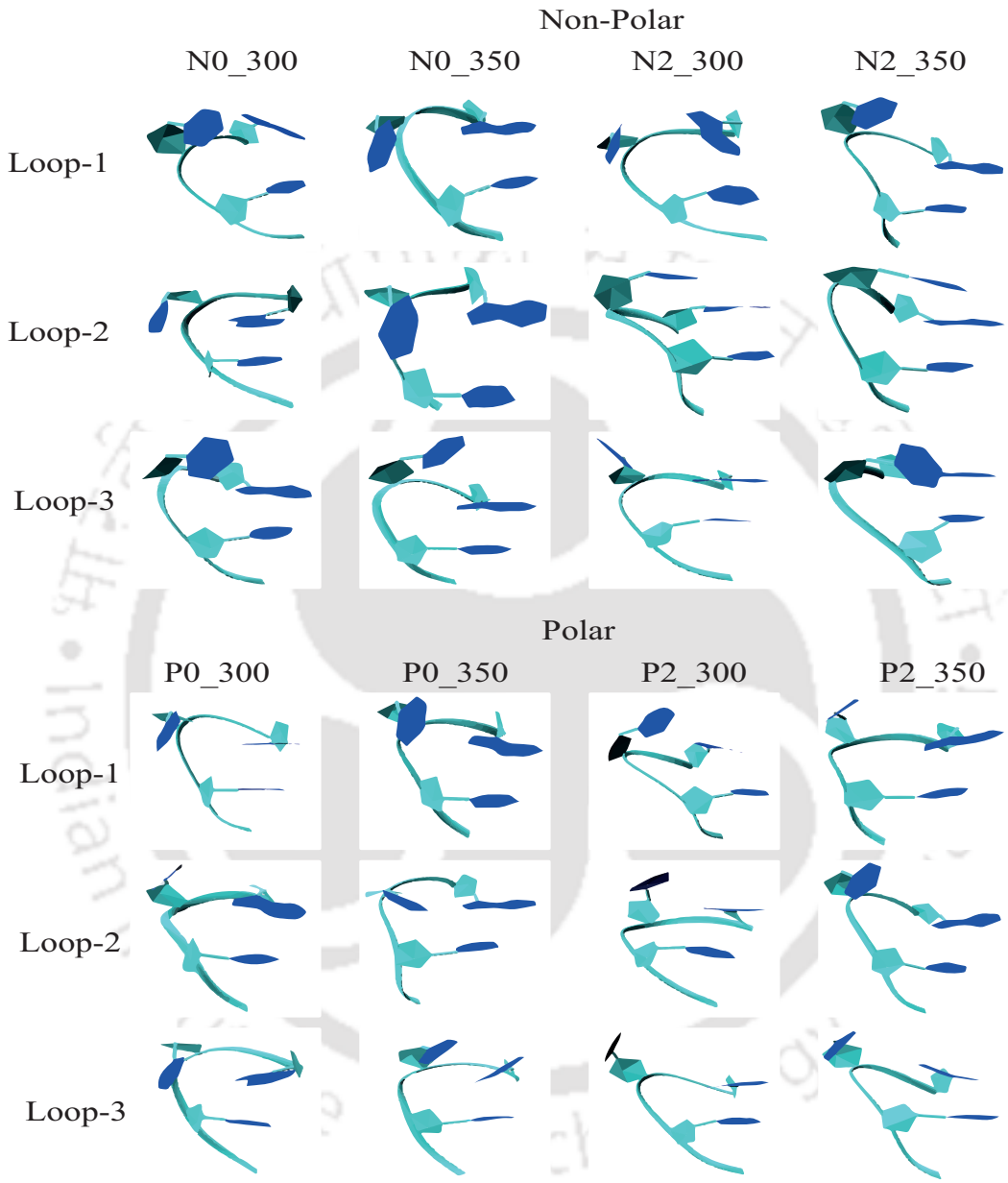


Figure 3.12: *Representative snapshots of the most populated structure of loop-1, loop-2 and loop-3 of human telomeric G-quadruplex DNA for different systems.*

In crystal structure of the parallel telomeric quadruplex DNA, T5-A7, T11-A13, and T17-A19 residues are involved in π-π stacking interactions in all three propeller loop regions. Now, to study the loop dynamics of the parallel telomeric quadruplex DNA, it is important to check whether these π-π stacking formation are maintained. Thus, at first we have performed widely [53, 62, 63] used the cluster structure analysis by Ester et al., Density

Based Spatial Clustering of Applications with Noise (DBSCAN) [64] clustering algorithm. We have considered all the heavy atoms of the parallel telomeric quadruplex DNA through out the 5μs trajectory for cluster structure analysis for all systems. Moreover, to perform the DBSCAN method, we have taken the minimum number of points (minpoints) is 4 and the distance cut off between these points (ϵ) is 0.35 Å. In **Fig. 3.12**, we have shown the most populated structure of loop-1, loop-2, and loop-3 regions individually. From **Fig. 3.12**, it looks apparently that all the above mentioned residues maintain its π-π stacking formation. Further, we need to check carefully whether the π-π stacking criteria are maintained. Following earlier studies [65–68], two criteria are set: (i) the cut off distance between two center of mass (COM) of the considered nucleobase ring is 5 Å and simultaneously, the angle cut off between two vector normals of the considered nucleobase

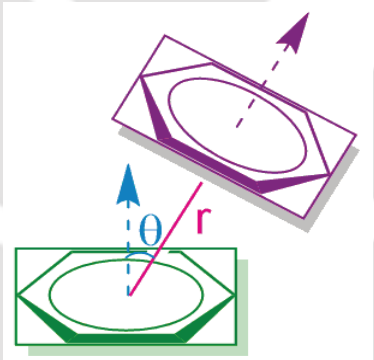


Figure 3.13: *Schematic representation of the orientational angles (θ) and COM distances (r) of the different aromatic planes of the considered nucleobase of human telomeric G-quadruplex DNA.*

ring is 30° (**Fig. 3.13**). In **Figures 3.14 and 3.15**, we have plotted the probability of the distance and angle between the considered nucleobase ring for all systems. Here, we have noticed that the distance criterion is maintained (all the distances are below 5 Å) by T5-A7, T11-A13, and T17-A19 residues of loop-1, loop-2, and loop-3 regions respectively, in both the non-polar and polar confinement conditions. However, taking into account of the highest angle probability, the scenario is different for non-polar and polar confinements. In case of non-polar confinement conditions, the highest probability of the angle for the T5-A7, T11-A13, and T17-A19 residues of loop-1, loop-2, and loop-3 regions are below 30°, excluding N0_300 system where the T5-A7, and T17-A19 residues of loop-1, and loop-3

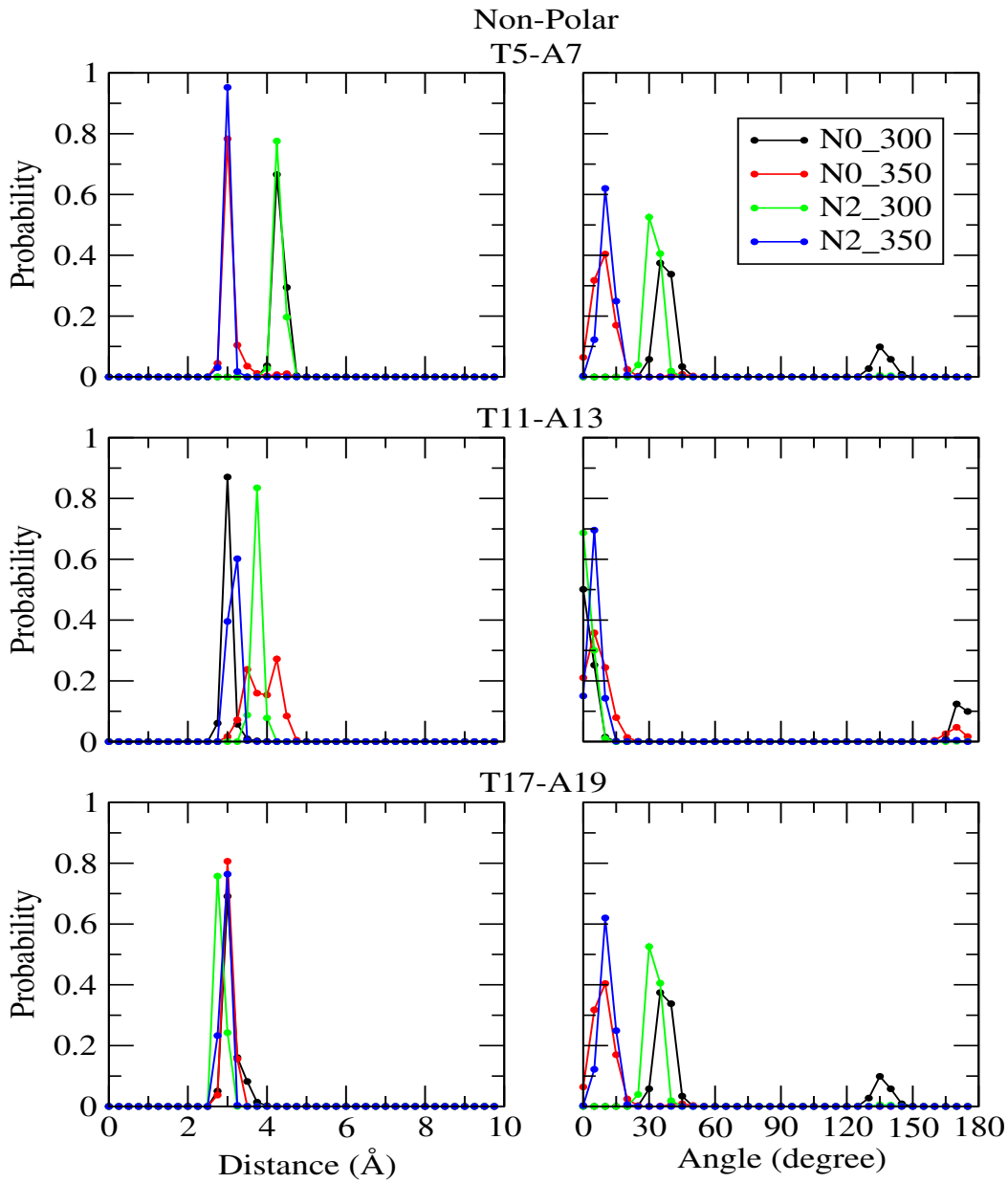


Figure 3.14: *Stacking probability of loop-1, loop-2 and loop-3 with respect to distance and angle between the corresponding planes under non-polar confinement.*

regions, exhibit the highest angle probability at 35°. On the other hand, under polar confinement conditions, the highest probability of the angle appears below 30° for all the systems, except for the P0_350 system. In P0_350 system, although the distance criterion is satisfied among all the three systems (namely P0_300, P2_300, and P2_350), in case of angle, the highest probability emerges at 40° for T5-A7, and T17-A19 residues of loop-1, and loop-3 regions.

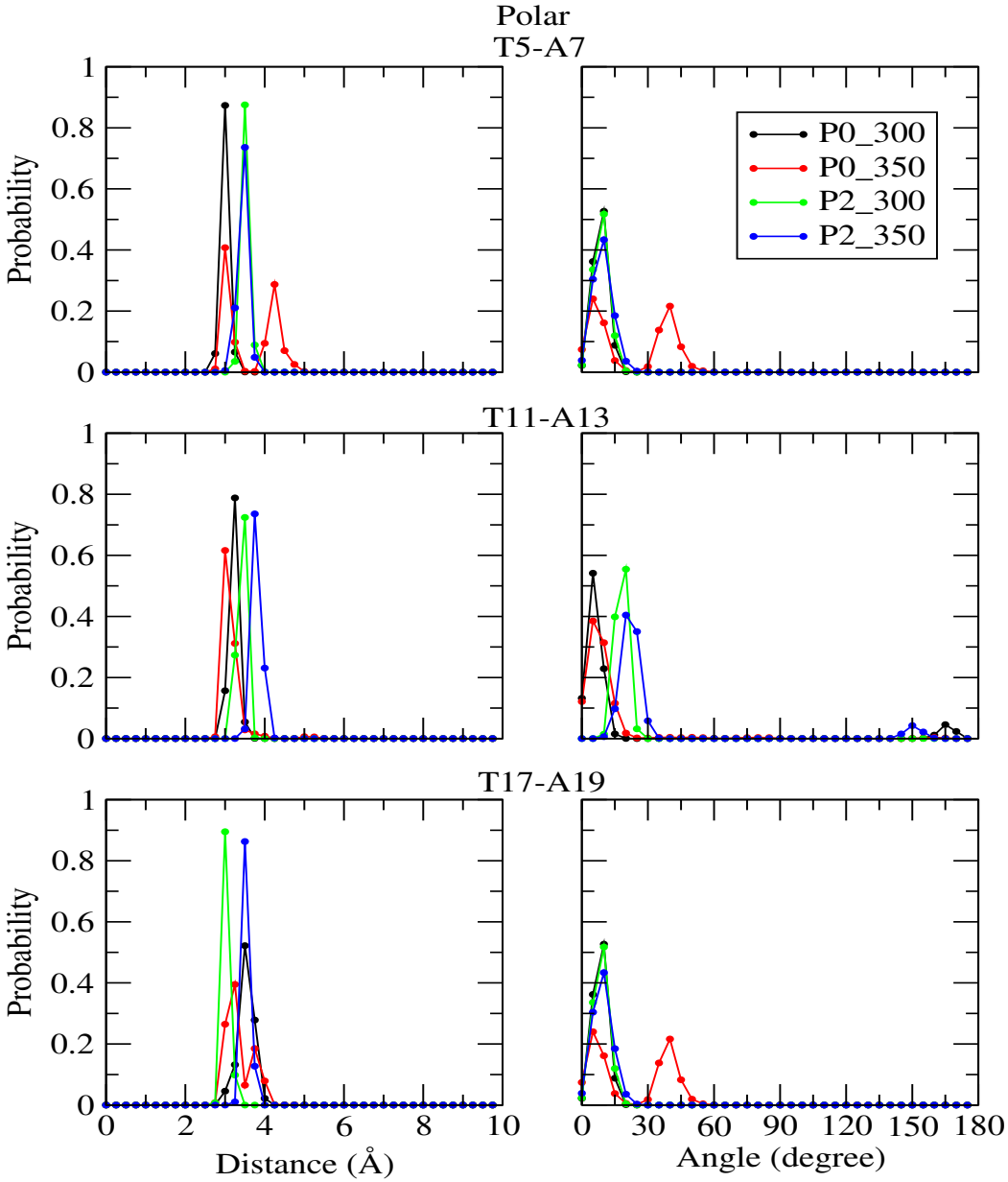


Figure 3.15: *Stacking probability of loop-1, loop-2 and loop-3 with respect to distance and angle between the corresponding planes under non-polar confinement.*

3.3.4 Role of the Cations on Dynamics of Parallel Telomeric Quadruplex DNA

Earlier studies [69–72] report that the cation plays a crucial role in the stabilization and formation of the G-quadruplex DNA. Thus, in this section we have studied the role of cations using the following parameters: (i) spatial distribution functions (SDFs), and (ii) first solvation shell.

3.3.4.1 Spatial Distribution Functions (SDFs)

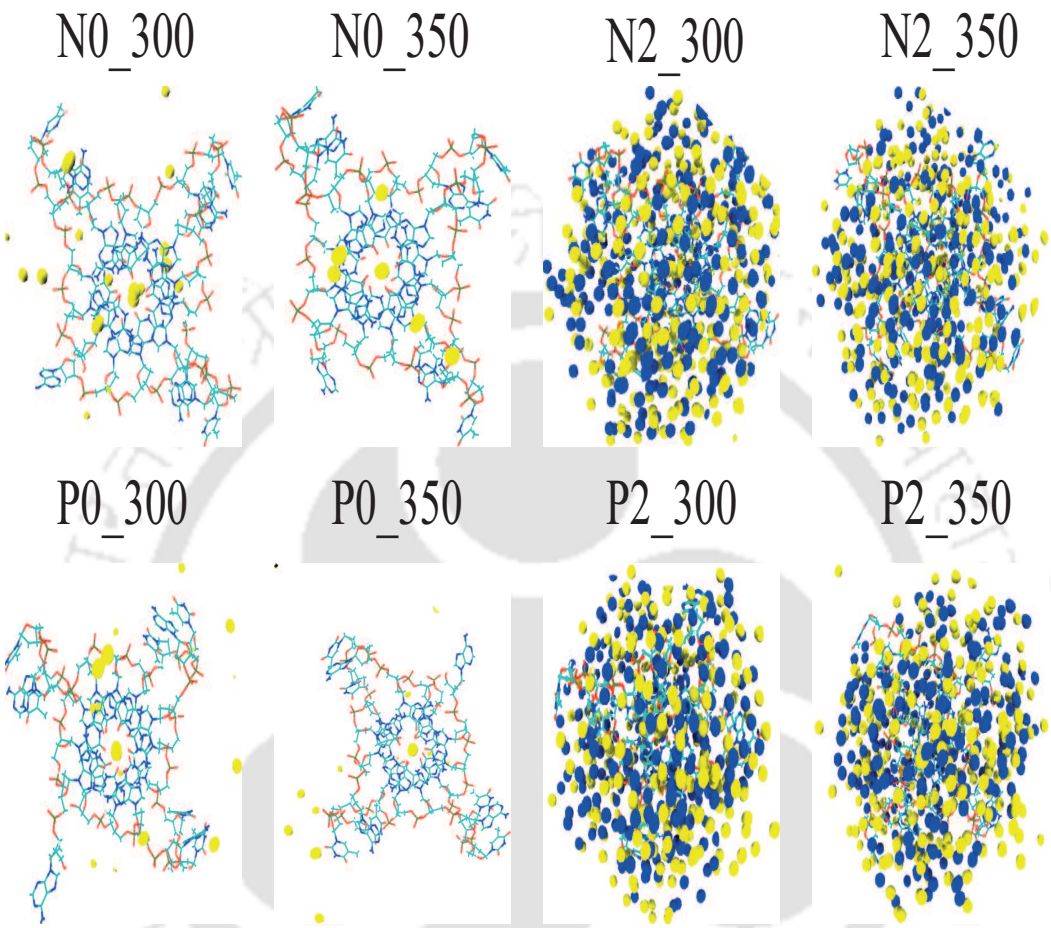


Figure 3.16: *Spatial density maps of K^+ and Cl^- around human telomeric G-quadruplex DNA for different systems. Here, yellow and blue represent K^+ and Cl^- ions respectively.*

Spatial distribution functions (SDFs) are a very useful method for visualizing the relationship between the ions and the quadruplex. To get more information about the role of cations on the parallel telomeric quadruplex DNA SDF is performed. Spatial distribution functions (SDFs) reveal the three dimensional probabilistic arrangement of molecules in the system. In **Fig. 3.16**, we have shown the distribution functions of the K^+ cations around the parallel telomeric quadruplex DNA using isovalue equals to 75 for all the systems. Here, high density of K^+ cations is observed in mainly two regions: (i) around the central core of G-quadruplex channel core and (ii) around the sugar-phosphate backbone regions for all the systems. Previously we have noticed that the K^+ cation does not take part in the ion-exchange with the outside ions and they are almost fixed at their position through out

the 5μs simulation run. Hence, the SDFs are high in the central core of the ion channel. In DNAs, sugar-phosphate backbone carries negative charge. So, the positively charged K^+ cations interact more with the sugar-phosphate backbone than the other regions. This is the reason behind the high density of K^+ cations at sugar-phosphate backbone regions. Furthermore, among loop and tetrad regions of the parallel telomeric quadruplex DNA, the density around the tetrad regions is higher than the loop regions. From **Fig. 3.16**, it is noticed that the density around the parallel telomeric quadruplex DNA is the highest for high salt concentration than at minimal salt concentration under both the non-polar and polar confinement conditions. On the other hand, at high salt concentration, the density of Cl^- ions is also higher in sugar-phosphate backbone regions due to Cl^- ions take part as counter ions of K^+ ions. Thus, more K^+ ions can accommodate in the backbone regions. Moreover, as expected with increasing temperature the density decreases for all the systems.

3.3.4.2 First Solvation Shell

From the preceding SDF analysis, we have observed the high density salt concentration near the backbone regions. In order to get a clearer view, we have estimated the number of K^+, Cl^-, and water molecules in First Solvation Shell (FSS) of individual residues of the human telomeric G-quadruplex DNA. The number of ions or molecules present in FSS, is estimated by determining the number of different species present within 3.5 Å of all the the heavy atoms of DNA. Here, it is noticed that the highest number of K^+ ions present in FSS of G-quadruplex DNA is for N2_300 system, followed by N2_350, N0_300, and finally N0_350 systems under non-polar confinement (**Fig. 3.17**). Moreover, higher number of K^+ ions is observed mainly in G2, G3, G8, G14, G20 and G21 residues of tetrad regions of the human telomeric G-quadruplex DNA. In case of polar confinement, the higher number of K^+ ions are also noticed in the tetrad regions, which is very similar to the non-polar confinement. Moreover, here it is worth mentioning that higher number of K^+ ions are also noticed in T11, T12 and A13 residues of loop-2 regions.

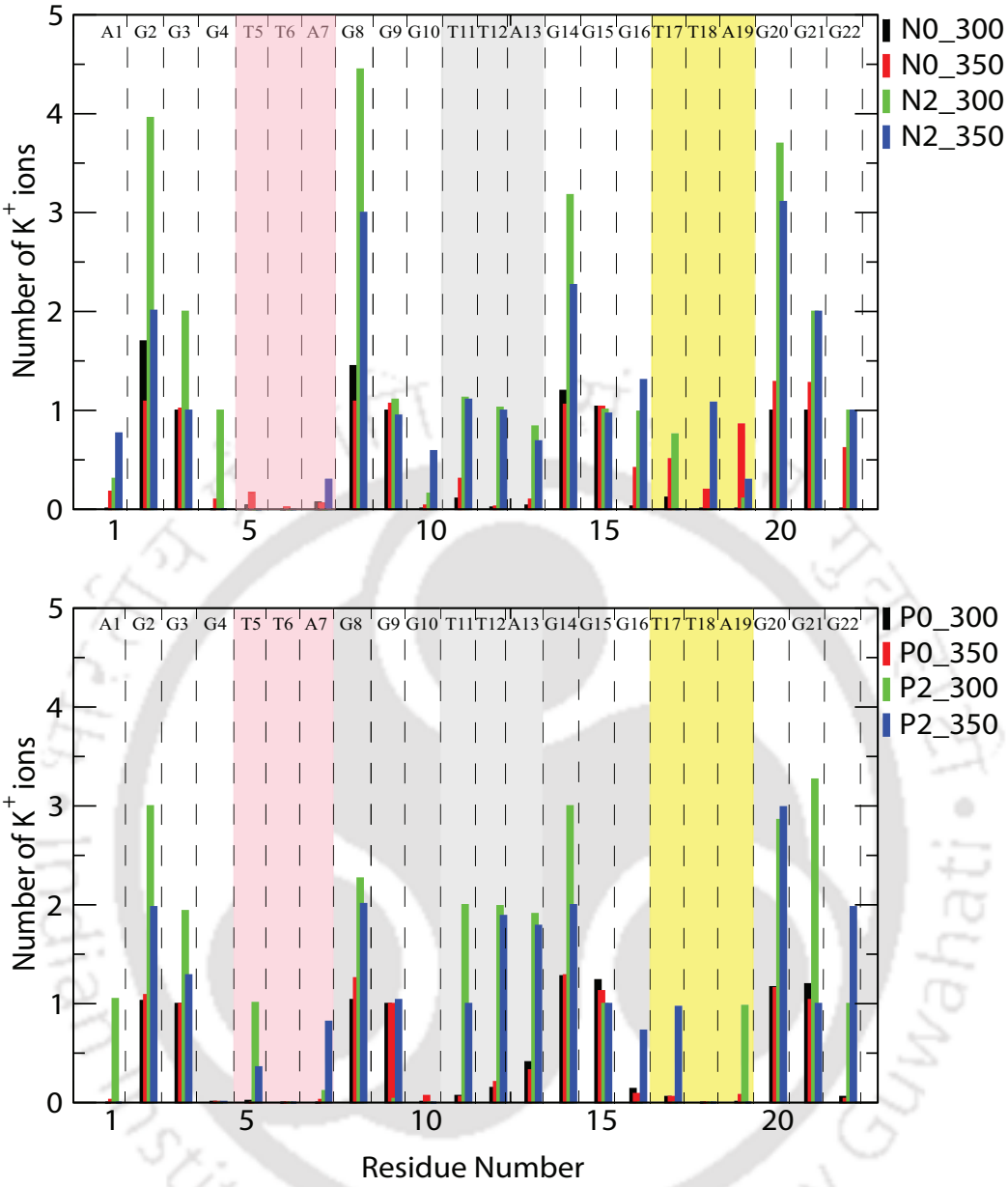


Figure 3.17: *Average number of K^+ ions present in the first solvation shell of each residue of the human telomeric G-quadruplex DNA for different systems. Loop-1, loop-2, and loop-3 regions are shown in pink, gray, and yellow, respectively.*

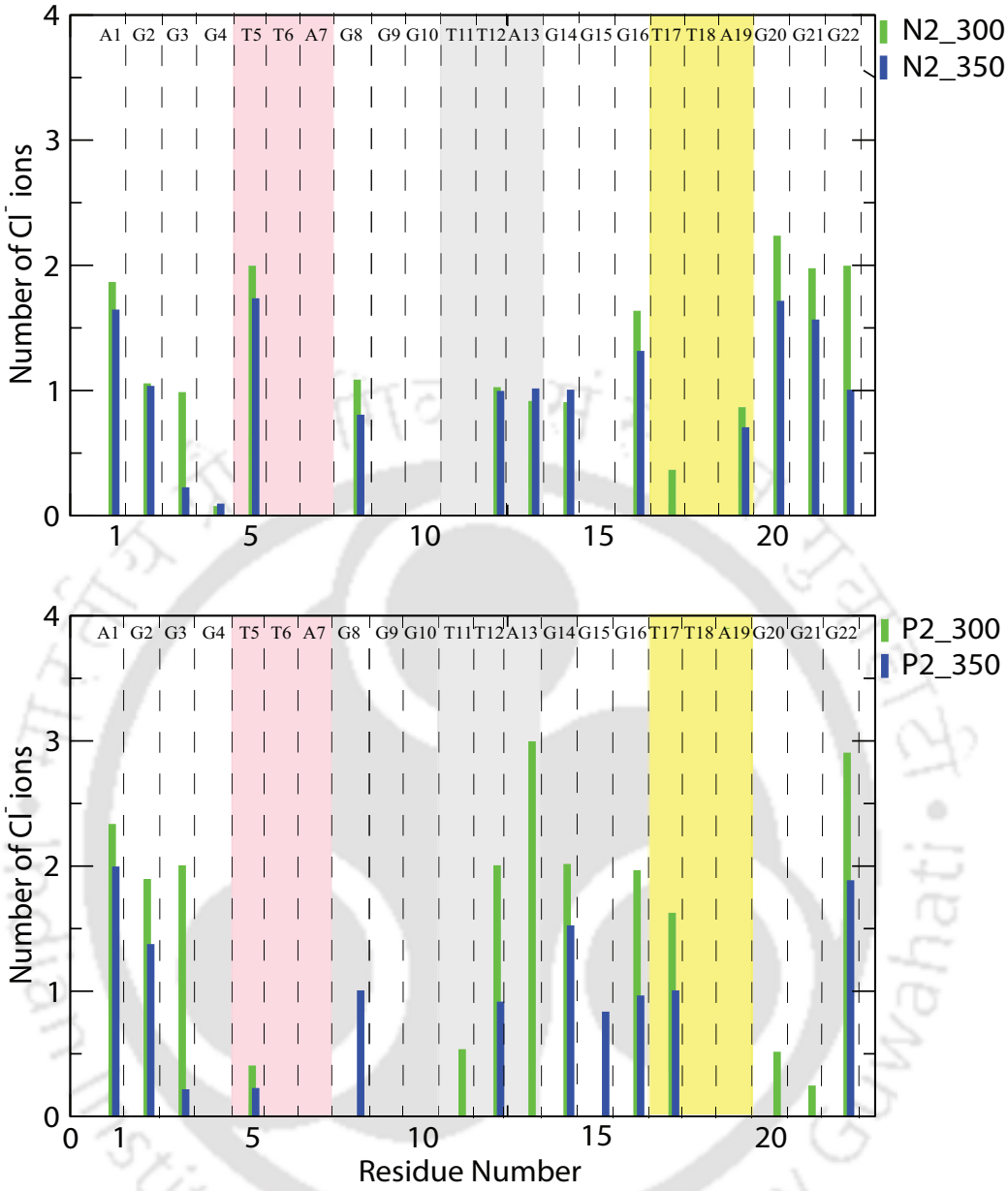


Figure 3.18: *Average number of Cl^- ions present in the first solvation shell of each residue of the human telomeric G-quadruplex DNA for different systems. Loop-1, loop-2, and loop-3 regions are shown in pink, gray, and yellow, respectively.*

The number of Cl^- ions present in FSS (**Fig. 3.18**) of the G-quadruplex DNA is calculated for the higher salt concentration system only. Here, we also notice that at higher temperature, the number of Cl^- ions in FSS decreases for both the non-polar and polar confinement. Besides, more Cl^- ions are present at the tetrad regions of G-quadruplex DNA, except for G4, G9, G10 and G15 residues, under non-polar confinement. Very similar scenario is noticed under polar conformation. Here, the number of Cl^- ions in FSS is higher at G2, G3, G14, G15, G16, and G22 residues of the tetrad regions, and also at T12 and

A13 of loop-2 and T17 of loop-3 regions of human parallel G-quadruplex DNA. **Fig. 3.19**, represents the number of water molecules present in the FSS of the human parallel G-quadruplex DNA. Here, slightly higher number of water molecules is present in the loop regions as compared to the tetrad regions under both the non-polar and polar confinement.

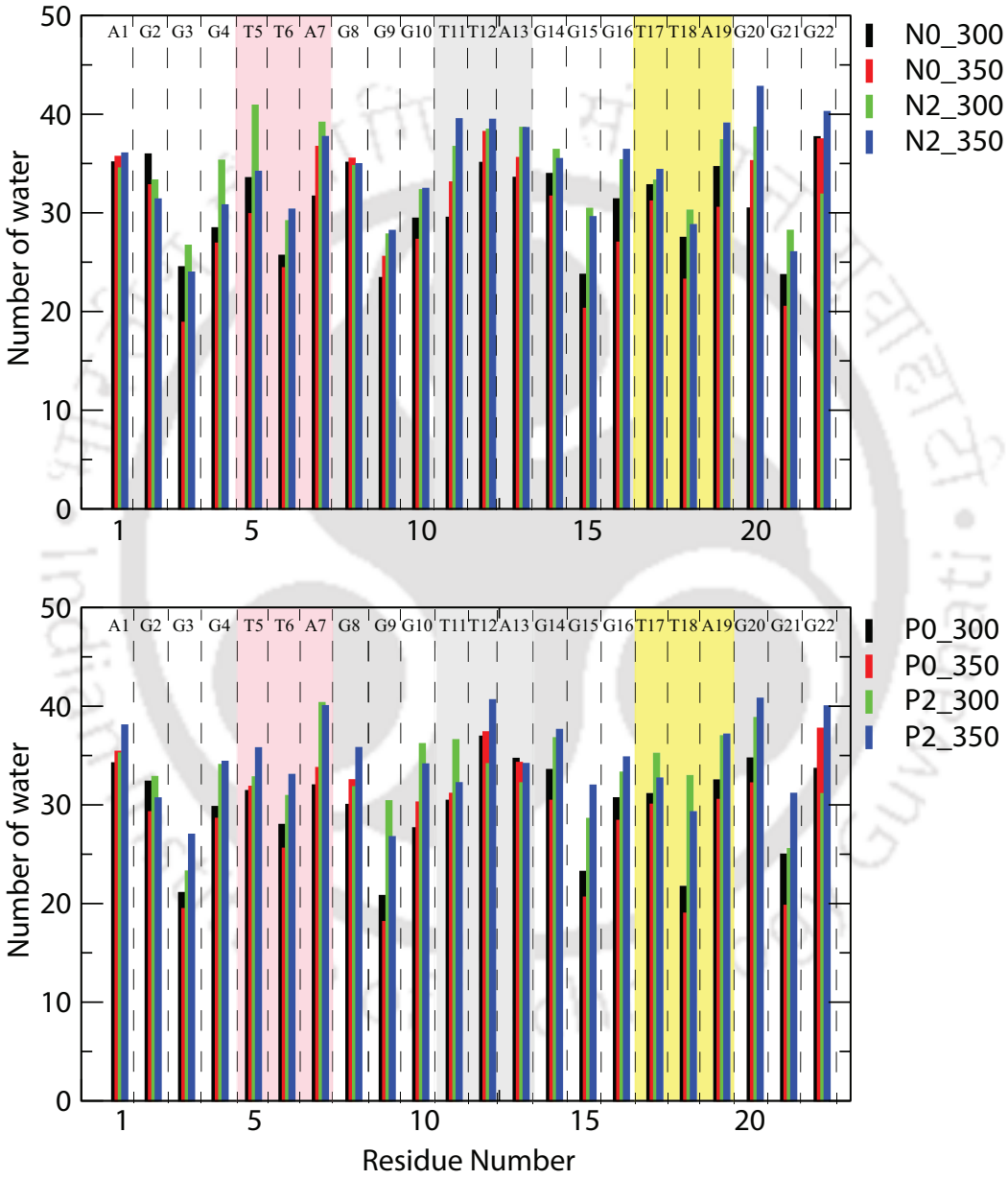


Figure 3.19: *Average number of water molecules present in the first solvation shell of each residue of the human telomeric G-quadruplex DNA for different systems. Loop-1, loop-2, and loop-3 regions are shown in pink, gray, and yellow, respectively.*

3.4 Summary and Conclusions

In this study, we have presented the conformational deviations of the human telomeric G-quadruplex DNA in different salt concentrations and temperatures under non-polar and polar confinement conditions, using 5 μs long run simulation for each system. From RMSD and 2D-RMSD analyses, we have noticed that the human telomeric G-quadruplex DNA deviates more from its crystal structure at minimal salt concentration than at the higher salt concentration under both the non-polar and polar confinement conditions. Moreover, among the sugar-phosphate backbone, tetrad and loop regions, more deviation from crystal structure is noticed mainly at the loop regions. Very similar outcome is procured from RMSF analysis. Moreover, the higher salt concentration provides more rigidity to the human telomeric G-quadruplex DNA. To understand the tetrad dynamics of the G-quadruplex DNA, we have estimated the number of hydrogen bonds of the tetrad and the distance between the central cations placed in the channel core of the human telomeric G-quadruplex DNA, which determines the occurrence of any ion exchange throughout the simulation. From hydrogen bond analysis, it is observed that the tetrads maintain the number of hydrogen bonds for all the systems. However, the number of hydrogen bond slightly decreases at higher salt concentration. Moreover, the distance between the two K^+ ions lies between 3.6 Å to 4.6 Å which depicts that no ion exchange occurs during the 5 μs simulation run for all the systems. Since the loop region fluctuates the most, it is very important to monitor the loop dynamics of the human telomeric G-quadruplex DNA. For this, we have performed the PCA analysis and also inspected the π-π stacking conformation of the loop regions individually. From PCA analysis, we have observed that the loop-1 and loop-3 show major dominant motions in non-polar and polar confinement conditions. Moreover, the higher dominant modes are observed at minimal salt concentration. With increasing temperature, the dominant modes of the loop regions increase. From the estimation of the distance and angle between the considered nucleobase, we have observed that the distance criterion of π-π stacking is maintained by all the considered nucleobase. However, higher angle probability is observed for T5-A7, and T17-A19 of loop-1 and loop-3 regions for minimal salt concentration systems under both the non-polar and polar confinement conditions. Next, we have performed SDF analysis to get a clear view about the role of ions on the human telomeric G-quadruplex DNA, where we observed that the density of K^+ ions is higher mainly in two regions: (i) the G-quadruplex channel core and (ii) the sugar-phosphate backbone regions. Here, we notice that, at high salt concentration, the

density is higher as compared to the minimal salt concentration. Moreover, it discloses the reason behind the rigidity of the backbone of the human telomeric G-quadruplex DNA. To get more information about the role of ions in conformational change of the human telomeric G-quadruplex DNA, we have estimated the number of K^+, Cl^-, and water molecules in the first solvation shell of the G-quadruplex DNA around its different residues. We have observed that higher number of K^+ ions mainly in G2, G3, G8, G14, G20 and G21 residues of tetrad regions of the human telomeric G-quadruplex DNA under both the non-polar and polar confinement conditions. It reveals that the K^+ ions are mainly responsible for the rigidity of the human telomeric G-quadruplex DNA. We believe that our study provides useful information about the conformational change of the human telomeric G-quadruplex DNA inside the non-polar as well as polar close confinement conditions.

References

[1] S. Rankin, A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian and S. Neidle, *J. Am. Chem. Soc.*, 2005, **127**, 10584–10589.

[2] T. Simonsson, M. Kubista and P. Pecinka, *Nucleic Acids Res.*, 1998, **26**, 1167–1172.

[3] H. Sun, A. Yabuki and N. Maizels, *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, 12444–12449.

[4] H. Sun, J. K. Karow, I. D. Hickson and N. Maizels, *J. Biol. Chem.*, 1998, **273**, 27587–27592.

[5] A. M. Olovnikov, *Exp. Gerontol.*, 1996, **31**, 443 – 448.

[6] A. G. Bodnar, M. Ouellette, M. Frolkis, S. E. Holt, C.-P. Chiu, G. B. Morin, C. B. Harley, J. W. Shay, S. Lichtsteiner and W. E. Wright, *Science*, 1998, **279**, 349–352.

[7] N. Kim, M. Piatyszek, K. Prowse, C. Harley, M. West, P. Ho, G. Coviello, W. Wright, S. Weinrich and J. Shay, *Science*, 1994, **266**, 2011–2015.

[8] E. H. Blackburn, E. S. Epel and J. Lin, *Science*, 2015, **350**, 1193–1198.

[9] J. W. Shim and L.-Q. Gu, *J. Phys. Chem. B*, 2008, **112**, 8354–8360.

[10] J. W. Shim, Q. Tan and L.-Q. Gu, *Nucleic Acids Res.*, 2008, **37**, 972–982.

[11] N. An, A. M. Fleming and C. J. Burrows, *J. Am. Chem. Soc.*, 2013, **135**, 8562–8570.

[12] N. An, A. M. Fleming, E. G. Middleton and C. J. Burrows, *Proc. Natl. Acad. Sci. U.S.A.*, 2014, **111**, 14325–14331.

[13] C. Ma, R. C.-T. Chan, C. T.-L. Chan, A. K.-W. Wong and W.-M. Kwok, *J. Phys. Chem. Lett.*, 2019, **10**, 7577–7585.

[14] S. Peng, B. Bie, H. Jia, H. Tang, X. Zhang, Y. Sun, Q. Wei, F. Wu, Y. Yuan, H. Deng and X. Zhou, *J. Am. Chem. Soc.*, 2020, **142**, 5049–5059.

[15] G. Borgohain and S. Paul, *J. Mol. Liq.*, 2017, **231**, 174 – 184.

[16] G. Borgohain and S. Paul, *J. Mol. Liq.*, 2017, **233**, 431 – 441.

[17] G. Brinkmann, O. D. Friedrichs, S. Lisken, A. Peeters and N. Van Cleemput, *MATCH Commun. Math. Comput. Chem*, 2010, **63**, 533–552.

[18] D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER 2018*, University of California, San Francisco, 2018.

[19] G. N. Parkinson, M. P. H. Lee and S. Neidle, *Nature*, 2002, **417**, 876–880.

[20] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

[21] L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, *J. Comput. Chem.*, 2009, **30**, 2157–2164.

[22] B. Islam, P. Stadlbauer, M. Krepl, M. Havrila, S. Haider and J. Sponer, *J. Chem. Theory Comput.*, 2018, **14**, 5011–5026.

[23] J. Šponer, G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R. A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka, N. G. Walter and M. Otyepka, *Chem. Rev.*, 2018, **118**, 4177–4338.

[24] B. Machireddy, G. Kalra, S. Jonnalagadda, K. Ramanujachary and C. Wu, *J. Chem. Inf. Model.*, 2017, **57**, 2846–2864.

[25] N. G. J. Richards and M. M. Georgiadis, *Acc. Chem. Res.*, 2017, **50**, 1375–1382.

[26] B. Islam, P. Stadlbauer, A. Gil-Ley, G. Pérez-Hernández, S. Haider, S. Neidle, G. Bussi, P. Banas, M. Otyepka and J. Sponer, *J. Chem. Theory Comput.*, 2017, **13**, 2458–2480.

[27] A. K. Sieradzan, P. Krupa and D. J. Wales, *J. Phys. Chem. B*, 2017, **121**, 2207–2219.

[28] P. Stadlbauer, L. Mazzanti, T. Cragnolini, D. J. Wales, P. Derreumaux, S. Pasquali and J. Šponer, *J. Chem. Theory Comput.*, 2016, **12**, 6077–6097.

[29] D. R. Gruber, J. J. Toner, H. L. Miears, A. V. Shernyukov, A. S. Kiryutin, A. A. Lomzov, A. V. Endutkin, I. R. Grin, D. V. Petrova, M. S. Kupryushkin, A. V. Yurkovskaya, E. C. Johnson, M. Okon, E. G. Bagryanskaya, D. O. Zharkov and S. L. Smirnov, *Nucleic Acids Res.*, 2018, **46**, 10827–10839.

[30] S. Haider, *J. Indian Inst. Sci*, 2018, **98**, 325–339.

[31] M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo and P. Jurečka, *J. Chem. Theory Comput.*, 2015, **11**, 5723–5736.

[32] R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka and T. E. Cheatham, *J. Chem. Theory Comput.*, 2016, **12**, 4114–4127.

[33] A. Prez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton and M. Orozco, *Biophysical J.*, 2007, **92**, 3817 – 3829.

[34] I. S. Joung and T. E. Cheatham, *J. Phys. Chem. B*, 2008, **112**, 9020–9041.

[35] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

[36] A. Waghe, J. C. Rasaiah and G. Hummer, *J. Chem. Phys*, 2002, **117**, 10789–10795.

[37] S. Pal and S. Paul, *Int. J. Biol. Macromol.*, 2019, **121**, 350 – 363.

[38] S. Pal and S. Paul, *J. Phys. Chem. C*, 2019, **123**, 11686–11698.

[39] S. Pal and S. Paul, *J. Phys. Chem. B*, 2020, **124**, 3123–3136.

[40] R. W. Pastor, B. R. Brooks and A. Szabo, *Mol. Phys.*, 1988, **65**, 1409–1419.

[41] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys*, 1984, **81**, 3684–3690.

[42] B. P. Uberuaga, M. Anghel and A. F. Voter, *J. Chem. Phys.*, 2004, **120**, 6363–6374.

[43] D. J. Sindhikara, S. Kim, A. F. Voter and A. E. Roitberg, *J. Chem. Theory Comput.*, 2009, **5**, 1624–1631.

[44] J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

[45] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

[46] D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.

[47] W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.

[48] M. Brehm and B. Kirchner, *J. Chem. Inf. Model.*, 2011, **51**, 2007–2023.

[49] S. Naskar, H. Joshi, B. Chakraborty, N. C. Seeman and P. K. Maiti, *Nanoscale*, 2019, **11**, 14863–14878.

[50] M. Havrila, P. Stadlbauer, B. Islam, M. Otyepka and J. Šponer, *J. Chem. Theory Comput.*, 2017, **13**, 3911–3926.

[51] S. Balasubramanian, S. Pal and B. Bagchi, *Phys. Rev. Lett.*, 2002, **89**, 115505.

[52] A. Chandra, *Phys. Rev. Lett.*, 2000, **85**, 768.

[53] S. Pal and S. Paul, *J. Phys. Chem. B*, 2020, **124**, 210–223.

[54] A. Chandra, *J. Phys. Chem. B*, 2003, **107**, 3899–3906.

[55] S. Paul and A. Chandra, *J. Chem. Theory Comput.*, 2005, **1**, 1221–1231.

[56] S. Pal, R. Roy and S. Paul, *J. Phys. Chem. B*, 2020, **124**, 7598–7610.

[57] M. Rebič, A. Laaksonen, J. Šponer, J. Uličný and F. Mocci, *J. Phys. Chem. B*, 2016, **120**, 7380–7391.

[58] S. Haider, G. N. Parkinson and S. Neidle, *Biophysical J.*, 2008, **95**, 296 – 311.

[59] A. Wolf and K. N. Kirschner, *J. Mol. Model.*, 2013, **19**, 539–549.

[60] S. M. Haider and S. Neidle, in *Innovations in Biomolecular Modeling and Simulations: Volume 2*, The Royal Society of Chemistry, 2012, vol. 2, pp. 33–52.

[61] S. Naskar, M. Gosika, H. Joshi and P. K. Maiti, *J. Phys. Chem. C*, 2019, **123**, 9461–9470.

[62] M. Shapovalov, S. Vucetic and R. L. Dunbrack, Jr., *PLOS Comput. Biol.*, 2019, **15**, 1–32.

[63] F. Heinkel, L. Abraham, M. Ko, J. Chao, H. Bach, L. T. Hui, H. Li, M. Zhu, Y. M. Ling, J. C. Rogalski, J. Scurll, J. M. Bui, T. Mayor, M. R. Gold, K. C. Chou, Y. Av-Gay, L. P. McIntosh and J. Gsponer, *Proc. Natl. Acad. Sci. U.S.A.*, 2019, **116**, 16326–16331.

[64] M. Ester, H.-P. Kriegel, J. Sander and X. Xu, Proc. of 2nd International Conference on Knowledge Discovery and, 1996, pp. 226–231.

[65] J. Šponer, J. E. Šponer, A. Mlădek, P. Jurečka, P. Banáš and M. Otyepka, *Biopolymers*, 2013, **99**, 978–988.

[66] P. Banáš, A. Mlădek, M. Otyepka, M. Zgarbová, P. Jurečka, D. Svozil, F. Lankaš and J. Šponer, *J. Chem. Theory Comput.*, 2012, **8**, 2448–2460.

[67] J. Šponer, K. E. Riley and P. Hobza, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2595–2610.

[68] H. Kruse, P. Banáš and J. Šponer, *J. Chem. Theory Comput.*, 2019, **15**, 95–115.

[69] N. Špačková, I. Berger and J. Šponer, *J. Am. Chem. Soc.*, 1999, **121**, 5519–5534.

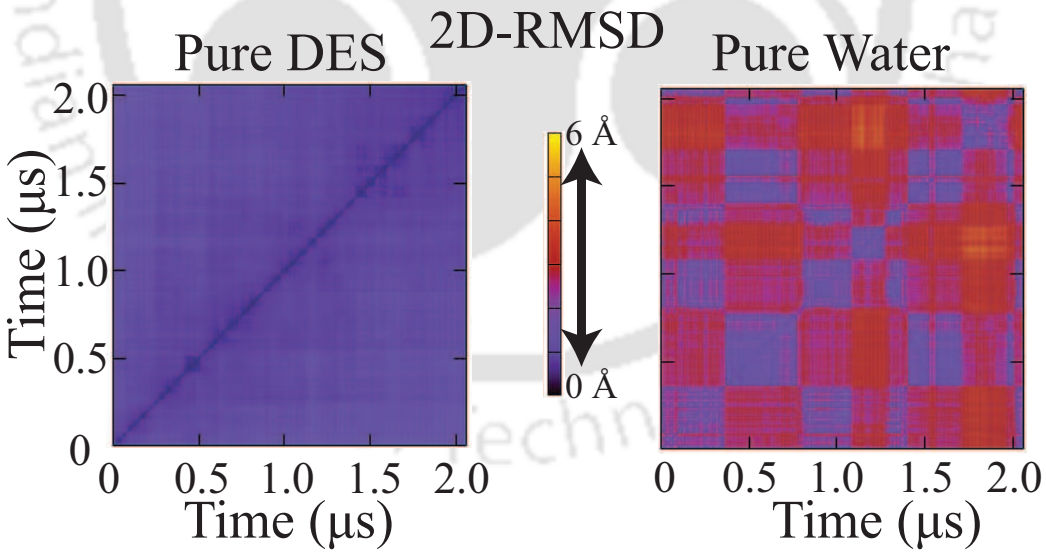
[70] S. Chowdhury and M. Bansal, *J. Phys. Chem. B*, 2001, **105**, 7572–7578.

[71] C. C. Hardin, T. Watson, M. Corregan and C. Bailey, *Biochemistry*, 1992, **31**, 833–841.

[72] F. M. Chen, *Biochemistry*, 1992, **31**, 3769–3776.

Chapter 4

Effect of Hydrated And Non-hydrated Choline Chloride-Urea Deep Eutectic Solvent (Reline) on Thrombin Binding G-quadruplex Aptamer (TBA)



Overview

Guanine rich quadruplex nucleic acid (G-DNA) sequence is highly polymorphic. The obtained structure of G-DNA is exquisitely impressible to its sequence and the chemical environment. Due to the controllable different structures G-DNA has acquired very much attention in various research areas such as nanotechnology, medicinal chemistry and molecular biology. However, the applications of G-DNA is mainly restricted to the aqueous media though, a large number of important chemical reactions, nanodevices etc. have also been carried out in purely water-free medium. Recently, Deep Eutectic Solvents (DESs) such as choline-urea (1:2) eutectic mixture, namely reline, has widely been used as a reaction media and also water-free storage media for biological system like different types of nucleic acids. Hence, it is very important to figure out the effect of the deep eutectic solvent with DNA. In this research work, we have discussed the interaction between reline with guanine rich quadruplex Thrombin Binding Aptamer (TBA) DNA at 300 K, for different reline concentrations. To understand the conformational behavior of quadruplex TBA in reline DES, we have performed total 10 μs all atom molecular dynamics simulations. Here, we notice that the structure of TBA deviates much more from its NMR structure at low reline concentrations. Noticeably, at high reline concentrations, the quadruplex TBA is much more rigid or less flexible than comparatively lower reline concentrations. Moreover, from the SDF study, the density of reline is higher near sugar-phosphate-backbone region than the others. Furthermore, at lower reline concentrations guanine-8 and thymine-9 of loop-2 stack onto each other which is not noticed at higher reline concentrations.

4.1 Introduction

Aptamers are comparatively very short sequences of DNA or RNA that have high potential to bind with specific targets such as thrombin [1, 2], nucleolin and STAT3 protein. Thrombin protein plays a very important role in the enzymatic blood coagulation cascade. Thrombin binding aptamer (TBA) with a sequence of 15-mer d(GGTTGGTGTGGTTGG) is the simplest G-quadruplex DNA and it binds with thrombin protein (**Fig. 4.1**). Thus, the important role of TBA in cell biology, attracts the scientific world in the field of molecular biology. The structure of TBA G-quadruplex DNA is already well established by NMR and X-ray crystallography studies [3–7]. These studies reveal that TBA contain two guanine tetrads (namely tetrad-1 and tetrad-2), two TT loops (i.e loop-1 and loop-2) and one TGT loop (loop-3) and it can fold into an anti-parallel chair like structure.

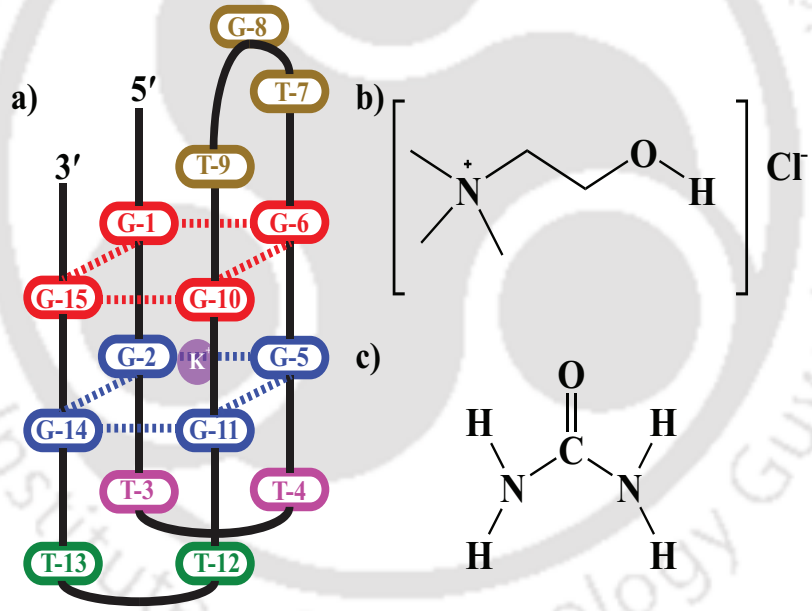


Figure 4.1: *Schematic representation of (a) TBA-G-quadruplex aptamer with K^+ ion (purple) (b) Choline chloride and (c) Urea molecules. Tetrad-1, tetrad-2, loop-1, loop-2, and loop-3 are shown in red, blue, magenta, brown, and green colors respectively. Guanine and Thymine are represented as G and T respectively.*

In the past decade, Ionic Liquids (ILs) analogues, popularly known as Deep Eutectic Solvents (DESs), have been used as a new class of solvents. These green solvents are obtained by the merging of two solid compounds (which have higher melting points than their mixture) at certain ratio. Mainly, the basic structure of DES is a combination of a molten salt and one or more hydrogen bond donor (HBD) [8]. Like ILs, DESs have unique

features, such as chemical, thermal and electro-chemical stability with a very low vapor pressures [9]. Due to these remarkable features, DESs have been used as a reaction media in various biological application fields like biocatalytic reactions [10], protein stabilization [11–13] and biosensors[14]. Furthermore, in small volume technologies, where small amount of water present vaporizes quickly, unlike ILs [15–20] the choice of DES as a solvent is a better option than water [21].

In recent years, reline, which is considered as the first generation of DESs, has widely been used in scientific and industrial applications. In 2003, Abbot et al. introduced [22] reline and its applications [23] to the scientific world. When choline chloride and urea (whose melting point is 302°C and 133°C respectively) are mixed at 1:2 molar ratios, the melting point of the corresponding eutectic mixture namely reline reduces to 12°C. The preparation of reline is very cheap as compared to the other ILs or DESs because of the naturally occurring and easily available precursors [24]. Due to its characteristics such as biodegradablity, non-toxicity [25, 26], biocompatiblity [27–29] and chemical inertness with water, reline is widely used as a biological storage medium [30].

In 2010, Hud et al. showed experimentally that in anhydrous choline-urea (reline) media, duplex, triplex and guanine rich TBA can exist [31]. Later, they demonstrated that the human telomere DNA can form propeller type of parallel guanine rich quadruplex structure in anhydrous reline DES medium [32] and double stranded DNA can fold into 2D origami structure in anhydrous and hydrated DES (glycholine) [33].

Zhao et al. showed that different types of guanine rich DNA's can form quadruplex structure and are stable in reline DES solvent [34, 35]. These studies indicated that the different type of nucleic acids are stable in DES media for longer periods than in water media. Hence the stability, formation and conformation of guanine rich quadruplex DNA in DES media, has received a lot of attention.

Since molecular dynamics (MD) simulations can provide useful information in molecular level to understand the behavior of guanine rich DNA in DES which, sometimes, cannot be achieved from conventional experiments, we have performed all atom molecular dynamics simulation. Until now, the conformational deviation and property of guanine rich DNA in DES have not been studied well from a theoretical approach. In this research work, we have studied the conformational behavior of guanine rich TBA in presence of anhydrous and as well as hydrated DES medium.

4.2 Models and Simulation Method

A series of all atom molecular dynamics was performed using AMBER18 [36] simulation package. In this study, the thrombin binding DNA aptamer d(GGTTGGTGTGGTTGG) (15-TBA) was taken in binary mixture of widely used non-hydrated and as well as hydrated Reline DES. The NMR structure of the thrombin binding DNA aptamer was obtained from the RCSB protein Data Bank (RCSB PDB) with PDB ID 148D (model-5) [5]. Initially a K^+ ion was placed in the core of the 15-TBA. The widely used latest AMBER DNA forcefield leaprc.nucleic.OL15 [37–48] was employed for 15-TBA. TIP3P [49] water model was employed in the simulations for hydrated systems. Since 15-TBA carries an electronic charge of -14e, 13 more K^+ ions were added to neutralize the system. The TIP3P specific Amber-adapted Joung and Cheatham force field parameter [50] was used for K^+ ions. The initial structure was built using PACKMOL [51]. For every system initially 15-TBA was placed in the center of the cubic box and then solvated with 250 choline chloride and 500 urea molecules. For hydrated reline systems, water molecules were adjusted to maintain the reline concentration. Then the topology and initial coordinates were converted according to truncated octahedron box using CPPTRAJ [52]. Since water plays a significant role in the stabilization of quadruplex DNA, the effect of water in hydrated DES was also studied. Therefore, the simulations were carried out by varying the ratio of DES to water. The number of water molecules added to maintain the reline (DES) concentration was estimated based on the percentage weight of DES over that weight of water (%w/w). The details of all systems are displayed in **Table 4.1**.

Table 4.1: *Number of molecules used in the simulation.*

System	Number Of Molecules				Reline conc. (wt %)
	Q-DNA	Urea	ChCL	Water	
S0	1	500	250	0	100
S1	1	500	250	36	99
S2	1	500	250	3605	50
S3	1	500	250	8411	30
W0	1	0	0	5000	0

Gaussian 09 [53] was used to optimize the geometries of the required ions and molecules. For the isolated choline cation and urea molecule, the respective optimized ge-

ometries were obtained at the HF/6-31G* [54, 55] level of theory and, using the restrained electrostatic potential (RESP) [56, 57] charge derivation method, partial charges were acquired. For these simulations, the general amber force field (GAFF) [58] was used for choline chloride and urea molecules [59–65]. In this work, we have used Perkins et al., force field parameters [62, 66] for choline chloride and urea molecules. Generally, electrostatic potentials are overestimated for ILs or DESs simulations. To control the electrostatic interactions a very well known method is often used, which involves the decrease of the charges of ionic components by a certain factor [62, 66]. This technique is also used for reline in simulation studies. Perkins et al. reported [61, 62] comparable results with experiments with a 20% reduction in charges with an AMBER forcefield [67, 68]. Therefore, the choline chloride charge was reduced by 20% .

For each system following steps were involved: (i) At first the system was energy minimized with 5000 steps via steepest descent followed by equal number of steps of conjugate gradients method. (ii) Then, each system was heated steadily from 0 to 300 K within 150 ps in canonical (NVT) ensemble using collision frequency 1.0 ps^{-1}. (iii) Next, the system was equilibrated for 250 ps at 300 K in NVT ensemble. For all the atoms of the quadruplex DNA (15-TBA) and the K^+ of the center core, 100 kcal mol^{-1} $Å^{-2}$ positional restraint was applied in the above three steps. (iv) Then a similar series of minimization and 250 ps equilibration (NVT) were performed in which the restraint force was reduced to 25, 20, 15, 10, 5, and finally 0 kcal mol^{-1} $Å^{-2}$, sequentially. (iv) Finally at 300 K, a 20 ns equilibration without any restrain force was performed in NPT ensemble followed by 2000 ns production run.

The long-ranged nonbonding electrostatic interactions was calculated by using particle mesh Ewald (PME) method [69] having a charge grid spacing of $\backsim$ 1 Å and a DSUM_TOL value of 10^{-5} Å for each of the systems. The cut-off radius was applied 12 Å for all nonbonding short-ranged interactions. SHAKE algorithm [70] with a tolerance of 10^{-5} Å was used to constrain the covalent bonds involving hydrogen atoms. The temperature was controlled by Langevin dynamics [71] with a collision frequency of 5 ps^{-1}. Finally, all MD simulations were carried out for 2000 ns at temperature of 300 K and pressure condition of 1 atm in isothermal-isobaric (NPT) ensemble without any restrain. Berendsen barostat with the collision frequency of 1.0 ps^{-1} [72, 73] was applied to maintain the physical pressure. The time integration step was set to 2 fs. The trajectories obtained from 2000 ns production run for all systems, were analyzed using CPPTRAJ program [52] of the AMBER18 toolkit. The snapshots were created using Visual Molecular Dynamics

(VMD) [74]. Using own python code inspired by Do_X3DNA package [75], all torsional angle plot was estimated. The spatial distribution functions (SDFs) for all systems were analyzed with the help of TRAVIS [76].

4.3 Results and Discussions

4.3.1 Root Mean Square Deviations (RMSDs)

In **Fig. 4.2** and **Table 4.2**, we have have shown the root mean square deviations (RMSDs) of the chosen quadruplex DNA (15-TBA). To begin with, we have considered all the heavy atoms for the calculation of the RMSD of TBA (**Fig. 4.2 (a)**), where its NMR structure is used as a reference structure.. First, we consider the RMSD and average RMSD values for whole quadruplex DNA. For all reline concentrations, as is evident from the RMSD values, the overall structure of TBA are quite stable. Although when the reline concentration decreases, the deviation from the NMR structure increases from 1.55 Å to 2.73 Å (system S0 to S3). For W0 system (which consists of only 15-TBA, K^+ and water) the RMSD values are very much similar to that of the lower reline concentration (i.e S3 system). From these observations, apparently, we can conclude that the reline concentration induces a change in the structure of quadruplex DNA. To find the region, where the quadruplex DNA is affected most in reline concentrations, we have divided the whole quadruplex DNA into three parts (i) sugar-phosphate-backbone, (ii) tetrad and (iii) loop. For the aforesaid regions, we have considered all the heavy atoms of sugar-phosphate, nucleoside base of tetrad and nucleoside base of loop **Fig. 4.2 (b)-(d)**. In case of sugar-phosphate-backbone regions, the RMSD values slightly increase from 1.06 Å to 1.73 Å as reline concentration decreases. For tetrad

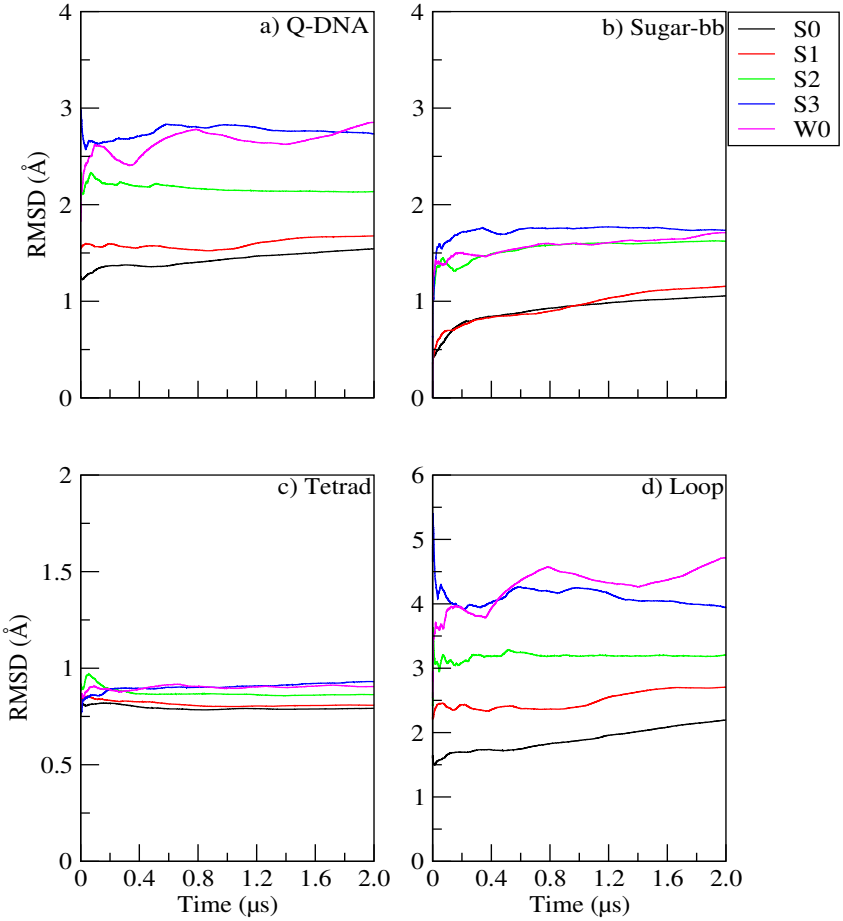


Figure 4.2: *Time evolution of the root mean square deviations (RMSDs) of all heavy atoms of (a) G-quadruplex-TBA, (b) sugar-phosphate backbone, (c) tetrad region and (d) loop region in different reline concentrations.*

Table 4.2: *Average root mean square deviations (RMSDs) of quadruplex TBA at different reline concentrations. The standard deviations are calculated by using block average method.*

RMSD (Å)	S0	S1	S2	S3	W0
Q-DNA	1.55 ± 0.14	1.68 ± 0.18	2.14 ± 0.18	2.73 ± 0.26	2.84 ± 0.51
Sugar-bb	1.06 ± 0.15	1.16 ± 0.24	1.62 ± 0.17	1.73 ± 0.21	1.70 ± 0.30
Tetrad	0.79 ± 0.04	0.80 ± 0.05	0.86 ± 0.06	0.93 ± 0.09	0.90 ± 0.10
Loop	2.21 ± 0.38	2.71 ± 0.40	3.21 ± 0.53	3.95 ± 0.70	4.69 ± 1.08

region, the values of RMSD at all concentrations are very much similar to each other (0.79 to 0.93 Å) but for loop region its value changes abruptly (2.21 to 3.95 Å). Considering the

RMSD values, among three regions of quadruplex DNA, the concentration of reline mostly induces the loop region whereas tetrad region is least affected. From the pairwise RMSD (**Fig. 4.3**) values, it is clear that the overall structure of quadruplex DNA deviates more from the initial structure at lower reline concentration. Thus, we can conclude that the decreasing reline concentration affects the conformational deviation of 15-TBA structure.

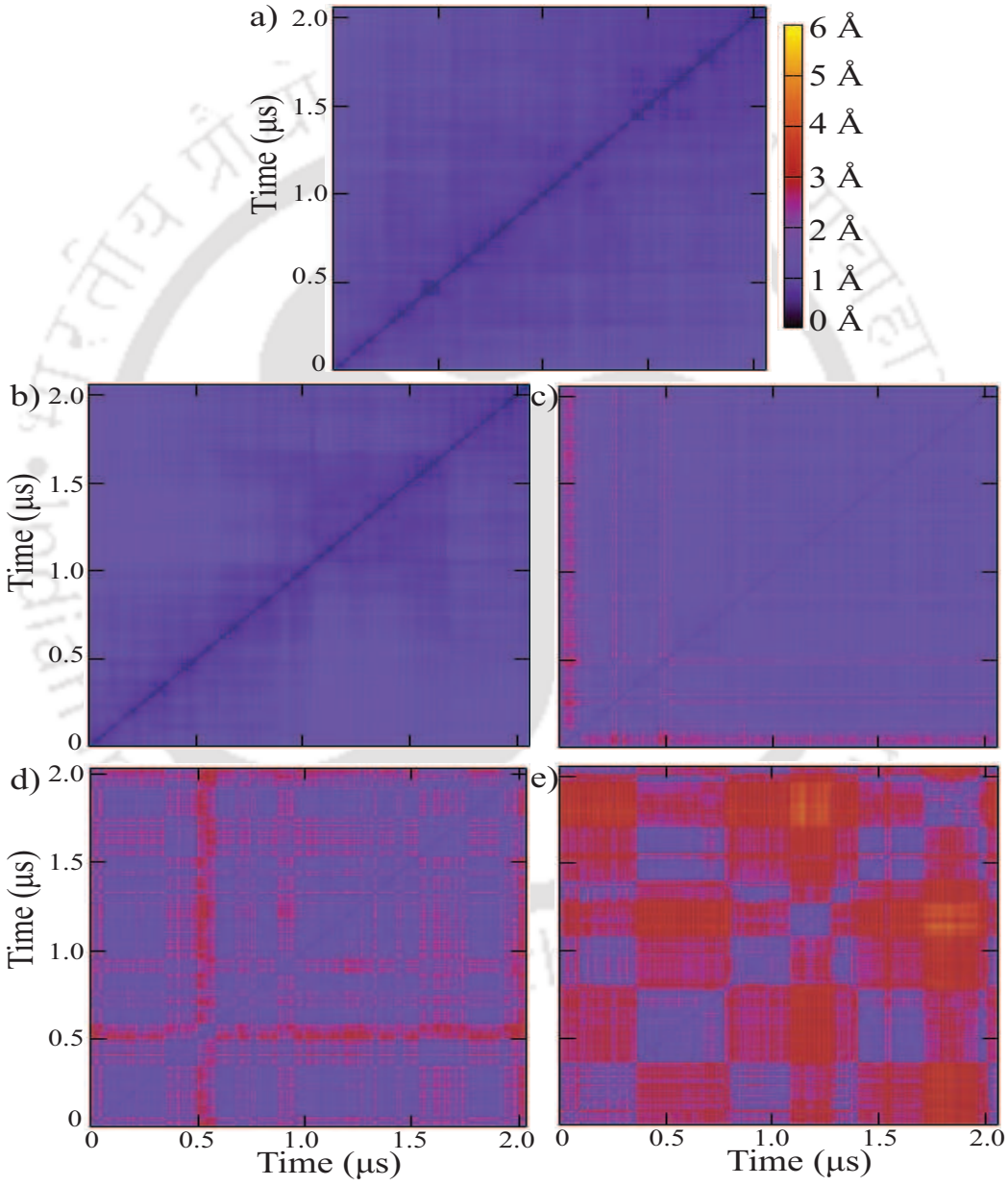


Figure 4.3: *Considering all heavy atoms of G-quadruplex TBA, 2D-RMSDs (pair-wise) of different systems (a) S0, (b) S1, (c) S2 (d) S3 and (e) W0 are illustrated.*

4.3.2 Root Mean Square Fluctuations (RMSFs)

For root mean square fluctuations (RMSFs) analysis, we have considered all the heavy atoms of the quadruplex TBA DNA. All atom RMSF values for 15-TBA at different reline concentrations are shown in **Fig. 4.4**. All atom positions of loop region (i.e loop-1, loop-2

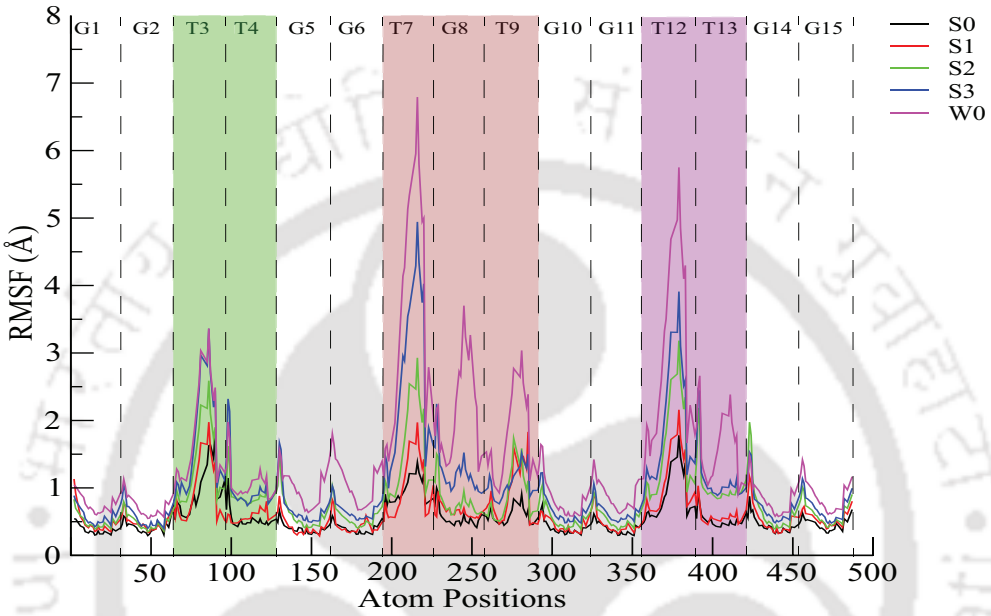


Figure 4.4: *Root mean square fluctuations (RMSFs) of all heavy atoms of G-quadruplex TBA for different systems. Loop-1, loop-2 and loop-3 regions are shown in green, brown and purple colors respectively.*

and loop-3) and tetrad region (namely, tetrad-1 and tetrad-2) of the 15-TBA are considered in order to understand the effect of reline concentrations. The NMR structure of 15-mer TBA is taken into consideration as the reference structure. From the RMSF plot, it is clearly seen that reline concentration induces the fluctuation of the studied 15-TBA. With decreasing reline concentrations, the RMSF values increase from system S0 to S3 (and also W0 system). Between the two regions loop and tetrad, the former region has a very high RMSF values. Among three loops (namely loop-1,loop-2 and loop-3), loop-2 has higher RMSF values than the other two. Overall, the RMSF values indicate that the 15-TBA structure is more flexible with decreasing reline concentration. These findings are in accordance with the calculated RMSD values discussed above.

4.3.3 Torsional Angles

To examine the flexibility of TBA structure we need to calculate four important angle parameters [77, 78]: (i) backbone torsional angles (α/γ), (ii) backbone torsional angles (β-δ), (iii) backbone torsional angles (ϵ-ζ), (iv) glycosidic torsional angles (χ). In wheel representation we have exhibited all the torsional angles of tetrad and loop residues in **Figures 4.5 and 4.6**. Here, we have considered the dihedral angle values from 0 to 360° instead of -180 to 180°. The distributions obtained from torsional angle analysis were clustered with 15° resolution and plotted with respect to corresponding probabilities.

Considering backbone torsional angles (α/γ) in tetrad residues the α/γ angles resemble with that of the NMR the structure at all reline concentrations. The α torsional angles of G-2, G-6, G-10, G-11 and G-15 are in the antiperiplanar (+ap∼180°) region and others are fall in +sc (synclinal) or g^+ (+sc∼60°) region. However, the α torsional angles of loop residues deviate more than that of the tetrad residues when reline concentration is decreased. In case of γ torsional angles similar pattern can be seen and the angles fall in the +sc or +ac region (+ac∼120°). Next we consider the β-δ torsional angle. For the tetrad residues, the β torsional angles of G-2, G-5 and G-6 residues fall in the +ap (antiperiplanar) region (+ap∼150-180°) and others fall in the -ap region (-ap∼180-210°). For the δ torsional angles, all the residues fall in the +ac region except for the G-15 residue (which falls in +ap region). In case of loop residues, β torsional angles for all the residues fall in the -ap region and T-4 and T-7 fall in the +ap while others fall in the +ac region and it is very much alike to the NMR structure. Here we have also observed that at low reline concentration, the torsional angle span over a large number of probable angles.

For ϵ torsional angles, G-5, G-6 and G-10 fall in the -ap region and others fall in the -ac region; for ζ torsional angles G-1, G-2, G-6 and T-13 fall in the -sc region, whereas G-10, G-11 and T-3 are in +ap region and others fall in +ac region For all the cases the angles are similar to that of the NMR structure. Now, we consider the most important glycosidic angle χ torsional angle which maintains the structure and conformation of the guanine-rich DNA. For loop residues all the χ torsional angles fall in the -ac region (very much alike to the NMR structure) except for the T-7, G-8, T-12 and T-13. But in case of tetrad, the χ torsional angle of the residues are very much comparable to the NMR structure.

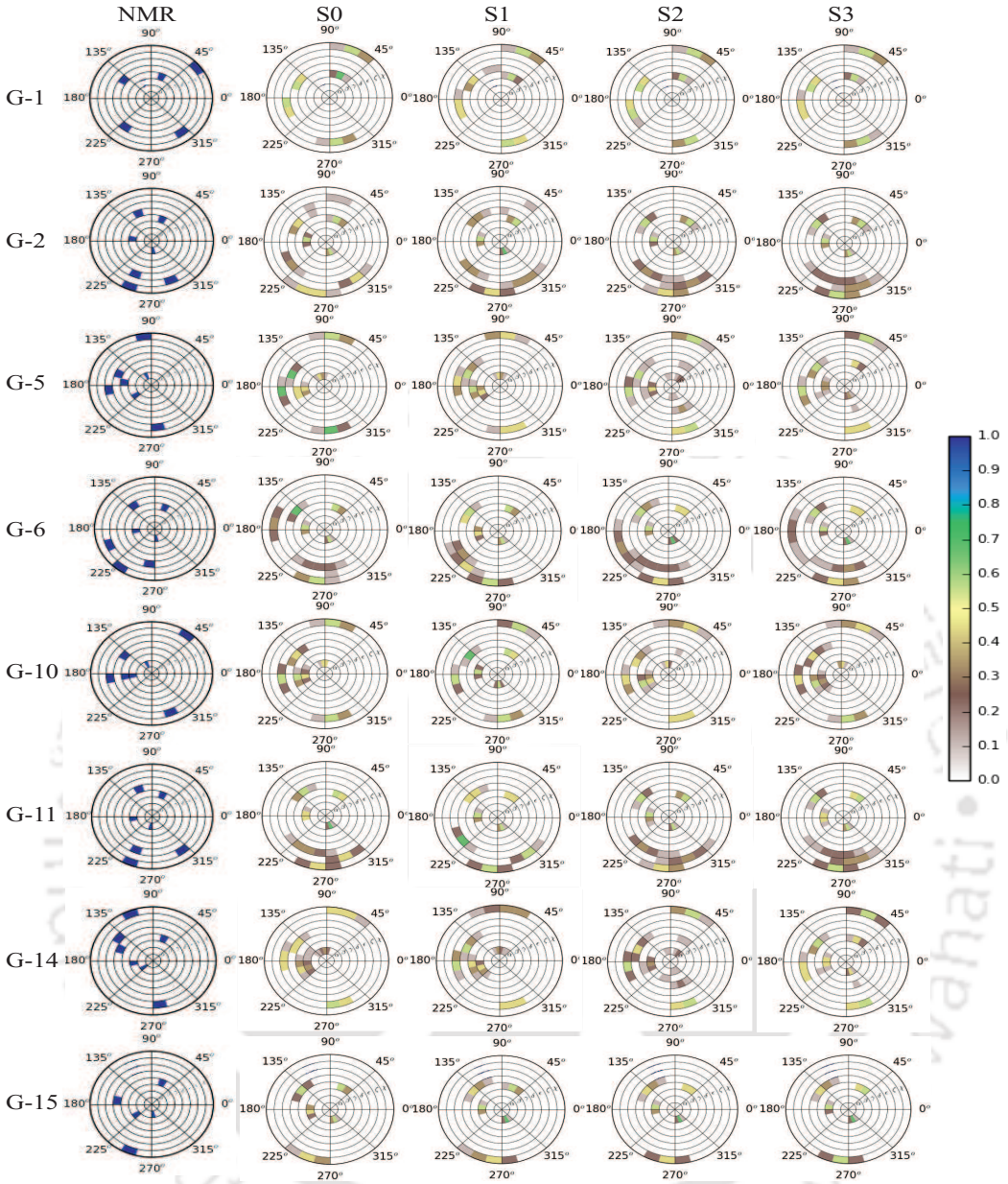


Figure 4.5: *All torsion angles of tetrad residues in a wheel representation for all the systems and the NMR structure. The color scale between 0 to 1 represents the probability of the corresponding angles. Guanine represents as G. From the center to edge of the wheel, α, β, γ, δ, ϵ, ζ and χ torsional angles are represented respectively.*

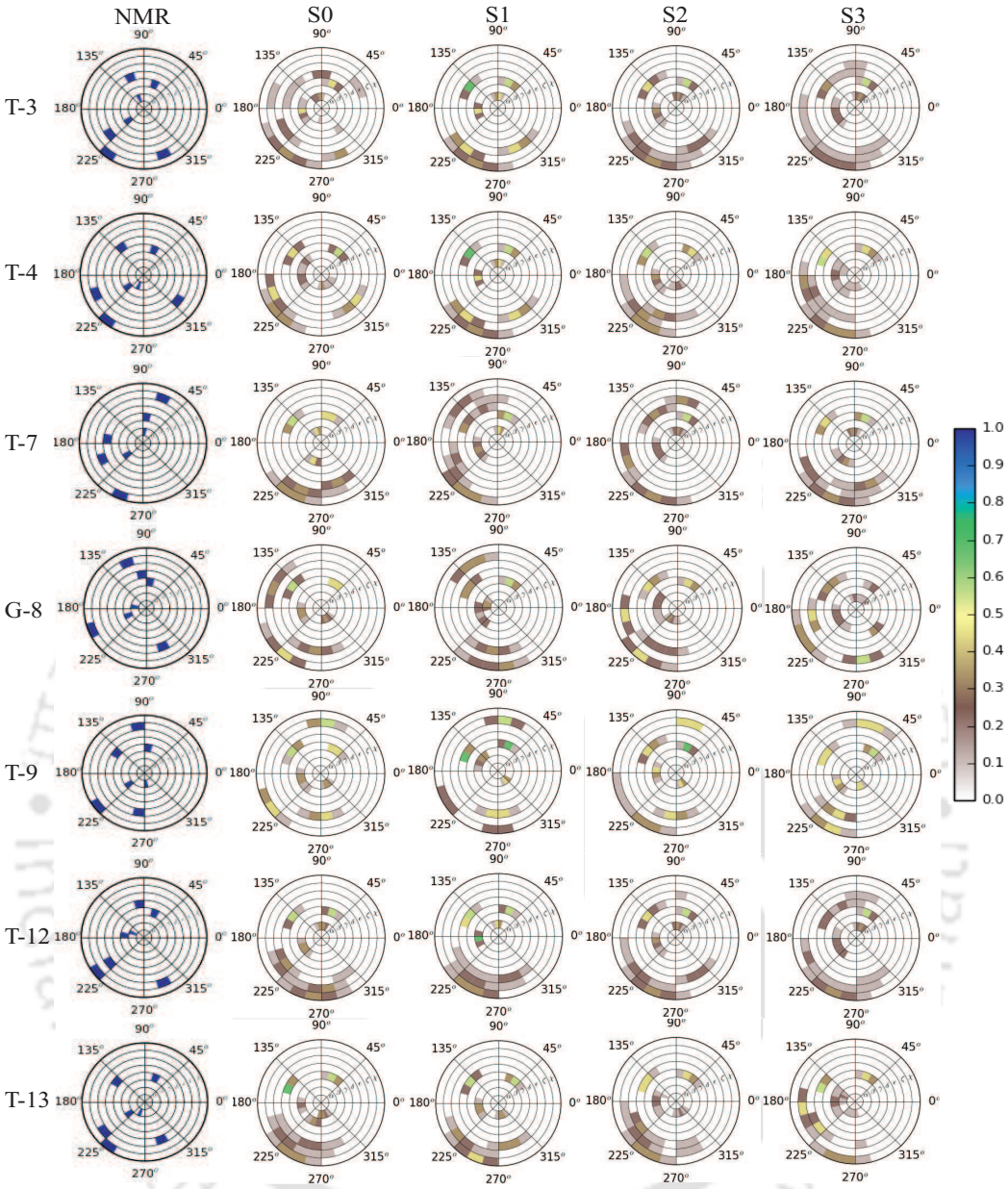


Figure 4.6: *All torsion angles of loop residues in a wheel representation for all the reline concentrations and the NMR structure. The color scale between 0 to 1 represents the probability of the respected angle. Guanine and thymine are represented as G and T respectively. From the center to edge in the wheel, α, β, γ, δ, ϵ, ζ and χ torsional angles are represented respectively.*

4.3.4 Radial Distribution Functions (RDFs)

In order to calculate the radial distribution functions (g(r)), we have divided the quadruplex TBA into three parts: (i) sugar-phosphate-backbone, (ii) guanine nucleoside bases of tetrads and (iii) thymine nucleoside bases of loops. In **Figures 4.7-4.9**, we have shown the radial distribution functions (RDFs) between the atoms of these three region and the N and O atoms of urea and O atom of choline. Here, we have only considered the atoms

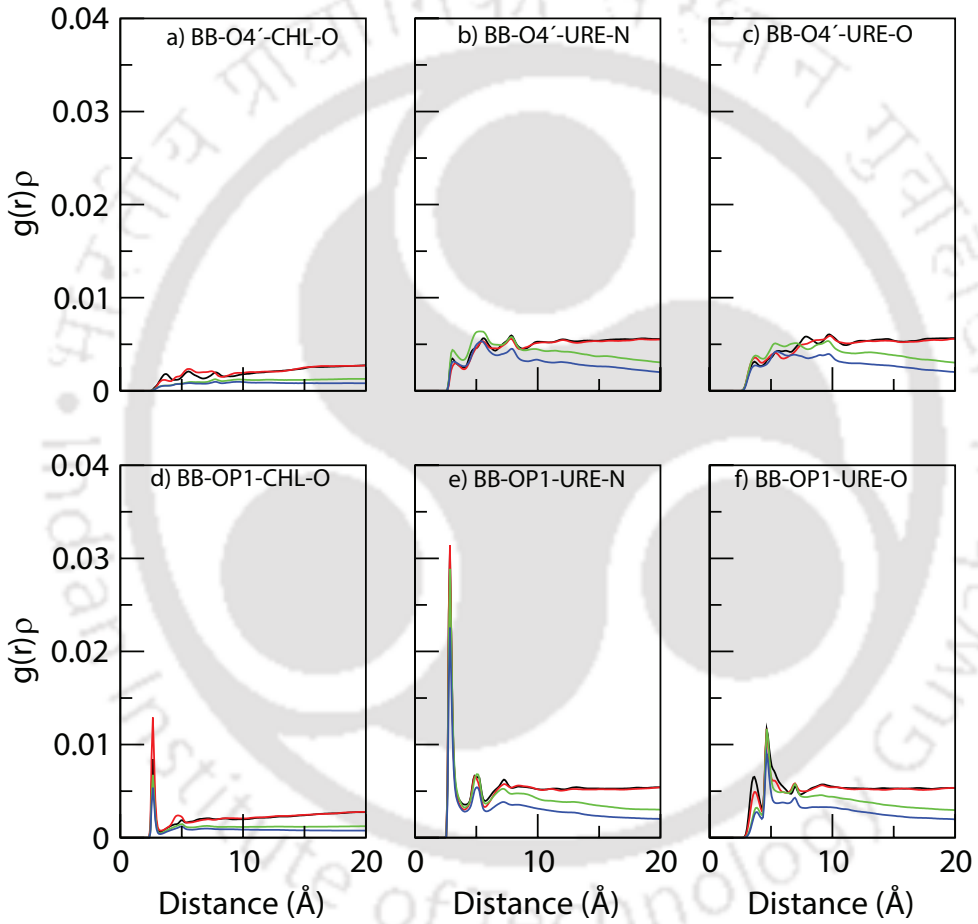


Figure 4.7: *Radial distribution functions (RDFs) of OP1 (a-c) and O4′ (d-f) atoms of sugar-phosphate-backbone with O atom of choline and N and O atoms of urea molecules respectively.*

which take part in hydrogen bond formation. Therefore, we have chosen N and O atoms of urea and O atom of choline and in case of quadruplex DNA, OP1 and O4′ atoms of sugar phosphate backbone, N1, N2, N3 and O6 atoms of guanine nucleoside base and N3, O2 and O4 atoms of thymine nucleoside base. Since the different systems have different densities, in order to calculate rdfs we have multiplied g(r) by the number density (ρ) of

the corresponding atom.

First, the g(r)ρ is considered for sugar-phosphate-backbone region (**Fig. 4.7**). With decreasing reline concentration, the peak height of the first maximum of g(r)ρ between OP1 and O4′ atoms of sugar-phosphate backbone with O atom of choline molecules decreases. Moreover, for the g(r)ρ between OP1 atom of sugar-phosphate backbone and O and N atoms of urea, same patterns of the g(r)ρ are noticed. However, for the g(r)ρ between O4′ atom with O and N atoms of urea the first peak heights remain almost unchanged for S0 to S2 systems and it is the lowest for S3 system. From the g(r)ρ, we find that choline and urea molecules interact with OP1 atom more favorably than with O4′ atom. However,

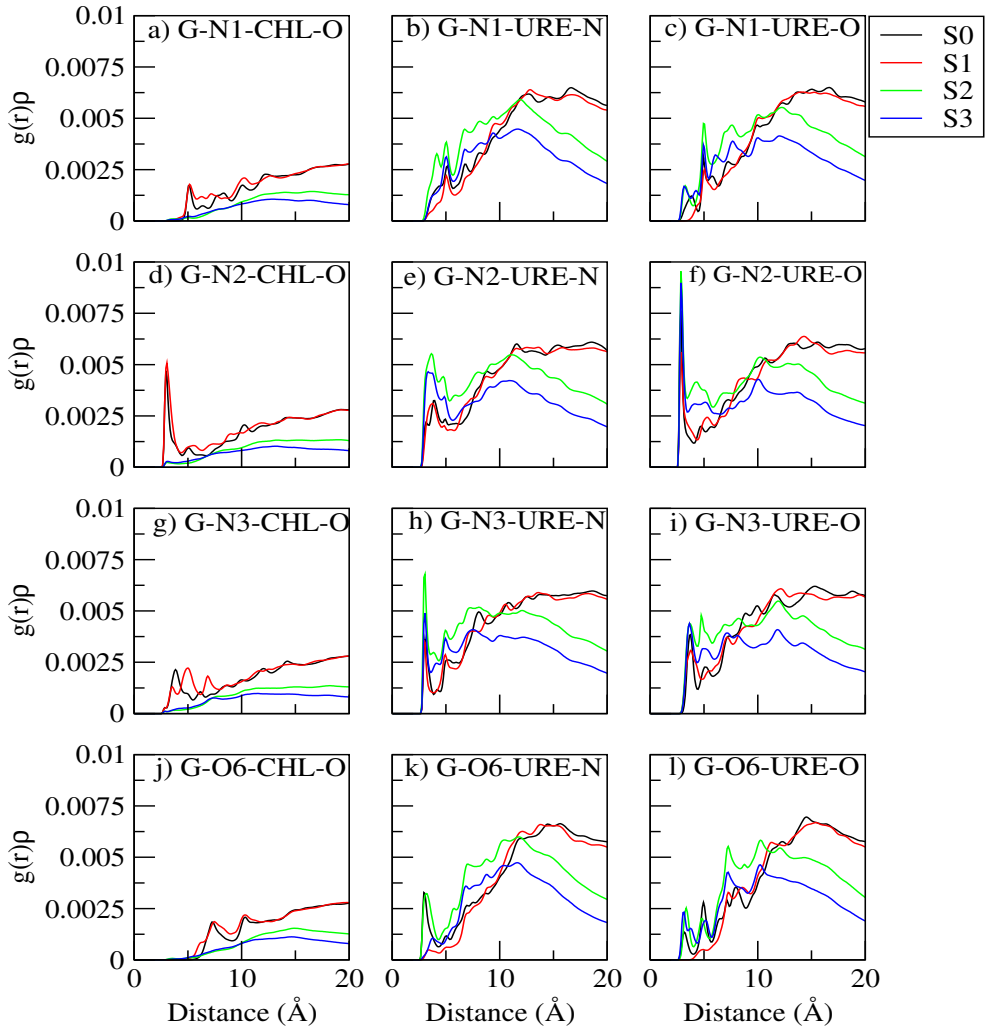


Figure 4.8: *Radial distribution functions (RDFs) of N1 (a-c), N2 (d-f), N3 (g-i), O6 (j-l) atoms of guanine with O atom of choline and N and O atoms of urea molecules respectively.*

it is also observed that the distribution functions involving O4′-urea-O and OP1-urea-O are not a strong function of reline concentration.

From the correlation functions involving tetrad regions of guanine nucleoside bases and urea and choline molecules (**Fig. 4.8**), it can be observed that only the interactions between N2 atom of guanine and choline and urea molecules are significant. Moreover, it is also noticed that the interactions between guanine residue and choline molecules decrease with decreasing reline concentrations. On the other hand, the interactions between guanine residue and urea molecules are not changed much as one moves from S0 to S3 system.

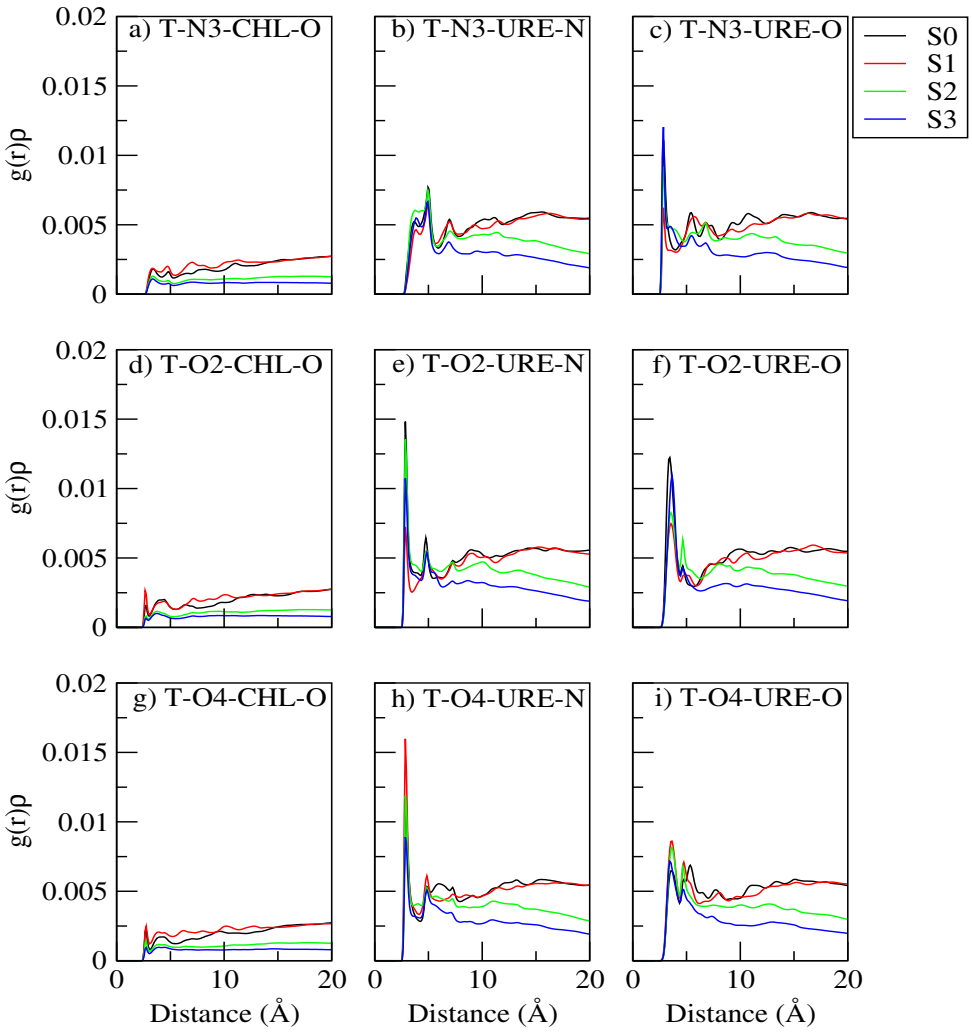


Figure 4.9: *All atom radial distribution functions (RDFs) of N3 (a-c), O2 (d-f), O4 (g-i) atoms of thymine with O atom of choline and N and O atoms of urea molecules respectively.*

Fig. 4.9 depicts the RDFs of thymine nucleoside base of loop residue and choline and urea molecules. From these RDFs, it is apparent that the interactions between O atom of choline and thymine nucleoside bases are negligibly small, Whereas, stronger interactions can be seen between urea molecules and thymine nucleoside bases. As, the reline concentration is decreased the first peak heights of distribution functions involving O2 and O4 atom of thymine and N atom of urea are decreased. For rest of the RDFs the peak heights do not follow any specific trend.

Taking into consideration of the rdf analysis, we conclude that, among three regions of quadruplex TBA, the interaction between sugar phosphate backbone region of quadruplex TBA with reline is more than the other regions. Moreover, between choline and urea molecules, urea interacts more favorably than choline with quadruplex TBA. With decreasing reline concentration, the interactions decreases only for choline molecules.

4.3.5 Spatial Distribution Functions (SDFs)

The spatial distribution functions (SDFs) depict probabilistic 3D structure of reline molecules and give more specific structuring information of quadruplex DNA. We have calculated SDF through 2000 ns trajectory using TRAVIS package [76] and for the plot we have applied CHIMERA visualization software [79] using 0.25 isovalue. In **Fig. 4.10**,

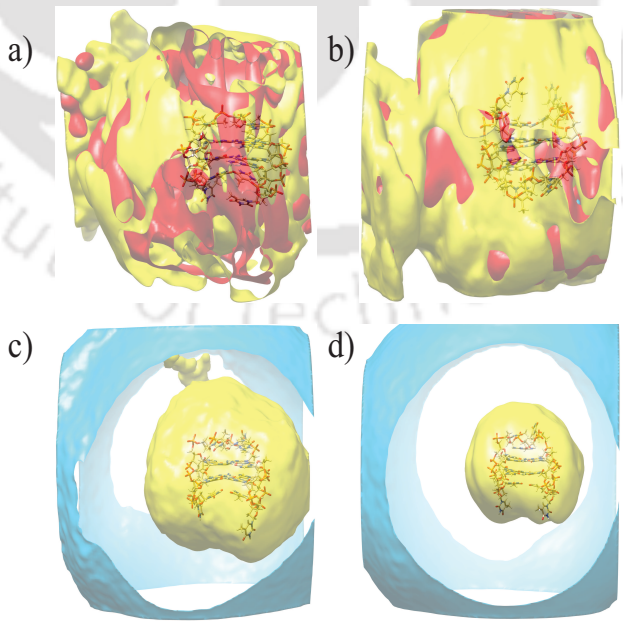


Figure 4.10: *Spatial density maps of choline, urea and water around quadruplex TBA at different reline concentrations. ((a)-(d)) are for systems S0, S1, S2 and S3 respectively. Here, red, yellow and sky-blue colors represent choline, urea and water.*

we have shown the SDFs at different reline concentrations. At high reline concentrations (i.e systems S0 and S1), the SDF of urea and choline around TBA shows a compact distribution of hydrogen bond donating urea and choline molecules around the reference TBA. It confirms that a strong hydrogen bond interaction is present between urea, choline and quadruplex DNA. At lower reline concentrations, the density of choline and urea around G-quadruplex TBA DNA decrease. On decreasing reline concentration, at first the choline molecules of reline are replaced by the water molecules and then the urea molecules are replaced by the water molecules. It indicates that when the number of water molecules are increased, the interactions between choline and urea with TBA are getting reduced.

4.3.6 Hydrogen Bonding

In reline solution, choline and urea molecules can form hydrogen bonds with like molecules as well as with guanine-rich-DNA. Hydrogen bonds play a key role in the stabilization of TBA molecule in reline solution. We have estimated the average number of hydrogen bonds formed by quadruplex-TBA with choline, urea and water molecules. For this, following previous studies [80–82] a set of geometric criteria are assigned. If the distance between the donor (D) and acceptor (A) is $\leq$ 3.5 Å and, simultaneously, the angle of D-H-A is $\leq$ 45° then a hydrogen bond is considered. From the **Table 4.3** and **Fig. 4.11**, it is clear

Table 4.3: *Average number of quadruplex TBA-choline, quadruplex TBA-urea and quadruplex TBA-water hydrogen bonds for different systems. Q-DNA, CHL, URE and WAT refer to quadruplex TBA DNA, choline, urea and water molecules respectively. The standard deviations of different values are estimated by block average method.*

H-bonds	S0	S1	S2	S3
Q-DNA-CHL	31.82 ± 6.83	36.48 ± 5.62	17.57 ± 4.89	15.08 ± 4.67
Q-DNA-URE	127.80 ± 8.87	117.60 ± 7.11	128.90 ± 12.0	99.95 ± 11.96
Q-DNA-WAT		1.19 ± 1.24	54.99 ± 8.38	76.09 ± 9.91

that at high reline concentration (for systems S0 and S1), the hydrogen bonds between choline and quadruplex DNA are almost same but it decreases when the reline concentration is decreased (for systems S2 and S3). The number of hydrogen bonds between urea and quadruplex TBA are almost same up to system S2 and then it decreases abruptly at very low reline concentration (i.e for S3 system). Similarly, the number of hydrogen bonds between water and TBA increases when reline concentrations decreases.

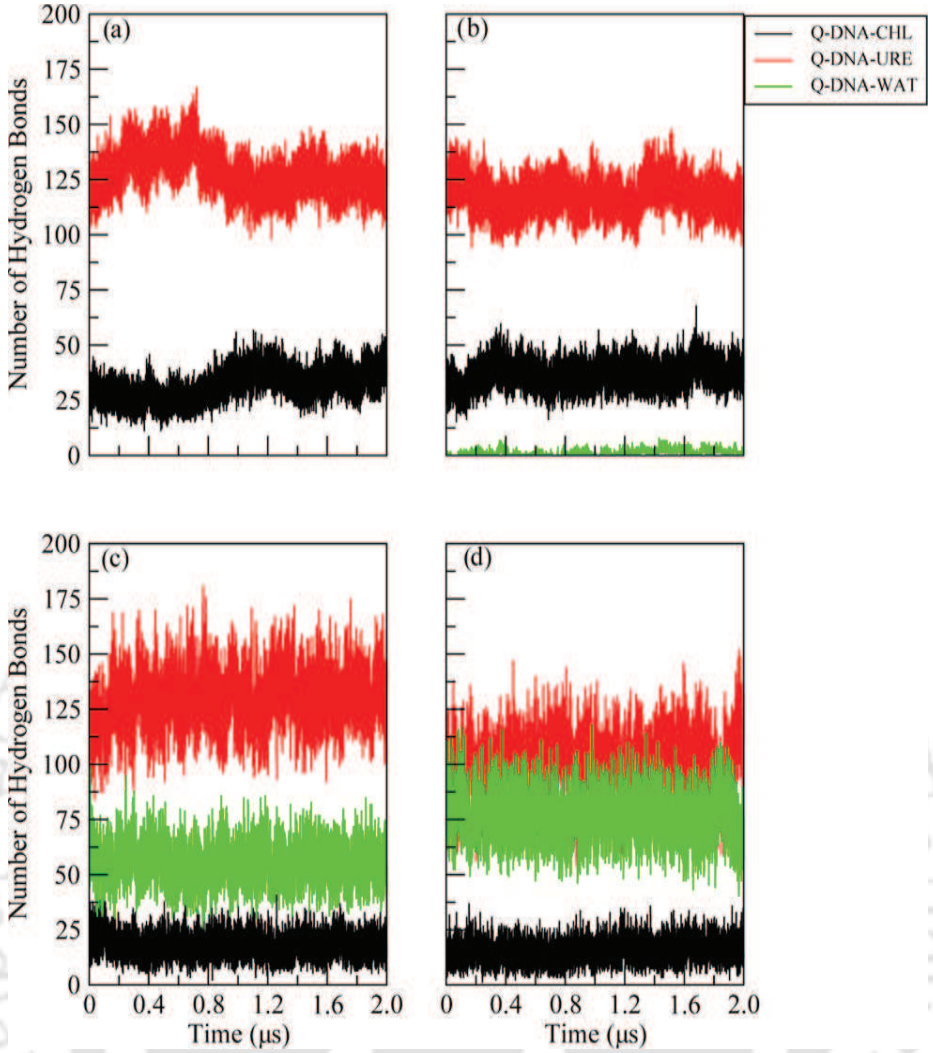


Figure 4.11: *Time evolution of hydrogen bonds between quadruplex TBA and choline, urea and water molecules at different reline concentrations. ((a)-(d)) refer to systems S0, S1, S2 and S3 respectively.*

To characterize the interaction of choline and urea with TBA in detail, we have shown in **Table 4.4** and **Fig. 4.12** how the number of choline and urea molecules that are

Table 4.4: *The average number of choline and urea molecules around 5.0 Å of quadruplex-TBA. The standard deviations are calculated by using block average method.*

Average Number		S0	S1	S2	S3
Choline	Tetrad	30.68 ± 2.94	32.96 ± 2.93	19.98 ± 2.57	15.99 ± 2.58
	Loop	38.68 ± 3.09	41.08 ± 3.35	25.66 ± 2.79	19.49 ± 2.79
Urea	Tetrad	72.12 ± 4.68	70.40 ± 6.03	69.71 ± 6.40	53.61 ± 5.92
	Loop	85.29 ± 7.46	77.71 ± 5.53	83.08 ± 6.93	62.65 ± 6.75

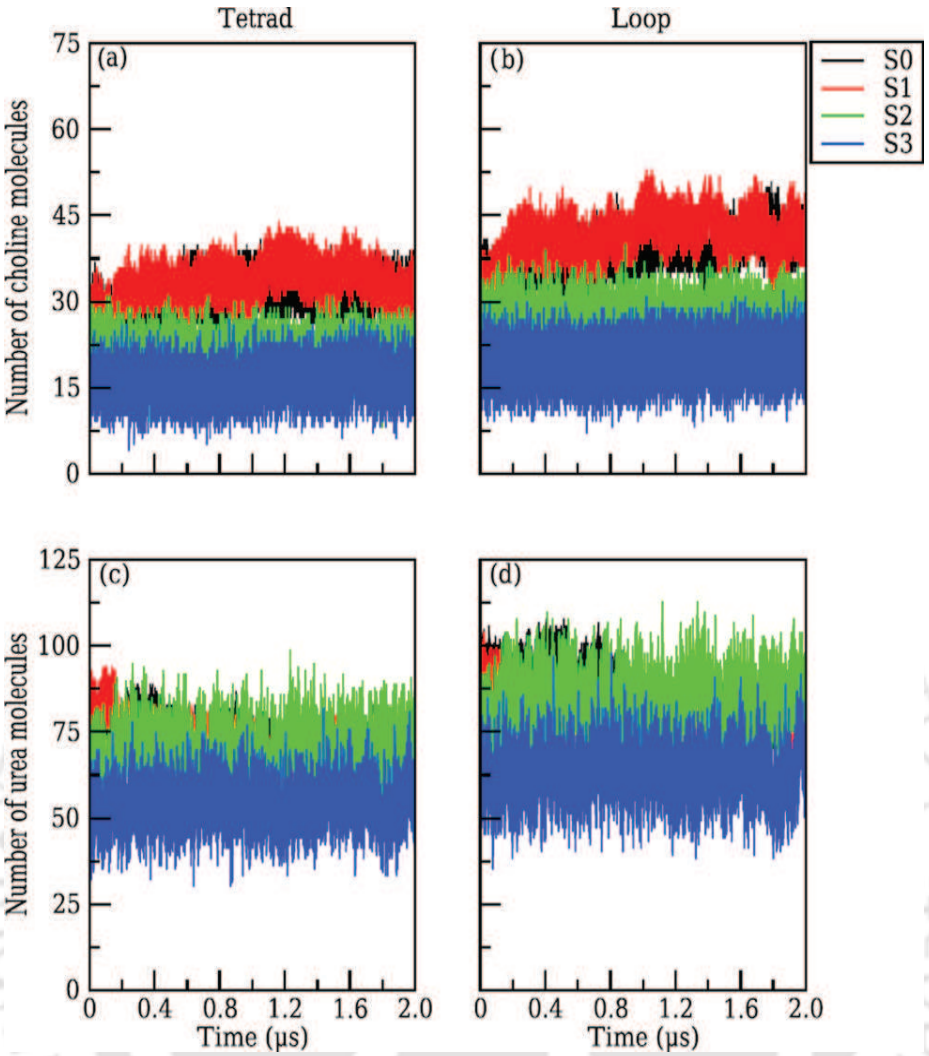


Figure 4.12: *Time evolution of number of choline and urea molecules within 5.0 Å of the TBA surface. Here ((a) and (c)) and ((b) and (d)) represent tetrad and loop regions.*

present within 5.0 Å of TBA surface change as simulation progresses. Considering first choline molecules, we have noticed that when reline concentration decreases choline molecules are replaced by the water. Similarly, for urea molecules, up to S2 system the average number of urea molecules are almost same but in the case of S3 system urea molecules are replaced by water abruptly. Thus, we can conclude that water molecules first replace the choline molecules and next it replaces the urea molecules with reducing reline concentration.

4.3.7 Loop Dynamics

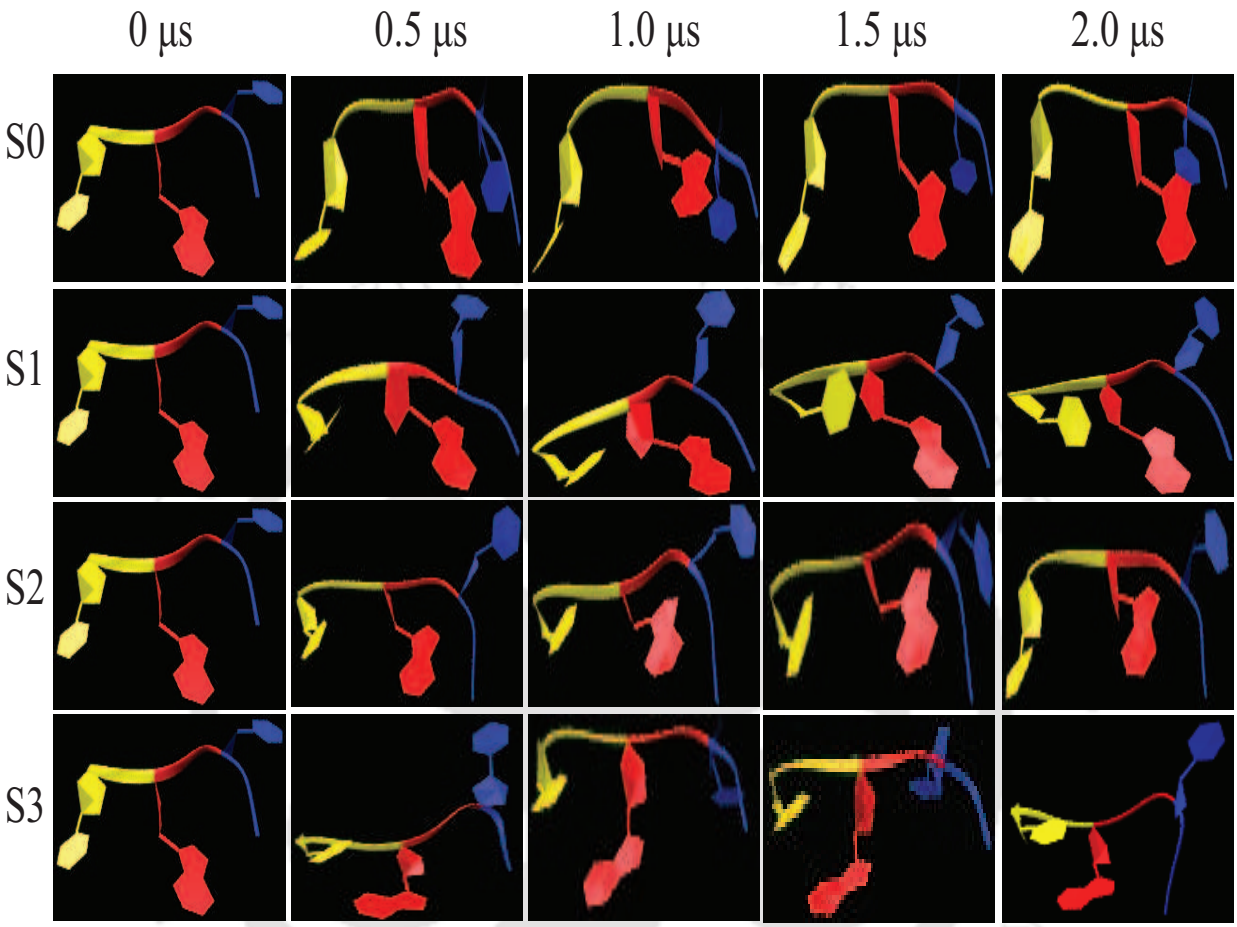


Figure 4.13: *Snapshots at different reline concentrations (S0-S3 systems) at 300 K temperature and at 0, 0.50, 1.0, 1.50 and 2.0 μs. In order to improve visual clarity choline, urea and water molecules are left off. Blue, red and yellow represent thymine-7, guanine-8 and thymine-9 residues respectively.*

In **Fig. 4.13**, we have demonstrated the progression of quadruplex TBA and its loop-2 of the guanine rich quadruplex TBA structure at different reline concentrations throughout the 2000 ns simulations at a time intervals of 500 ns. From the snapshots at different time intervals, it can be seen that, the final structure of loop-2 at all reline concentrations more or less deviates from the initial NMR structure. Moreover, at low reline concentration for S3 system, for the guanine-8 and thymine-9 nucleoside bases of loop-2, π-π stacking interaction is observed. It is to be noted that in the initial NMR structure, there is no such stacking interaction. Abiding with previous study [83], if the distance between center of mass of the two nucleoside bases is $\leq$ 5.0 Å and, simultaneously, the highest probability of

the angle between two nucleoside plane is considered less than or equal to 20° π-π stacking interaction is taken to consideration. In **Table 4.5**, for all systems the average

Table 4.5: *The average distance between center of mass of heavy atoms of G-8 and T-9 nucleoside bases. The standard deviations are estimated by block averaging method.*

System	Average Distance (Å)
S0	11.17 ± 0.48
S1	7.67 ± 1.06
S2	8.70 ± 0.79
S3	4.88 ± 0.47

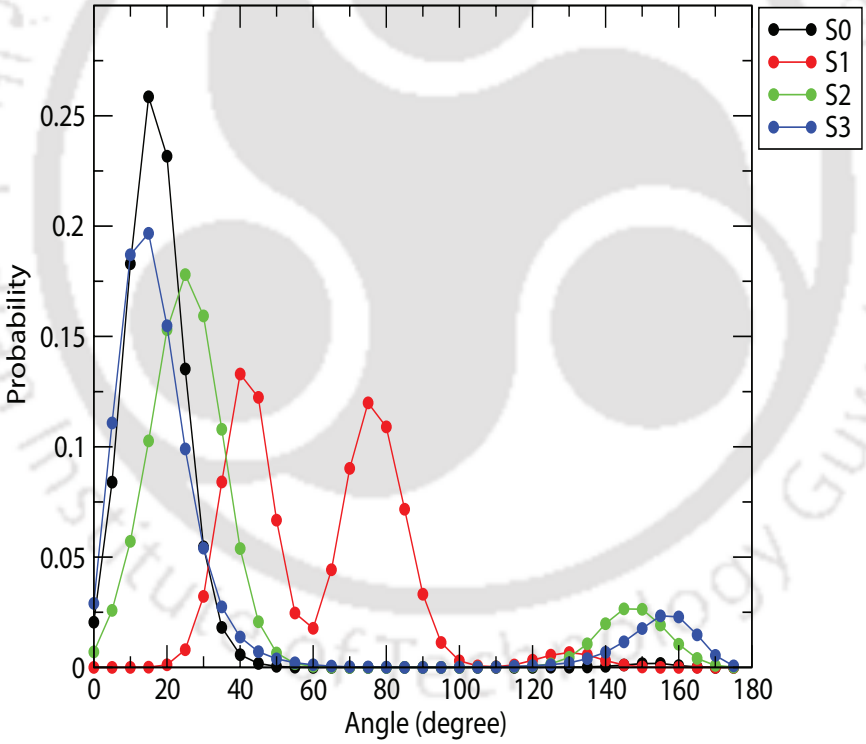


Figure 4.14: *Stacking probability with respect to angle between the corresponding plane of the Guanine-8 and Thymine-9 residues at all reline concentrations.*

distances between guanine-8 and thymine-9 nucleoside bases are presented and in **Fig. 4.14**, the stacking probabilities of the corresponding planes of these nucleoside bases are shown. The average distance of the previously mentioned nucleoside bases of loop-2 is 4.88 Å for system S3 and the highest probability of the angle between two nucleoside plane is

less than 20° for S3 system. Moreover, among all the systems, these nucleoside bases of loop-2 of S3 system fulfil both the criteria. Thus, it can be proposed that at low reline concentration (i.e S3 system), guanine-8 and thymine-9 nucleoside bases of loop-2 stack to each other which is not noticed at high reline concentrations.

4.3.8 Limitations

In this work, we have used the recent force field parameters of different atomic sites of quadruplex DNA and deep eutectic solvents. Moreover, it is not a priory guaranteed that the latest AMBER force field parameters are fully balanced with the reline environment and force field. Due to the lack of experimental as well as simulation benchmark data, it is very difficult to resolve this issue. However, the studied quadruplex TBA has no flanking bases that could trap the loops in long-living states and also loop-1 and loop-2, which are facing each other, are very short in length. Thus, we can predict that sampling is a less severe issue for the presently studied system.

4.4 Summary and Conclusions

Classical all-atom molecular dynamics simulations can provide significant information about the structural changes thrombin binding aptamer (TBA) at different concentrations of reline DES media. From RMSD, RMSF and torsional angle analyses, we have noticed that the structure and conformation of guanine rich TBA is more rigid or less flexible than that in the lower reline concentration. This is attributed to the fact that at high concentration the density of reline around the quadruplex TBA is very much higher than that at a lower reline concentration. We also observed that at low reline concentration the choline molecules around the quadruplex DNA are replaced by water molecules. Thus, the interaction and hydrogen bond formation between choline and quadruplex TBA DNA are getting affected at low reline concentrations. Moreover, the density of reline is comparably much higher near the phosphate regions. It is also observed that guanine-8 and thymine-9 nucleoside bases of loop-2 of S3 system stack to each other which is not noticed at high reline concentrations. Overall, the conformational and structural features of TBA in anhydrous and hydrated reline media agree well with experimental findings [31, 34, 35]. Our study proposes that for long term nucleic acid storage, reline is a good choice of solvent as a storage medium.

References

[1] L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas and J. J. Toole, *Nature*, 1992, **355**, 564.

[2] T. Rydel, K. Ravichandran, A. Tulinsky, W. Bode, R. Huber, C. Roitsch and J. Fenton, *Science*, 1990, **249**, 277–280.

[3] I. Russo Krauss, A. Merlino, A. Randazzo, E. Novellino, L. Mazzarella and F. Sica, *Nucleic Acids Res.*, 2012, **40**, 8119–8128.

[4] H. Ranpura and P. H. Bolton, *Biophys. J.*, 2014, **106**, 65a.

[5] P. Schultze, R. F. Macaya and J. Feigon, *J. Mol. Biol.*, 1994, **235**, 1532 – 1547.

[6] R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe and J. Feigon, *Proc. Natl. Acad. Sc.*, 1993, **90**, 3745–3749.

[7] V. M. Marathias and P. H. Bolton, *Nucleic Acids Res*, 2000, **28**, 1969–1977.

[8] A. Paiva, R. Craveiro, I. Aroso, M. Martins, R. L. Reis and A. R. C. Duarte, *ACS Sustain. Chem. Eng.*, 2014, **2**, 1063–1071.

[9] K. R. Seddon, *J. Chem. Technol. Biotechnol.*, 1997, **68**, 351–356.

[10] S. V. Muginova, A. Z. Galimova, A. E. Polyakov and T. N. Shekhovtsova, *J. Anal. Chem.*, 2010, **65**, 331–351.

[11] K. Fujita, D. R. MacFarlane, M. Forsyth, M. Yoshizawa-Fujita, K. Murata, N. Nakamura and H. Ohno, *Biomacromolecules*, 2007, **8**, 2080–2086.

[12] N. Byrne, L.-M. Wang, J.-P. Belieres and C. A. Angell, *Chem. Commun.*, 2007, 2714–2716.

[13] R. Madeira Lau, M. J. Sorgedrager, G. Carrea, F. van Rantwijk, F. Secundo and R. A. Sheldon, *Green Chem.*, 2004, **6**, 483–487.

[14] J. L. L. Carter, C. C. Santini, L. J. Blum and B. Doumche, *RSC Adv.*, 2015, **5**, 23715–23718.

[15] A. Chandran, D. Ghoshdastidar and S. Senapati, *J. Am. Chem. Soc.*, 2012, **134**, 20330–20339.

[16] D. Ghoshdastidar, D. Ghosh and S. Senapati, *J. Phys. Chem. B*, 2016, **120**, 492–503.

[17] D. Ghoshdastidar and S. Senapati, *Nucleic Acids Res.*, 2018, **46**, 4344–4353.

[18] M. Kulkarni and A. Mukherjee, *RSC Adv.*, 2016, **6**, 46019–46029.

[19] S. Satpathi, M. Kulkarni, A. Mukherjee and P. Hazra, *Phys. Chem. Chem. Phys.*, 2016, **18**, 29740–29746.

[20] D. Saha, M. Kulkarni and A. Mukherjee, *Phys. Chem. Chem. Phys.*, 2016, **18**, 32107–32115.

[21] M. Armand, F. Endres, D. R. MacFarlane, H. Ohno and B. Scrosati, *Nat. Mater.*, 2009, **8**, 621.

[22] A. P. Abbott, G. Capper, D. L. Davies, R. K. Rasheed and V. Tambyrajah, *Chem. Commun.*, 2003, 70–71.

[23] A. P. Abbott, G. Capper, D. L. Davies, K. J. McKenzie and S. U. Obi, *J. Chem. Eng. Data*, 2006, **51**, 1280–1282.

[24] Q. Zhang, K. De Oliveira Vigier, S. Royer and F. Jérme, *Chem. Soc. Rev.*, 2012, **41**, 7108–7146.

[25] Q. Wen, J.-X. Chen, Y.-L. Tang, J. Wang and Z. Yang, *Chemosphere*, 2015, **132**, 63 – 69.

[26] M. Hayyan, M. A. Hashim, M. A. Al-Saadi, A. Hayyan, I. M. AlNashef and M. E. Mirghani, *Chemosphere*, 2013, **93**, 455 – 459.

[27] M. Gutiérrez, M. Ferrer, L. Yuste, F. Rojo and F. delMonte, *Angew. Chem*, **49**, 2158–2162.

[28] R. Esquembre, J. M. Sanz, J. G. Wall, F. del Monte, C. R. Mateo and M. L. Ferrer, *Phys. Chem. Chem. Phys.*, 2013, **15**, 11248–11256.

[29] H. G. Morrison, C. C. Sun and S. Neervannan, *Int. J. Pharm.*, 2009, **378**, 136 – 139.

[30] H. Tateishi-Karimata and N. Sugimoto, *Nucleic Acids Res.*, 2014, **42**, 8831–8844.

[31] I. Mamajanov, A. Engelhart, H. Bean and N. Hud, *Angew. Chem.*, **49**, 6310–6314.

[32] F. M. Lannan, I. Mamajanov and N. V. Hud, *J. Am. Chem. Soc.*, 2012, **134**, 15324–15330.

[33] I. Gállego, M. A. Grover and N. V. Hud, *Angew. Chem.*, **54**, 6765–6769.

[34] C. Zhao, J. Ren and X. Qu, *Langmuir*, 2013, **29**, 1183–1191.

[35] C. Zhao and X. Qu, *Methods*, 2013, **64**, 52 – 58.

[36] D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER 2018*, University of California, San Francisco, 2018.

[37] M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo and P. Jurečka, *J. Chem. Theory Comput.*, 2015, **11**, 5723–5736.

[38] R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka and T. E. Cheatham, *J. Chem. Theory Comput.*, 2016, **12**, 4114–4127.

[39] B. Islam, P. Stadlbauer, M. Krepl, M. Havrila, S. Haider and J. Sponer, *J. Chem. Theory Comput.*, 2018, **14**, 5011–5026.

[40] J. Šponer, G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R. A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka, N. G. Walter and M. Otyepka, *Chem. Rev.*, 2018, **118**, 4177–4338.

[41] B. Machireddy, G. Kalra, S. Jonnalagadda, K. Ramanujachary and C. Wu, *J. Chem. Inf. Model.*, 2017, **57**, 2846–2864.

[42] M. Havrila, P. Stadlbauer, B. Islam, M. Otyepka and J. Šponer, *J. Chem. Theory Comput.*, 2017, **13**, 3911–3926.

[43] N. G. J. Richards and M. M. Georgiadis, *Acc. Chem. Res.*, 2017, **50**, 1375–1382.

[44] B. Islam, P. Stadlbauer, A. Gil-Ley, G. Pérez-Hernández, S. Haider, S. Neidle, G. Bussi, P. Banas, M. Otyepka and J. Sponer, *J. Chem. Theory Comput.*, 2017, **13**, 2458–2480.

[45] A. K. Sieradzan, P. Krupa and D. J. Wales, *J. Phys. Chem. B*, 2017, **121**, 2207–2219.

[46] P. Stadlbauer, L. Mazzanti, T. Cragnolini, D. J. Wales, P. Derreumaux, S. Pasquali and J. Šponer, *J. Chem. Theory Comput.*, 2016, **12**, 6077–6097.

[47] D. R. Gruber, J. J. Toner, H. L. Miears, A. V. Shernyukov, A. S. Kiryutin, A. A. Lomzov, A. V. Endutkin, I. R. Grin, D. V. Petrova, M. S. Kupryushkin, A. V. Yurkovskaya, E. C. Johnson, M. Okon, E. G. Bagryanskaya, D. O. Zharkov and S. L. Smirnov, *Nucleic Acids Res.*, 2018, **46**, 10827–10839.

[48] S. Haider, *J. Indian Inst. Sci.*, 2018, **98**, 325–339.

[49] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

[50] I. S. Joung and T. E. Cheatham III, *J. Phys. Chem. B*, 2008, **112**, 9020–9041.

[51] L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, *J. Comput. Chem.*, **30**, 2157–2164.

[52] D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.

[53] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian09 Revision A.02*, 2016, Gaussian Inc. Wallingford CT.

[54] G. A. Petersson, A. Bennett, T. G. Tensfeldt, M. A. AlLaham, W. A. Shirley and J. Mantzaris, *The J. Chemical Physics*, 1988, **89**, 2193–2218.

[55] G. A. Petersson and M. A. AlLaham, *J. Chem. Phys.*, 1991, **94**, 6081–6090.

[56] J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model.*, 2006, **25**, 247 – 260.

[57] C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.

[58] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, **25**, 1157–1174.

[59] K. G. Sprenger, V. W. Jaeger and J. Pfaendtner, *J. Phys. Chem. B*, 2015, **119**, 5882–5895.

[60] Y. Zhang and E. J. Maginn, *J. Phys. Chem. B*, 2012, **116**, 10036–10048.

[61] S. L. Perkins, P. Painter and C. M. Colina, *J. Phys. Chem. B*, 2013, **117**, 10250–10260.

[62] S. L. Perkins, P. Painter and C. M. Colina, *J. Chem. Eng. Data*, 2014, **59**, 3652–3662.

[63] B. Doherty and O. Acevedo, *J. Phys. Chem. B*, 2018, **122**, 9982–9993.

[64] M. H. Mamme, S. L. C. Moors, H. Terryn, J. Deconinck, J. Ustarroz and F. De Proft, *J. Phys. Chem. Lett.*, 2018, **9**, 6296–6304.

[65] E. O. Fetisov, D. B. Harwood, I.-F. W. Kuo, S. E. E. Warrag, M. C. Kroon, C. J. Peters and J. I. Siepmann, *J. Phys. Chem. B*, 2018, **122**, 1245–1254.

[66] D. Shah and F. S. Mjalli, *Phys. Chem. Chem. Phys.*, 2014, **16**, 23900–23907.

[67] C. Schröder, *Phys. Chem. Chem. Phys.*, 2012, **14**, 3089–3102.

[68] H. Liu and E. Maginn, *J. Chem. Phys.*, 2011, **135**, 124507.

[69] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

[70] J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

[71] R. W. Pastor, B. R. Brooks and A. Szabo, *Mol. Phys.*, 1988, **65**, 1409–1419.

[72] B. P. Uberuaga, M. Anghel and A. F. Voter, *J. Chem. Phys.*, 2004, **120**, 6363–6374.

[73] D. J. Sindhikara, S. Kim, A. F. Voter and A. E. Roitberg, *J. Chem. Theory Comput.*, 2009, **5**, 1624–1631.

[74] W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.

[75] R. Kumar and H. Grubmüller, *Bioinformatics*, 2015, **31**, 2583–2585.

[76] M. Brehm and B. Kirchner, *J. Chem. Inf. Model.*, 2011, **51**, 2007–2023.

[77] M. Rebič, A. Laaksonen, J. Šponer, J. Uličny and F. Mocci, *J. Phys. Chem. B*, 2016, **120**, 7380–7391.

[78] X. Cang, J. Šponer and T. E. Cheatham III, *Nucleic Acids Res.*, 2011, **39**, 4499–4512.

[79] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, **25**, 1605–1612.

[80] S. Das and S. Paul, *J. Phys. Chem. B*, 2017, **121**, 8774–8785.

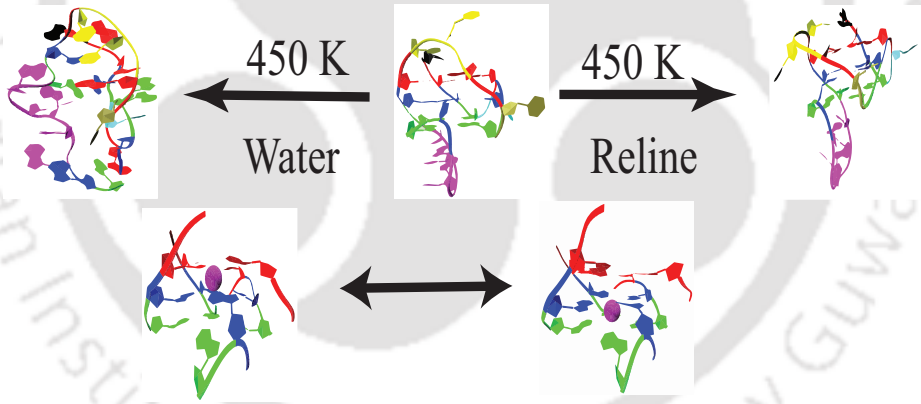
[81] B. Sharma and S. Paul, *J. Phys. Chem. B*, 2016, **120**, 9019–9033.

[82] G. Borgohain and S. Paul, *Comput. Theor. Chem.*, 2018, **1131**, 78 – 89.

[83] S. Pal and S. Paul, *Int. J. Biol. Macromol.*, 2019, **121**, 350 – 363.

Chapter 5

Understanding the Role of Reline on Temperature Induced Conformational Changes of c-KIT G-quadruplex DNA



Overview

The non-canonical guanine rich DNAs have drawn particular attention to the scientific world due to its controllable diverse and polymorphic structures. Apart from biological and medical significance, G-quadruplex DNAs are widely used in various fields such as nanotechnology, nanomachine, biosensors and biocatalyst. So far, the applications of the G-quadruplex DNA are mainly limited in the water medium. Recently, a new generation of solvent named deep eutectic solvent (DES), has become very popular and has been widely used as a reaction medium of biocatalytic reactions and long-term storage medium for nucleic acids, even at high temperature. Hence, it is essential to understand the role of DES on temperature-induced conformational changes of a G-quadruplex DNA. In this research work, we have explored the temperature-mediated conformational dynamics of c-KIT oncogene promoter G-quadruplex DNA in reline medium in the temperature range of 300 K to 500 K, using a total of 10 μs unbiased all-atom molecular dynamics simulation. Here, from RMSD, RMSF, R_g and principal component analyses, we notice that the c-KIT G-quadruplex DNA is stable up to 450 K in reline medium. However, it unfolds in water medium at 450 K. It is found that the hydrogen bonding interactions between c-KIT G-quadruplex DNA and reline play a key role in the stabilization of the G-quadruplex DNA even at high temperature. Furthermore, in this work we have observed a very interesting and distinctive phenomenon of the central cation of the G-quadruplex DNA. It's position was seen to fluctuate between the two tetrad cores i.e., the region between tetrad-1 and tetrad-2 and that between tetrad-2 and tetrad-3 and vice versa at 450 K and 500 K in reline medium which is absent in water medium at 450 K. Moreover, the rate of its oscillation is increased when temperature is increased.

5.1 Introduction

The proto-oncogene c-KIT G-quadruplex DNA controls key transduction cascades for the growth and proliferation of cell [1]. The c-KIT protein-tyrosine kinase plays a vital role in maintaining normal cell growth and therefore the mutation or over-expression of structurally essential areas of the protein lead to impaired activity, conducting oncogenic cellular transformations [2–4]. In many varieties of cancer, such as prostate [5] cancer, adenocarcinoma lung cancer [6], seminoma [7], melanoma [8] and gastrointestinal stromal tumours (GIST) [9, 10], c-KIT over-expression or mutation is observed. Their effects may be relatively high in a large number of cases. Therefore, many small drug molecules such as carbazole derivatives[11], isoalloxazines [12], quinazolones [13], bistriazoles [14] etc. which stabilize the G-quadruplex DNA, have drawn the attention in the emerging therapeutic field. In addition to the biological significance, it has also become a promising candidate for catalyst, nano-machine and nano-structure development [15–20], due to the controllable diverse structures and robust physiochemical existence of G-quadruplex DNA.

The proto-oncogene human c-KIT G-quadruplex DNA consists of 22 mer d(AGGGAGGGCGCTGGGAGGGAGGG) sequence (**Figures 5.1 (a) and (b)**). Various NMR and X-ray crystallographic studies [21, 22] have established the structure of c-KIT G-quadruplex DNA and have revealed that c-KIT DNA contains three G-tetrads (namely tetrad-1, tetrad-2 and tetrad-3) along with four loop regions (namely loop-1, loop-2, loop-3, loop-4). Moreover, loop-1 and loop-2 contain adenosine-5 (A5) and cytosine-9 (C9) nucleoside bases respectively, and loop-3 has cytosine-11 (C11) and thymine-12 (T12) nucleoside bases. Furthermore, c-KIT G-quadruplex DNA has one long loop (namely loop-4) which consists of A16-G17-G18-A19-G20 nucleoside bases. Taking into account these recent studies [11, 23, 24] we have used NMR structure of c-KIT G-quadruplex DNA of PDB ID:2O3M [21].

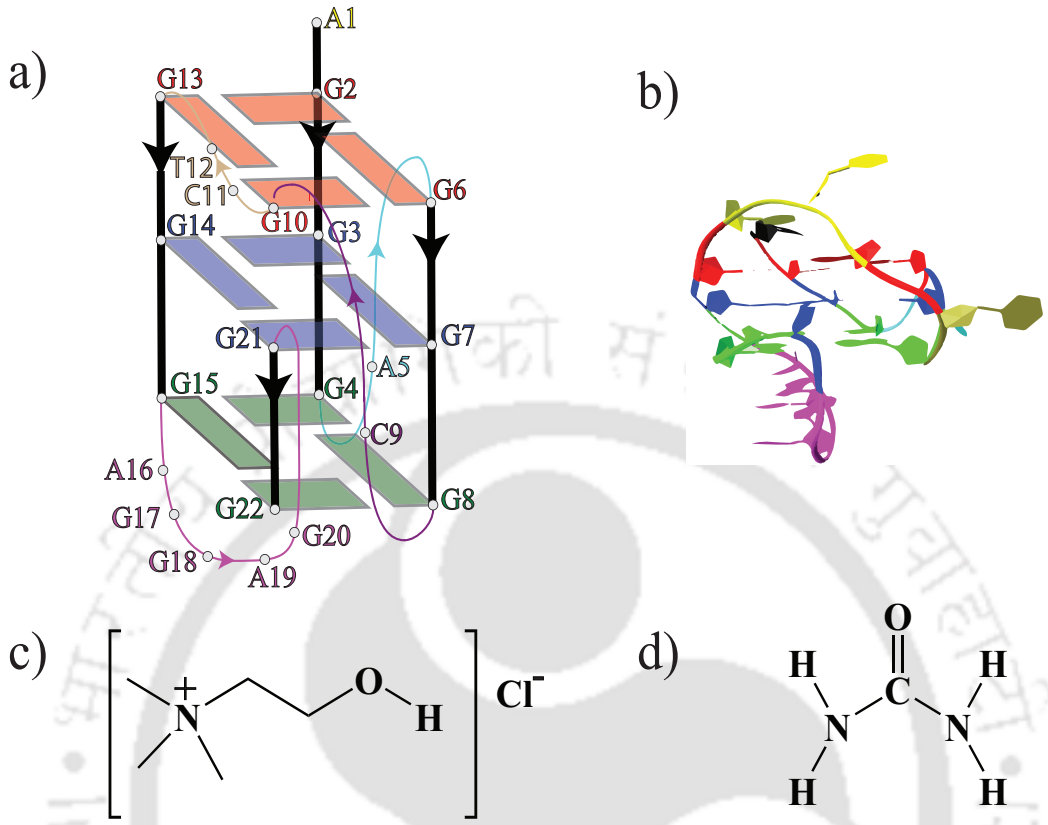


Figure 5.1: *Schematic representation of (a) c-KIT G-quadruplex DNA, new ribbons representation of (b) c-KIT G-quadruplex DNA, structural formula of (c) Choline chloride and (d) Urea molecules. In new ribbons representation, Tetrad-1, tetrad-2, tetrad-3, loop-1, loop-2, loop-3 and loop-4 are shown in red, blue, green, cyan, tan, yellow, and magenta colors respectively. Adenine, Guanine, Thymine and Cytosine are represented as A, G, T and C respectively.*

For the past two decades, ionic liquids have received the attention of researchers due to their wide range of applications, especially as the next-generation green solvent medium. However, the utilization of ionic liquids (ILs) in the industrial field is very limited due to their high cost. Therefore, a cheap and eco-friendly new class of green solvent analogues to the ILs, namely deep eutectic solvents (DESs) has been the spotlight of the scientific world. DESs are the mixture of a salt and organic compound, which have a significantly lower melting point than the corresponding pure compounds. Most of the DESs share similar features such as chemical, thermal and electro-chemical stability along with low vapor pressure with ILs [25]. DESs are used in numerous biological applications such as in biocatalytic reaction [26], protein stabilization [27–29], biosensors [30], as a reaction medium, because of the aforementioned unique characteristics.

Among many DESs, the mixture of choline chloride (salt) and urea (organic compound), also known as reline, is most popular as a solvent medium [31] as it is (i) cheap, (ii) easy to prepare, (iii) chemically inert with water [32], (iv) non-toxic [33], (iv) biodegradable [34], (v) biocompatible [35], and (vi) biorenewable [36]. In 2003, Abbott et al. discovered that when choline chloride (hydrogen bond acceptor) and urea (hydrogen bond donor) are mixed at 1:2 molar ratio, it forms eutectic mixture which has a lower melting point (i.e., 12 °C) than its parent compounds (the melting points of urea and choline chloride are 302 °C and 133 °C respectively) and is also applicable in various fields [37, 38].

In 2010, Vijayaraghavan et al. experimentally reported [39] that the double-helical B-DNA structure retained its native structure in choline lactate (CL) IL at different temperatures for a long time. In the same year, Hud and his coworker reported that there are four different types of nucleic acids such as DNA-duplex , RNA-duplex, oligonucleotide and G-quadruplex DNA which are stable in anhydrous reline (DES) medium [40]. Later they have also demonstrated the folding/unfolding kinetics of the G-quadruplex DNA in reline medium [41]. In 2015, another pioneer work has been done by them, which showed that DNA nano-structure such as DNA origami and a DNA tail system can be folded and are stable in anhydrous "glycholine" (mixture of choline chloride and urea at molar ratio of 1:4) [42]. Zhao et al. experimentally found [43, 44] that ten different G-quadruplex DNA's can be formed intra-molecular, inter-molecular, and higher-order G-quadruplex structure in reline (DES) medium and some of them such as tetrahymena telomeric, c-MYC, c-KIT, KRAS and A_4G_6 are "ultrastable" beyond 95 °C in reline DES medium. In contrast, other important studies [45–49] also suggested that DES is a good choice in comparison with water, as a nucleic acid storage. It is to be mentioned that there is no such direct evidence about the conformational dynamics of G-quadruplex in DES medium at high temperatures [50]. Additionally, the reason behind the stability of G-quadruplex DNA in reline medium at high temperature is not elucidated yet.

To overcome this issue and also to understand the conformational dynamics of G-quadruplex DNA at elevated temperatures, we have performed all-atom molecular dynamics simulation of c-KIT G-quadruplex DNA in reline medium. In our previous work (Chapter 4) [52], we have shown that in anhydrous and hydrated high reline concentration, thrombin binding G-quadruplex aptamer (TBA) is more rigid as compared to the lower concentration of reline. In this work, we examine the change of conformational dynamics of c-KIT DNA as temperature changes from 300 K to 500 K temperature at 50 K interval.

5.2 Models and Simulation Method

A set of all-atom molecular dynamics (MD) simulation was carried out using AMBER18 [53] simulation package. In this work, human c-KIT oncogene 5′-d(AGGGAGGGCGCTGGGAGGAGGG)-3′ was taken in a binary mixture of reline deep eutectic solvent (DES). Five different temperatures ranging from 300 K to 500 K at 1 atm physical pressure were considered. The initial coordinates of the human c-KIT oncogene were taken from the NMR structure of PDB ID:2O3M [21]. Earlier studies established that the central cations increase the stability of the G-quadruplex DNA [54–57]. Moreover, Bergues-Pupo et. al. observed that in absence of central cations, the structure of G-quadruplex DNA is very unstable or deviate much more than its native structure in water medium [58]. In this work, our aim is to check the thermal stability of G-quadruplex in reline medium. Thus, initially, we had not placed any cations in the central core of tetrad channels of the c-KIT G-quadruplex DNA. Since, c-KIT NMR structure carries an electronic charge of -21e, to neutralize the system, 21 K^+ ions were added. For K^+ ions, TIP3P specific Amber-adapted Joung and Cheatham force-field [59] parameters were employed. Considering recent studies [60–68], parmOL15 [69–71] was applied for c-KIT oncogene. Namely R300-R500 systems were made using following steps: Initially, the neutralized c-KIT oncogene was placed in the center of a cubic box and then 250 choline chloride and 500 urea molecules were added randomly to solvate the DNA system. For every system, initial structure was constructed by PACKMOL [72] package. Here, it is worth mentioning

Table 5.1: *Number of G-quadruplex DNA, urea, choline chloride (ChCl) and water molecules used in different systems.*

System	Number Of Molecules				Reline conc. (wt %)	Temperature (K)
	Q-DNA	Urea	ChCl	Water		
R300	1	500	250	0	100	300
R350	1	500	250	0	100	350
R400	1	500	250	0	100	400
R450	1	500	250	0	100	450
R500	1	500	250	0	100	500
W450	1	0	0	5000	0	450

that, to compare the stability of c-KIT oncogene between reline (DES) and water medium, another system (namely W450 system) was also built, where the neutralized c-KIT onco-

gene was solvated by 5000 TIP3P [73] water molecules. After that, initial coordinates and topology for each system, were converted into truncated octahedron box by CPPTRAJ [74]. In **Table 5.1**, we have mentioned the details of all the systems.

The geometries of isolated choline cation and urea molecule were optimized using abinitio HF/6-31G* [75, 76] level of theory by Gaussian 09 [77] package. For the partial charges of choline cation and urea molecule, restrained electrostatic potential (RESP) [78, 79] method was employed. For this study, the general amber force field (GAFF) [80] parameters, which were previously reported by Perkins et al., were applied [81, 82] for choline chloride and urea molecules. It should be mentioned here that General AMBER Force Field (GAFF) can precisely predict many ILs' as well as DESs' thermodynamic and transportation properties [83]. Electrostatic potential for IL or DES simulations is generally overestimated. To overcome this issue, a well-known technique [84–86] is often used to regulate electrostatic interactions that includes a certain factor in the reduction of ionic charges. In our study, we reduce the partial charges of choline chloride by 20%, as reported by previous studies [81, 82, 86, 87].

The following steps were implemented for each system: (i) At first, the energy was minimized for 5000 steps via steepest descent method followed by an equal number of steps in conjugate gradient method to eliminate bad contacts between quadruplex c-KIT DNA and reline solvent. (ii) Next, the system was heated gradually from 0 K to the desired temperature at an interim of 50 K for 25 ps in canonical ensemble (NVT) with a collision frequency of 1 ps^{-1}. (iii) Then, the system was equilibrated for 250 ps at the desired temperature in canonical ensemble. During these steps, all atoms of the c-KIT G-quadruplex DNA were fixed with a harmonic restraint of force constant 100 kcal mol^{-1} $Å^{-2}$, to their energy-minimized configurations. (iv) After that, a similar sequence of minimization and equilibration (NVT) were followed with a reduced positional restraint force of 25, 20, 15, 10, 5 and eventually 0 kcal mol^{-1} $Å^{-2}$. (v) Finally, to maintain the density, a 20 ns NPT equilibration was performed and this is followed by 2 μs production run with a 5 ps^{-1} collision frequency at coveted temperature and 1 atm physical pressure.

During this step, the temperature regulation was guided by Langevin dynamics [88]. To control the physical pressure Berendsen barostat [89–91] with a pressure relaxation time of 2 ps was employed. SHAKE algorithm [92] was implemented with a tolerance of 10^{-5} Å to restrain the covalent bonds involving hydrogen atoms. A 2 fs time integration step was applied. To calculate the long-ranged electrostatic interactions, particle mesh Ewald method [93] was employed. A cut off radius of 12 Å was used for all non bonded

short-ranged interaction. Similar methods was applied in our previous works [51, 52]. All the trajectories were analyzed by the CPPTRAJ [74] of Amber-Tools19 [53] and Visual Molecular Dynamics (VMD) [94], which was also used as a visualization tool. For all torsional angles and spatial density maps, in house python code inspired by Do_X3DNA [95] and TRAVIS [96] were used respectively.

We performed the principal component analysis (PCA) on 131 backbone atoms of G-quadruplex c-KIT DNA with the help of CPPTRAJ [74] of AMBER18 [53] package.

5.3 Results and Discussions

5.3.1 Structural Changes of c-KIT G-quadruplex DNA at Different Temperatures

5.3.1.1 Root Mean Square Deviations (RMSDs)

The root mean square deviation is a very important measurement to monitor the structural deviation of protein and DNA from its native state or a reference state. Previous studies [97, 98] have shown that the RMSD can be used as a reaction coordinate in folding and unfolding of the DNA. In **Fig. 5.2**, we have presented the rms fit RMSD values of G-quadruplex c-KIT DNA for different systems (namely systems R300 to R500 and W450). It is worth mentioning that the RMSD was calculated by considering all the heavy atoms of the G-quadruplex DNA, where the NMR structure was taken as a reference. We observe that for R300-R450 systems, the RMSD values of c-KIT DNA varies from 2 Å to 3 Å.

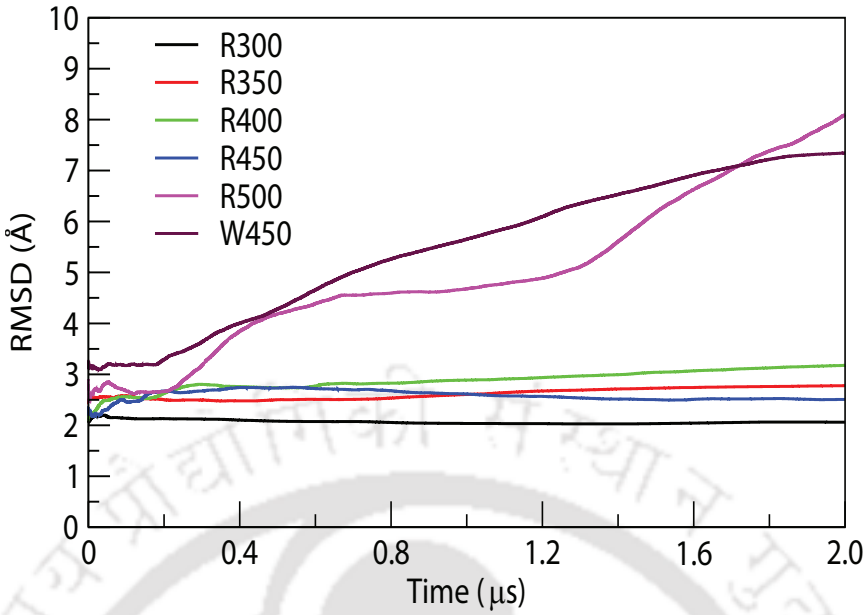


Figure 5.2: *Time progression of the root mean square deviations (RMSDs) of all heavy atoms of c-KIT G-quadruplex DNA for different systems.*

From this observation, we can predict that, for those particular systems, the c-KIT DNA maintains its native state at a temperature range of 300 K to 450 K. But, at 500 K (R500 system), the drastic change of RMSD value (8.37 Å) indicates that G-quadruplex c-KIT DNA deviates to a great extent from its native state after 200 ns. Moreover, we see a similar scenario for W450 system in which the G-quadruplex DNA is immersed in pure water at 450 K temperature. To gather a thorough information, we have divided the overall DNA into three regions, i.e., (i) sugar phosphate backbone region, (ii) tetrad and (iii) loop regions. From **Fig. 5.3 (a)-(c)**, we have observed that among these three regions, the loop regions are more flexible in all the systems. Specifically, the RMSD values of sugar phosphate backbone varies from 1.08 Å to 2.08 Å and is very similar to each other for R300-R450 systems. For tetrad and loop regions, the RMSD values varies from 0.97 Å to 1.83 Å and 2.52 Å to 3.71 Å respectively for R300-R450 systems. Moreover, for R500 and W450 systems, after 200 ns, the G-quadruplex c-KIT DNA starts to unfold. It suggests that G-quadruplex c-KIT DNA is stable up to 450 K in reline medium but it unfolds at 450 K in pure water medium.

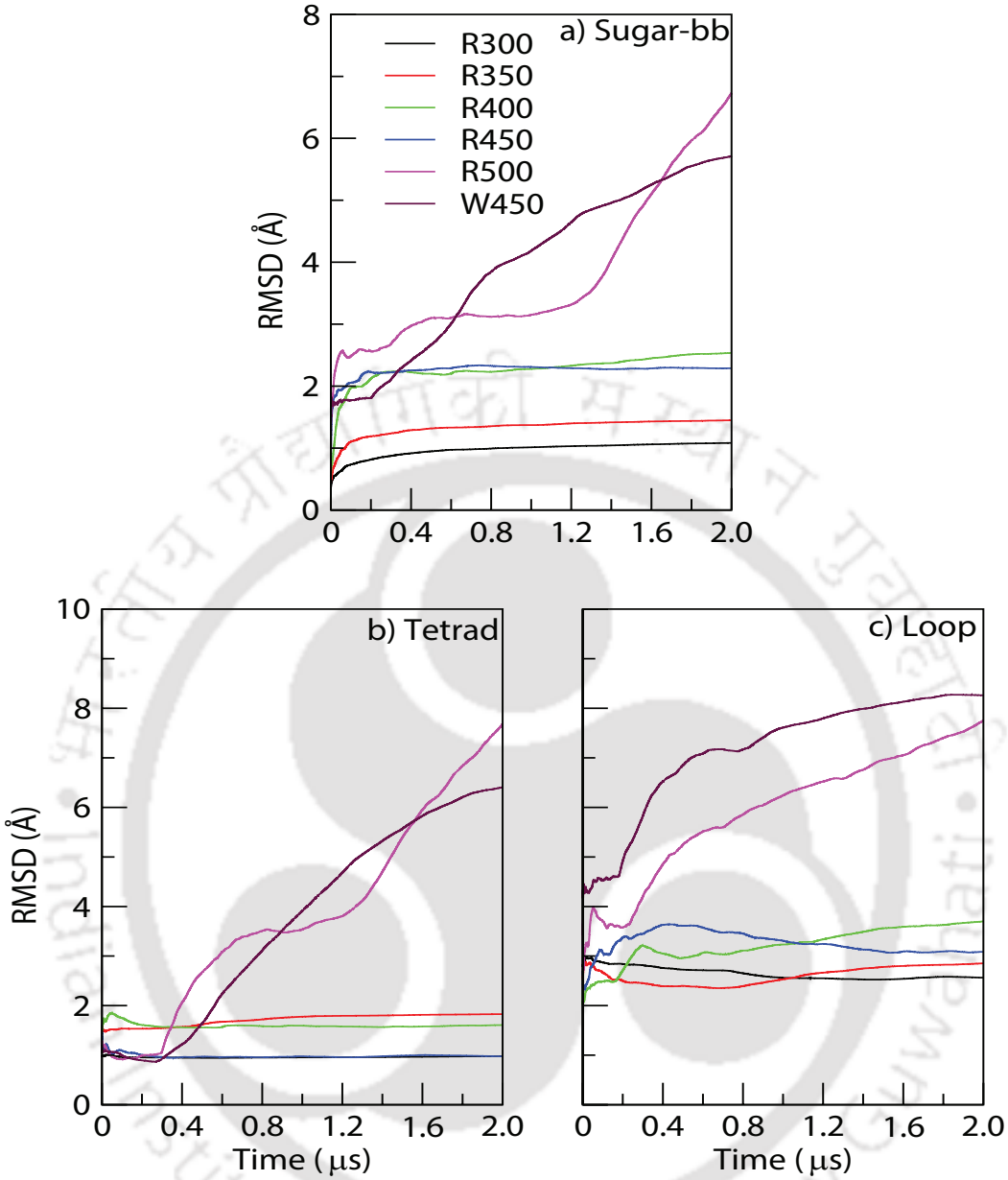


Figure 5.3: *Time progression of the root mean square deviations (RMSDs) of all heavy atoms of: (a) sugar-phosphate backbone, (b) tetrad region and (c) loop region in different reline concentrations.*

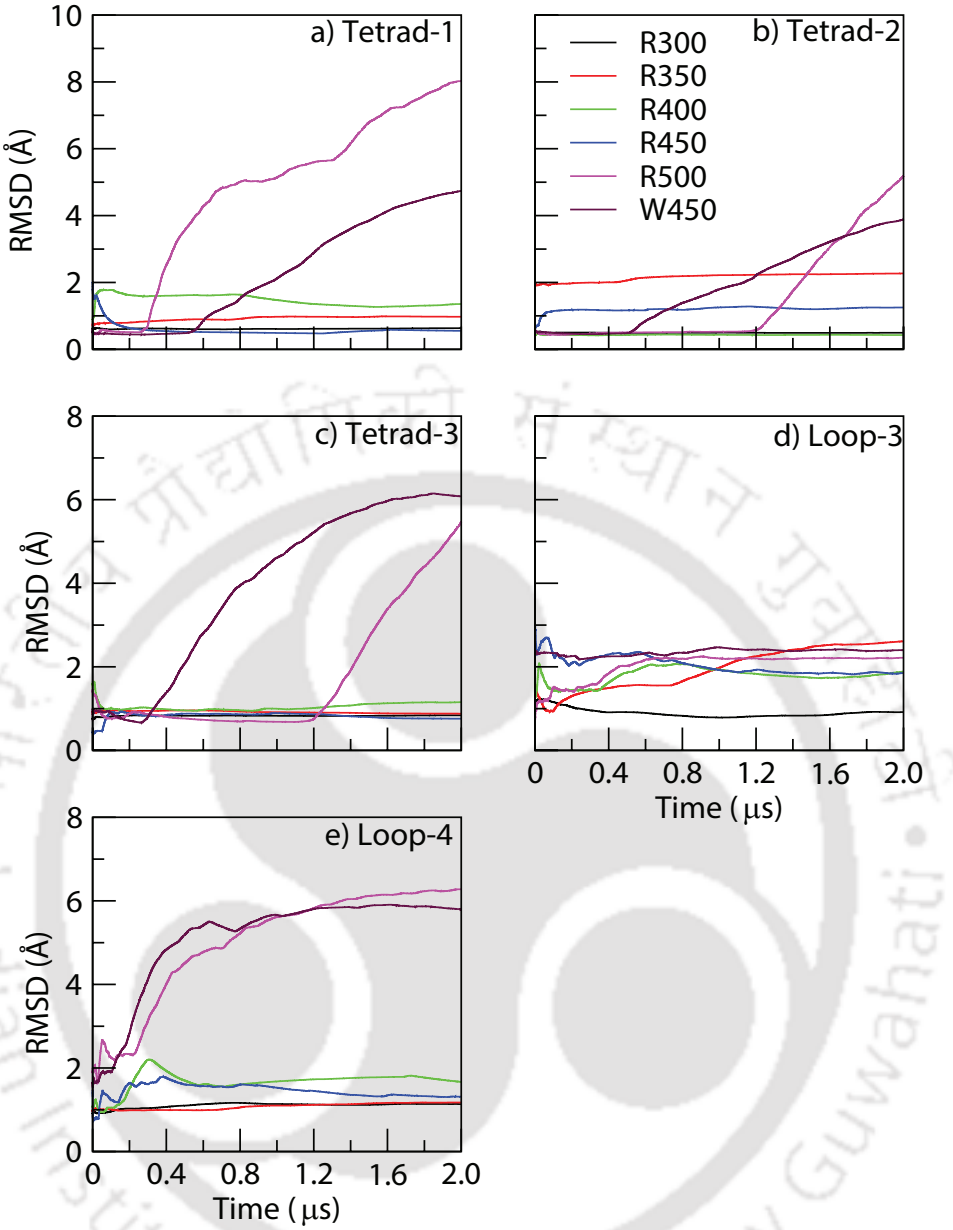


Figure 5.4: *Considering all heavy atoms RMSD of the (a-c) tetrad-1, tetrad-2, tetrad-3 and (d-e) loop-3 and loop-4 regions of c-KIT G-quadruplex DNA with respect to the NMR structure in simulations for different systems.*

In **Fig. 5.4**, we have shown the changes in tetrad-1, tetrad-2, tetrad-3, loop-3 and loop-4 regions for different systems. We have also estimated the pairwise RMSD (**Fig. 5.5**) of all the heavy atoms of c-KIT DNA over a period of 2 μs.

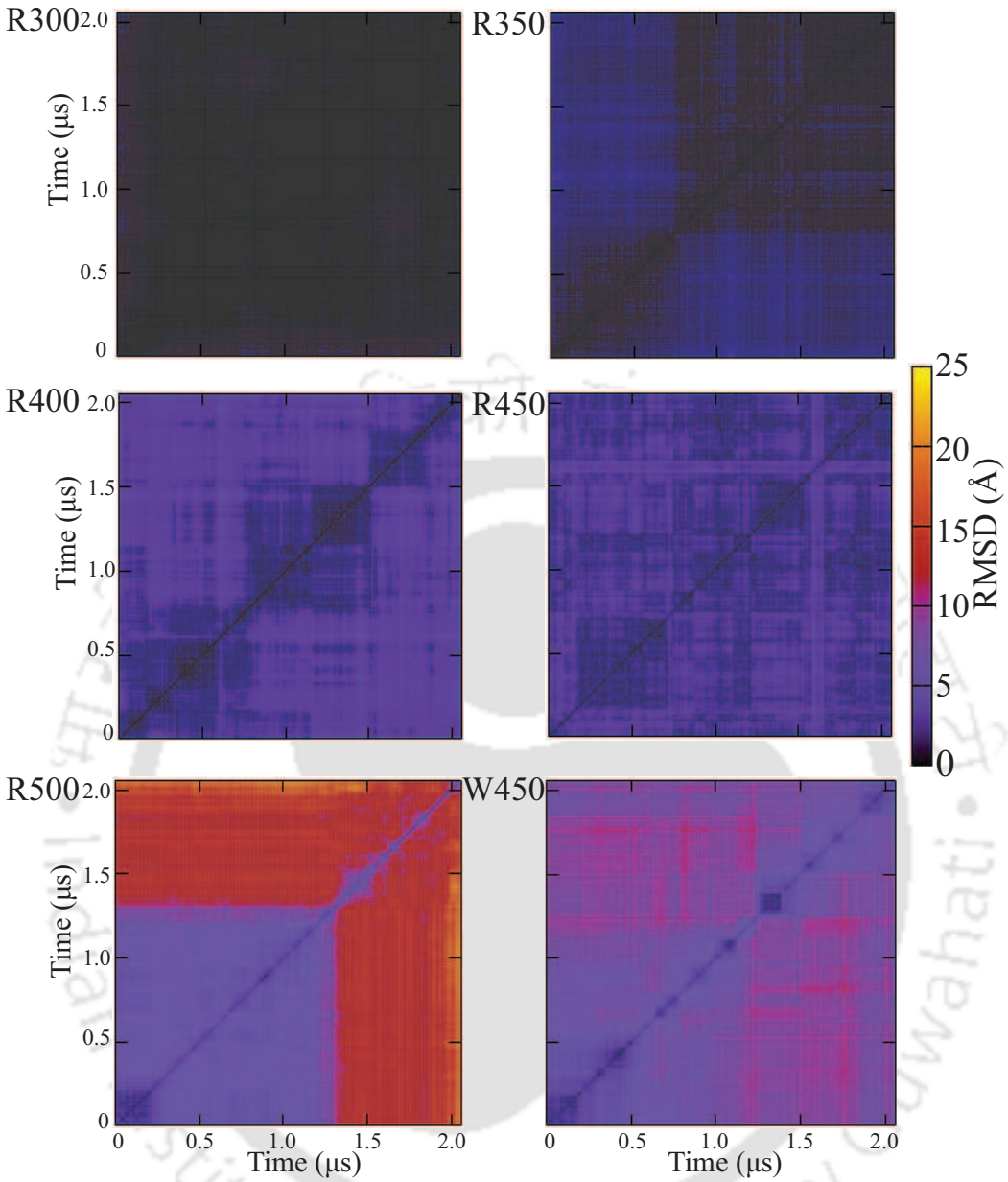


Figure 5.5: *Taking into account all heavy atoms of c-KIT G-quadruplex DNA, pairwise 2D-RMSDs (pair-wise) of different systems.*

5.3.1.2 Root Mean Square Fluctuations (RMSFs)

To perceive the conformational behavior of a specific residue of G-quadruplex c-KIT DNA, we need to perform root mean square fluctuation (RMSF) analysis. From the RMSF values of the specified residues, we can identify the residues responsible for fluctuations. In **Fig. 5.6**, the RMSFs of G-quadruplex c-KIT DNA for all systems is presented. We have noticed that mainly the loop regions (namely, loop-1, loop-2, loop-3 and loop-4) fluctuate more than the tetrad regions for R300-R450 systems. Precisely, Adenine-5 (loop-1), Cytosine-9 (loop-

2), Cytosine-11 (loop-3) and Guanine-18 (loop-4) fluctuate with increasing temperature (300 K to 450 K), but drastic fluctuations are noticed for all residues of G-quadruplex c-KIT DNA at 500 K in reline medium and also at 450 K in water medium.

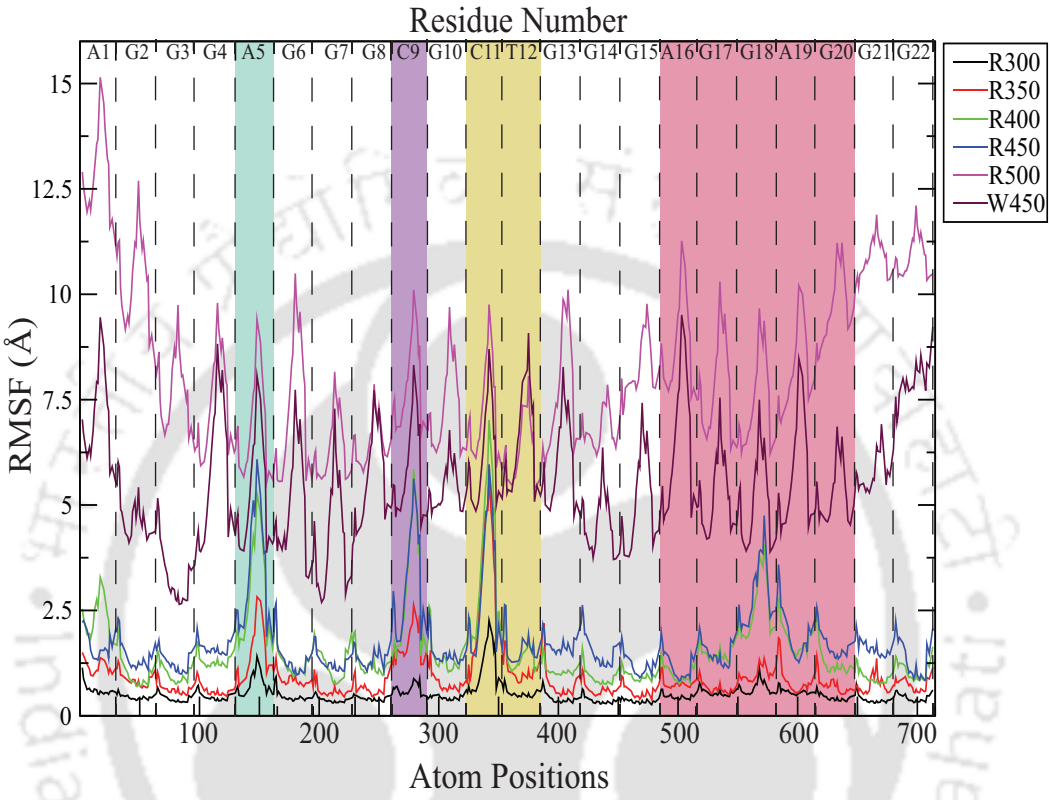


Figure 5.6: *Root mean square fluctuations (RMSFs) of all heavy atoms of c-KIT G-quadruplex DNA for different systems. Loop-1, loop-2, loop-3 and loop-4 regions are shown in cyan, purple, yellow and magenta colors respectively. Adenine, Guanine, Thymine and Cytosine are represented as A, G, T and C respectively.*

5.3.1.3 Radius of Gyrations (R_gs)

Radius of gyration is an essential method for checking the compactness of biomacromolecules [99, 100]. The radius of gyration (R_g) is the root mean square distance of various atoms of a molecule from the axis of rotation. In **Fig. 5.7**, we have shown the estimated R_g of the G-quadruplex c-KIT DNA considering the O6 atom of guanine residue of tetrads region. Here, we have noticed that the R_g values differ from 3.74 Å to 4.61 Å for R300-R450 systems. It indicates that the G-quadruplex c-KIT DNA maintains its compactness when the temperature of the system is changed from 300 K to 450 K in reline medium. However, extreme change in R_g values are observed for systems R500 (10.93 Å) and W450 (8.65 Å).

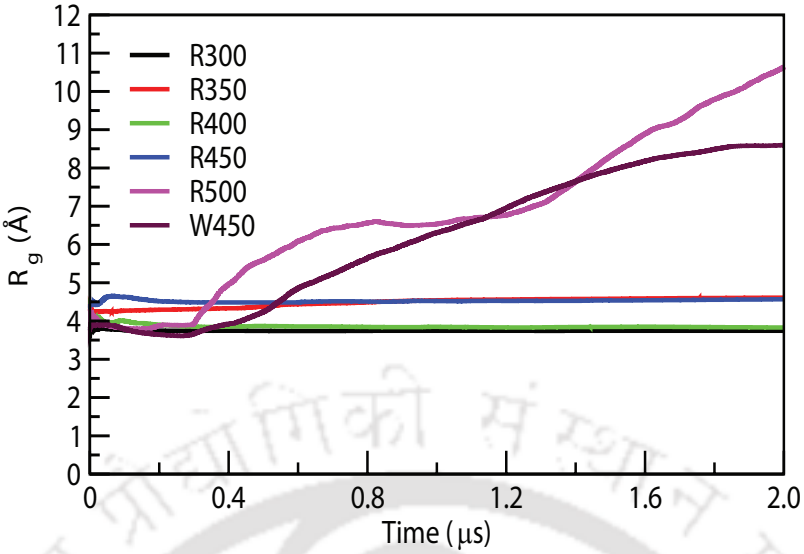


Figure 5.7: *Time evolution of the R_gs of O6 atom of guanine residues of all tetrads for different systems.*

5.3.1.4 Principal Component Analysis

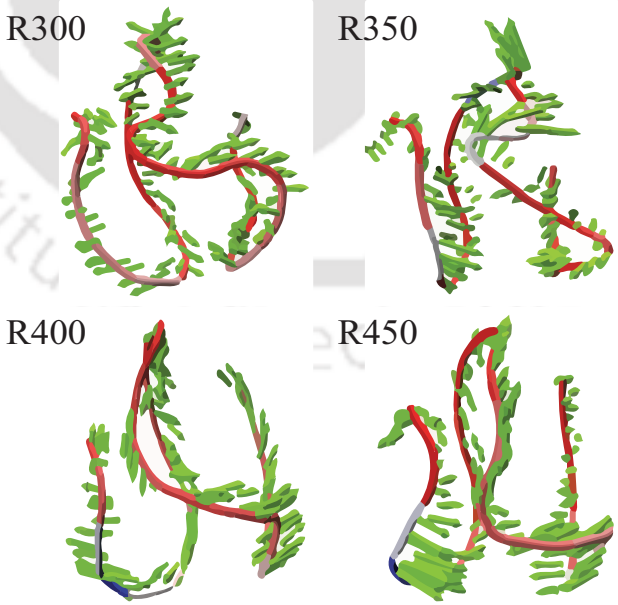


Figure 5.8: *Showing the dominant motions of c-KIT G-quadruplex DNA along with the first eigenvector in porcupine plot for different systems. The models are shown as a backbone trace. The arrows attached to each backbone atom indicate the direction of the eigenvector and the size of each arrow shows the magnitude of the corresponding eigenvalue.*

In **Fig. 5.8**, we present the dominant modes for R300-R450 systems along with the first eigenvector. Here, we notice that the motions of the backbone atoms increase with increasing temperature (R300-R450 systems). The cumulative fraction of eigenvalues of the first 50 eigenvectors are shown in **Fig. 5.9**. In this **Fig. 5.9**, we see

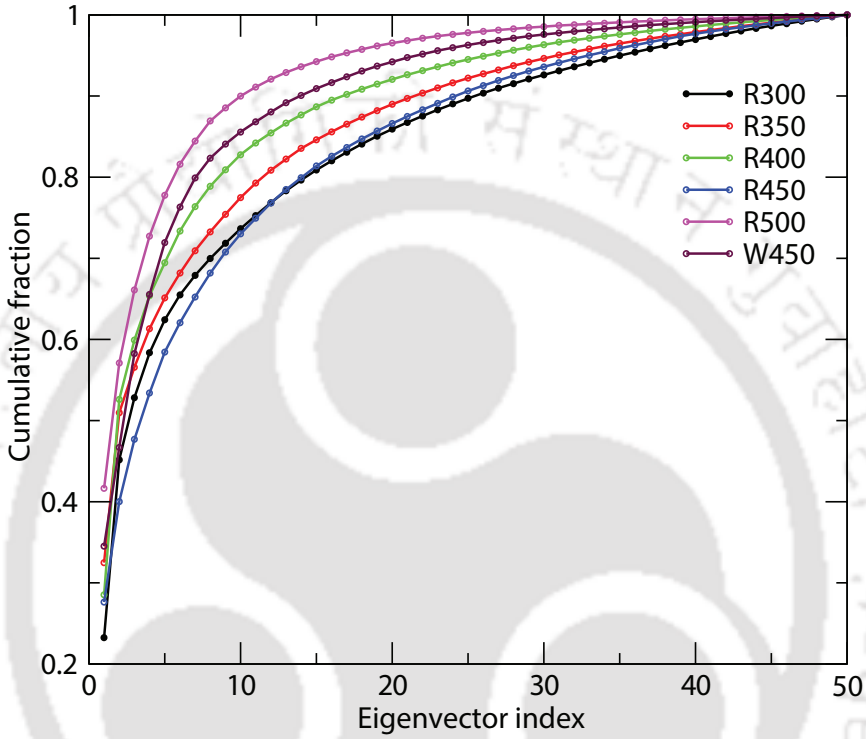


Figure 5.9: *Cumulative fraction of the eigenvalues taken from principal component analysis (PCA) for G-quadruplex-c-KIT DNA for different systems.*

that the majority of conformational changes i.e., more than 50% is identified by the first three eigenvectors. Moreover, the concerted motions are mostly contributed by the first principal component (PC1) for the systems i.e., 23%, 32%, 28%, 27% and 41% respectively for R300-R500 systems and 34% for W450 system. **Fig. 5.9** also shows that the overall motion (≈100%) is made by the first 50 eigenvalues. To explore the conformational changes of G-quadruplex c-KIT DNA, we show the projection of the Cartesian coordinates along with first three principal components (namely PC1, PC2 and PC3) in **Fig. 5.10**.

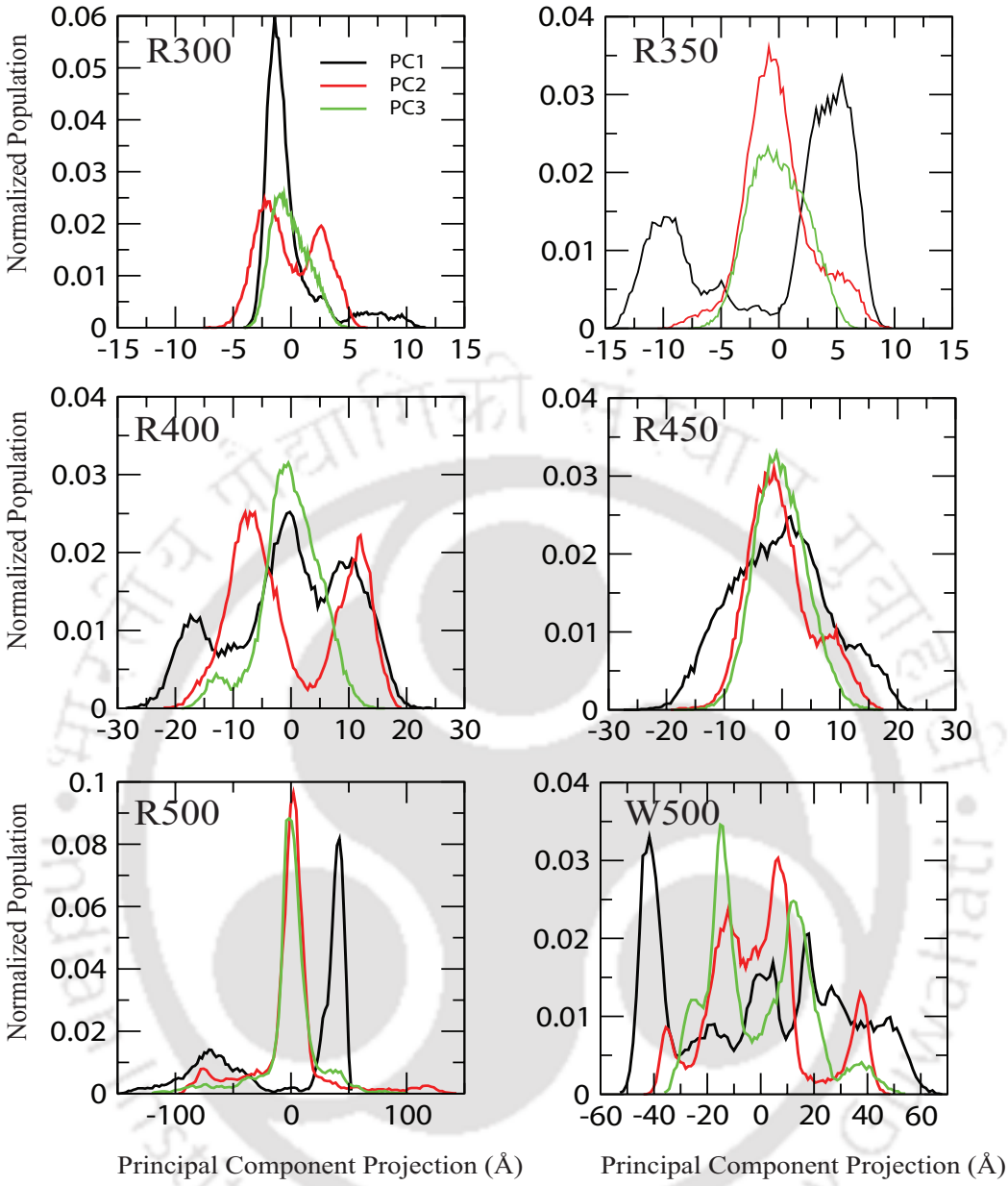


Figure 5.10: *Considering first three principal components (PCs) , the projection of the Cartesian coordinates for G-quadruplex-c-KIT DNA for different systems.*

It implies that the principal component projection of the trajectory is increased with increasing temperature. For further investigation, we have projected two principal components i.e., PC1 and PC2 in terms of free energy change following the equation:

$$\Delta G_i = -k_B T \left[\ln P_i - \ln P_{\max}\right] \tag{5.1}$$

In **Fig. 5.11**, we have shown the projection of two parameters i.e., PC1 and PC2 to formulate the free energy landscapes for all systems. Here, for R350 and R400 systems,

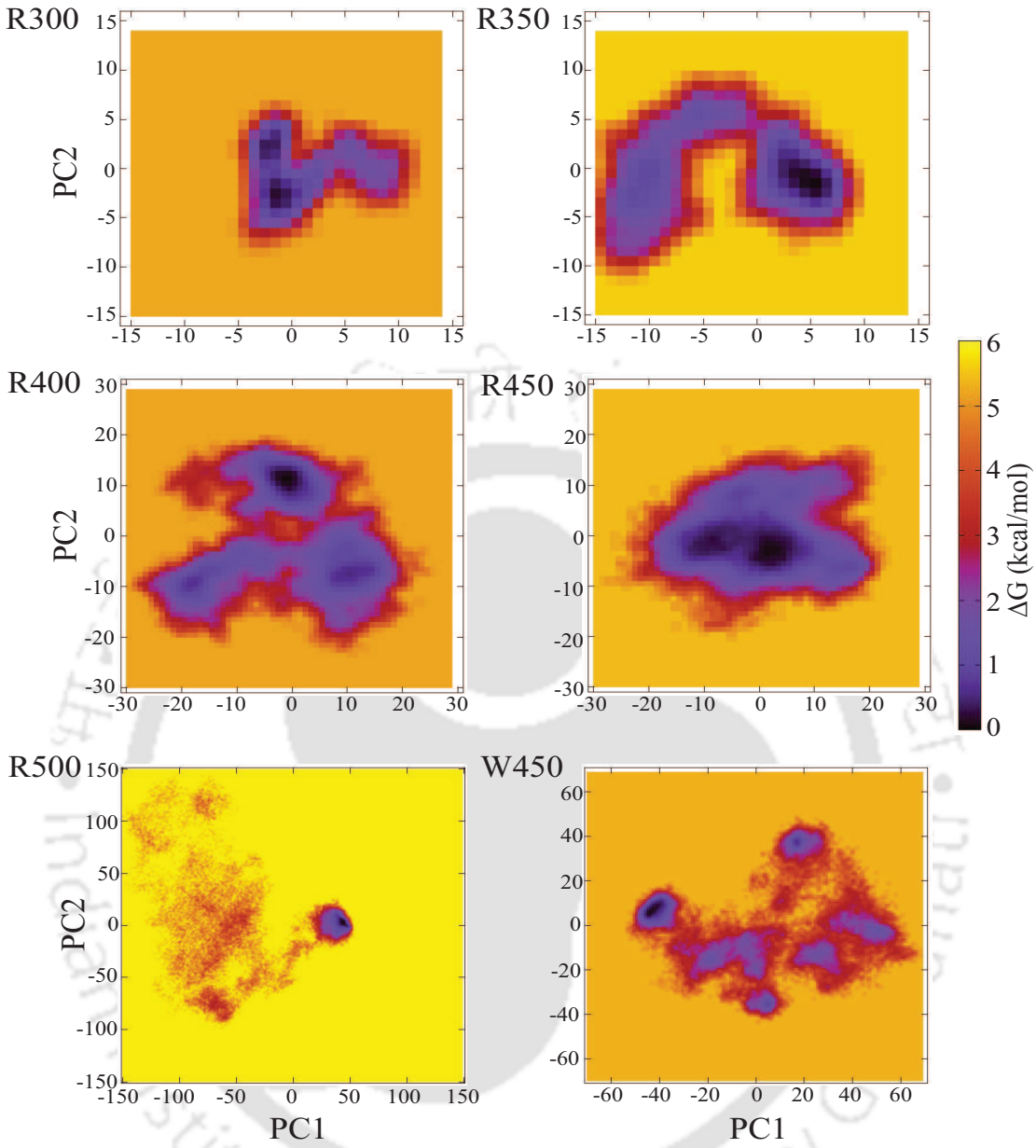


Figure 5.11: *Free energy landscapes (FELs) plotted as a function of principal components of PC1 and PC2 for the backbone atoms of G-quadruplex-c-KIT DNA for different systems.*

the free energy basin/well has appeared at PC2 value of 0 and PC1 value of -5 which implies that the most populated conformational structure maintains its native state. Moreover, with increasing temperature, the eigenvector of projection PC1 and PC2 is slightly increased but remains close to the native state. However, an extreme change is observed at 500 K in reline medium (R500 system) and at 450 K (W450 system) in pure water medium. Here, free energy well/basin is observed at 50 eigenvector of PC1 for R500 system and -40 eigenvector of PC1 for W450 system, which depicts that the most populated state for those systems are far from the native state.

From RMSD analysis, it is found that the structure of c-KIT DNA is similar to its native state as temperature increases from 300 K to 450 K in reline medium. However, it unfolds at 500 K in reline medium as well as at 450 K in water medium. RMSF values indicate that with increasing temperature, the loop regions not only fluctuate more than the tetrad but are also more flexible. These results correspond to the calculated RMSD values discussed above. Moreover, R_g values suggest that the compactness of G-quadruplex c-KIT DNA is lost when the temperature is increased to 500 K (for R500 system) in reline and 450 K in water medium (for W450 system).

5.3.1.5 Torsional Angle Analysis

Considering these studies [101, 102], it is very important to check flexibility of c-KIT G-quadruplex DNA using torsional angle analysis. We have used four torsional angle parameters i.e., (i) backbone torsional angles (α/γ), (ii) backbone torsional angles (β/δ), (iii) backbone torsional angles (ϵ/ζ), and, (iv) glycosidic torsional angle (χ) to calculate the torsional angles of c-KIT G-quadruplex DNA. In the wheel representation, we have shown the torsional angles of all the residues (**Figures 5.12-5.16**). The torsional angle distributions were clustered at a resolution of 15° with corresponding probabilities.

The α torsional angles of most of the residues of tetrad regions fall in synclinal ($-$sc 270$-$360°) or gauche (g$-$) region or close to g$-$ regions i.e., anticlinal ($-$ac 210$-$270°) except for G-6 of R450 systems. Moreover, it depicts that the α backbone torsional angles are very similar or close to the NMR structure. On the other hand, the γ torsional angles of most of the residues have been found in the +sc (30$-$90°) or g(+) except of G-6 (γ torsional angles fall in g$-$ region) of R300 and R350 systems and for G-8 (trans (t) region) of R300 systems. Overall, most of the α/γ torsional angles are found in g$-$/g+ regions for tetrad residues which is very similar to the NMR reference structure, although the probability decreases with increasing temperature. Similarly, in case of loop regions, α/γ torsional angles fall in g$-$/g+ regions or very close to a$-$/a+ regions, although there are some residues like A-19 and G-22 where α torsional angles are observed in t regions (for R300 system).

Now, another backbone torsional angles, which are the β/δ torsional angles, are being taken into account. The β torsional angles are found in t regions for most of the residues and it is very much similar to the NMR structure. Nonetheless, for some of the residues of R300 systems like G-4, A-5 (also R350 system), C-9, G-10, G-14, G-15 (also R350 system), A-19 and G-22, the β torsional angles slightly deviate from t region.

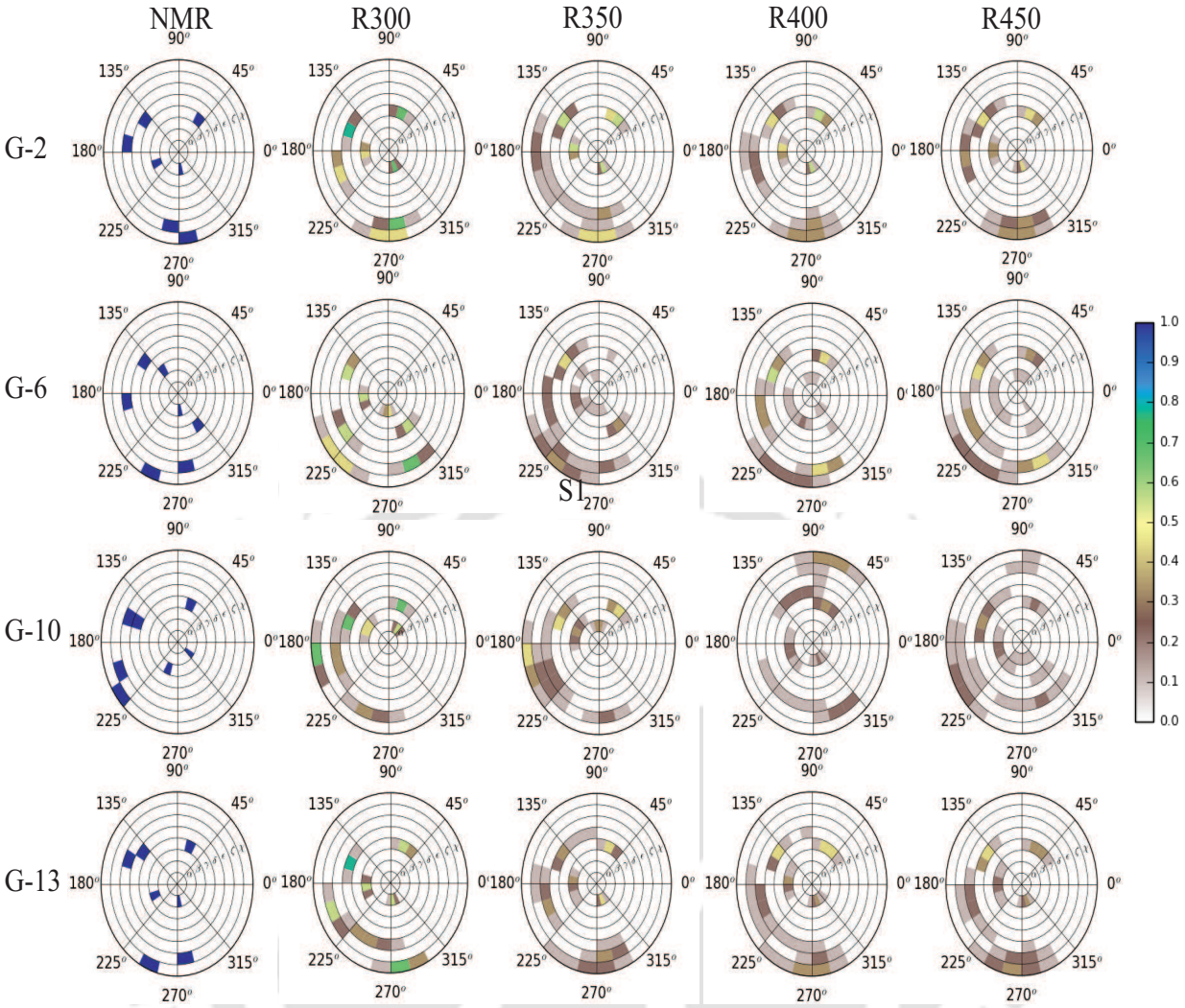


Figure 5.12: *Wheel representations of all torsion angles of tetrad-1 residues for different systems and the NMR structure. The color scale between 0 and 1 refers the probability of the corresponding angle. Guanine is represented as G . In the wheel plot, from the center to edge, α, β, γ, δ, ϵ, ζ, and χ torsional angles are depicted.*

Considering backbone torsional angles ϵ/ζ for all residues, we have found that for most of the residues, ϵ/ζ torsional angles are in the t region with BI conformer and these are very similar or close to the NMR structure. Here, it is worth mentioning that for some of the residues like G-8, G-14 and G-18, ϵ/ζ torsional angles fall in the t region with BII conformer.

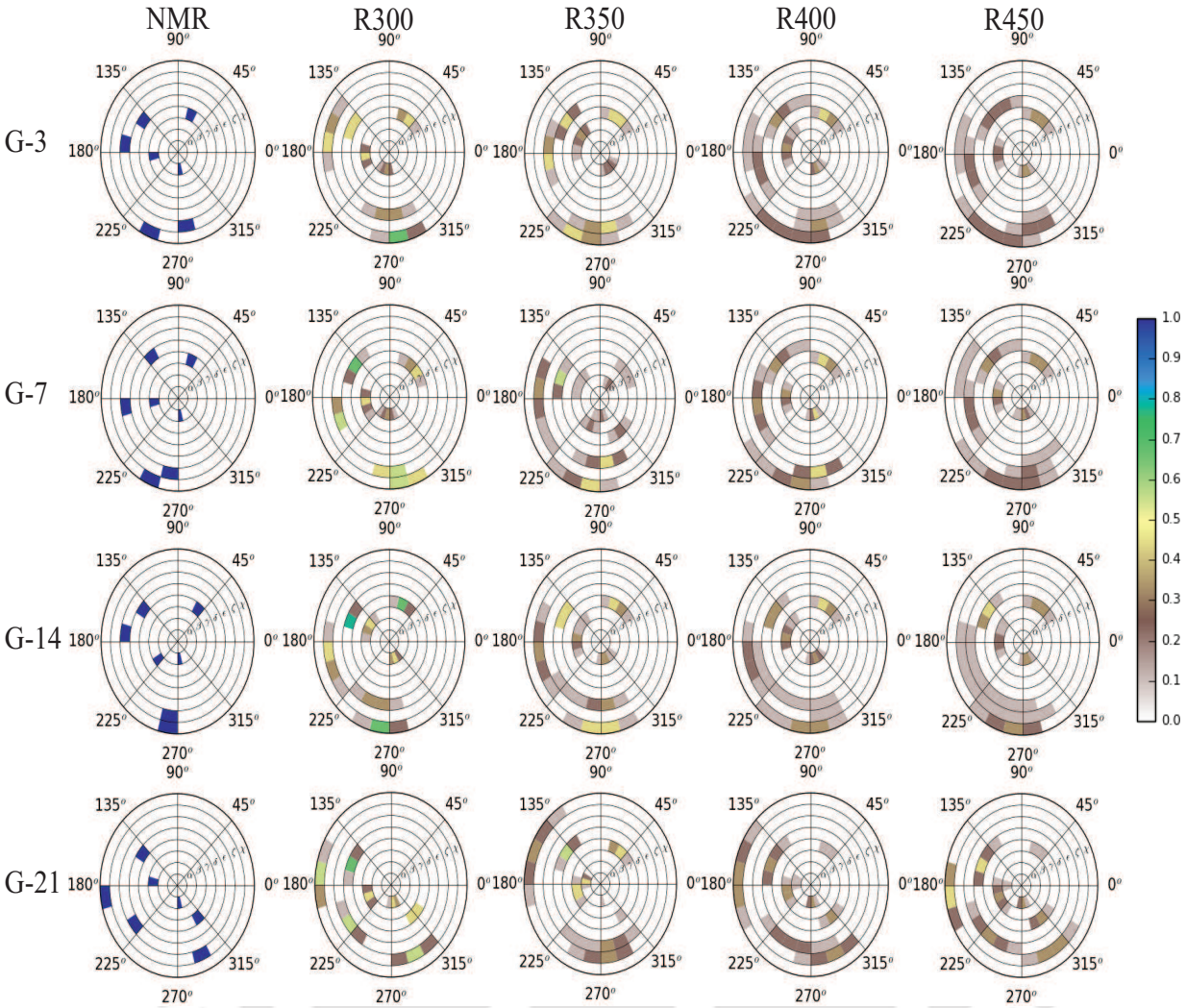


Figure 5.13: *Wheel representations of all torsion angles of tetrad-2 residues for different systems and the NMR structure. The color scale between 0 and 1 refers the probability of the corresponding angle. Guanine is represented as G . In the wheel plot, from the center to edge, α, β, γ, δ, ϵ, ζ, and χ torsional angles are depicted.*

Next, the most significant glycosidic angle, χ torsional angle, that preserves the guanine-rich DNA's structure and conformation is now considered. Here, we have noticed that for all residues the χ torsional angles fall either in a- or close to a- i.e, t region with anti conformation. Moreover, these angles are very much comparable to that of the reference NMR structure, although, there are some major exceptions like A-19 residue where χ torsional angles fall in g- region with syn conformation for R300-R450 systems.

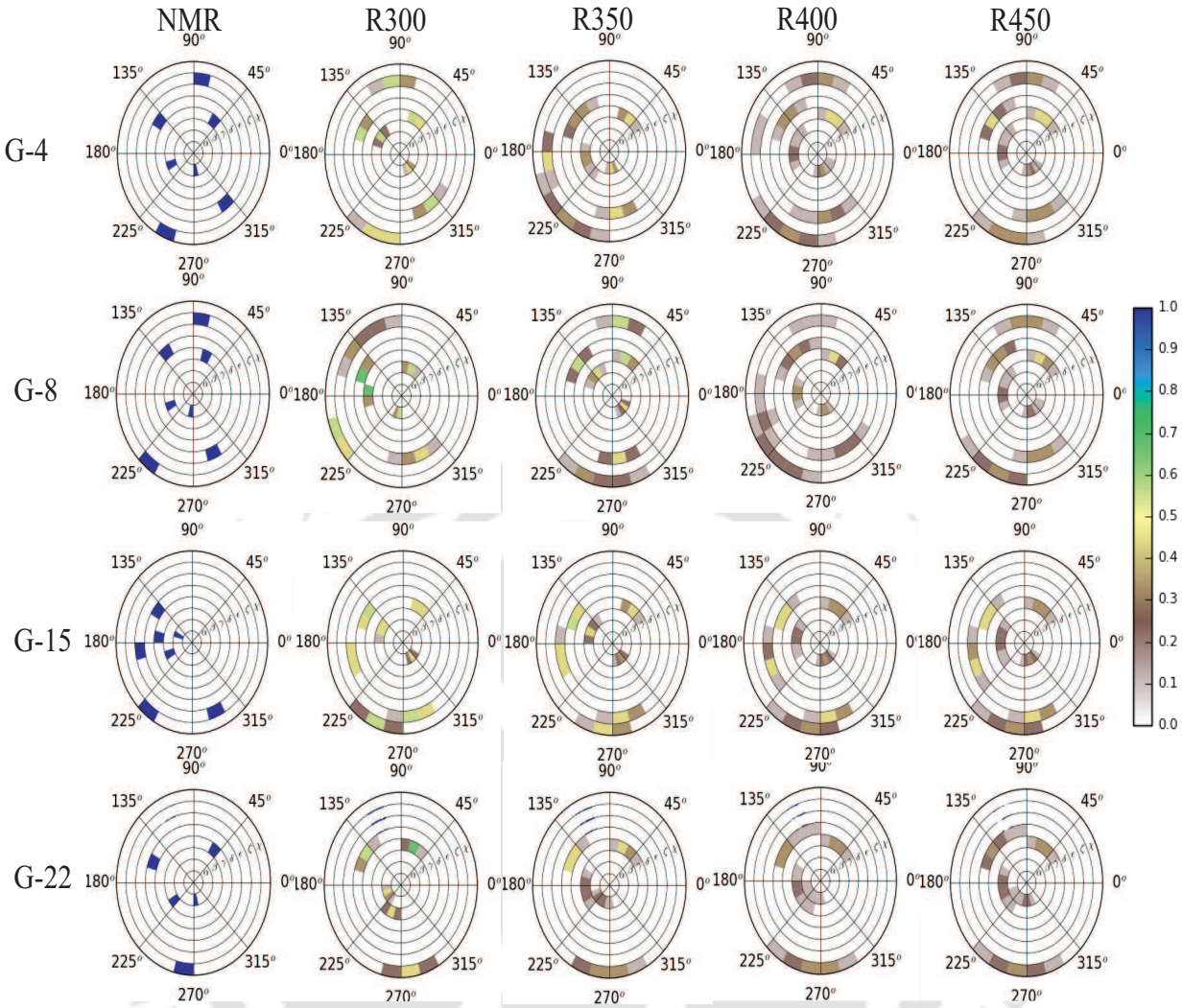


Figure 5.14: *Wheel representations of all torsion angles of tetrad-3 residues for different systems and the NMR structure. The color scale between 0 and 1 refers the probability of the corresponding angle. Guanine is represented as G . In the wheel plot, from the center to edge, α, β, γ, δ, ϵ, ζ, and χ torsional angles are depicted.*

Overall, it concludes that most of the torsional angles are very similar or very close to the reference NMR structure but the probability of a particular torsional angle decreases with increasing temperature.

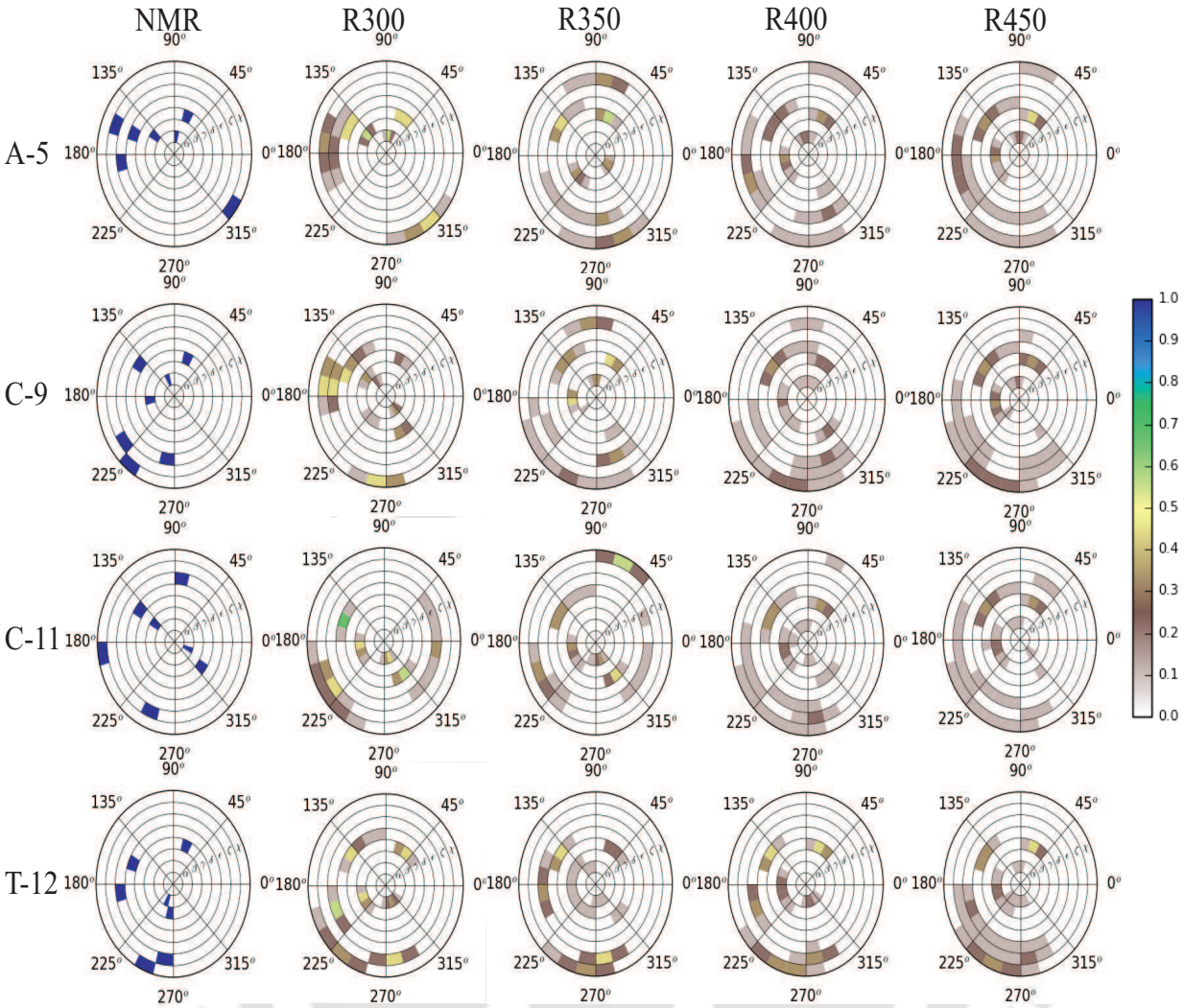


Figure 5.15: *Wheel representations of all torsion angles of loop-1, loop-2 and loop-3 residues for different systems and the NMR structure. The color scale between 0 and 1 refers the probability of the corresponding angle. Adenine, cytosine and thymine are represented as A, C, and T respectively . In the wheel plot, from the center to edge, α, β, γ, δ, ϵ, ζ, and χ torsional angles are depicted.*

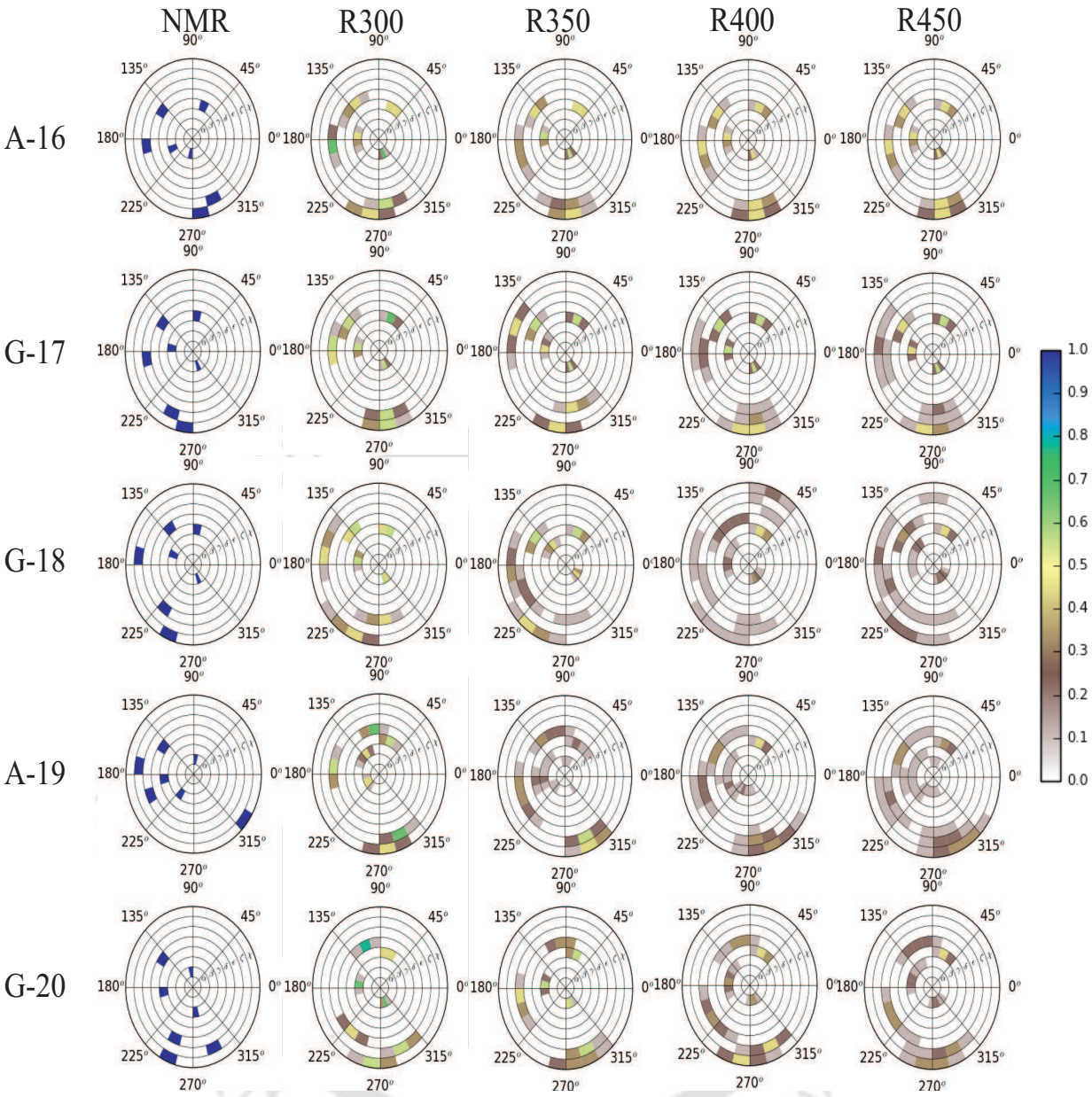


Figure 5.16: *Wheel representations of all torsion angles of loop-4 residues for different systems and the NMR structure. The color scale between 0 and 1 refers the probability of the corresponding angle. Adenine and guanine, are represented as A and G respectively. In the wheel plot, from the center to edge, α, β, γ, δ, ϵ, ζ, and χ torsional angles are depicted.*

5.3.2 Loop Dynamics of c-KIT G-quadruplex DNA at Different Temperatures

From the previous analyses such as RMSD and RMSF, it is very clear that loop-4, which consists of A-16, G-17, G-18, G-19 and G-20 residues, deviates or fluctuates more than loop-2. Moreover, from **Fig. 5.3**, it is clear that loop-3 deviates from the NMR structure in the range of 1-2.5 Å. On the other hand, in case of loop-4, the deviations are very small i.e., 1-2 Å for R300-R450 systems, but it increases significantly for R500 and W450 systems. Considering these informations, we focus the loop dynamics study on loop-4 of the c-KIT G-quadruplex DNA.

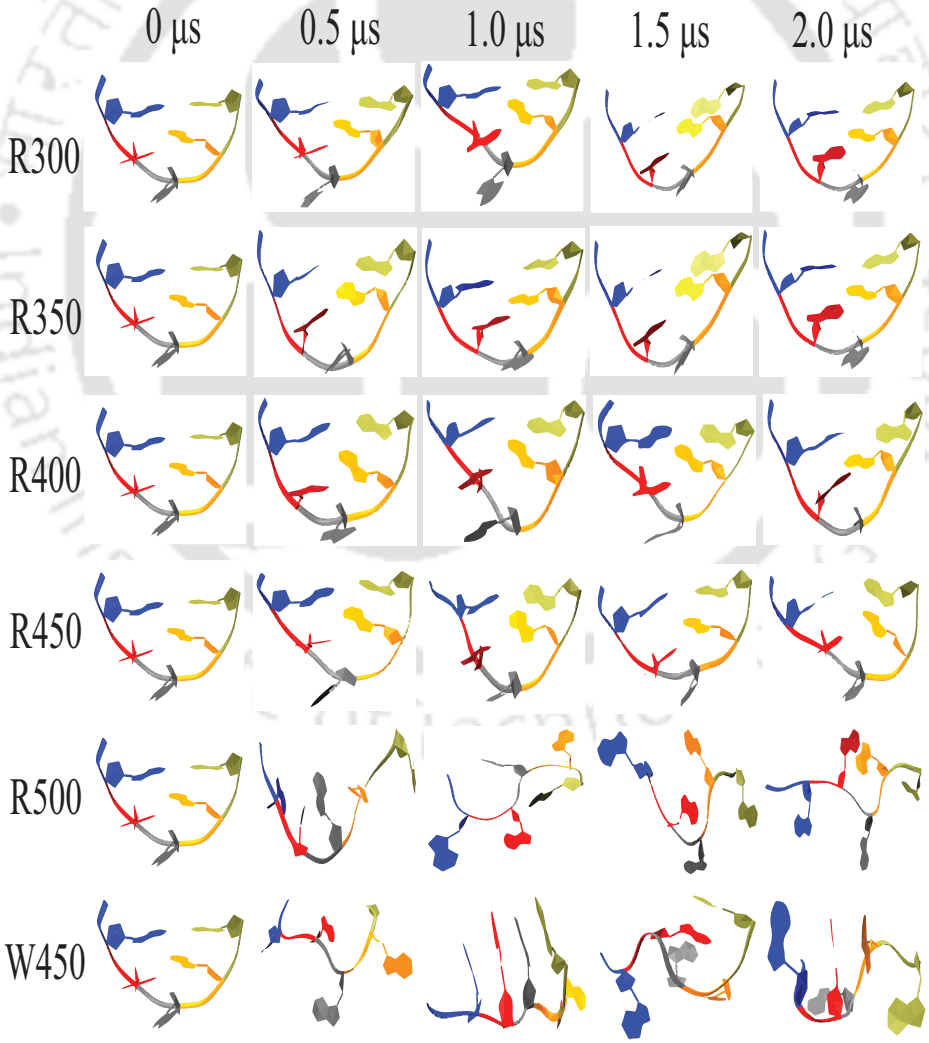


Figure 5.17: *Snapshots of loop-4 of c-KIT G-quadruplex DNA at 500 ns time interval for different systems. Here A16, G17, G18, A19 and G20 represent in blue, red , gray, orange and tan colors respectively.*

In **Fig. 5.17**, we have shown the snapshots of loop-4 with time progression. From the snapshots, it is quite clear that with increasing temperature from 300 K to 450 K in reline medium, the nucleoside bases are arranged in stacking alignments with each other. We have calculated the π-π stacking probability of nucleoside bases. Considering our previous studies as well as other pioneering works [103–106], we have set two criteria for π-π stacking interactions which are: (i) the distance between the center of mass of the two corresponding planes is $\leq$ 5.0 Å and simultaneously (ii) the angle between two corresponding planes is $\leq 20^{\circ}$. **Table 5.2** clearly indicates that A16-G17, G17-G18 and A19-G20 residues are involved π-π stacking interactions in the reference NMR structure.

Table 5.2: *Stacking distance and angle between the corresponding planes of the tetrad residues for reference NMR structure.*

Residue	Distance (Å)	Angle (°)
G2-G3	3.67	1.28
G3-G4	3.94	0.07
G6-G7	3.65	174.91
G7-G8	4.28	6.27
G10-G21	4.09	8.47
G21-G22	4.52	172.84
G13-G14	4.10	3.47
G14-G15	3.69	5.32

Fig. 5.18 depicts that the highest probability of distance and angle of loop-4 i.e., A16-G17, G17-G18 and A19-G20 nucleoside bases are within $\leq$ 5.0 Å and $\leq 20^{\circ}$ respectively for the temperature range of 300 K to 450 K in reline medium thus fulfilling both the criteria for π-π stacking interactions. However, at 500 K in reline medium and 450 K in water medium i.e., for R500 and W450 systems, no π-π stacking interactions are noticed. This observation concludes that π-π stacking interactions also play very crucial role to stabilize the c-KIT G-quadruplex DNA. Moreover, the π-π stacking interactions of the nucleoside bases in reline solution are stable up to 450 K, but no such interaction is observed at 500 K in reline (R500 system) and at 450 K in water medium (W450 system).

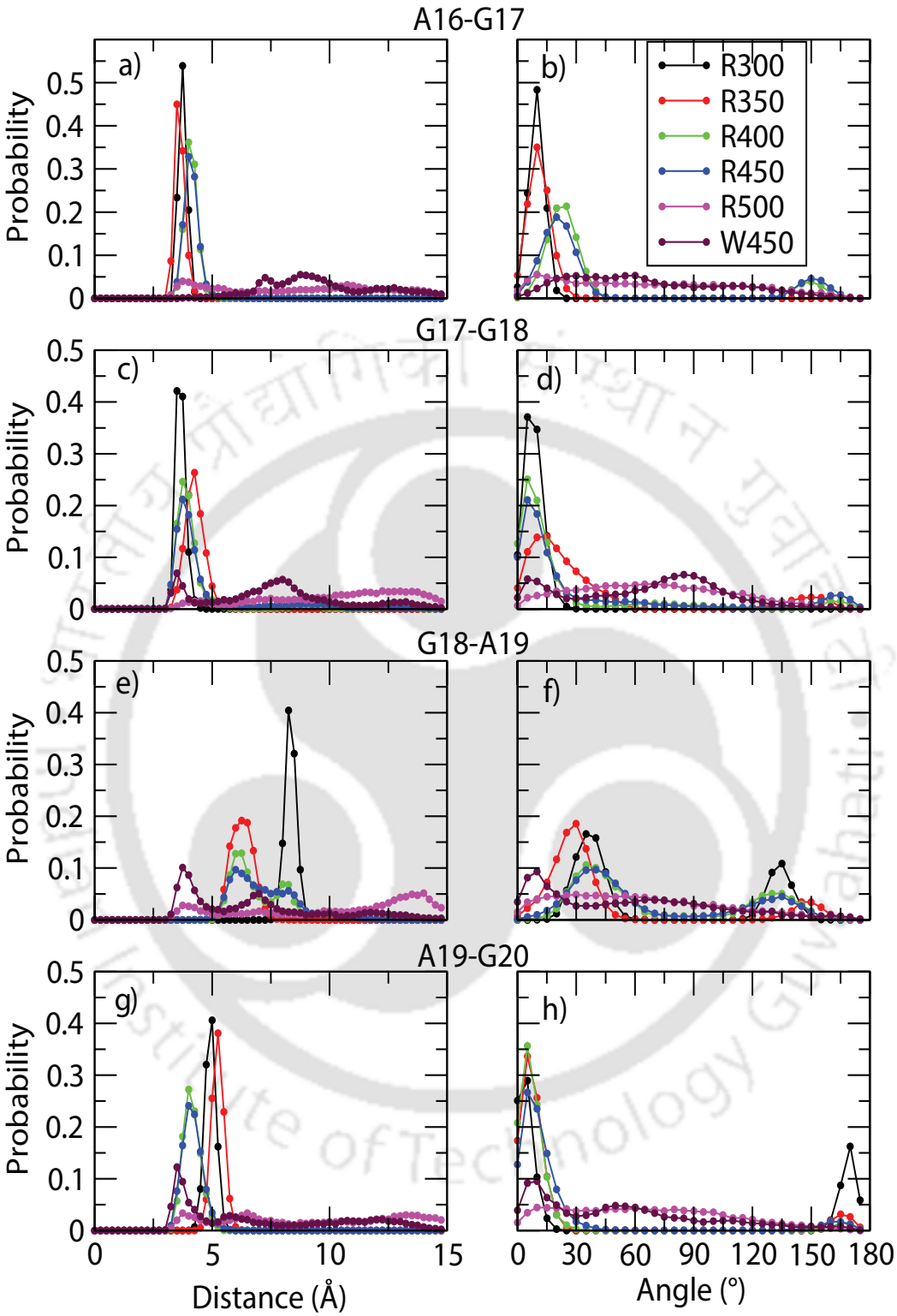


Figure 5.18: *Stacking probability with respect to distance and angle between the corresponding planes of the loop-4 residues for different systems.*

In **Fig. 5.19**, we have presented the dominant motion of the PCA of loop-4 for R300-R450 systems.

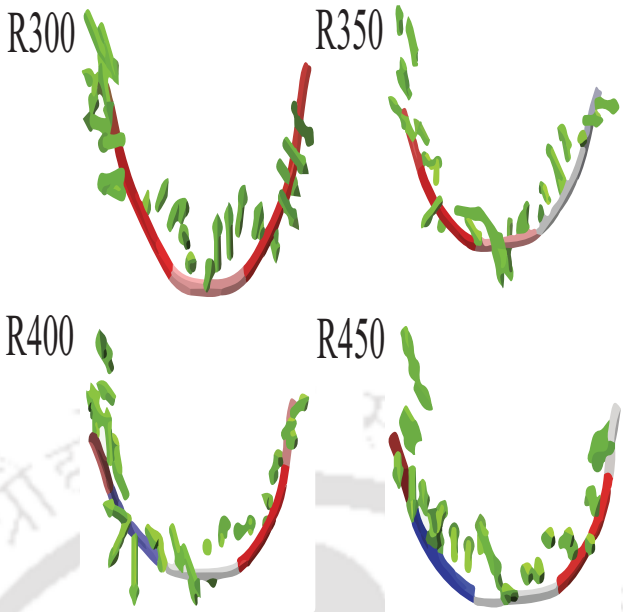


Figure 5.19: *Showing the dominant motions of loop-4 of c-KIT G-quadruplex DNA along with the first eigenvector in porcupine plot for different systems.*

5.3.3 Understanding the Role of Reline on c-KIT G-quadruplex DNA at Different Temperatures

5.3.3.1 Radial Distribution Functions (RDFs)

In our previous study (Chapter 4) [52], to relate the conformational change of G-quadruplex DNA in non-hydrated and as well as hydrated choline chloride-urea (i.e., reline) mixture, we had calculated site-site radial distribution functions (RDF) of sugar phosphate backbone, tetrad and loop regions separately. We found that choline cation and urea mainly interact with the backbone region of G-quadruplex DNA due to the presence of negatively charged oxygen(OP1/OP2), which is attached to the phosphorus atom in the backbone region. In **Fig. 5.20**, we present the site-site RDFs between oxygen(OP1/OP2) atom of sugar phosphate backbone with not only the oxygen atom of choline but also with the nitrogen and oxygen atoms of urea. Focusing on the interaction between sugar phosphate backbone and choline/urea we notice the following: (i) In the g(r) between oxygen (OP1/OP2) atom of sugar phosphate backbone and that of choline (**Fig. 5.20 (a)**), a sharp first peak appears at 2.65 Å with a small second peak at 5 Å. Moreover, with increasing temperature (i.e., R300-R500 systems), the peak height follows a decreasing trend. (ii) The site-site RDFs between oxygen (OP1/OP2) atom of sugar phosphate backbone and nitrogen and oxygen atoms of urea individually, (**Fig. 5.20 (b)**) reveal a steep peak at 2.85 Å which

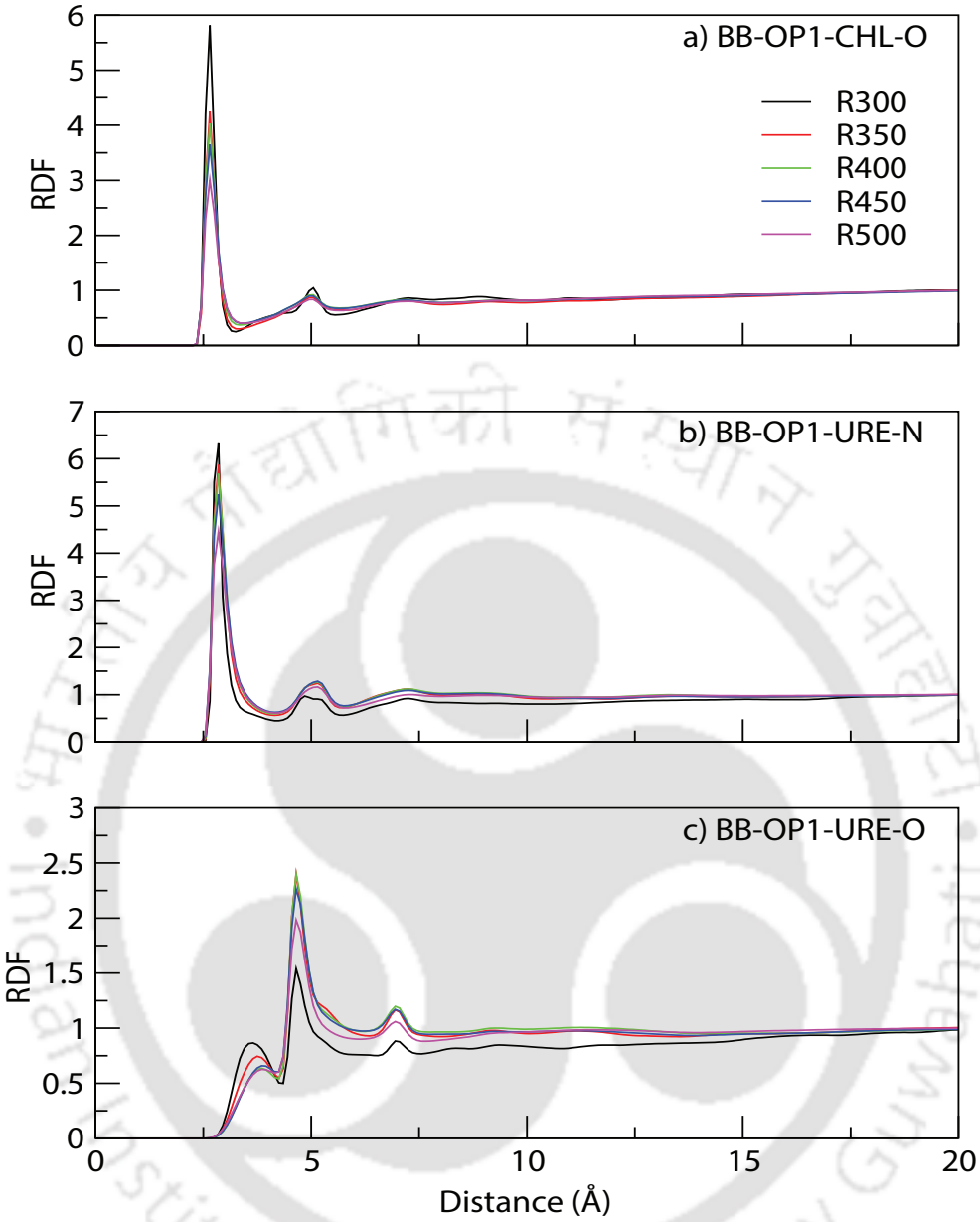


Figure 5.20: *Radial distribution functions (RDFs) of OP1 atom of sugar-phosphate-backbone with O atom of choline (a) and N and O atoms (b and c) of urea molecules respectively.*

diminishes when the temperature is increased. (iii) **Fig. 5.20** (c) displays the RDFs involving oxygen of sugar phosphate backbone and nitrogen atomic sites of urea. The decreasing trend of the sharp peak which appears at 3.75 Å implies that the interaction declines with increasing temperature. Overall, the RDFs implies that the presence of strong interactions between sugar phosphate backbone of G-quadruplex c-KIT DNA, which decreases when temperature increases, as expected.

In **Fig. 5.21**, we have shown the running coordination number of OP1 atom

of sugar-phosphate-backbone around the O atom of choline and N and O atoms of urea molecules for all systems. Moreover, we have also observed that the decreasing trend of running coordination number with increasing temperature.

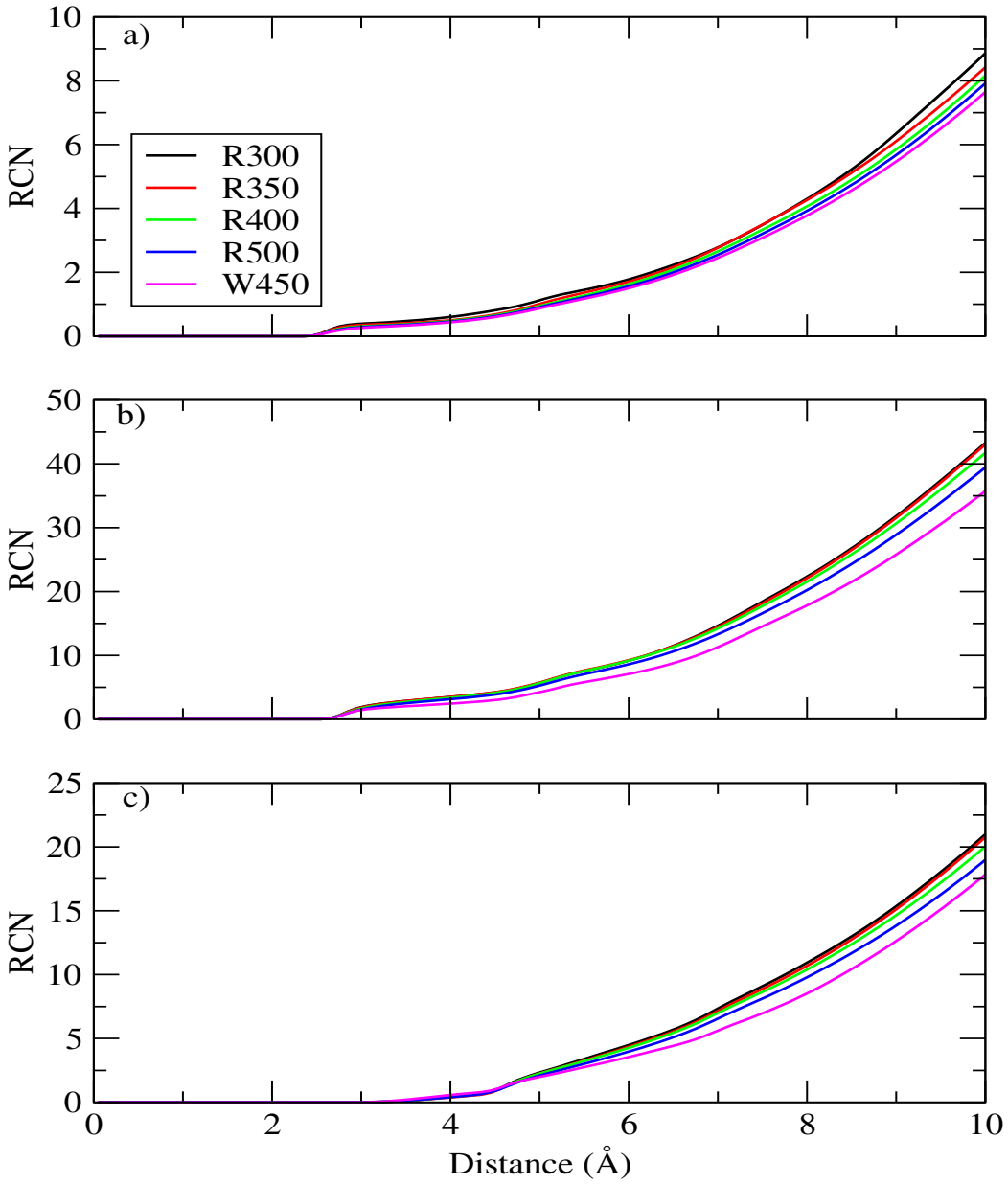


Figure 5.21: *Running Coordination Numbers (RCNs) of OP1 atom of sugar-phosphate-backbone with O atom of choline (a) and N and O atoms (b and c) of urea molecules respectively.*

5.3.3.2 Spatial Distribution Functions (SDFs)

The spatial distribution functions (SDFs) demonstrate the three-dimensional probabilistic arrangement of choline, urea molecules and K^+ ion around the DNA and provide more specific structural details on G-quadruplex DNA. Thus, we have estimated the SDF over 2 μs simulation time for R300-R450 systems and 200 ns for R500 and W450 systems. In

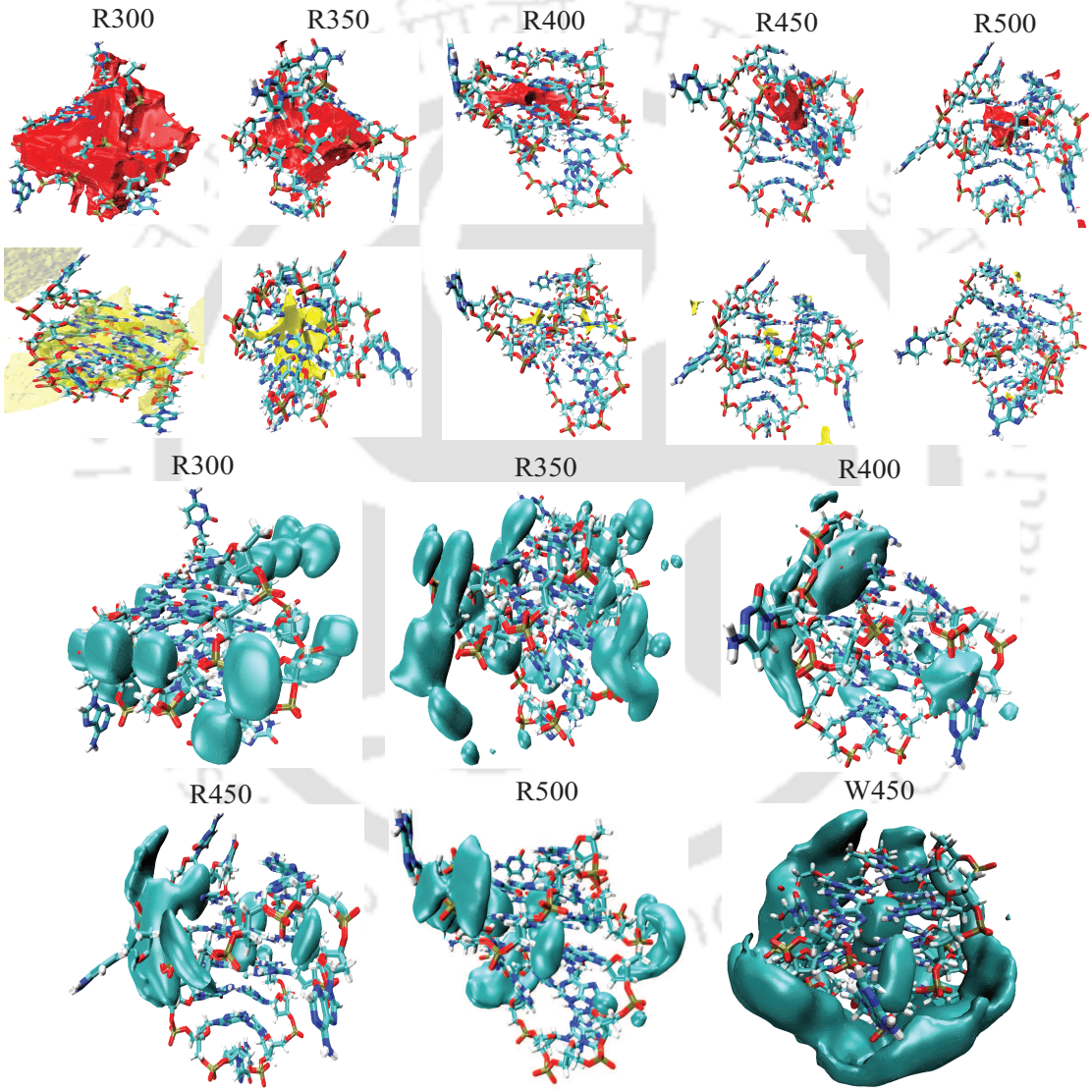


Figure 5.22: *Spatial density maps of choline, urea, and K^+ around c-KIT G-quadruplex DNA for different systems. Here, red, yellow and cyan represent choline, urea, and K^+ respectively.*

Fig. 5.22, we have shown the density distribution of choline, urea and K^+ around the c-KIT G-quadruplex DNA for different systems. Moreover, for choline and urea for density

distributions we have used isovalue 5×10^{-5} and for K^+, isovalue 1 is used. Here, we have noticed a compact distribution of hydrogen bond interacting urea and choline molecules around the reference c-KIT G-quadruplex DNA. Furthermore, the density of choline and urea around the c-KIT G-quadruplex DNA decreases with increasing temperature. On the other hand, the density of K^+ is mainly observed in phosphate regions of the backbone for all systems and in the ion channel regions for R300, R450, R500 and W450 systems. It decreases slightly with increasing temperature. Surprisingly, the K^+ distribution is denser in W450 system than any other system in reline medium. This observation indicates that the K^+ cations are less mobile in reline medium as compared to that in water medium. The density and as well as the viscosity of reline is higher than that of water [107]. Hence, the motion of ions are restricted in DESs, which were simplified by physical properties in hole theory [108–110]. According to the Abbott et al., hole theory [37], the ionic materials have vacant spaces when they are melted. The average size of these holes help in the easy movement of the small ions in DES s rather than the ions with large radius like K^+. So, the density of K^+ around the c-KIT G-quadruplex DNA is very high in water medium as compared to the reline medium.

5.3.3.3 First Solvation Shell

Considering previous studies [111, 112], the first solvation shell (FSS) of biomolecules can be described as the number of the solvent molecules present at a distance of 3.5 Å from the biomolecule surface. In more details, if any heavy atom of a solvent is present within 3.5 Å of any heavy atom of a residue of G-quadruplex DNA, it will be considered as a molecule which is present in a first solvation shell. **Fig. 5.23** shows the average first shell solvation number of reline mixture with each residue of G-quadruplex c-KIT DNA. It must be mentioned here that for R500 system, we have considered the trajectories before the DNA starts unfolding i.e., up to 150 ns of the simulation. Each residue of c-KIT DNA is found to be solvated with the minimum number of solution species in reline medium when temperature is the highest i.e, 500 K (R500 system). Consequently, the highest average first shell solvation number is observed at a low temperature i.e., 300 K or R300 system. Interestingly, we notice that among all residues of G-quadruplex c-KIT-DNA, loop regions are solvated more than the tetrad regions. This is apparent because the loop regions expand more than the tetrad regions with increasing temperature, as a result the former is more exposed to the solvent. This observation further confirms that at relatively high temperature (up to 450 K) the reline medium makes G-quadruplex c-KIT DNA more rigid,

which leads to stabilize the G-quadruplex DNA more in reline than in water medium.

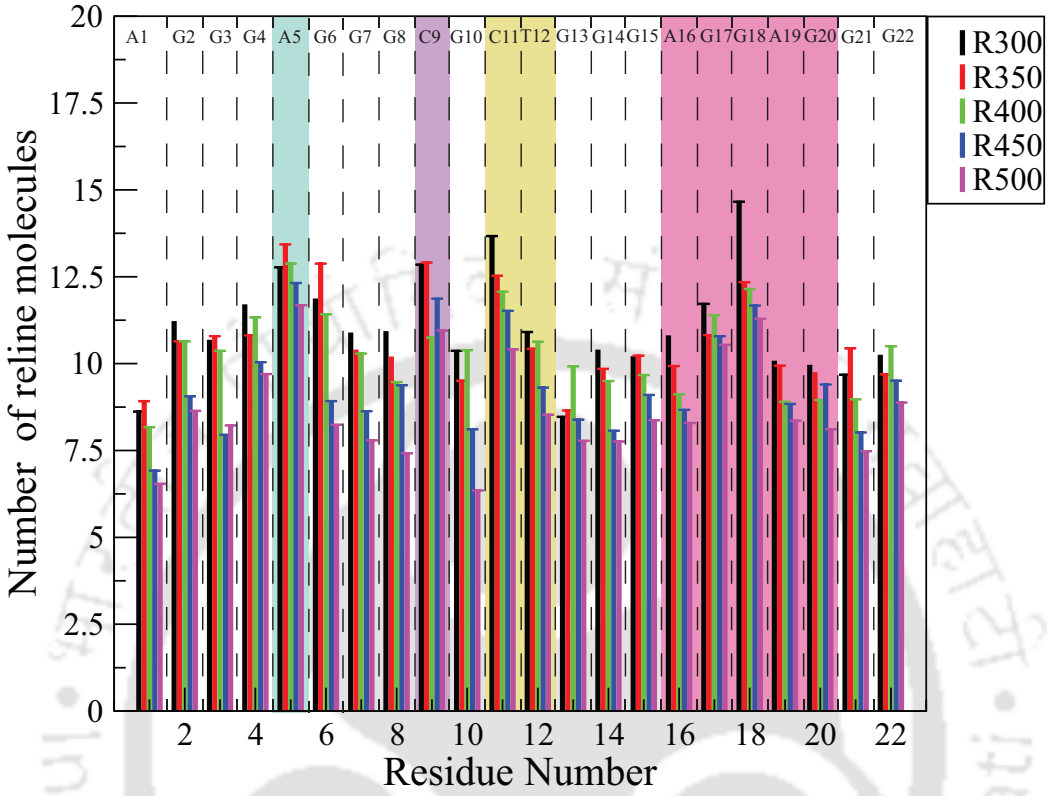


Figure 5.23: *The average number of the reline molecules present in the first solvation shell between each residue of the c-KIT G-quadruplex DNA for different systems. The standard error is calculated by block averaging method. Loop-1, loop-2, loop-3 and loop-4 regions are shown in cyan, purple, yellow and magenta colors respectively. Adenine, Guanine, Thymine and Cytosine are represented as A, G, T and C respectively.*

5.3.3.4 Hydrogen Bonding

In our previous study [52], we had seen that hydrogen bonds play a very important role in driving the G-quadruplex DNA to be rigid in reline medium. Thus, we have calculated the average number of hydrogen bonds between the backbone and reline medium. It is worth mentioning here that we have determined the number of hydrogen bonds of R500 system, before the G-quadruplex DNA starts unfolding i.e., up to 150 ns of the simulation run. We have set two requirements to determine the hydrogen bonds as followed by preceding researches [113–119], where (i) the distance cut-off between donor (D) and acceptor (A) is less than or equal to 3.5 Å and at the same time (ii) the donor(D)-acceptor(A)-hydrogen(H) angle cut-off is less than or equal to 45°.

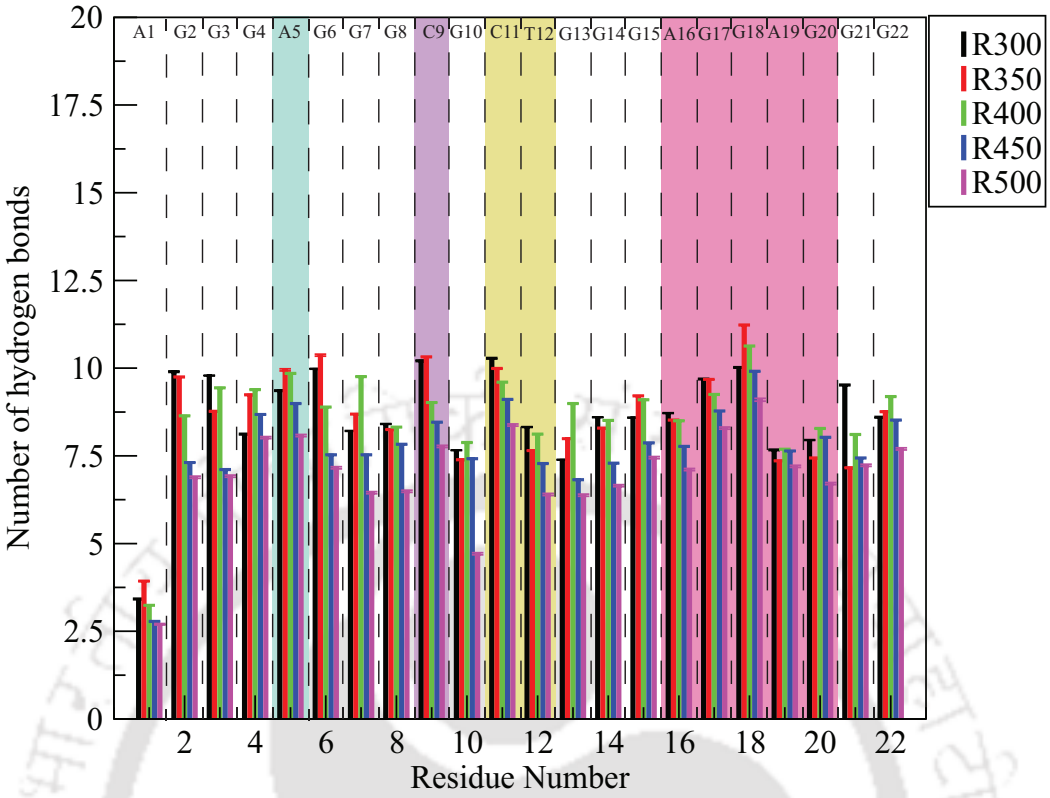


Figure 5.24: *The average number of hydrogen bonds between each residue of the c-KIT G-quadruplex DNA with reline medium for different systems. The standard errors are calculated by block averaging method. Loop-1, loop-2, loop-3 and loop-4 regions are shown in cyan, purple, yellow and magenta colors respectively. Adenine, Guanine, Thymine and Cytosine are represented as A, G, T and C respectively.*

In **Fig. 5.24**, we present the average number of hydrogen bonds between each residues of the c-KIT DNA and the reline molecules. Here, we have noticed that the average hydrogen bonds are comparable for the R300-R450 systems. Greater number of hydrogen bonds is perceived in the loop regions rather than the tetrad regions. However, for R500 systems, the average number of hydrogen bonds is lower for all the residues compared to that of the other systems. Furthermore, this trend is also observed in the previously discussed section (first solvation shell). This implies that the interaction between G-quadruplex c-KIT DNA and reline is favored due to the formation of stable hydrogen bonds. Thus, it suggests that when the temperature is up to 450 K (R300 to R450 systems), reline urges G-quadruplex DNA to be rigid, which is not observed in water medium at 450 K (W450 system). Nonetheless, at relatively high temperature i.e., 500 K (R500 system), the stable hydrogen bonds between the nucleic acid and the reline molecules collapses.

5.3.3.5 Role of Intra-tetrads Hydrogen Bonds and Solvation of the Backbone of c-KIT G-quadruplex DNA

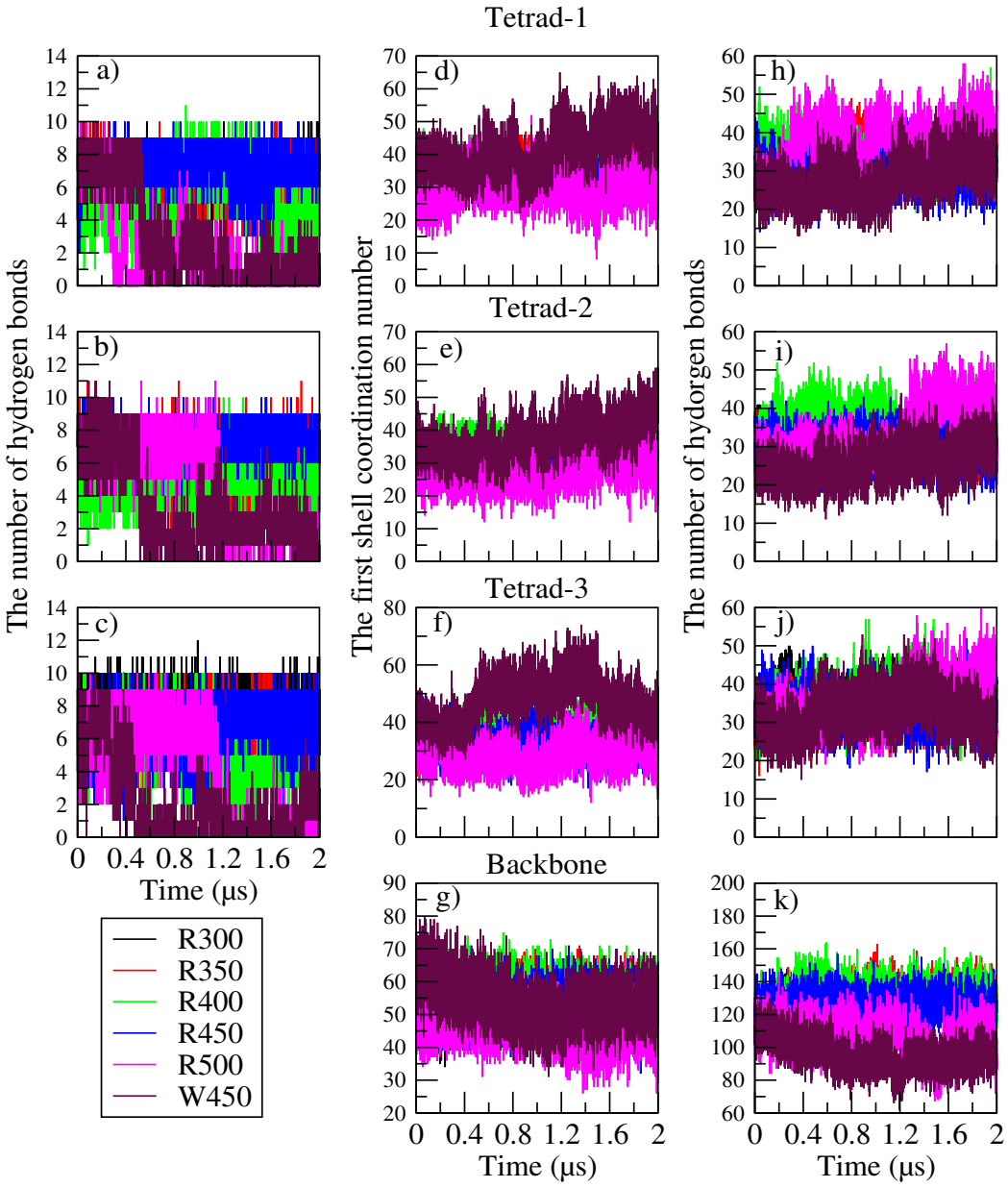


Figure 5.25: *The number of hydrogen bonds of different tetrads (a-c) with time progression. (d-g) and (h-k) represents the number of first solvation shell of solvent molecules and hydrogen bonds between solvents and different tetrads and backbone respectively.*

To understand the change in the conformational dynamics of c-KIT G-quadruplex DNA, we have determined the time evolution of average number of intra-tetrads hydrogen bonds since the guanine residues of the corresponding tetrads play a crucial role in the stabilization

of the c-KIT G-quadruplex DNA. In **Fig. 5.25 (a)-(c)**, we have shown these hydrogen bonds for all the three tetrads with time progression. The intra-hydrogen bonds between the guanine residues of the corresponding tetrads are very much similar and are stable through out the simulation for systems R300 to R450. Interestingly, a marked difference is noticed for R500 and W450 systems. For W450 system i.e., c-KIT G-quadruplex DNA at 450 K in water medium, the number of intra-hydrogen bonds of the tetrads is very similar with that of the other systems up to 500 ns simulation time, after which the stable intra-hydrogen bonds suddenly decreases with time for all the tetrads and this decrease is observed for all tetrads simultaneously. It indicates that stable intra-hydrogen bonds of all the three tetrads i.e., tetrad-1, tetrad-2 and tetrad-3 break simultaneously on in a concerted mechanism, which leads to the commencement of the unfolding of the c-KIT G-quadruplex DNA in water medium at 450 K. In contrast, for R500 system i.e., c-KIT G-quadruplex DNA at 500 K in reline medium, the intra-hydrogen bonds of tetrad-1 are stable up to 300 ns, after which it starts to collapse. Meanwhile, the intra-hydrogen bonds of tetrad-2 and tetrad-3 remain intact up to 1.2 μs, after which they start to perish. This observation depicts that, after collapsing the intra-hydrogen bonds of the tetrad-1, reline molecules help to retain the essential tetrad-2 and tetrad-3 of the c-KIT G-quadruplex DNA for quite a long time. In another way, we can assume that reline slows the thermal unfolding or denaturing process of the c-KIT G-quadruplex DNA. Moreover, to investigate the role of solvent in the conformational dynamics of the tetrads and the backbone of the c-KIT G-quadruplex DNA, we have calculated the number of reline molecules in the first solvation shell of c-KIT G-quadruplex DNA and the number of hydrogen bonds between solvent and the tetrads and the backbone separately. The number of reline molecules in the first solvation shell of the tetrads and the backbone follows a decreasing trend with increasing temperature, which is expected. Additionally, it increases dramatically once c-KIT quadruplex DNA starts to unfold for both the systems R500 and W450 systems. This is quite obvious because, in the unfolded state, DNA is more exposed to the solvent than its folded state. From **Fig. 5.25 (h)-(k)**, we find that, with increasing temperature up to 450 K, the hydrogen bonds between c-KIT G-quadruplex DNA (tetrads and backbone) and reline decreases which causes the c-KIT G-quadruplex DNA to be more flexible, although the DNA maintains its native state in reline medium. In contrast, for W450 system i.e., 450 K in water medium, the structural rigidity of the c-KIT G-quadruplex DNA is not maintained due to the less number of hydrogen bonds between water and c-KIT G-quadruplex DNA compared to that in reline medium, which accounts for the unfolding of the c-KIT G-quadruplex DNA from

its native state. Moreover, further heating, at 500 K, reline cannot preserve the rigidness of the c-KIT G-quadruplex DNA which stimulates the unfolding of the c-KIT G-quadruplex DNA. From hydrogen bond analysis, it is very clear that the hydrogen bonds in reline stabilize the G-quadruplex fold much more as compared to that in water.

5.3.3.6 π-π Stacking Interactions

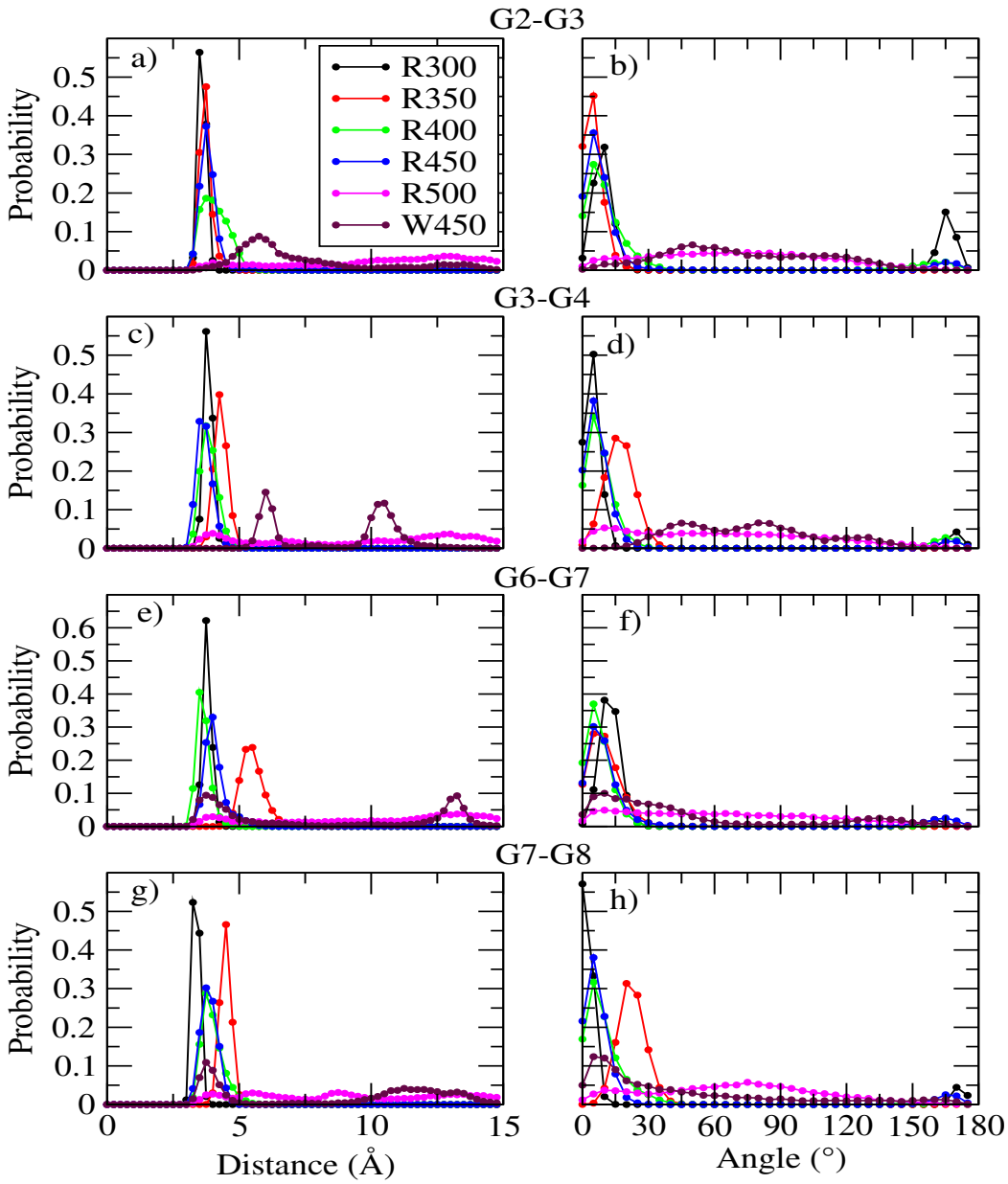


Figure 5.26: *Stacking probability with respect to distance and angle between the corresponding planes of the tetrad residues for different systems.*

If the hydrogen bond was the sole cause of stability of c-KIT G-quadruplex DNA, why did not the DNA not unfold even more to increase its surface area, further increasing the number of hydrogen bonds? To find the reason behind this, we need to focus on the

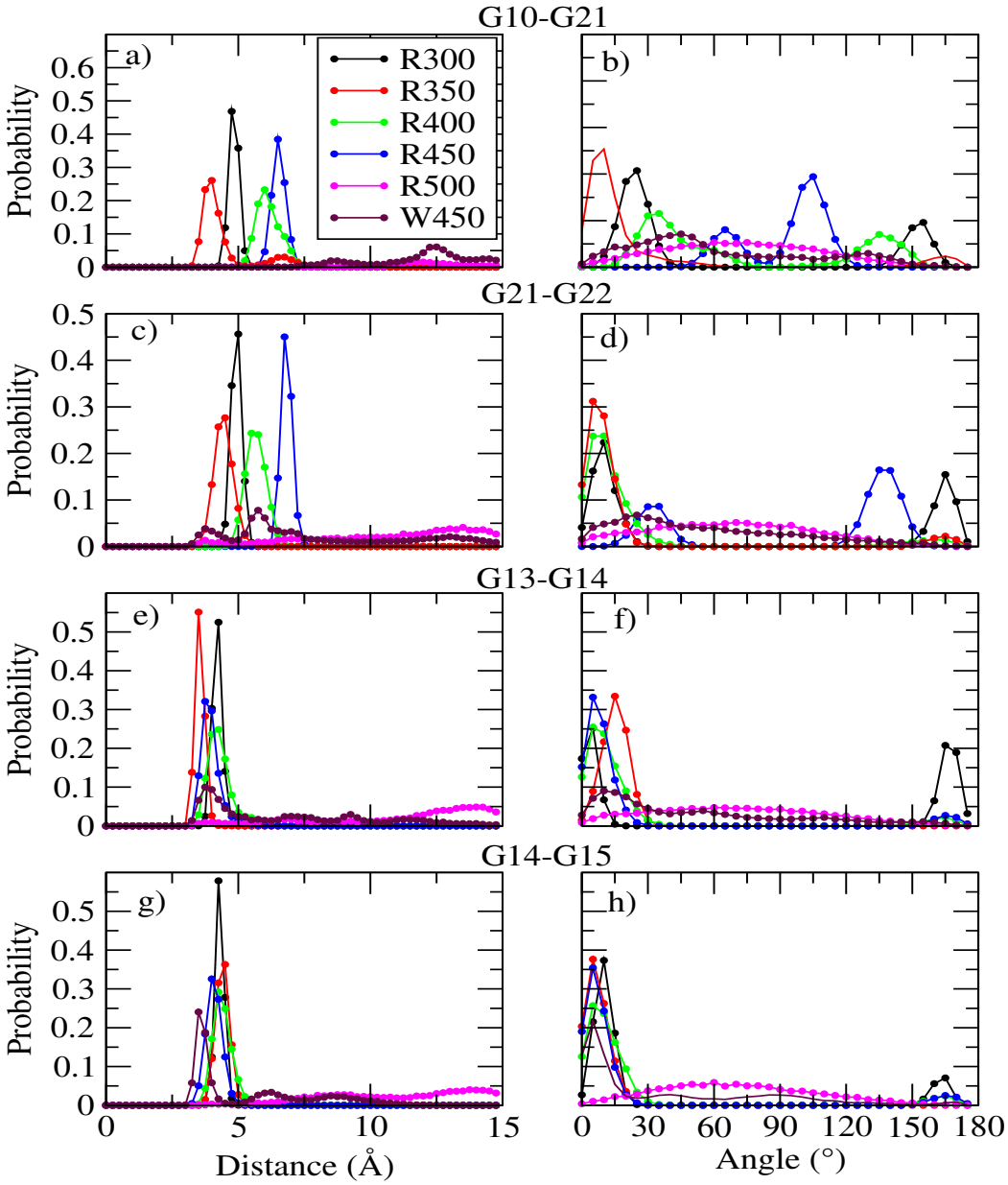


Figure 5.27: *Stacking probability with respect to distance and angle between the corresponding planes of the tetrad residues for different systems.*

property, apart from favorable hydrogen bond formation, which helps in stabilization of the G-quadruplex DNA i.e., π-π stacking interactions of tetrads. The π-π stacking interaction plays a very important role to stabilize the tetrad, which leads to the stabilization of

the folded G-quadruplex DNA. In **Figures 5.26 and 5.27**, we have shown the distance and angle probability of tetrad residues respectively. Taking into account the previously mentioned stacking criteria, all the tetrads residues fulfill both the criteria of π-π stacking interactions in the temperature range of 300 K to 450 K. However, there are no π-π stacking interactions present when the temperature is increased i.e, for R500 and W450 systems. It is worth mentioning that G10-G21 and G21-G22 slightly deviate from the stacking interaction criteria at a temperature of 400 and 450 K in reline medium. Overall, this observation concludes that the stable π-π stacking interactions are disrupted at 500 K in reline and 450 K in water medium, which leads to unfolding of c-KIT G-quadruplex DNA.

5.3.3.7 Role of the Central Cations

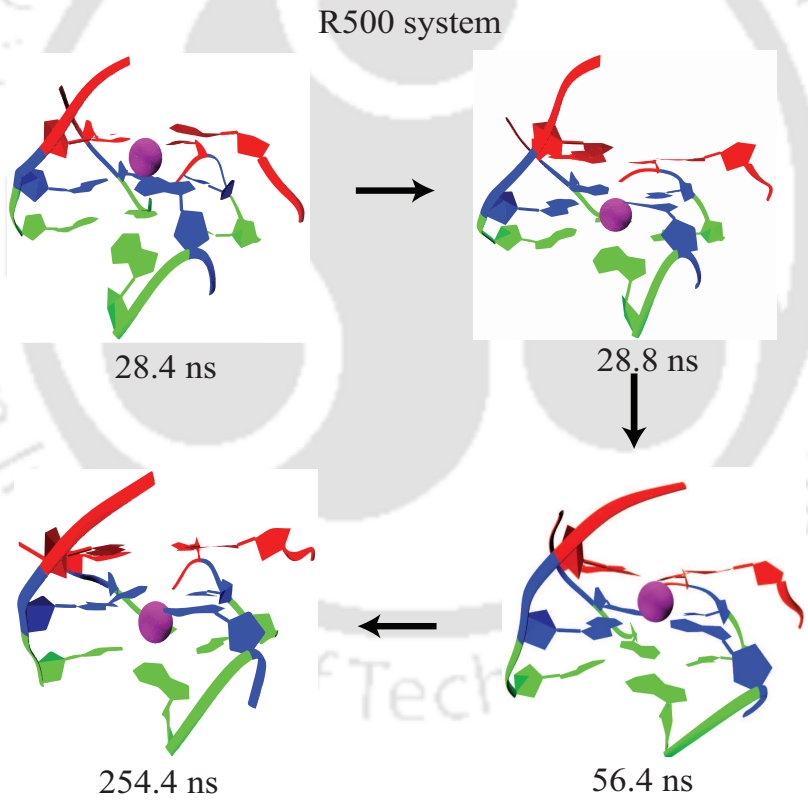


Figure 5.28: *Snapshots of K^+ central cation oscillation from the ion channel of tetrad-1 and tetrad-2 to tetrad-2 and tetrad-3 and vice versa for R500 systems. Here the tetrad-1, tetrad-2, tetrad-3 and K^+ ion represent in red, blue, green and purple colors respectively.*

The central cations play a major role in the formation and stabilization of the G-quadruplex DNA, which is why it is important to discuss further about the involvement of the cations. In the beginning of the simulation, there are no central cations present in the central core of

the G-quadruplex tetrad channels. With time progression, only one K^+ (namely K1) enters between tetrad-1 and tetrad-2 for R300, R450 and R500 systems in the equilibration state at 0.8, 1.1 and 2.3 ns respectively. Here, it is worth mentioning that, for R350 and R400 systems, no cation enters in the core of c-KIT G-quadruplex throughout the simulation. Due to this reason, the RMSD values of c-KIT G-quadruplex DNA for R350 and R400 systems are slightly more than its native state, although the G-quadruplex is very stable throughout the 2.0 μs simulation. This is why the RMSD, RMSF and R_g values for these systems do not show any trend with increasing temperature. It also concludes that, in absence of central cation, c-KIT G-quadruplex is stable up to 400 K in reline medium. For R300 system, K1 is stable at a fixed position in between tetrad-1 and tetrad-2 throughout the 2.0 μs simulation. In this regard we note that Bergues-Pupo et al. observed that, in absence of central cations, the structure of G-quadruplex DNAs is very unstable i.e., it deviates much more than its native structure in water medium [58] at 300 K. However, they have also reported that, if one ion is captured in the simulation run, the G-quadruplex returns from its deviated state to a stable equilibrium state at 300 K. Here, in case of R450 and R500 systems, we have observed a very interesting phenomenon. For R450 system, the cation which enters the channel core i.e., namely K1, is not fixed at a certain point throughout the 2.0 μs simulation. It fluctuates its position between the two tetrad cores i.e., the place between tetrad-1 and tetrad-2 and that between tetrad-2 and tetrad-3. We have observed that it oscillates twice between the regions throughout the whole simulation. For R500 systems, the K1 oscillates in its position (from tetrad-1 to tetrad-2 or vice versa) four times till 255 ns (in **Fig. 5.28**) and after that, the c-KIT G-quadruplex DNA starts to unstabilize. To measure this 'up down' mechanism of central cation, we have calculated the distance between the center of mass of O6 atom of guanine residues of tetrad-1 and the K1 atom and that between tetrad-2 and K1 atom. We have shown the results in **Figures 5.29 and 5.30**. From these results, we can also conclude that, with increasing temperature the rate of fluctuation of position of K1 is increased. This unique phenomenon of the K^+ central cation in G-quadruplex DNA is not reported yet.

On the other hand, in case of W450 system, the first K^+ cation (K1) enters in between tetrad-1 and tetrad-2 at 2.8 ns i.e., at the starting of production run. Then, at 68.08 ns the K1 shifts to the region between tetrad-2 and tetrad-3, leaving the region between tetrad-1 and tetrad-2 vacant. Additionally, the second K^+ cation (namely K2) enters in the vacant region between tetrad-1 and tetrad-2 at 68.4 ns. Interestingly, at

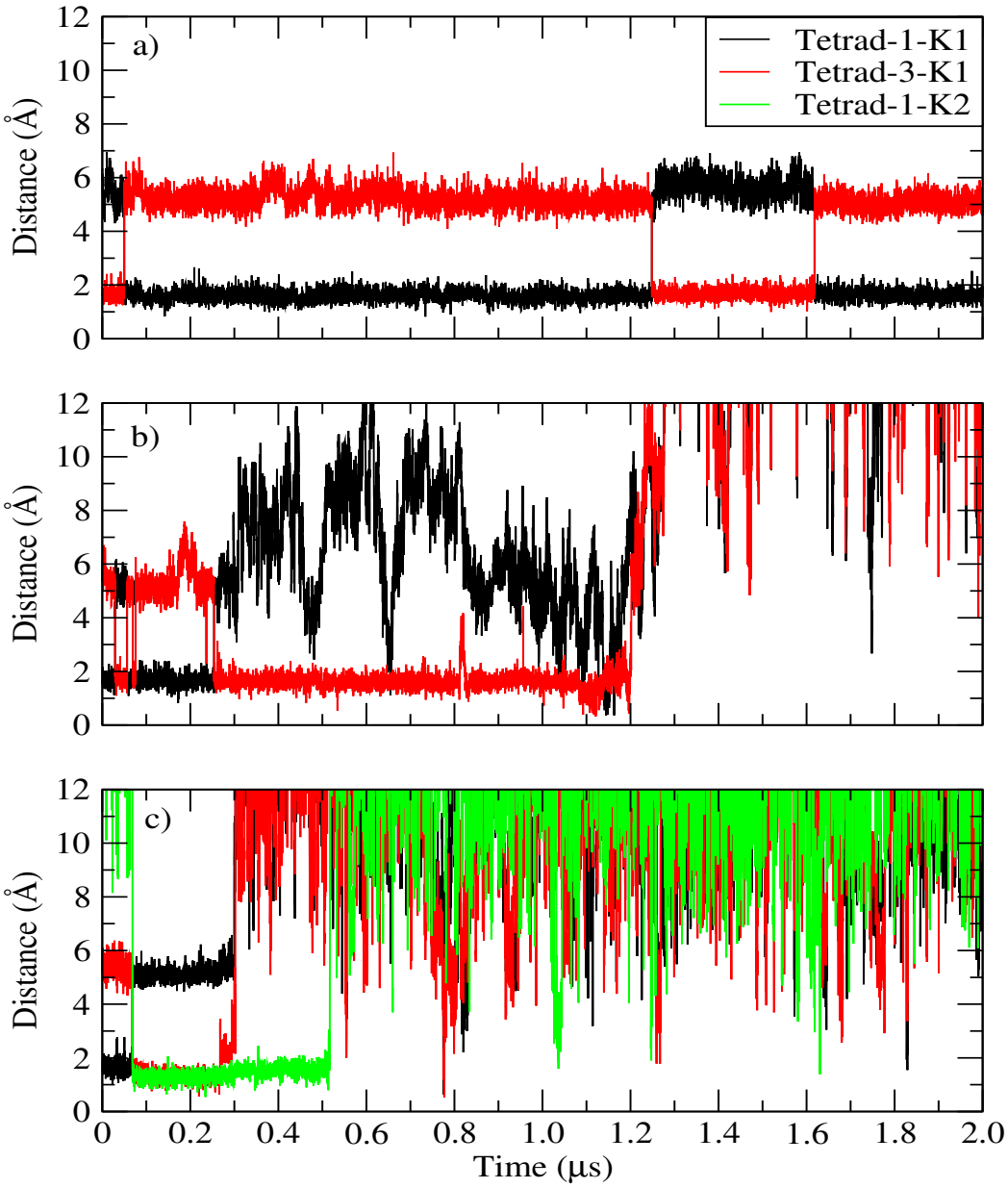


Figure 5.29: *Distance between the captured K^+ and the center of mass of O6 atoms of guanine residues of tetrad-1 and tetrad-3 for (a) R450, (b) R500 and (c) W500 systems. Here, K1 and K2 represent the first and second cation captured by the tetrad ion channel.*

299.6 ns, the K1, which is present between tetrad-2 and tetrad-3, exits the ion channel and simultaneously tetrad-3 starts to break (G4 residue starts to fluctuate). Finally, at 512 ns all the tetrads becomes unstable for W450 system, even though K2 cation is present between tetrad-1 and tetrad-2, at that moment. Here, it is worth mentioning that the 'up down' mechanism of central cation is not observe in water medium at 450 K. Overall,

from this observation, we also conclude that the c-KIT G-quadruplex DNA is stable in presence of one central K^+ cation in the tetrad channel in reline medium at 450 K, but it becomes unstable in water medium even though two central K^+ cations are present inside the channel core of the tetrads.

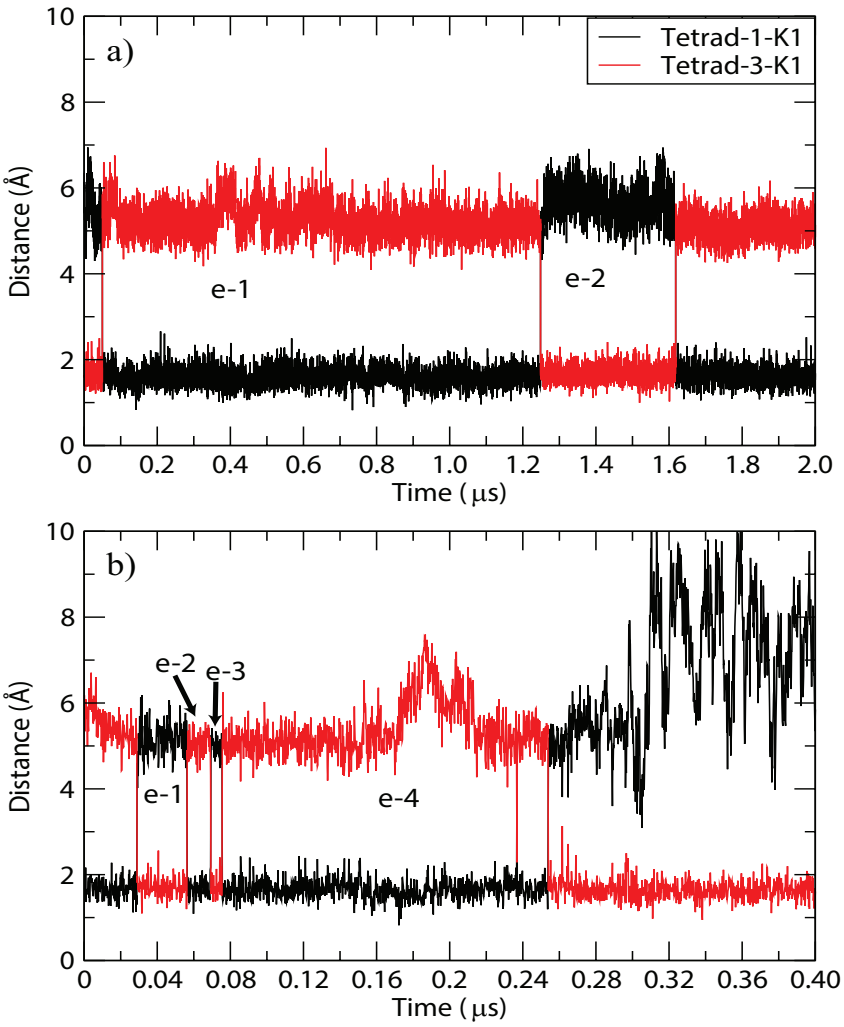


Figure 5.30: *Distance between the captured K^+ and the center of mass of O6 atoms of guanine residues of tetrad-1 and tetrad-3 for (a) R450, and (b) R500 systems. Here, K1 represent the first cation captured by the tetrad ion channel and 'e' represent for exchange number.*

5.3.4 Denaturation of c-KIT G-quadruplex DNA

In this study, it is noticed that reline protects the c-KIT DNA from high temperature denaturation. Here, it is important to study the in-details dynamics of the tetrads of c-KIT G-quadruplex DNA in reline as well water medium in the temperature range of 300

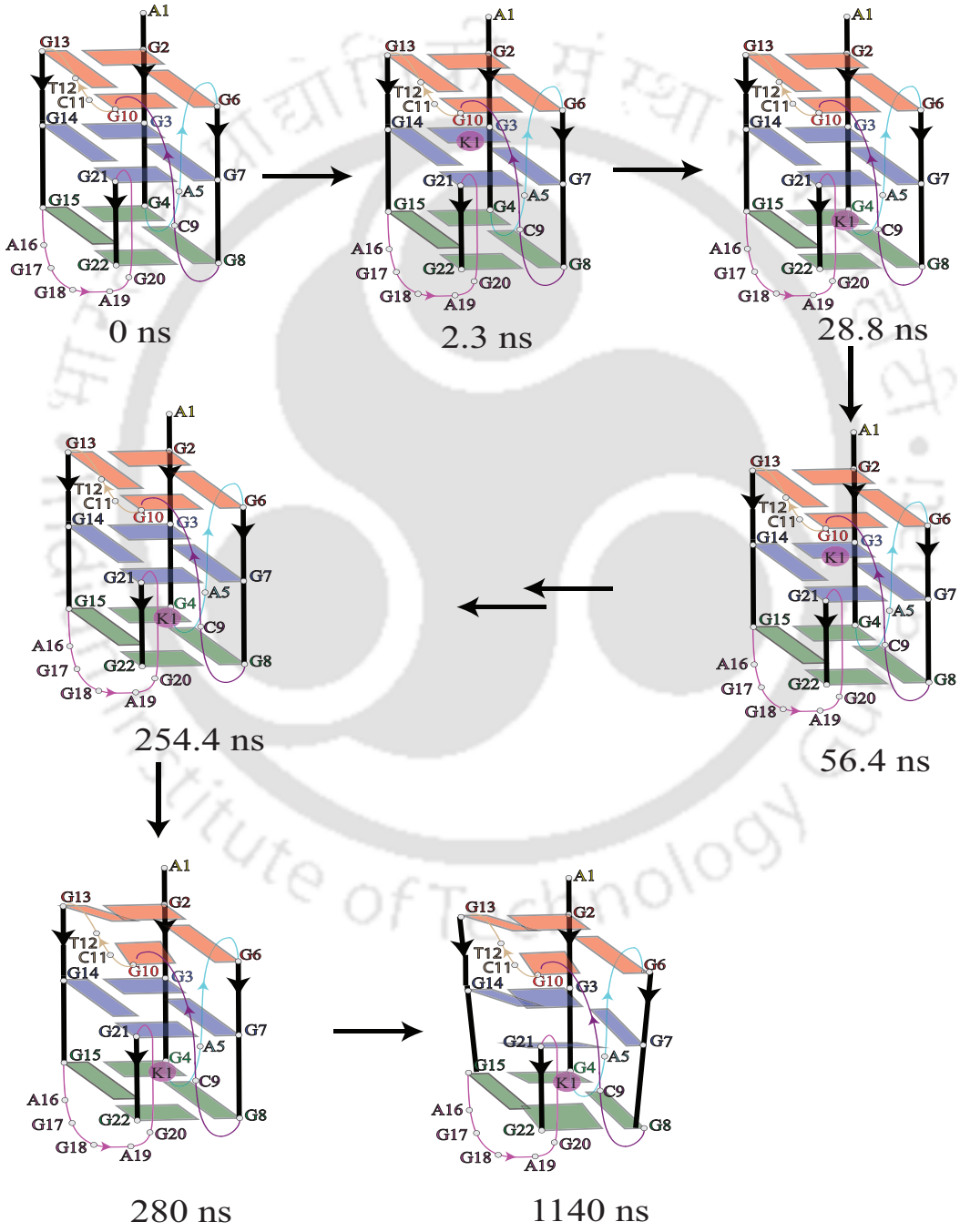


Figure 5.31: *Schematic representation of the denaturation pathway of the c-KIT G-quadruplex DNA in reline medium at 500 K.*

K to 500 K. To investigate the denaturation pathway of c-KIT G-quadruplex DNA, we have observed that initially the denaturation pathway is very similar in both the media. The denaturation triggers from the tetrad where no central cation is present i.e., tetrad-1 for R500 system and tetrad-3 for W450 system. However, the rest of the denaturation pathway is very different from each other. In **Figures 5.31 and 5.32**, we have shown the

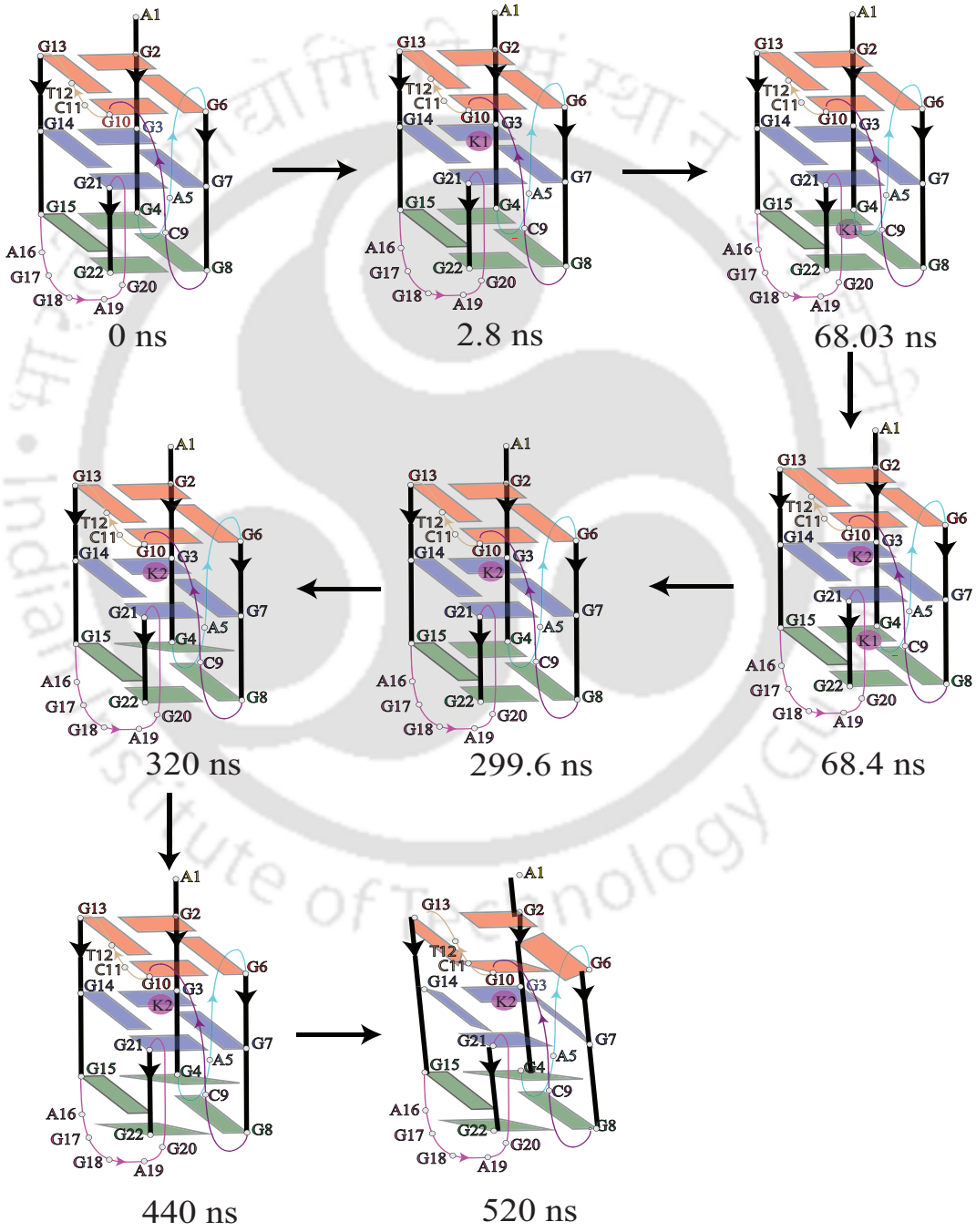


Figure 5.32: *Schematic representation of the denaturation pathway of the c-KIT G-quadruplex DNA in water medium at 450 K.*

schematic diagram of the denaturation pathway of the c-KIT G-quadruplex DNA at 500 K in reline and at 450 K in water medium respectively. R500 system (**Fig. 5.31**), at 280 ns, G10 and G13 residues of tetrad-1 start to break the tetrad geometry, although the other tetrads maintain their stability up to 1120 ns. After that, the tetrad-1 and tetrad-2 of the c-KIT G-quadruplex DNA starts to bifurcate. Finally, no tetrad conformation is observed at 1200 ns even though the central cation is present.

On the other hand, the c-KIT G-quadruplex DNA is stable up to 299.6 ns in presence of one central cation in W450 system (**Fig. 5.32**), after which, the G4 residues of tetrad-3 start to fluctuate and break the π-π stacking formation with G3. However, the other residues of the tetrad regions still remain intact. Then, at 440 ns, all the guanine residues of tetrads-3 start to break the tetrad orientation and simultaneously, the other tetrad regions (tetrad-1 and tetrad-2) also begin to lose their tetrad formation. Finally, at 520 ns, we do not observe any tetrad formation.

5.3.5 Limitations

In this study, the latest force field parameters for different atomic sites have been used for G-quadruplex DNA and DESs. In fact, the current AMBER force-field parameters are not assured in a priori to be entirely compatible with the reline system and force field. This issue is very difficult to solve because of the lack of experimental and simulation results.

5.4 Summary and Conclusions

All-atom molecular dynamics simulations are carried out to understand the effect of temperature on c-KIT G-quadruplex DNA in natural deep eutectic solvent (NADES) i.e., reline medium. The root mean square deviation, root mean square fluctuation, radius of gyration and free energy landscapes analyses provide the information of conformational deviation of the c-KIT G-quadruplex DNA in the temperature range of 300 K to 450 K in reline medium and also its unfolding at 500 K in reline medium and 450 K in water medium. From those analyses, we observe that the c-KIT DNA is stable at temperature of 450 K in reline medium. However, it unfolds when the solvent is water at the same temperature. To find the reason for the stability of the c-KIT G-quadruplex DNA in reline medium up to a temperature of 450 K temperature, we have determined the number of reline molecules in the first solvation shell of DNA and also the average number of hydrogen bonds between them. It is found that the hydrogen bonds between reline and c-KIT G-quadruplex DNA

is responsible for its stability at high temperature (up to 450 K). Moreover, from spatial distribution functions, we have observed that the compact distribution of hydrogen bonded choline and urea molecules around c-KIT G-quadruplex DNA decreases with increasing temperature. However, the density of K^+ ion slightly decreases when temperature is increased. The π-π stacking interactions of tetrads play a very crucial role in the stabilization of G-quadruplex formation. In this work, we have also observed that π-π stacking interactions of tetrads are very stable in reline medium up to 450 K, but in water medium, no such π-π stacking interactions are observed at high temperature (for W450 system). This is because of the unique characteristics of reline such as low vapour pressure, high thermal and chemical stability at high temperature. We have observed that, in reline medium, c-KIT G-quadruplex is stable up to 450 K but it unfolds or is unstable at 500 K even though, it unfolds in water medium at 450 K, which leads us to conclude that reline provides the ultrastability to c-KIT G-quadruplex DNA at elevated temperature in comparison with water medium, which is also validated by previous experimental work [43, 44]. Additionally, the tetrad dynamics of c-KIT G-quadruplex DNA in water medium suggests that the stable hydrogen bonds between guanine residues of the corresponding tetrads are collapsed by a concerted mechanism. However, the story is completely different in reline medium. When the temperature is elevated to 500 K, we notice that, after breaking the stable intra-tetrad hydrogen bonds of tetrad-1, other tetrads (namely tetrad-2 and tetrad-3) are stable for a long time, although they too collapse eventually. Hence, it can be assumed that reline, not only protects the c-KIT G-quadruplex DNA, but also it slows the thermal denaturing process. The central cations also play a major role to form and stabilize the G-quadruplex DNA. On investigating the role of central cations, we have observed that the central cation K^+ oscillates its position in between tetrad-1 and tetrad-2 and tetrad-2 and tetrad-3 at 450 K and 500 K in reline medium. Moreover, the rate of oscillations of the position of the cation is increased when temperature rises from 450 K to 500 K. However, this phenomenon is absent in water medium at 450 K. This distinctive 'up down' phenomenon of the central cation of the ion channel in reline medium is not yet reported.

References

[1] Y. Yarden, W. J. Kuang, T. Yang-Feng, L. Coussens, S. Munemitsu, T. J. Dull, E. Chen, J. Schlessinger, U. Francke and A. Ullrich, *EMBO J.*, 1987, **6**, 3341–3351.

[2] J. Lennartsson and L. Rönnstrand, *Physiol. Rev.*, 2012, **92**, 1619–1649.

[3] R. Roskoski, *Biochem. Biophys. Res. Commun.*, 2005, **337**, 1 – 13.

[4] C. E. Edling and B. Hallberg, *Int. J. Biochem. Cell Biol.*, 2007, **39**, 1995 – 1998.

[5] R. Simak, P. Capodieci, D. Cohen, W. Fair, H. Scher, J. Melamed, M. Drobnjak, W. Heston, U. Stix, G. Steiner *et al.*, *Histol. Histopathol.*, 2000, **15**, 365–374.

[6] P. Micke, J. G. Hengstler, H. Albrecht, A. Faldum, F. Bittinger, K. Becker, R. Wiewrodt, B. Fischer and R. Buhl, *Tumor Biol.*, 2004, **25**, 235–242.

[7] A. McIntyre, B. Summersgill, B. Grygalewicz, A. J. Gillis, J. Stoop, R. J. van Gurp, N. Dennis, C. Fisher, R. Huddart, C. Cooper *et al.*, *Cancer Res.*, 2005, **65**, 8085–8089.

[8] K. S. Smalley, K. L. Nathanson and K. T. Flaherty, *Cancer Res.*, 2009, **69**, 3241–3244.

[9] S. Sakurai, T. Fukasawa, J.-M. Chong, A. Tanaka and M. Fukayama, *Jpn. J. Cancer Res.*, 1999, **90**, 1321–1328.

[10] M. Sattler and R. Salgia, *Leuk. Res.*, 2004, **28**, 11–20.

[11] A. Głuszyńska, B. Juskowiak, M. Kuta-Siejkowska, M. Hoffmann and S. Haider, *Molecules*, 2018, **23**, 1134.

[12] M. Bejugam, S. Sewitz, P. S. Shirude, R. Rodriguez, R. Shahid and S. Balasubramanian, *J. Am. Chem. Soc.*, 2007, **129**, 12926–12927.

[13] X. Wang, C.-X. Zhou, J.-W. Yan, J.-Q. Hou, S.-B. Chen, T.-M. Ou, L.-Q. Gu, Z.-S. Huang and J.-H. Tan, *ACS Med. Chem. Lett.*, 2013, **4**, 909–914.

[14] T. P. Garner, H. E. L. Williams, K. I. Gluszyk, S. Roe, N. J. Oldham, M. F. G. Stevens, J. E. Moses and M. S. Searle, *Org. Biomol. Chem.*, 2009, **7**, 4194–4200.

[15] S. Roe, D. J. Ritson, T. Garner, M. Searle and J. E. Moses, *Chem. Commun.*, 2010, **46**, 4309–4311.

[16] P. Alberti and J.-L. Mergny, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 1569–1573.

[17] J. J. Li and W. Tan, *Nano Letters*, 2002, **2**, 315–318.

[18] C. Zhao, L. Wu, J. Ren and X. Qu, *Chem. Commun.*, 2011, **47**, 5461–5463.

[19] Y. Xiao, V. Pavlov, T. Niazov, A. Dishon, M. Kotler and I. Willner, *J. Am. Chem. Soc.*, 2004, **126**, 7430–7431.

[20] K. Dutta, T. Fujimoto, M. Inoue, D. Miyoshi and N. Sugimoto, *Chem. Commun.*, 2010, **46**, 7772–7774.

[21] A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle and D. J. Patel, *J. Am. Chem. Soc.*, 2007, **129**, 4386–4392.

[22] D. Wei, G. N. Parkinson, A. P. Reszka and S. Neidle, *Nucleic Acids Res.*, 2012, **40**, 4691–4700.

[23] A. M. Salsbury and J. A. Lemkul, *J. Phys. Chem. B*, 2019, **123**, 148–159.

[24] R. Rocca, F. Moraca, G. Costa, C. Talarico, F. Ortuso, S. Da Ros, G. Nicoletto, C. Sissi, S. Alcaro and A. Artese, *ACS Med. Chem. Lett.*, 2018, **9**, 848–853.

[25] K. R. Seddon, *J. Chem. Technol. Biotechnol.*, 1997, **68**, 351–356.

[26] S. V. Muginova, A. Z. Galimova, A. E. Polyakov and T. N. Shekhovtsova, *J. Anal. Chem.*, 2010, **65**, 331–351.

[27] K. Fujita, D. R. MacFarlane, M. Forsyth, M. Yoshizawa-Fujita, K. Murata, N. Nakamura and H. Ohno, *Biomacromolecules*, 2007, **8**, 2080–2086.

[28] N. Byrne, L.-M. Wang, J.-P. Belieres and C. A. Angell, *Chem. Commun.*, 2007, 2714–2716.

[29] R. Madeira Lau, M. J. Sorgedrager, G. Carrea, F. van Rantwijk, F. Secundo and R. A. Sheldon, *Green Chem.*, 2004, **6**, 483–487.

[30] J. L. L. Carter, C. C. Santini, L. J. Blum and B. Doumche, *RSC Adv.*, 2015, **5**, 23715–23718.

[31] Q. Zhang, K. De Oliveira Vigier, S. Royer and F. Jérme, *Chem. Soc. Rev.*, 2012, **41**, 7108–7146.

[32] C. Ma, A. Laaksonen, C. Liu, X. Lu and X. Ji, *Chem. Soc. Rev.*, 2018, **47**, 8685–8720.

[33] F. Ilgen, D. Ott, D. Kralisch, C. Reil, A. Palmberger and K
relax.

[34] Y. Yu, X. Lu, Q. Zhou, K. Dong, H. Yao and S. Zhang, *Chem.: Eur. J.*, 2008, **14**, 11174–11182.

[35] K. D. Weaver, H. J. Kim, J. Sun, D. R. MacFarlane and G. D. Elliott, *Green Chem.*, 2010, **12**, 507–513.

[36] M. Di Gioia, M. Nardi, P. Costanzo, A. De Nino, L. Maiuolo, M. Oliverio and A. Procopio, *Molecules*, 2018, **23**, 1891.

[37] E. L. Smith, A. P. Abbott and K. S. Ryder, *Chem. Rev.*, 2014, **114**, 11060–11082.

[38] Q. Zhang and J. M. Shreeve, *Chem. Rev.*, 2014, **114**, 10527–10574.

[39] R. Vijayaraghavan, A. Izgorodin, V. Ganesh, M. Surianarayanan and D. MacFarlane, *Angew. Chem.*, 2010, **49**, 1631–1633.

[40] I. Mamajanov, A. Engelhart, H. Bean and N. Hud, *Angew. Chem.*, **49**, 6310–6314.

[41] F. M. Lannan, I. Mamajanov and N. V. Hud, *J. Am. Chem. Soc.*, 2012, **134**, 15324–15330.

[42] I. Gállego, M. A. Grover and N. V. Hud, *Angew. Chem.*, **54**, 6765–6769.

[43] C. Zhao, J. Ren and X. Qu, *Langmuir*, 2013, **29**, 1183–1191.

[44] C. Zhao and X. Qu, *Methods*, 2013, **64**, 52 – 58.

[45] S. Satpathi, M. Kulkarni, A. Mukherjee and P. Hazra, *Phys. Chem. Chem. Phys.*, 2016, **18**, 29740–29746.

[46] K. Jumbri, M. B. Abdul Rahman, E. Abdulmalek, H. Ahmad and N. M. Micaelo, *Phys. Chem. Chem. Phys.*, 2014, **16**, 14036–14046.

[47] A. Chandran, D. Ghoshdastidar and S. Senapati, *J. Am. Chem. Soc.*, 2012, **134**, 20330–20339.

[48] D. Ghoshdastidar and S. Senapati, *Nucleic Acids Res.*, 2018, **46**, 4344–4353.

[49] A. Garai, D. Ghoshdastidar, S. Senapati and P. K. Maiti, *J. Chem. Phys.*, 2018, **149**, 045104.

[50] H. Tateishi-Karimata and N. Sugimoto, *Nucleic Acids Res.*, 2014, **42**, 8831–8844.

[51] S. Pal and S. Paul, *Int. J. Biol. Macromol.*, 2019, **121**, 350 – 363.

[52] S. Pal and S. Paul, *J. Phys. Chem. C*, 2019, **123**, 11686–11698.

[53] D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER 2018*, University of California, San Francisco, 2018.

[54] S. Chowdhury and M. Bansal, *J. Phys. Chem. B*, 2001, **105**, 7572–7578.

[55] P. Stadlbauer, M. Krepl, I. Cheatham, Thomas E., J. Koča and J. Šponer, *Nucleic Acids Res.*, 2013, **41**, 7128–7143.

[56] N. Špačková, I. Berger and J. Šponer, *J. Am. Chem. Soc.*, 2001, **123**, 3295–3307.

[57] M. Cavallari, A. Calzolari, A. Garbesi and R. Di Felice, *J. Phys. Chem. B*, 2006, **110**, 26337–26348.

[58] A. E. Bergues-Pupo, J. R. Arias-Gonzalez, M. C. Morn, A. Fiasconaro and F. Falo, *Nucleic Acids Research*, 2015, **43**, 7638–7647.

[59] I. S. Joung and T. E. Cheatham III, *J. Phys. Chem. B*, 2008, **112**, 9020–9041.

[60] B. Islam, P. Stadlbauer, M. Krepl, M. Havrila, S. Haider and J. Sponer, *J. Chem. Theory Comput.*, 2018, **14**, 5011–5026.

[61] J. Šponer, G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R. A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka, N. G. Walter and M. Otyepka, *Chem. Rev.*, 2018, **118**, 4177–4338.

[62] B. Machireddy, G. Kalra, S. Jonnalagadda, K. Ramanujachary and C. Wu, *J. Chem. Inf. Model.*, 2017, **57**, 2846–2864.

[63] N. G. J. Richards and M. M. Georgiadis, *Acc. Chem. Res.*, 2017, **50**, 1375–1382.

[64] B. Islam, P. Stadlbauer, A. Gil-Ley, G. Pérez-Hernández, S. Haider, S. Neidle, G. Bussi, P. Banas, M. Otyepka and J. Sponer, *J. Chem. Theory Comput.*, 2017, **13**, 2458–2480.

[65] A. K. Sieradzan, P. Krupa and D. J. Wales, *J. Phys. Chem. B*, 2017, **121**, 2207–2219.

[66] P. Stadlbauer, L. Mazzanti, T. Cragnolini, D. J. Wales, P. Derreumaux, S. Pasquali and J. Šponer, *J. Chem. Theory Comput.*, 2016, **12**, 6077–6097.

[67] D. R. Gruber, J. J. Toner, H. L. Miears, A. V. Shernyukov, A. S. Kiryutin, A. A. Lomzov, A. V. Endutkin, I. R. Grin, D. V. Petrova, M. S. Kupryushkin, A. V. Yurkovskaya, E. C. Johnson, M. Okon, E. G. Bagryanskaya, D. O. Zharkov and S. L. Smirnov, *Nucleic Acids Res.*, 2018, **46**, 10827–10839.

[68] S. Haider, *J. Indian Inst. Sci*, 2018, **98**, 325–339.

[69] M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo and P. Jurečka, *J. Chem. Theory Comput.*, 2015, **11**, 5723–5736.

[70] R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka and T. E. Cheatham, *J. Chem. Theory Comput.*, 2016, **12**, 4114–4127.

[71] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton and M. Orozco, *Biophysical J.*, 2007, **92**, 3817 – 3829.

[72] L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, *J. Comput. Chem.*, **30**, 2157–2164.

[73] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

[74] D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.

[75] G. A. Petersson, A. Bennett, T. G. Tensfeldt, M. A. AlLaham, W. A. Shirley and J. Mantzaris, *J. Chem. Phys.*, 1988, **89**, 2193–2218.

[76] G. A. Petersson and M. A. AlLaham, *J. Chem. Phys.*, 1991, **94**, 6081–6090.

[77] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian09 Revision A.02*, 2016, Gaussian Inc. Wallingford CT.

[78] J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model.*, 2006, **25**, 247 – 260.

[79] C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.

[80] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, **25**, 1157–1174.

[81] S. L. Perkins, P. Painter and C. M. Colina, *J. Phys. Chem. B*, 2013, **117**, 10250–10260.

[82] S. L. Perkins, P. Painter and C. M. Colina, *J. Chem. Eng. Data*, 2014, **59**, 3652–3662.

[83] K. G. Sprenger, V. W. Jaeger and J. Pfaendtner, *J. Phys. Chem. B*, 2015, **119**, 5882–5895.

[84] M. H. Mamme, S. L. C. Moors, H. Terryn, J. Deconinck, J. Ustarroz and F. De Proft, *J. Phys. Chem. Lett.*, 2018, **9**, 6296–6304.

[85] E. O. Fetisov, D. B. Harwood, I.-F. W. Kuo, S. E. E. Warrag, M. C. Kroon, C. J. Peters and J. I. Siepmann, *J. Phys. Chem. B*, 2018, **122**, 1245–1254.

[86] D. Shah and F. S. Mjalli, *Phys. Chem. Chem. Phys.*, 2014, **16**, 23900–23907.

[87] C. Schröder, *Phys. Chem. Chem. Phys.*, 2012, **14**, 3089–3102.

[88] R. W. Pastor, B. R. Brooks and A. Szabo, *Mol. Phys.*, 1988, **65**, 1409–1419.

[89] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

[90] B. P. Uberuaga, M. Anghel and A. F. Voter, *J. Chem. Phys.*, 2004, **120**, 6363–6374.

[91] D. J. Sindhikara, S. Kim, A. F. Voter and A. E. Roitberg, *J. Chem. Theory Comput.*, 2009, **5**, 1624–1631.

[92] J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

[93] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

[94] W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.

[95] R. Kumar and H. Grubmüller, *Bioinformatics*, 2015, **31**, 2583–2585.

[96] M. Brehm and B. Kirchner, *J. Chem. Inf. Model.*, 2011, **51**, 2007–2023.

[97] S. Naskar, H. Joshi, B. Chakraborty, N. C. Seeman and P. K. Maiti, *Nanoscale*, 2019, **11**, 14863–14878.

[98] M. Havrila, P. Stadlbauer, B. Islam, M. Otyepka and J. Šponer, *J. Chem. Theory Comput.*, 2017, **13**, 3911–3926.

[99] Z. Jiang, L. You, W. Dou, T. Sun and P. Xu, *Polymers*, 2019, **11**, 282.

[100] M. Kaburagi, H. Yamada, T. Miyakawa, R. Morikawa, M. Takasu, T. A. Kato and M. Uesaka, *Polymer J.*, 2016, **48**, 189–195.

[101] M. Rebič, A. Laaksonen, J. Šponer, J. Uličny and F. Mocci, *J. Phys. Chem. B*, 2016, **120**, 7380–7391.

[102] X. Cang, J. Šponer and T. E. Cheatham III, *Nucleic Acids Res.*, 2011, **39**, 4499–4512.

[103] J. Šponer, J. E. Šponer, A. Mládek, P. Jurečka, P. Banáš and M. Otyepka, *Biopolymers*, 2013, **99**, 978–988.

[104] P. Banáš, A. Mládek, M. Otyepka, M. Zgarbová, P. Jurečka, D. Svozil, F. Lankaš and J. Šponer, *J. of Chem. Theo. and Comput.*, 2012, **8**, 2448–2460.

[105] J. Šponer, K. E. Riley and P. Hobza, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2595–2610.

[106] H. Kruse, P. Banáš and J. Šponer, *J. Chem. Theo. Comput.*, 2019, **15**, 95–115.

[107] V. Agieienko and R. Buchner, *J. Chem. Eng. Data*, 2019, **64**, 4763–4774.

[108] A. P. Abbott, R. C. Harris and K. S. Ryder, *J. Phys. Chem. B*, 2007, **111**, 4910–4913.

[109] A. P. Abbott, *ChemPhysChem*, 2004, **5**, 1242–1246.

[110] A. P. Abbott, *ChemPhysChem*, 2005, **6**, 2502–2505.

[111] G. Borgohain, B. Mandal and S. Paul, *Phys. Chem. Chem. Phys.*, 2017, **19**, 13160–13171.

[112] S. Paul and S. Paul, *J. Phys. Chem. B*, 2015, **119**, 9820–9834.

[113] S. Balasubramanian, S. Pal and B. Bagchi, *Phys. Rev. Lett.*, 2002, **89**, 115505.

[114] A. Chandra, *Phys. Rev. Lett.*, 2000, **85**, 768.

[115] S. Pal and S. Paul, *J. Phys. Chem. B*, 2020, **124**, 210–223.

[116] A. Chandra, *J. Phys. Chem. B*, 2003, **107**, 3899–3906.

[117] S. Paul and A. Chandra, *J. Chem. Theory Comput.*, 2005, **1**, 1221–1231.

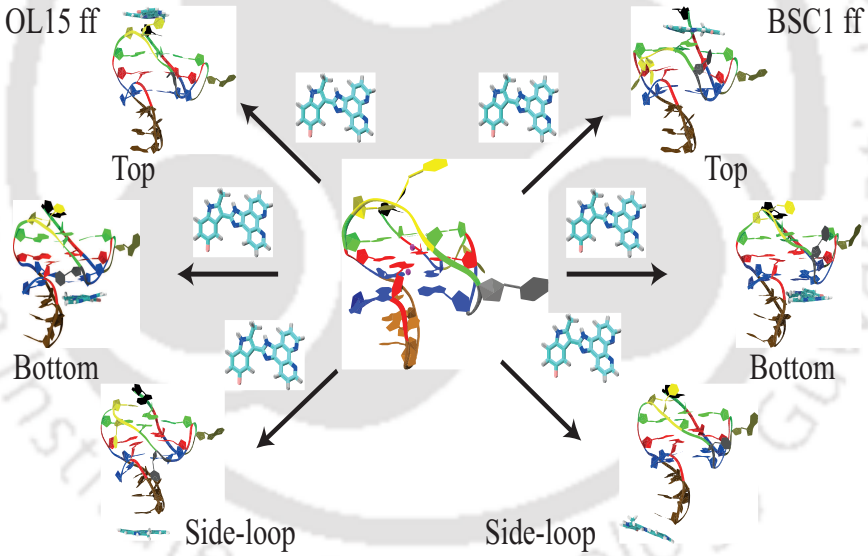
[118] S. Das and S. Paul, *J. Phys. Chem. B*, 2017, **121**, 8774–8785.

[119] G. Borgohain and S. Paul, *Comput. Theor. Chem.*, 2018, **1131**, 78 – 89.

[120] A. Pandey, R. Rai, M. Pal and S. Pandey, *Phys. Chem. Chem. Phys.*, 2014, **16**, 1559–1568.

Chapter 6

Investigation of Binding Modes and Pathway of APTO-253 on c-KIT G-quadruplex DNA



Overview

The stability of c-KIT G-quadruplex DNA by ligands has been a significant concern in the growing field of cancer therapy. Thus, it is very important to understand the mechanism behind the high binding affinity of the small drug molecules on the c-KIT G-quadruplex DNA. In this study, we have investigated the binding mode and pathway of APTO-253 ligand on the c-KIT G-quadruplex DNA employing a total of 10 μs all atom molecular dynamics simulation and further 8.82 μs simulation via umbrella sampling method using both OL15 and BSC1 latest force fields for DNA structures. From the cluster structure analysis, mainly three binding pathways i.e., top, bottom and side loop stacking modes are identified. Moreover, RMSD, RMSF and 2D-RMSD values indicate that the c-KIT G-quadruplex DNA and APTO-253 molecules are stable throughout the simulation run. Furthermore, the number of hydrogen bonds of each tetrads and the distance between the two central K^+ cations confirm that the c-KIT G-quadruplex DNA maintains its conformation in the process of complex formation with the APTO-253 ligand. The binding free energies and the minimum values in potential of mean forces suggest that the binding processes are energetically favorable. Further, we have found that the bottom stacking mode is the most favorable binding mode among all the three modes for OL15 force field. However, for BSC1 force field, both the top and bottom binding modes of APTO-253 ligand in c-KIT G-quadruplex DNA are comparable to each other. To investigate the driving force for the complex formation, we have noticed that the van der Waals (vdW) and π-π stacking interactions are mainly responsible. Our detailed studies provide an useful information for the discovery of novel drugs in the field of stabilization of G-quadruplex DNAs.

6.1 Introduction

In the promoter region of oncogene c-KIT, a type III receptor tyrosine kinase (RTK) is present which regulates the stem cell factors (SCF), as a result, triggers proliferation, differentiation and survival in haemopoietic precursor cells [1–3]. Destabilization of the c-KIT causes over expressions or mutations which lead to several diseases [4] such as gastrointestinal stromal tumors (GIST) [5, 6], mastocytosis, acute myelogenous leukemia (AML) or hematopoitic malignancy [7–9], prostate [10], adenocarcinoma lung cancers [11], pancreatic cancers [12], seminomas [13], melanoma [14]. In certain malignant growth types, for example, GIST, the fundamental driver of sub-atomic occasions in tumorogenesis is the initiating transformations in c-KIT. For this situation, focusing on c-KIT with Gleevec (imatinib mesylate) expands endurance by around 70%-80% following 2 years contrasted and disease treatment without c-KIT focusing on drugs [15, 16].

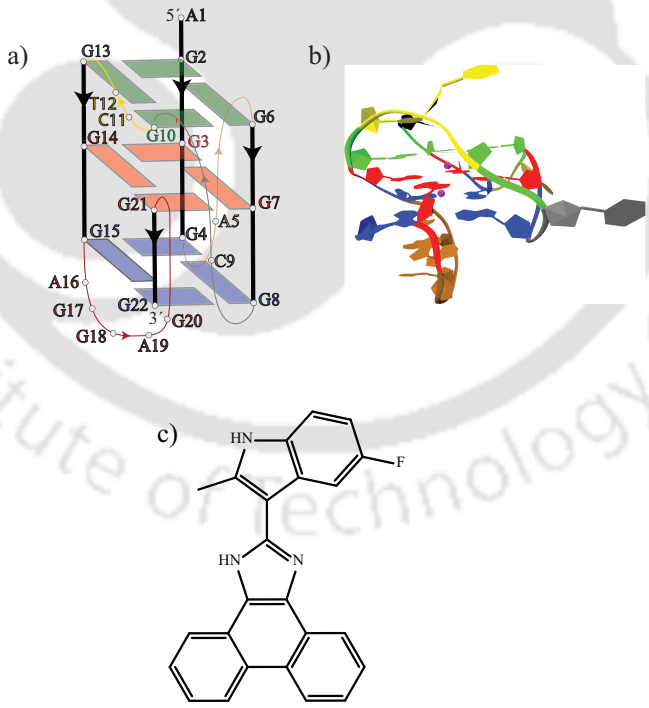


Figure 6.1: *Schematic representation of (a) c-KIT G-quadruplex DNA, new ribbons representation of (b) c-KIT G-quadruplex DNA, structural formula of (c) APTO-253 molecule, In new ribbons representation, tetrad-1, tetrad-2, tetrad-3, loop-1, loop-2, loop-3, and loop-4 are shown in green, red, blue, tan, gray, yellow, and ochre colors, respectively. In CPK representation, K^+ cations are shown in purple color. Adenine, guanine, thymine, and cytosine are represented as A, G, T, and C, respectively.*

Numerous studies [17, 18] confirmed that the c-KIT oncogene contains of 22 mer d(AGGGAGGGCGCTGGGAGGGAGGG) sequence (**Fig. 6.1 (a) and (b)**). Moreover, it has three tetrad regions (namely tetrad-1 (G2-G6-G10-G13), tetrad-2 (G3-G7-G21-G14) and tetrad-3 (G4-G8-G22-G15)) and four loop regions (namely loop-1 (A5), loop-2 (C9),loop-3 (C11, T12), and loop-4 (A16-G17-G18-A19-G20)).

In 2007, Balasubramanian and his co-worker developed six isoalloxazines ligands which are capable to induce and stabilize the c-KIT quadruplex DNA, but not duplex DNA and hence, possibly distinguish between quadruplex and duplex DNA [19]. Another pioneering work by Niedle and his colleagues introduced that the naphthalene diimide derivatives stabilized the c-KIT quadruplex DNA strongly and prevents the tumor for further growth in the GIST tumor cells [20]. Recently, Głuszyńska et al. reported that carbazole derivatives stabilize parallel c-MYC as well as c-KIT quadruplex DNA [21, 22]. Moreover, many small drugs/ligands stabilize the c-KIT quadruplex DNA, such as bisquinoline-pyrrole oligoamides [23], quinazolones [24], bistriazoles [25, 26], alkaloids and their derivatives [27, 28], oxazole-based peptide macrocycles [29], biaryl polyamides [30], indenoisoquinolines [31] and indenopyrimidines [32] , benzo[a]phenoxazines [33] , indolylmethyleneindanones [34] , terpyridines and their Pt(II) complexes [35] , triarylpyridines [36], cationic porphyrins [37] and pentaheteroaryls [38] , phthalocyanines [39] etc..

Nowadays, in drug development, computational methods have become important, as they offer comprehensive structural knowledge that may be missing in experimental findings. Among computational methods, all atom molecular dynamics has recently become a very popular method to find out the ligand binding mode and binding possibility on different types of DNA in therapeutic fields [40–46]. Recently, Wu and his coworker reported that the binding mode and selectivity of the daunomycin ligand on a parallel DNA quadruplex ($d(TGGGGT)_4$) and DNA duplex [47] by all atom molecular dynamics simulation. Moreover, in their another work [48], they analyzed how RHPS4 ligand binds with the telomeric G-quadruplex and duplex DNA using the same method. Furthermore, they have also shown the binding mode of different ligands, Telomestatin, TMPyP4, BSU6037, and BRACO19 on telomeric GQuadruplex DNA [49, 50]. These pioneer works provide the useful informations about drug binding on different types of G-quadruplex DNA in the field of drug discovery fields.

APTO-253 is currently in clinical development for the diagnosis of acute myeloid leukemia (AML). It is a small-molecule and a derivative of 2-indolyl imidazole [4,5-d] phenanthroline (**Fig. 6.1 (c)**), which is commonly cytotoxic to different cell lines of

human tumor cells, including strong tumours, tumors, and lymphomas [51]. APTO-253 has been advanced to a phase I test in solid tumor patients((NCT123226) [52]. In 2018, Rice and his colleagues discovered that APTO-253 stabilizes the guanine rich DNA i.e., G-quadruplex DNA using a modified FRET assay [53, 54]. Moreover, they observed that TMPyP4 and CX-5461 stabilized telomere, MYC, KIT as well as rRNA and duplex DNA [53]. In contrast, APTO-253 stabilized only G4 motifs i.e., telomere, MYC and KIT DNA but not rRNA and duplex DNA. It indicates that APTO-253 is more selective towards G4-DNA than TMPyP4 and CX-5461. Additionally, they showed that very low concentrations of APTO-253 decreased the expression of MYC and KIT in AML cell lines (In **Table 6.1** we present the IC_{50} values of different cell lines). They suggested that the stabilization of guanine-rich DNA by APTO-253 resulted in single and double strand breaks at replication forks and telomers; followed by cell cycle arrest apoptosis of cancer cell via DDR

Table 6.1: *APTO-253 IC_{50} values in leukemia and lymphoma cell lines.*

Disease Type	Cell Lines	$IC_{50}(\mu M)$ Mean
MCL	**Jeko-1**	0.057
MCL	**GRANTA-519**	0.082
Burkitt's	**Raji**	0.1
AML	**MOLM-13**	0.14
MCL	**Mino**	0.23
AML	**MV4-11**	0.24
AML	**EOL-1**	0.3
AML	**THP1**	0.34
Burkitt's	**Ramos**	0.35
AML	**HL-60**	0.46
AML	**SKM-1**	0.48
AML	**KG-1**	0.51
DLBCL	**SUDHL-6**	0.51
T-ALL	**Jurkat**	0.52
AML	**Nomo-1**	1.45
AML	**HEL92.1.7**	1.75

(DNA-damage response) pathway. Hence, the stabilization of guanine rich quadruplex DNA at oncogene and telomere regions is the key process to follow DDR pathway. Surprisingly, the mechanism of binding process and possible binding sites of the APTO-253 on G-quadruplex DNA, has not yet been elucidate.

To solve this dilemma and also to understand the details of the dynamics of the

c-KIT quadruplex DNA, we have performed all atom molecular dynamics simulation. In this work we have exclusively shown that the probable binding mode and binding site of the APTO-253 ligand on c-KIT quadruplex DNA. Starting from an unbound ligand and c-KIT quadruplex DNA, different binding modes are observed for the complex using different characterized methods. From RMSD, RMSF, and 2D-RMSD analyses, it is observed that c-KIT quadruplex DNA maintains its native state during the complex formation with APTO-253 ligand. Furthermore, we have found that the APTO-253 ligand mainly binds with c-KIT quadruplex DNA in three different modes i.e., top, bottom and side loop modes (mainly with respect to the position of the DNA binding with the ligand). The binding free energy values indicate that all these complex formations are energetically favorable which is further confirmed by potential mean forces values of umbrella sampling method.

6.2 Models and Simulation Method

6.2.1 System Preparation and Force Fields

The APTO-253 molecule was constructed using Gaussview [55]. At first, the geometry of the APTO-253 molecule was optimized at the ab initio HF/6-31G* [56, 57] level of theory with the help of Gaussian 16 [58] package. Then, the partial charges of this molecule were obtained using restrained electrostatic potential (RESP) [59, 60] charge derivation method in the ANTECHAMBER module in AMBER18 package [61]. The other force field parameters were taken from the general amber force field (GAFF) [62]. Those basis set and force fields were widely used in recent works [22, 40, 42, 48] on drug binding of different types of quadruplex DNA.

In this study, the initial coordinates of human c-KIT oncogene 5′-d(AGGGAGGGCGCTGGGAGGAGGG)-3′ was extracted from the NMR structure of PDB ID:2O3M [17]. Initially, there were no metal ion present in this NMR structure. So, two K^+ ions were placed in the channel of the center core of the c-KIT quadruplex DNA (one was placed in between tetrad-1 and tetrad-2 and other was placed in between tetrad-2 and tetrad-3). Since, c-KIT quadruplex DNA carries -21e electronic charge, another 19 K^+ ions were added to neutralize the DNA. TIP3P [63] water was used to solvate the DNA and drug molecule. TIP3P water specific AMBER-adapted Joung and Cheatham force-field [64] parameters were applied for K^+ ion . Following recent studies [49, 65–74], for c-KIT quadruplex DNA, leaprc.DNA.OL15 [75, 76] , which includes parmbsc0 [77] and

OL15 [75] corrections to the ff10 forcefield and leaprc.DNA.bsc1 [78], the updated version of the bsc0, were employed.

Initially, the neutralized DNA was placed in the center of the sphere and the drug molecule APTO-253 was placed at a distance of 40 Å from the center of the c-KIT quadruplex DNA. Then three layers of the sphere with radius of 20 Å, 30 Å, and 40 Å, consisting of 1000, 2000 and 3000 more water molecules were added respectively to immersed the drug molecule and DNA systems. Finally, 4000 water molecules were added to solvate overall systems. Each system was constructed with the help of PACKMOL [79] software. Then, the initial coordinates and topology for each system, were transformed into truncated octahedron box by CPPTRAJ [80].

Considering our previous works [81–83] to run the unbiased all-atom molecular dynamics simulation, the following steps were implemented for each of the system: (i) At first, energy minimization was done by 5000 steps of steepest descent method, followed by the same number of the steps of conjugate gradient method to eliminate bad contacts between quadruplex c-KIT DNA and water. (ii) After that, each system was heated very slowly from 0 to 300 K at an interim of 50 K for 25 ps in canonical (NVT) ensemble with 1 ps^{-1} collision frequency. (iii) Then, each system was equilibrated for 250 ps at the temperature of 300 K in canonical (NVT) ensemble. Throughout these above mentioned steps, all atoms of the quadruplex DNA including K^+ ions placed in center of the channel core and the drug molecule APTO-253 were fixed with a harmonic restraint of force constant 100 kcal mol^{-1} $Å^{-2}$. (iv) Next, the positional restrain were slowly removed to 25, 20, 15, 10, 5 and finally 0 kcal mol^{-1} $Å^{-2}$ with a series of minimization and equilibration. (v) Finally, for each system a 20 ns NPT equilibration was carried out to maintain the system density and this was followed by 1000 ns production run with a 5 ps^{-1} collision frequency at 1 atm physical pressure and 300 K temperature. The details of all systems studied are mentioned in **Table. 6.2**.

Table 6.2: *The number of molecules, drug initial state and force fields used in different systems.*

No. of				Drug	Force	Total
Q-DNA	**Ligand**	**Water**	**Run**	**Initial State**	**Field**	**Time (μs)**
0	1	2000	1	N/A	N/A	1
1	1	10000	1 × 5	free	OL15	5
1	1	10000	1 × 5	free	bsc1	5

In these steps, to control the temperature, Langevin dynamics [84] was employed and to maintain the physical pressure, Berendsen barostat [85–87] with a pressure relaxation time of 2 ps was implemented. With a tolerance of 10^{-5} Å SHAKE algorithm [88] was applied to restrain the covalent bonds involving hydrogen atoms. A 2 fs time integration step was implemented. A cut off radius of 12 Å was employed for all non bonded short-ranged interactions and for the long-ranged electrostatic interactions, particle mesh Ewald method [89] was applied. CPPTRAJ [80] of Amber-Tools19 [61] and Visual Molecular Dynamics (VMD) [90] were employed to analyze the trajectories and also as a visualization tool.

6.2.2 Convergence of Simulations

The root mean square deviation (RMSD) and root mean square fluctuation (RMSF) were calculated considering all heavy atoms of the quadruplex DNA and APTO-253. Here, it is worth mentioning that the initial structures of both quadruplex DNA and APTO-253 were considered as the reference structure. The cumulative RMSD value of the quadruplex DNA below 3 Å reveals that the quadruplex DNA was very stable through out the simulation. A cutoff atomic distance of 3.5 Å was used to measure the atomic contacts between the heavy atoms of the quadruplex DNA and the ligand molecule. To define the binding stability of the quadruplex DNA and ligand molecule, the contact number 10 was considered (see **Figures 6.2-6.3**).

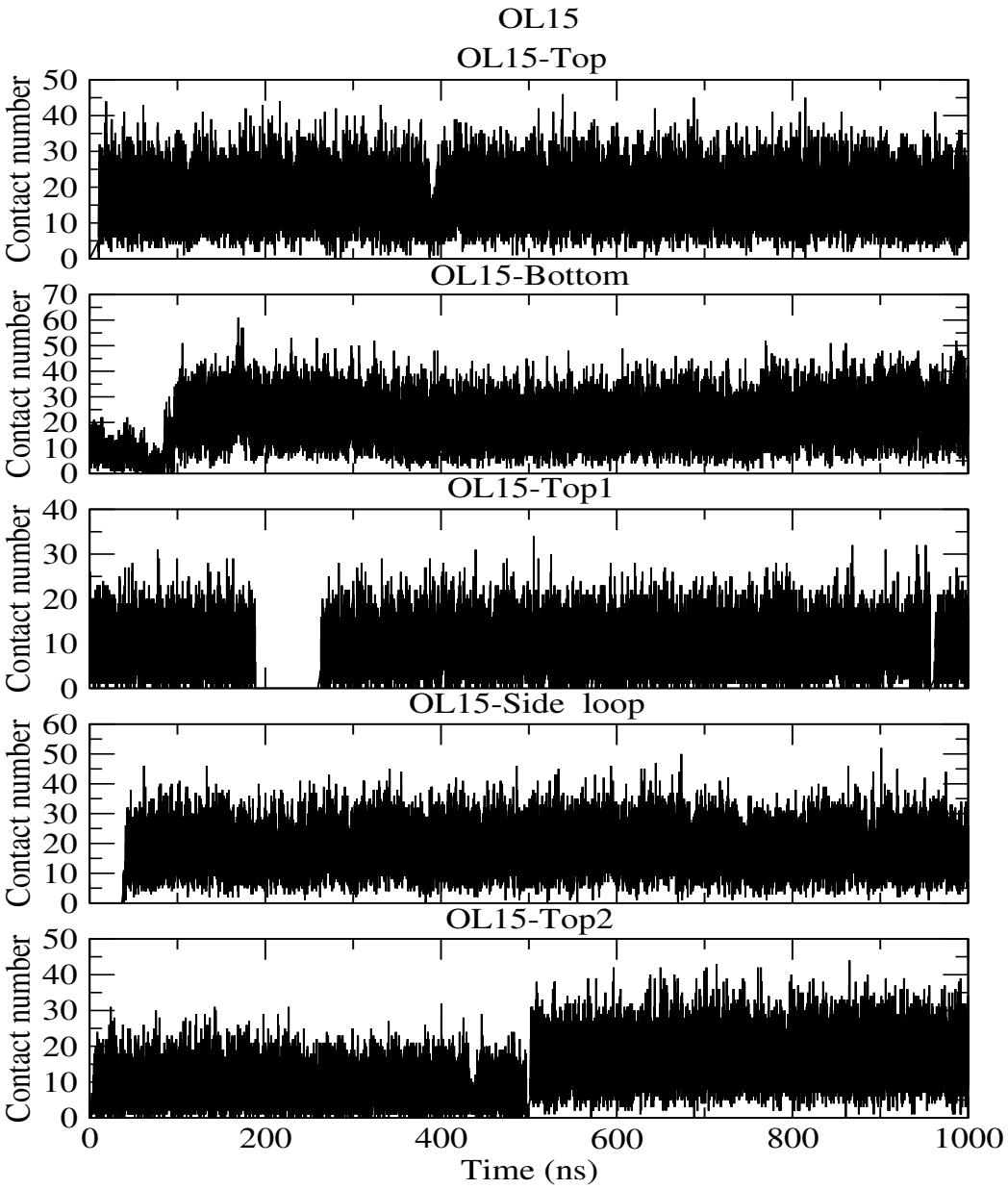


Figure 6.2: *Contact number of APTO-253 around 3.5 Å distance of c-KIT quadruplex DNA for different systems.*

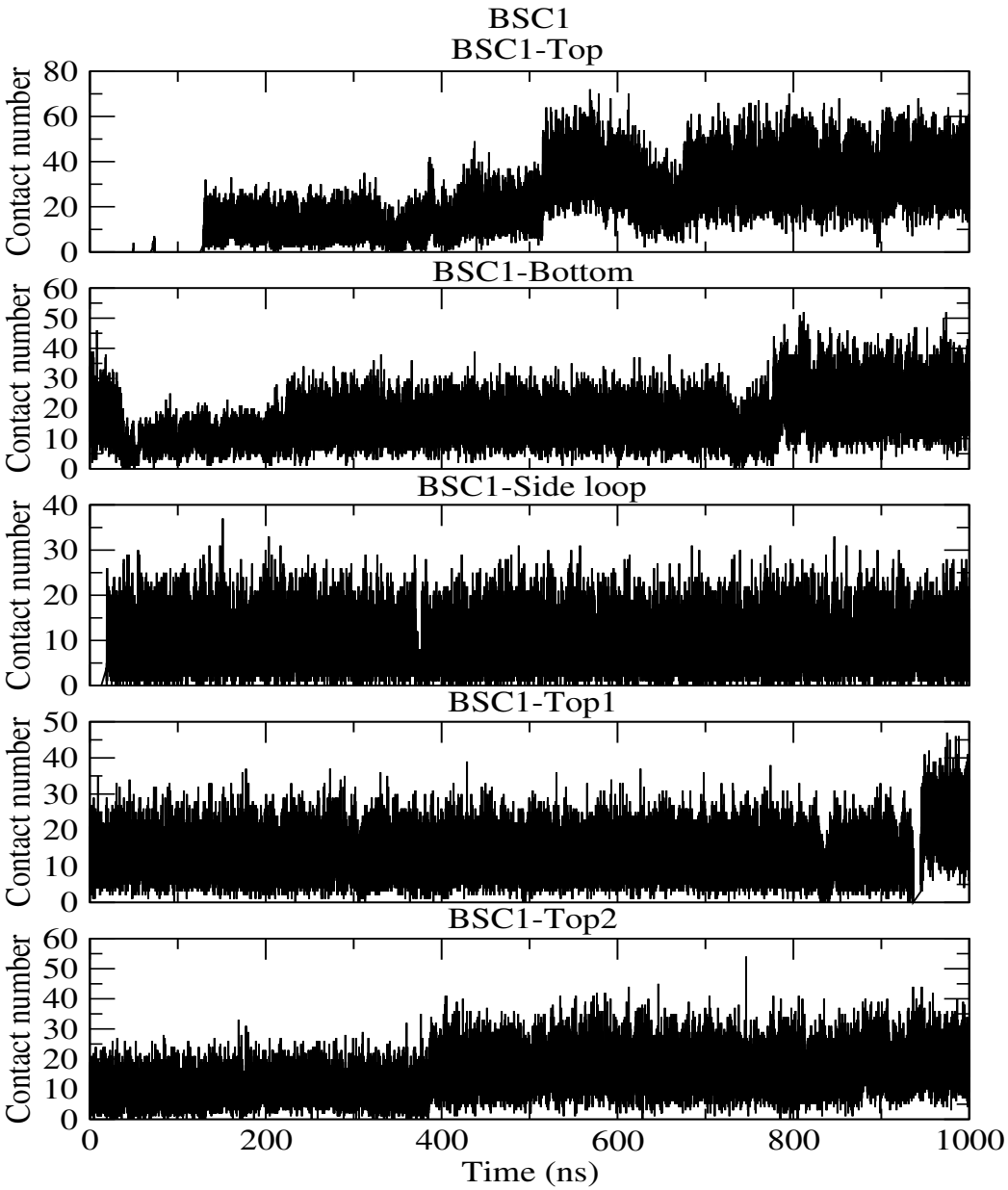


Figure 6.3: *Contact number of APTO-253 around 3.5 Å distance of c-KIT quadruplex DNA for different systems.*

6.2.3 Binding Mode Identification

To identify the binding mode of the drug molecule APTO-253 on c-KIT quadruplex DNA, we carried out the widely employed [91–93] Ester et al., Density Based Spatial Clustering of Applications with Noise (DBSCAN) [94] clustering algorithm. For the cluster analysis, we considered (for forming a cluster), the minimum number of points (minpoints) to be 25 and the distance cutoff between these points (ϵ) was 1.0 Å. The centroid structure was considered for the structural family.

6.2.4 Order Parameters to Characterize DNA-drug Binding Pathway

To investigate the details of drug binding on quadruplex DNA, we estimated the following order parameters: (i) RMSD and 2D-RMSD (ii) RMSF (iii) hydrogen bond analysis, (iv) distance, (v) MM-GBSA binding energy (ΔE), and (vi) π-π stacking. Root mean square fitted RMSD and RMSF of quadruplex DNA and the drug molecule were calculated. Considering previous studies [93, 95–99], for hydrogen bond analysis we had imposed two criteria (a) a distance cutoff between donor (D) and acceptor (A) of 3.5 Å and simultaneously, (b) a maximum of donor(D)-acceptor(A)-hydrogen(H) angle cutoff of 45°. The hydrogen bonds of three tetrad regions (namely tetrad-1, tetrad-2, and tetrad-3) were calculated to check the stability of the quadruplex DNA. The distances were calculated between center of mass (COM) of the quadruplex DNA and COM of the drug molecule for all systems and also between two K^+ ions that were present in the ion channel core of the quadruplex DNA. Following some pioneer works [48–50, 100], we estimated the MM-GBSA [101] (Molecular Mechanics Generalized Born-Surface Area) binding energy between ligand and quadruplex DNA. To find the MM-GBSA binding energy, we employed the python code MMPBSA.py [102] of MM-GBSA module of AMBER18 package [61] and along with the GB1 model with mBondi radii, a salt concentration of 0.1 M and a surface tension of 0.0072 kcal.$Å^{-2}$. MM-GBSA binding energy (ΔE) for each system was estimated by using the following equations,

$$\Delta E = E_{\text{complex}} - E_{\text{DNA\_free}} - E_{\text{lig\_free}} \tag{6.1}$$

$$\Delta E = \Delta E_{\text{vdW}} + \Delta E_{\text{EEL}} + \Delta E_{\text{ESURF}} + \Delta E_{\text{EGB}} + \Delta E_{\text{conformation}} \tag{6.2}$$

$$\Delta E_{\text{x}} = E_{\text{x\_complex}} - E_{\text{x\_DNA\_complex}} - E_{\text{x\_lig\_complex}} \tag{6.3}$$

where x = vdW, ESURF and EGB

$$\Delta E_{\text{Conformation}} = E_{\text{DNA\_complex}} + E_{\text{lig\_complex}} - E_{\text{DNA\_free}} - E_{\text{lig\_free}} \tag{6.4}$$

The above equations have the following terms, van der Waals interaction energy (vdW), electrostatic interaction (EEL), Non-polar solvation energy (ESURF), polar solvation energy (EGB) and the change of the conformation energy for quadruplex DNA and ligand molecules. Recent studies [101, 103–108] have shown that GB models predict the hydration-free energy quite accurately. Following previous reports [81, 82, 109–113], in order to consider the π-π stacking interactions are operative, we have set two criteria: (i) the cutoff distance between the COM of APTO-253 ligand and the COM of respective nucleobase ring, is 5 Å and at the same time (ii) the angle cutoff between the considered ring is 20°.

6.2.5 Umbrella Sampling

To find the possible binding mechanism of interaction between ligand and quadruplex DNA, umbrella sampling method [114] was used. Moreover, different initial state were employed to calculate the potentials mean forces (PMFs). Here, the Z-component distance (r) was used as reaction coordinate (ξ) and it was measured between the center of mass (COM) of specific residue of nucleobase ring and the ligand molecule. In this US process, simulations were performed in the presence of a biasing window potential w(ξ) with the goal of increasing sampling in the vicinity of the specified value ξ by considering a spring constant value of 1.0 kcal/mol at 300 K. The reaction coordinate was changed from 3 to 15 Å by a small increment of 0.25 Å. A total of 49 biased simulations were performed with i = 0,1,..,48 in the phase space configuration X by a harmonic potential:

$$w_i(\xi) = 1/2K\,(\xi(X) - \xi_i)^2 \tag{6.5}$$

where K is the force constant for harmonic restraint. To estimate the PMF, for each system, following procedure was followed: (a) each system was energy minimized as above mentioned protocol. Then, a 10 ns equilibration followed by a 20 ns production run at 300 K and 1 atm pressure, was performed for each window, leading to a total of a 1.47 μs simulation run for 49 windows. Then, the results were analyzed using the weighted histogram analysis method (WHAM) [115] to obtain the PMFs. Then, the PMF value was calculated P(ξ) from the following equation:

$$\mathrm{PMF}(\xi) = -k_{\mathrm{B}}T \ln P(\xi) \tag{6.6}$$

6.3 Results

6.3.1 Binding Mode and Binding Site Identification of ligand with c-KIT quadruplex DNA

For each system, the simulation has initiated from an unbound state of ligand and c-KIT quadruplex DNA which eventually results in complex formation. Thus, it is very important to investigate how the ligand molecule binds with the c-KIT quadruplex DNA during the

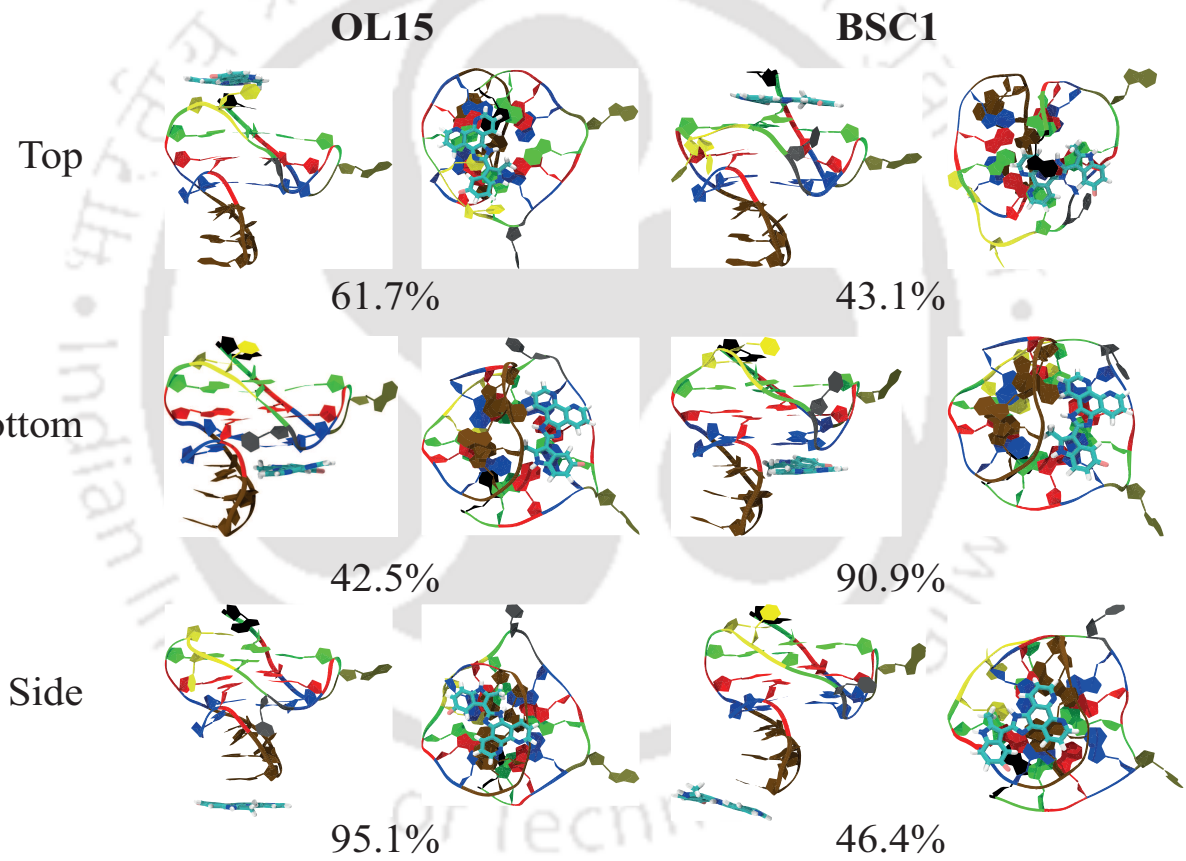


Figure 6.4: *Representative snapshots with percentage of the most populated structure for different systems.*

1000 ns production. So, we have performed cluster analysis (mentioned above in Method section) of each system for the last 200 ns trajectory. Here it is worth mentioning that 10 structural families (for both the OL15 and BSC1 force fields of c-KIT quadruplex DNA) of ligand-DNA complex were observed with more than 40% population (**Fig. 6.4**). Based on the position of the binding of the APTO-253 ligand with c-KIT quadruplex DNA, these

10 structural families can be merged in three binding mode categories: (i) top stacking (ii) bottom stacking (iii) side loop binding. In **Fig. 6.4**, we have presented the binding modes of the APTO-253 ligand with c-KIT quadruplex DNA.

6.3.2 Top Stacking Mode of APTO-253 on c-KIT quadruplex DNA

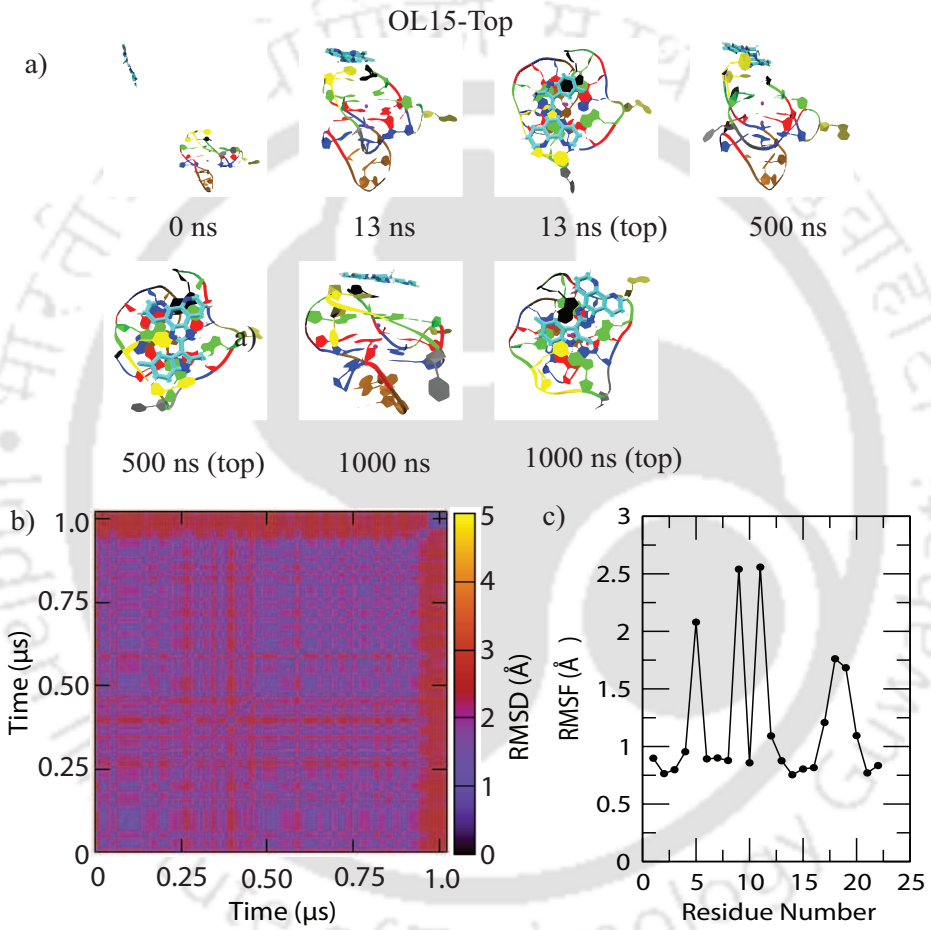


Figure 6.5: *(a) Snapshots representations of the complex formation of the c-KIT quadruplex DNA with time progress, (b) taking into account all heavy atoms of c-KIT G-quadruplex DNA, pairwise 2D-RMSDs , and (c) Root-mean-square fluctuations (RMSFs) of all heavy atoms of c-KIT G-quadruplex DNA.*

Here, we have discussed about how the APTO-253 ligand binds with the c-KIT quadruplex DNA in top stacking mode in OL15 force field. Initially, the APTO-253 ligand is placed 40 Å away from the c-KIT quadruplex DNA. However, at 13 ns, the ligand molecule binds with the c-KIT quadruplex DNA in the top position, i.e., it binds with A1 and T12 residues of the c-KIT quadruplex DNA and this binding is stable through out the 1000 ns simulation

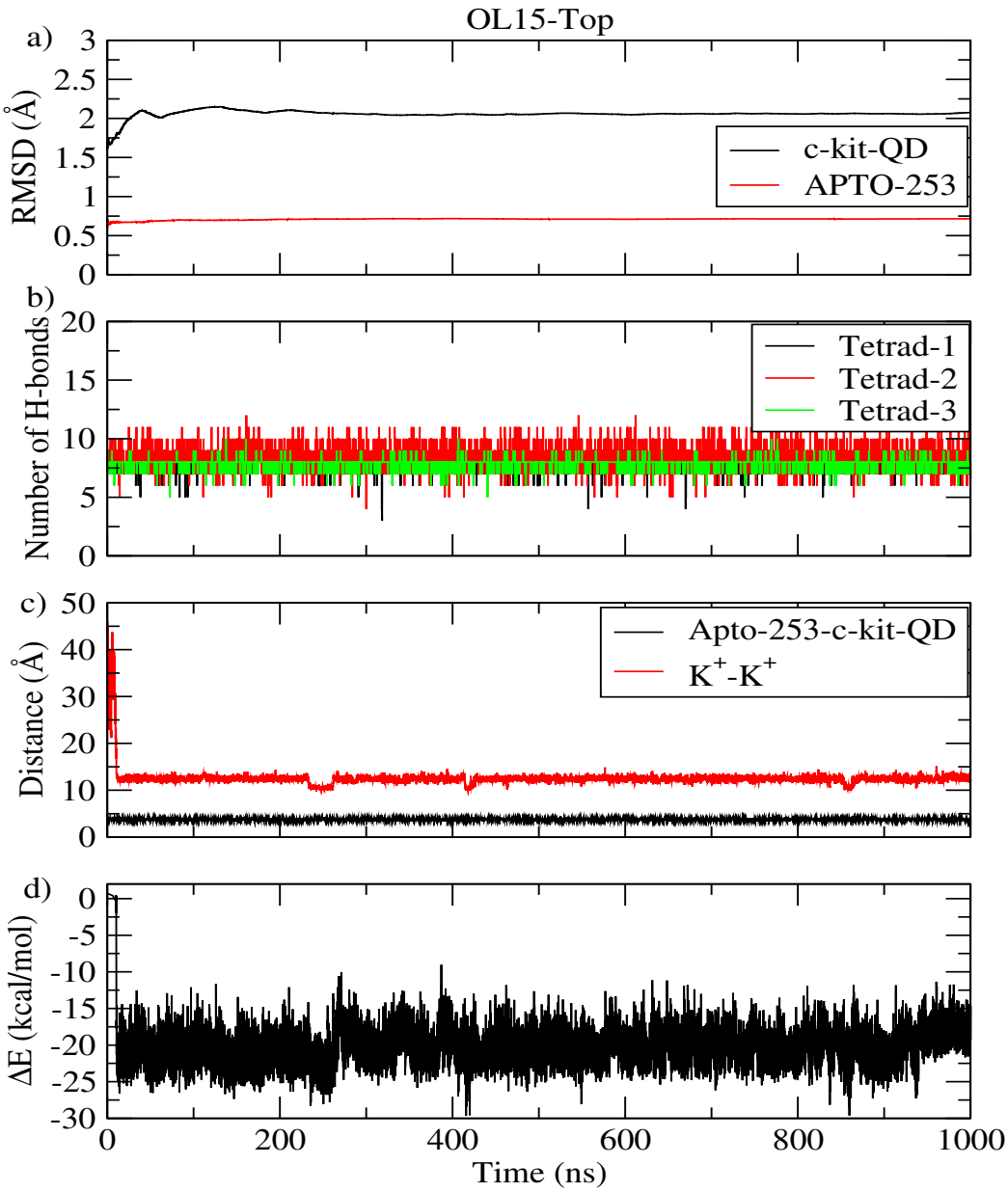


Figure 6.6: *(a) Time progression of the root-mean-square deviations (RMSDs) of all heavy atoms of c-KIT G-quadruplex DNA, (b) number of hydrogen bonds for tetrads, (c) the distance between center of masses of c-KIT G-quadruplex DNA and APTO-253, and the distance between the two K^+ central cations, and (d) the binding free energy of complex formation of APTO-253 ligand and c-KIT G-quadruplex DNA with time progression.*

run (**Fig. 6.5 (a)**). From **Fig. 6.6** (a), RMSD values of 2.08 Å and 0.74 Å of the c-KIT quadruplex DNA and APTO-253 respectively are noticed. Moreover, the 2D-RMSD and RMSF plots (**Fig. 6.5 (b) and (c)**) also display the low 2D-RMSD (below 3 Å) and RMSF (below 2.5 Å) values of c-KIT quadruplex DNA. These observations indicate that both the

molecules are very stable in complex mode. Furthermore, the hydrogen bonding between tetrads (namely tetrad-1, tetrad-2 and tetrad-3) residues are also maintained (nearly 8 for each tetrads) beautifully. The central cations play a very important role in conserving the structure of c-KIT quadruplex DNA [116, 117]. Thus, we have estimated the distance between two K^+ central cations which are placed in the central core of the ion channel as a function of simulation time. From **Fig. 6.6 (c)**, we observe that this distance is maintained at 3.38 Å which leads us to believe that the c-KIT quadruplex DNA is stable even after binding with the APTO-253. Moreover, distance between the center of masses of c-KIT quadruplex DNA and APTO-253 is initially at 40 Å, since they are placed 40 Å away from each other. However, with progression of time this distance is reduced. Next, we have calculated the binding free energy between APTO-253 ligand and c-KIT quadruplex DNA. The binding free energy (ΔE) value of -18.18 kcal/mol designates that the top stacking binding mode of APTO-253 ligand on c-KIT quadruplex DNA is energetically favorable (**Fig. 6.6 (d)**).

Next, we have demonstrated the top stacking mode of APTO-253 on c-KIT quadruplex DNA in BSC1 force field. Interestingly, here, APTO-253 binds in a different manner. From **Fig. 6.7 (a)**, we have seen that initially, the distance between APTO-253 and c-KIT quadruplex DNA is 40 Å but after 130 ns, it binds with the C9 residue, which is stable upto 380 ns. After that, the APTO-253 binds with the tetrad-1 residues i.e., G2, G6, G10 and G13 residues of c-KIT quadruplex DNA which continues upto 600 ns. After 600 ns, A1 residue of loop-3 binds with APTO-253 in 'sandwich' manner (A1-(top)-APTO-253-(middle)-tetrad-1-(bottom)) and this complex formation is stable upto 1000 ns production run. Now, it is more important to check the stability of the quadruplex DNA and APTO-253 molecules. **Fig. 6.8 (a)**, represents that the RMSD values are 2.57 Å and 0.71 Å for the c-KIT quadruplex DNA and APTO-253 molecules respectively. Moreover, the 2D-RMSD value of the c-KIT quadruplex DNA is below 3.5 Å (**Fig. 6.7 (b)**). However, slightly higher RMSF value (below 4 Å (**Fig. 6.7 (c)**) of T12 residue indicate that it fluctuates more than the other residues due to the stable 'sandwich' formation. Furthermore, the number hydrogen bonds (8 for each tetrads) is also maintained (**Fig. 6.8 (b)**). The distance between two central K^+ cations is 3.64 Å which demonstrates that the stability of the c-KIT quadruplex DNA in ligand binding complex formation (**Fig. 6.8 (c)**). Moreover, the binding free energy (ΔE) of the c-KIT quadruplex DNA and APTO-253 is -34.34 kcal/mol (**Fig. 6.8 (d)**), which indicates that the complex formation is energetically very much favorable for this case too.

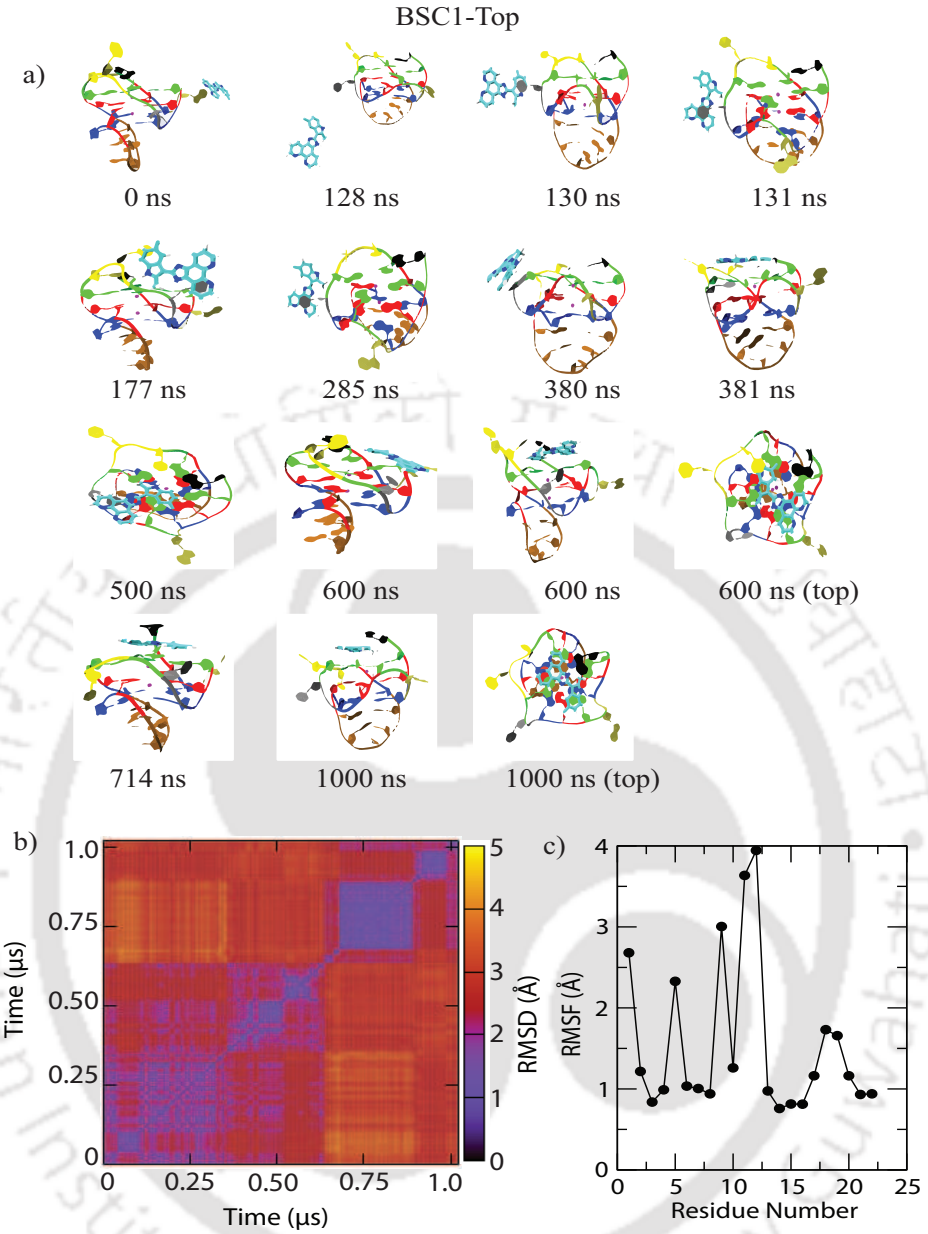


Figure 6.7: *(a) Snapshots representations of the complex formation of the c-KIT quadruplex DNA with time progress, (b) taking into account all heavy atoms of c-KIT G-quadruplex DNA, pairwise 2D-RMSDs , and (c) Root-mean-square fluctuations (RMSFs) of all heavy atoms of c-KIT G-quadruplex DNA.*

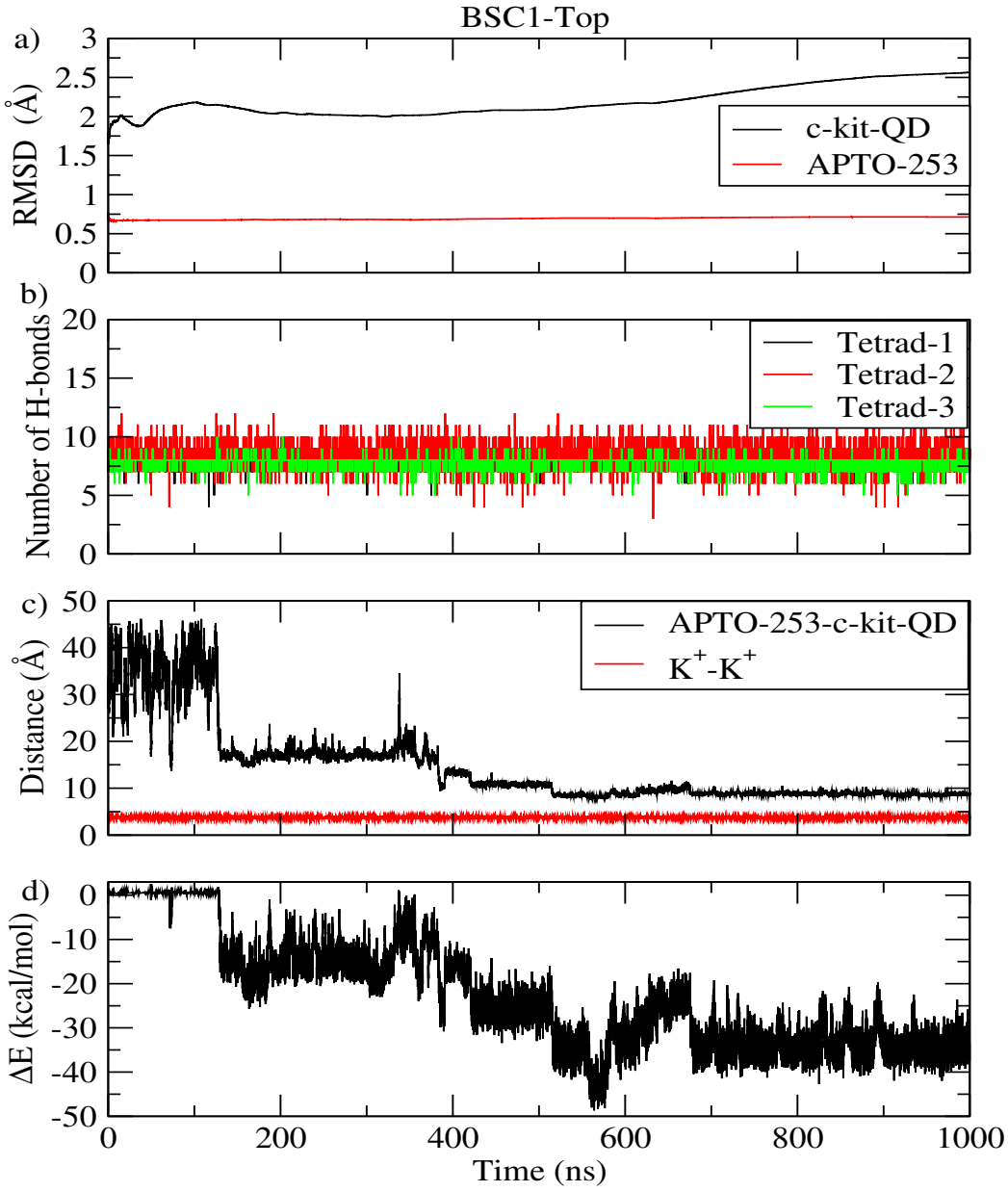


Figure 6.8: *(a) Time progression of the root-mean-square deviations (RMSDs) of all heavy atoms of c-KIT G-quadruplex DNA, (b) number of hydrogen bonds for tetrads, (c) the distance between center of masses of c-KIT G-quadruplex DNA and APTO-253, and the distance between the two K^+ central cations, and (d) the binding free energy of complex formation of APTO-253 ligand and c-KIT G-quadruplex DNA with time progression.*

6.3.3 Bottom Stacking Mode of APTO-253 on c-KIT quadruplex DNA

In this section we have discussed about the bottom stacking mode of APTO-253 on c-KIT quadruplex DNA for both OL15 and BSC1 force fields. At first, we have examined the OL15 system. Here, we have noticed that the APTO-253 ligand binds with the loop-4

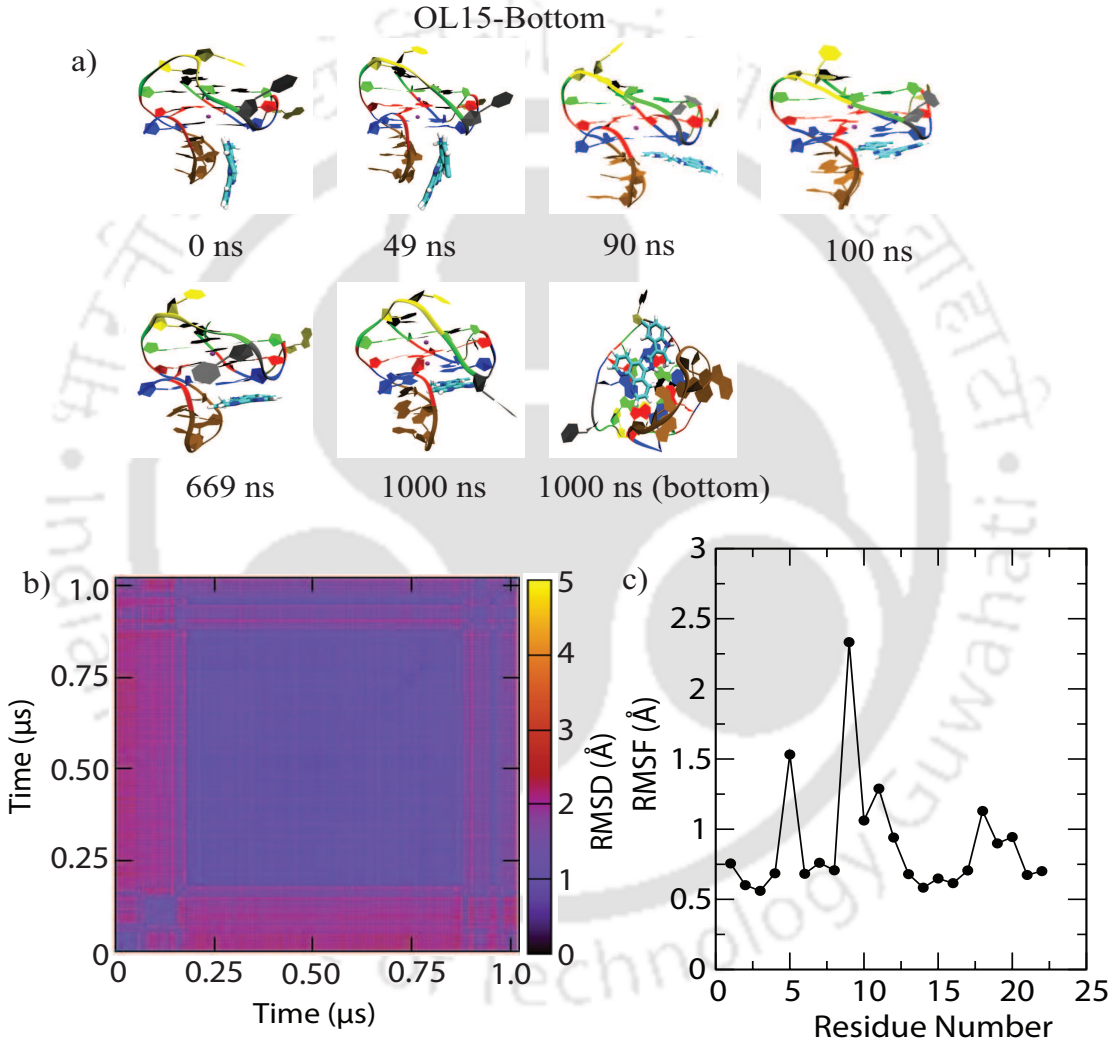


Figure 6.9: *(a) Snapshots representations of the complex formation of the c-KIT quadruplex DNA with time progress, (b) taking into account all heavy atoms of c-KIT G-quadruplex DNA, pairwise 2D-RMSDs of different systems, and (c) Root-mean-square fluctuations (RMSFs) of all heavy atoms of c-KIT G-quadruplex DNA for different systems.*

residues of c-KIT quadruplex DNA (**Fig. 6.9 (a)**) at the onset of the production run. This indicates that the binding occurs during equilibration and this side loop binding mode continues upto 50 ns of production. After that, the APTO-253 ligand shifts to the

G4 residue of the tetrad-3 region of the c-KIT quadruplex DNA at 90 ns. However, this formation is stable only upto 100 ns. After 100 ns of the production run, it binds to two residues i.e., G4 and G8 and this bottom stacking mode is stable upto 1000 ns simulation

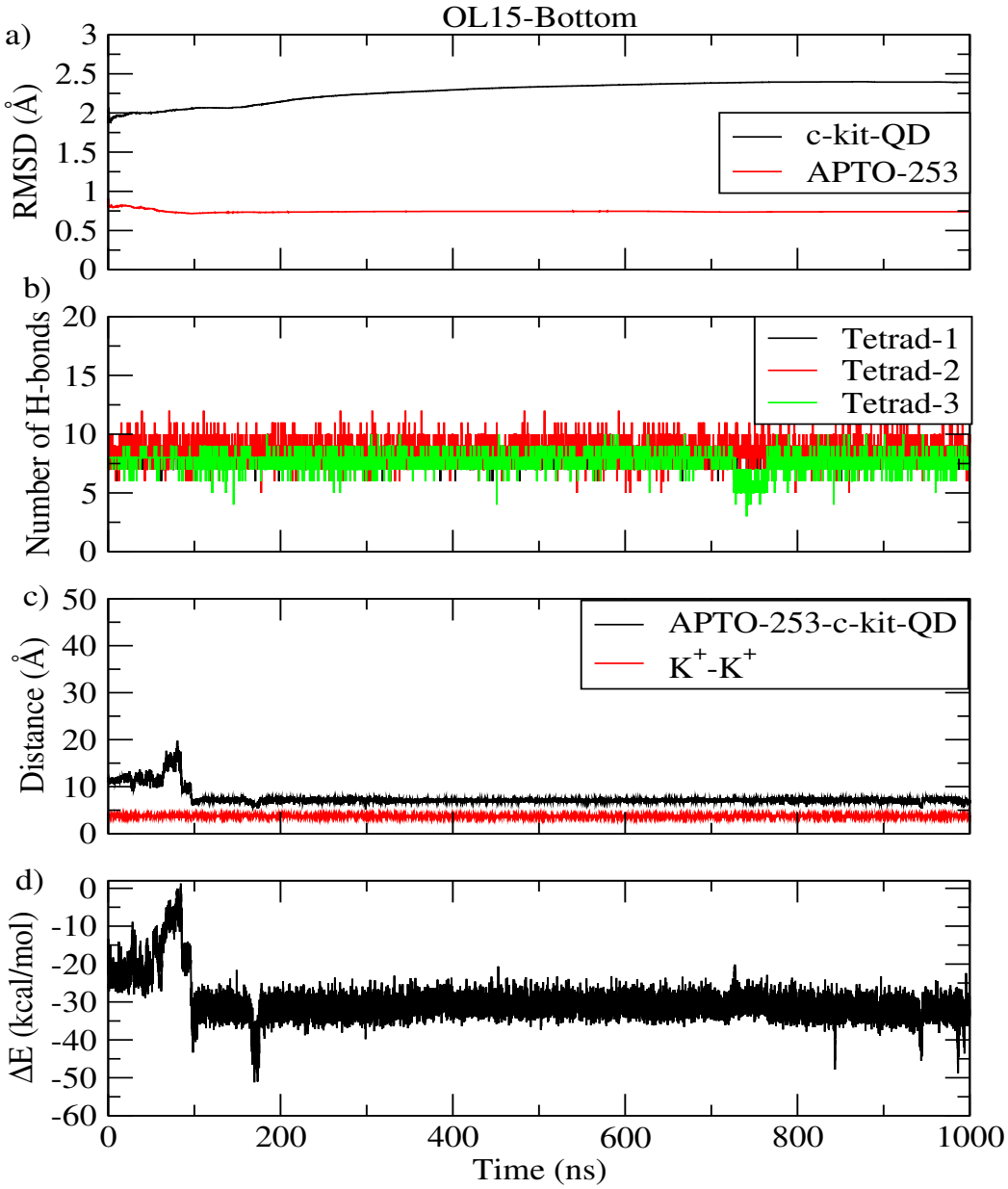


Figure 6.10: *(a) Time progression of the root-mean-square deviations (RMSDs) of all heavy atoms of c-KIT G-quadruplex DNA, (b) number of hydrogen bonds for tetrads, (c) the distance between center of masses of c-KIT G-quadruplex DNA and APTO-253, and the distance between the two K^+ central cations, and (d) the binding free energy of complex formation of APTO-253 ligand and c-KIT G-quadruplex DNA with time progression.*

run (**Fig. 6.9 (a)**). In **Fig. 6.10 (a)**, the low RMSD values of the c-KIT quadruplex DNA (2.39 Å) and APTO-253 (0.74 Å) are noticed. Moreover, the lower 2D-RMSD and RMSF values (below 3 Å) support the previous observations (**Fig. 6.9 (b) and (c)**). From **Fig. 6.10 (b)**, it can be observed that the number of hydrogen bonds of all tetrads are pretty similar (nearly 8 for each tetrads), which leads us to conclude that the tetrad formation is stable throughout the simulation run. Moreover, the distance between two K^+ central cations (shown in **Fig. 6.10 (c)**) is 3.71 Å. From the above investigations, it is clear that both the c-KIT quadruplex DNA and APTO-253 are very stable in bottom stacking mode. Next, we have estimated the binding free energy (ΔE) of this complex formation, which is shown in **Fig. 6.10 (d)**. The value of ΔE is -34.24 kcal/mol, which leads us to conclude that this binding is also favorable.

Next, the BSC1 force field system is considered. In **Fig. 6.11 (a)**, we have represented the snapshots (without water molecules) of the system. Here, we have observed that the APTO-253 molecule binds with the A1 and T12 residues of the c-KIT quadruplex DNA in top stacking mode at the starting of the production run which further indicates that the binding occurs during equilibration step. However, this complex formation is disrupted after 40 ns and the ligand molecule binds with C9 residue of the loop-2 region (side loop binding mode) which continues upto 60 ns simulation run. After that, the APTO-253 ligand molecule is shifted to the loop-4 region of the c-KIT quadruplex DNA at 80 ns. This arrangement pursues upto 780 ns. Finally, the ligand molecule binds with G4 and G8 of the tetrad-3 region of the c-KIT quadruplex DNA. This complex formation is continued upto 1000 ns simulation run. In this simulation run, we have observed all types of stacking modes. However, interestingly, the lower RMSD values 2.28 Å and 0.73 Å of c-KIT quadruplex DNA and APTO-253 ligand molecule respectively (**Fig. 6.12 (a)**) denote that the deviation is very small from the initial structure for both the molecules. Moreover, the lower 2D-RMSD and RMSF values (below 3 Å for both the cases) also guided the previous observation (**Fig. 6.11 (b) and (c)**). From hydrogen bond analysis, the number of hydrogen bonds of each tetrads are nearly 8 (**Fig. 6.12 (b)**), which leads us to conclude that the c-KIT quadruplex DNA is stable through out the simulation run. Furthermore, the distance between two central K^+ ion is 3.66 Å (**Fig. 6.12 (c)**). From overall observations, it can be concluded that the c-KIT quadruplex DNA is very stable through out the simulation run. A value of -30.91 kcal/mol for the binding free energy (**Fig. 6.12 (d)**) points out that the bottom stacking is again favorable.

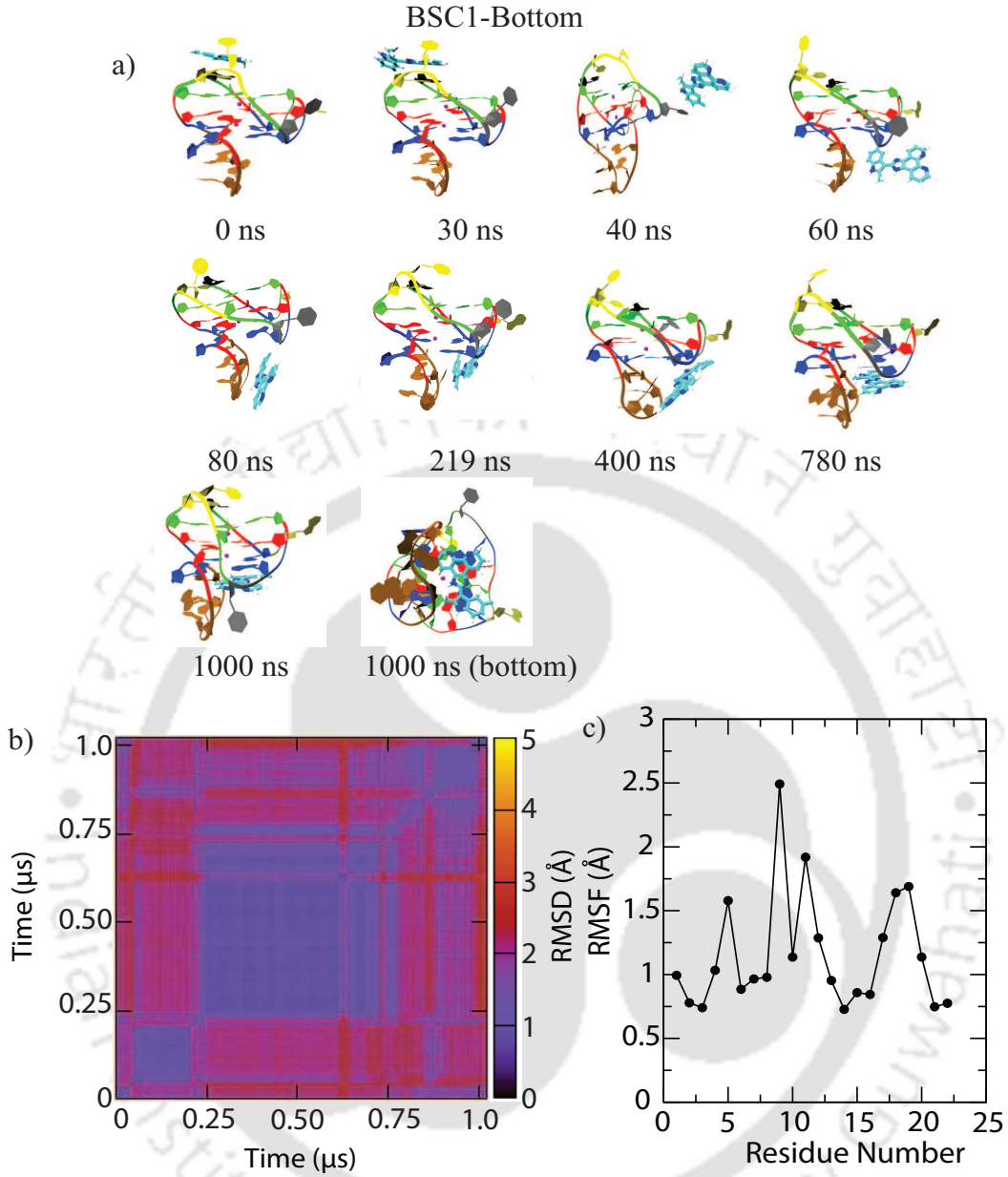


Figure 6.11: *(a) Snapshots representations of the complex formation of the c-KIT quadruplex DNA with time progress, (b) taking into account all heavy atoms of c-KIT G-quadruplex DNA, pairwise 2D-RMSDs of different systems, and (c) Root-mean-square fluctuations (RMSFs) of all heavy atoms of c-KIT G-quadruplex DNA for different systems.*

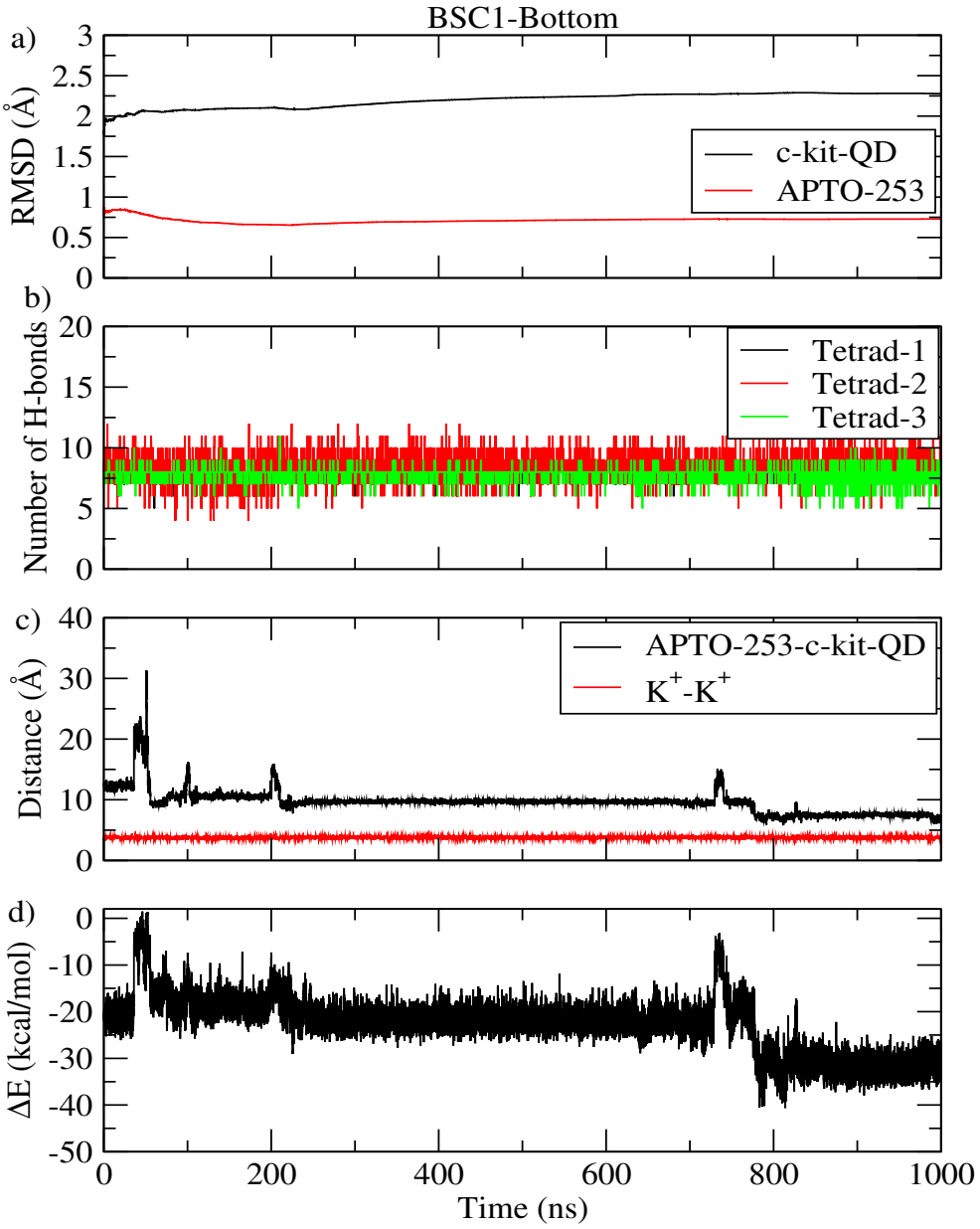


Figure 6.12: *(a) Time progression of the root-mean-square deviations (RMSDs) of all heavy atoms of c-KIT G-quadruplex DNA, (b) number of hydrogen bonds for tetrads, (c) the distance between center of masses of c-KIT G-quadruplex DNA and APTO-253, and the distance between the two K^+ central cations, and (d) the binding free energy of complex formation of APTO-253 ligand and c-KIT G-quadruplex DNA with time progression.*

6.3.4 Side Loop Stacking Mode of APTO-253 on c-KIT quadruplex DNA

Apart from the top and bottom stacking binding modes, we have also observed another binding mode i.e., side loop binding mode. For OL15 force field, (**Fig. 6.13 (a)**), we have noticed that APTO-253 molecule binds with A5 residue of loop-1 region of the c-KIT quadruplex DNA at the starting of the production run. This indicates that the binding occurs during 20 ns equilibration step. However, this complex formation is stable only upto 190 ns. After that, at 280 ns, it shifts to the loop-4 region and binds with G18 residue of the c-KIT quadruplex DNA. This side loop binding mode is noticed till the end of the simulation run. In **Fig. 6.14 (a)**, we have presented the RMSD values of both the c-KIT

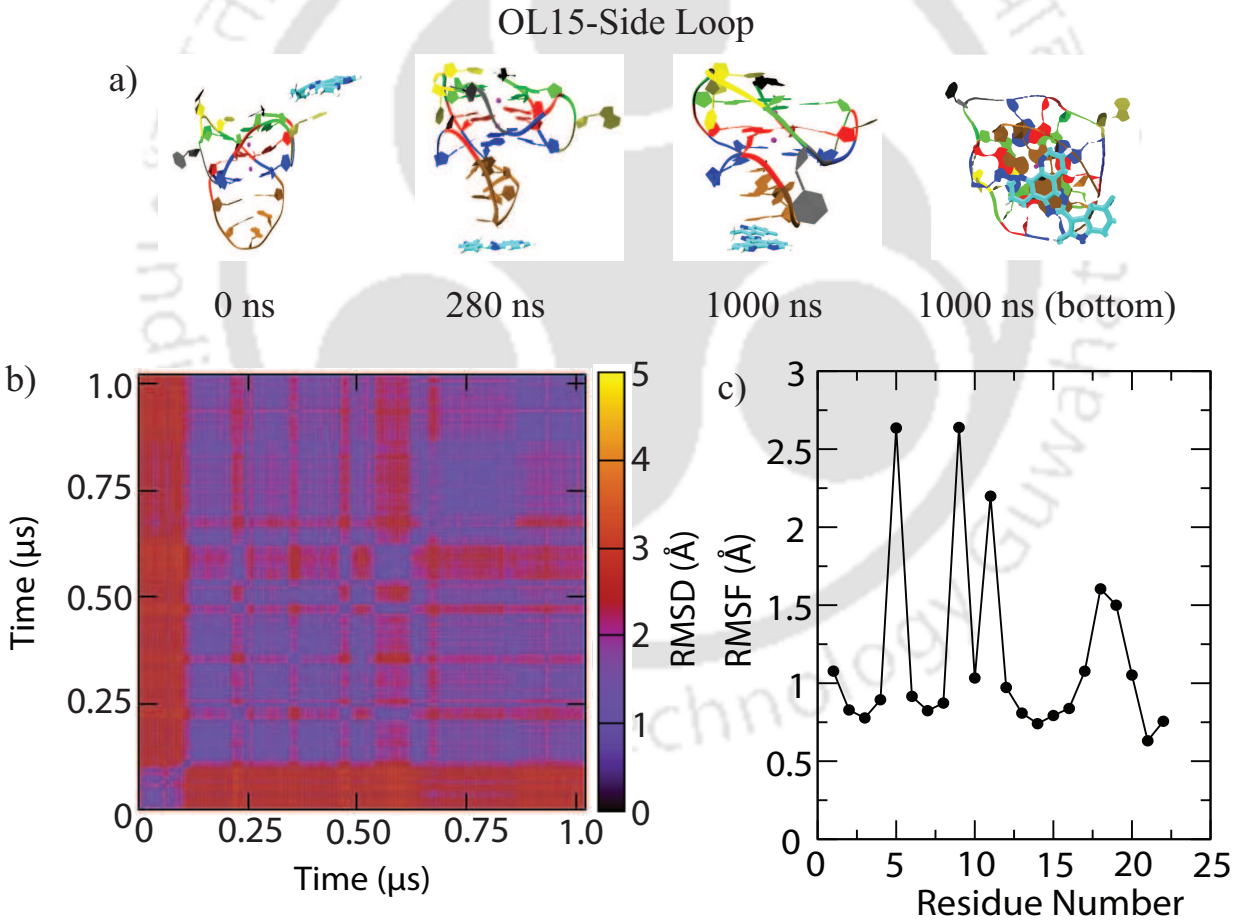


Figure 6.13: *(a) Snapshots representations of the complex formation of the c-KIT quadruplex DNA with time progress, (b) taking into account all heavy atoms of c-KIT G-quadruplex DNA, pairwise 2D-RMSDs of different systems, and (c) Root-mean-square fluctuations (RMSFs) of all heavy atoms of c-KIT G-quadruplex DNA for different systems.*

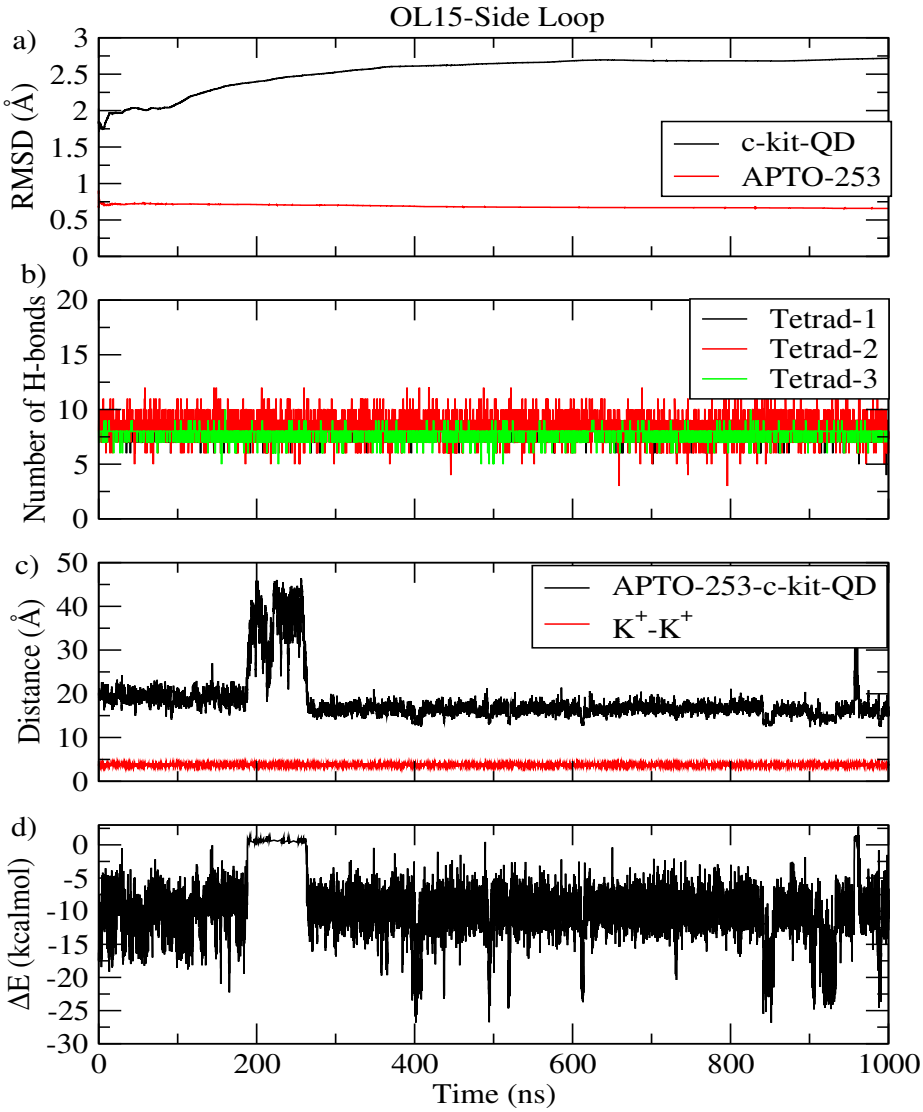


Figure 6.14: *(a) Time progression of the root-mean-square deviations (RMSDs) of all heavy atoms of c-KIT G-quadruplex DNA, (b) number of hydrogen bonds for tetrads, (c) the distance between center of masses of c-KIT G-quadruplex DNA and APTO-253, and the distance between the two K^+ central cations, and (d) the binding free energy of complex formation of APTO-253 ligand and c-KIT G-quadruplex DNA with time progression.*

quadruplex DNA and APTO-253 molecules. Here, the RMSD values of the c-KIT quadruplex DNA and the APTO-253 ligand molecules are 2.72 Å and 0.65 Å respectively. Moreover, from **Fig. 6.13 (b) and (c)**, the lower (below 3 Å) 2D-RMSD and RMSF values are noticed. Next, we have calculated the number of hydrogen bonds for all tetrads regions. The number of hydrogen bonds (**Fig. 6.14 (b)**) for each tetrad is 8 which again signifies

that the tetrad residues have maintained their arrangements through out the simulation run. Moreover, the distance between the two central K^+ ions which are placed in the central core of the ion channels, is 3.74 Å(**Fig. 6.14 (c)**). From the above observations, it is clear that the c-KIT quadruplex DNA and APTO-253 molecules very slightly deviate from their initial structures. Now, in **Fig. 6.14 (d)**, the MM-GBSA free binding energy of the APTO-253 ligand and the c-KIT quadruplex DNA complex formation are presented. A value of −9.46 kcal/mol for the MM-GBSA free binding energy (ΔE) indicates that the side loop binding mode of the complex formation is favorable.

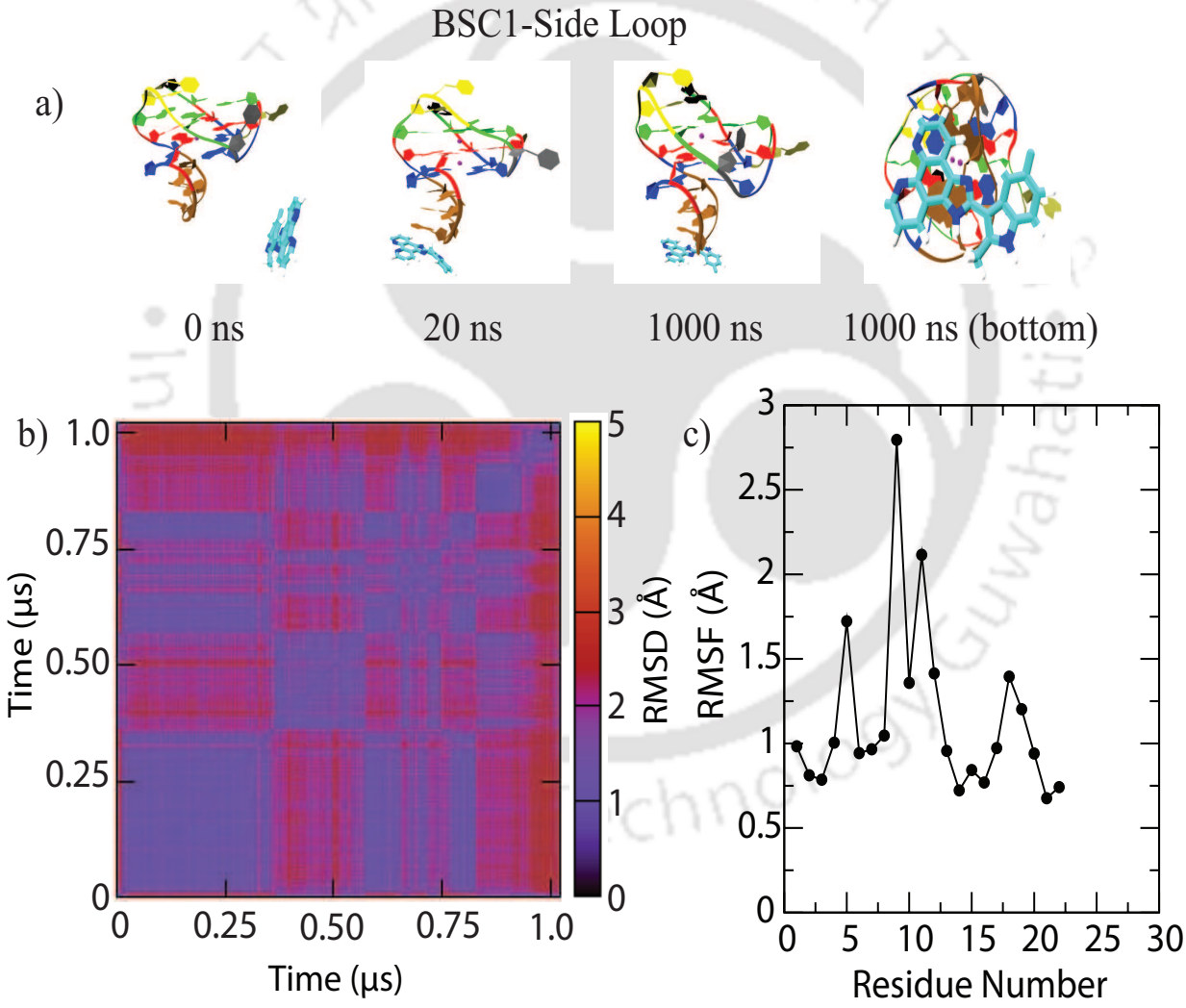


Figure 6.15: *(a) Snapshots representations of the complex formation of the c-KIT quadruplex DNA with time progress, (b) taking into account all heavy atoms of c-KIT G-quadruplex DNA, pairwise 2D-RMSDs of different systems, and (c) Root-mean-square fluctuations (RMSFs) of all heavy atoms of c-KIT G-quadruplex DNA for different systems.*

Now, the BSC1 force field system is considered. In **Fig. 6.15 (a)**, the time evolution snapshots of the complex are presented. At the starting of the simulation run, the distance between the COM of APTO-253 and the COM of the c-KIT quadruplex DNA is nearly 40 Å. With simulation progression, the APTO-253 ligand molecule shifts to the

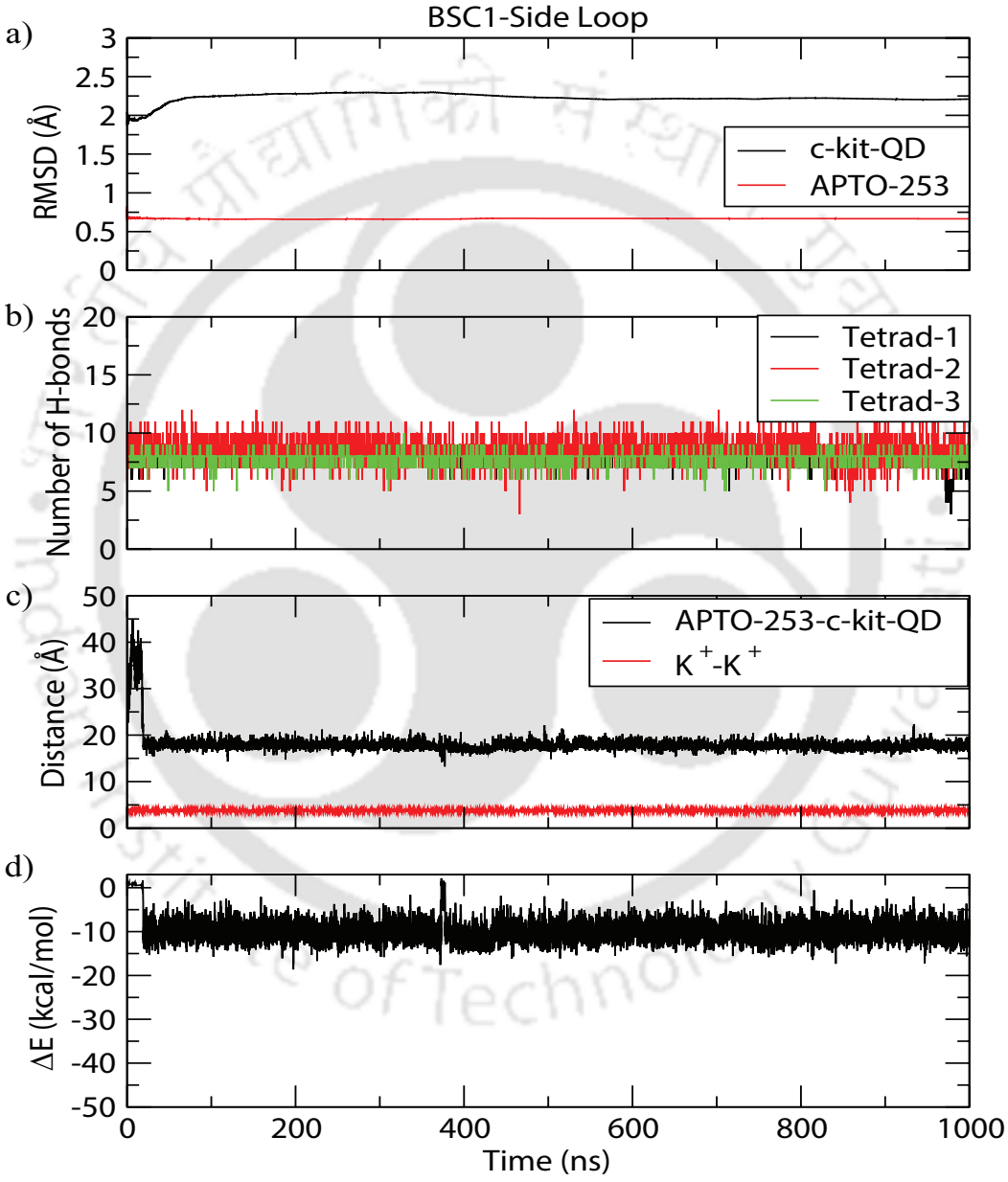


Figure 6.16: *(a) Time progression of the root-mean-square deviations (RMSDs) of all heavy atoms of c-KIT G-quadruplex DNA, (b) number of hydrogen bonds for tetrads, (c) the distance between center of masses of c-KIT G-quadruplex DNA and APTO-253, and the distance between the two K^+ central cations, and (d) the binding free energy of complex formation of APTO-253 ligand and c-KIT G-quadruplex DNA with time progression.*

loop-4 region of the c-KIT quadruplex DNA and binds with the G18 residue of loop-4 at 20 ns. Again, this side loop binding mode of the APTO-253 on the c-KIT quadruplex DNA is noticed through out the 1000 ns simulation run. RMSD deviations from the initial structures of both the c-KIT quadruplex DNA and the APTO-253 molecules are again analyzed. The RMSD values are 2.21 Å and 0.66 Å for the c-KIT quadruplex DNA and the APTO-253 molecules respectively (**Fig. 6.16 (a)**). From **Fig. 6.15 (b) and (c)**, the values of 2D-RMSD and RMSF below 3 Å confirm that the c-KIT quadruplex DNA have very low deviation from the initial structure through out the 1000 ns simulation run. Next, the number of hydrogen bonds for each tetrad of the c-KIT quadruplex DNA (**Fig. 6.16 (b)**) is calculated to be 8 and the distance between the two central K^+ cations is 3.73 Å (**Fig. 6.16 (c)**) which leads us to conclude that the c-KIT quadruplex DNA and APTO-253 ligand molecules is stable though out the 1000 ns simulation run in the side loop binding mode. Furthermore, we have presented the MM-GBSA free binding energy of this binding mode in **Fig. 6.16 (d)**, which has a value of −10.3 kcal/mol, indicating that the complex formation is favorable.

Other simulations results are shown in **Figures 6.17-6.24**.

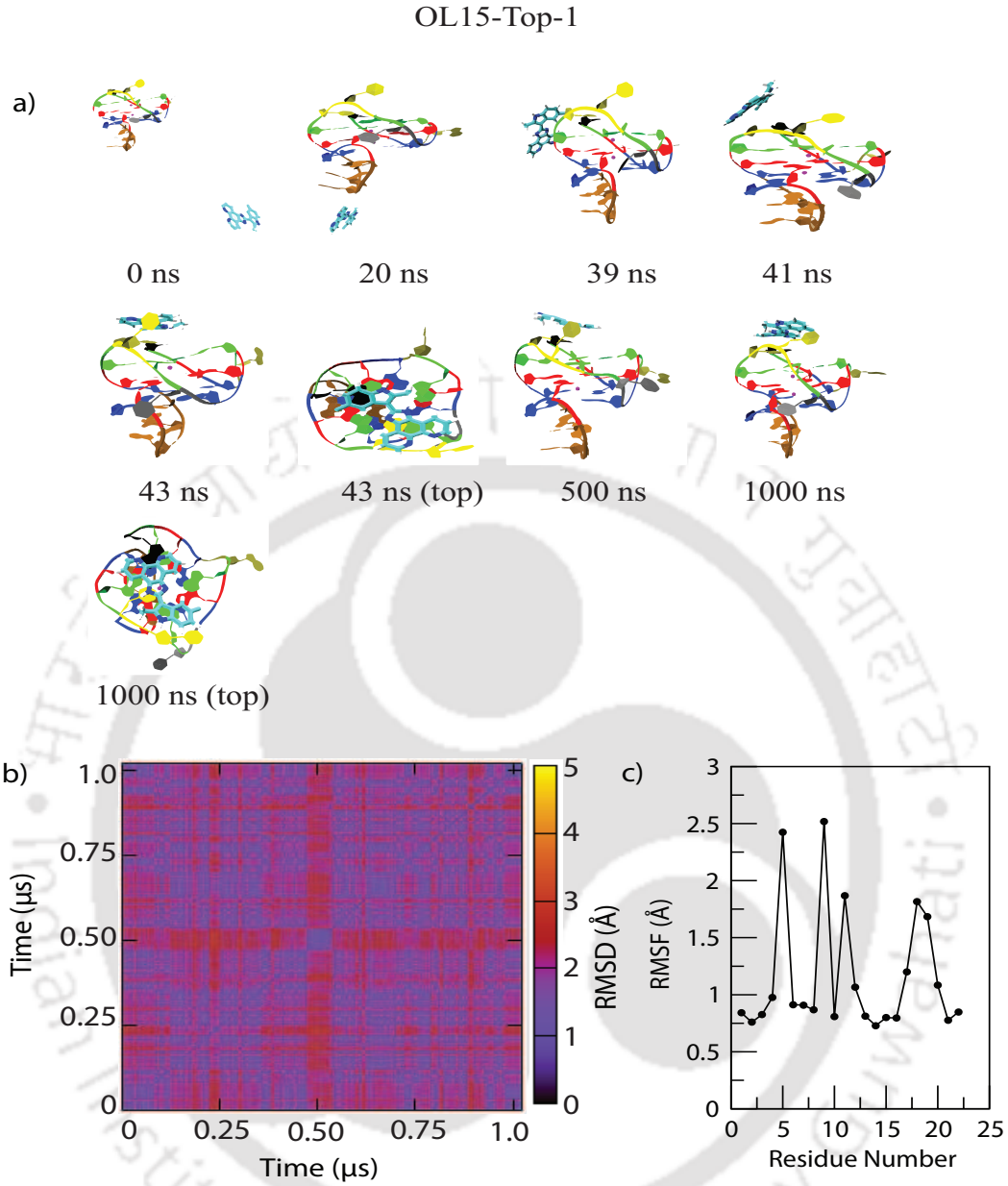


Figure 6.17: *(a) Snapshots representations of the complex formation of the c-KIT quadruplex DNA with time progress, (b) taking into account all heavy atoms of c-KIT G-quadruplex DNA, pairwise 2D-RMSDs of different systems, and (c) Root-mean-square fluctuations (RMSFs) of all heavy atoms of c-KIT G-quadruplex DNA for different systems.*

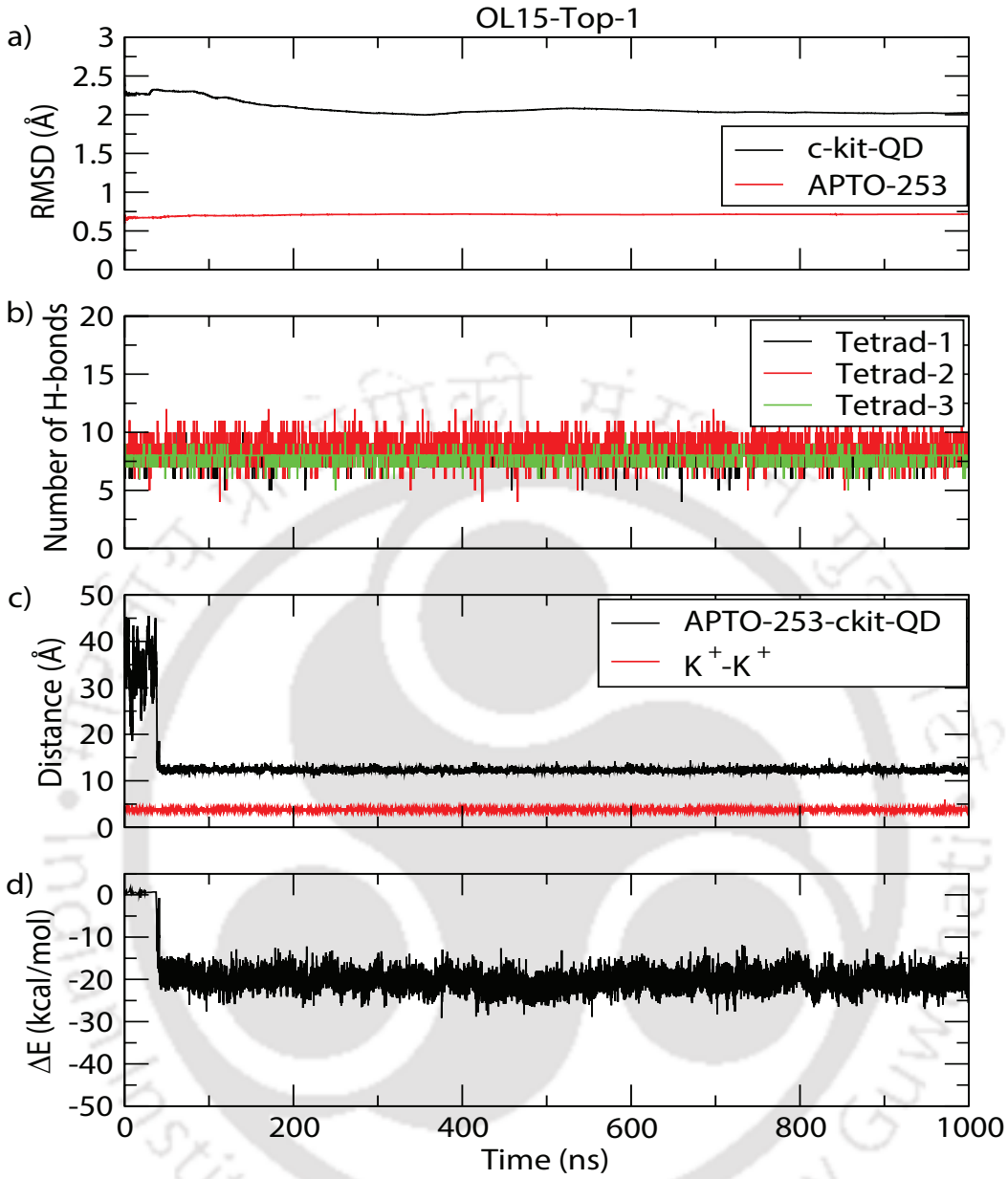


Figure 6.18: *(a) Time progression of the root-mean-square deviations (RMSDs) of all heavy atoms of c-KIT G-quadruplex DNA, (b) number of hydrogen bonds for tetrads, (c) the distance between center of masses of c-KIT G-quadruplex DNA and APTO-253, and the distance between the two K^+ central cations, and (d) the binding free energy of complex formation of APTO-253 ligand and c-KIT G-quadruplex DNA with time progression.*

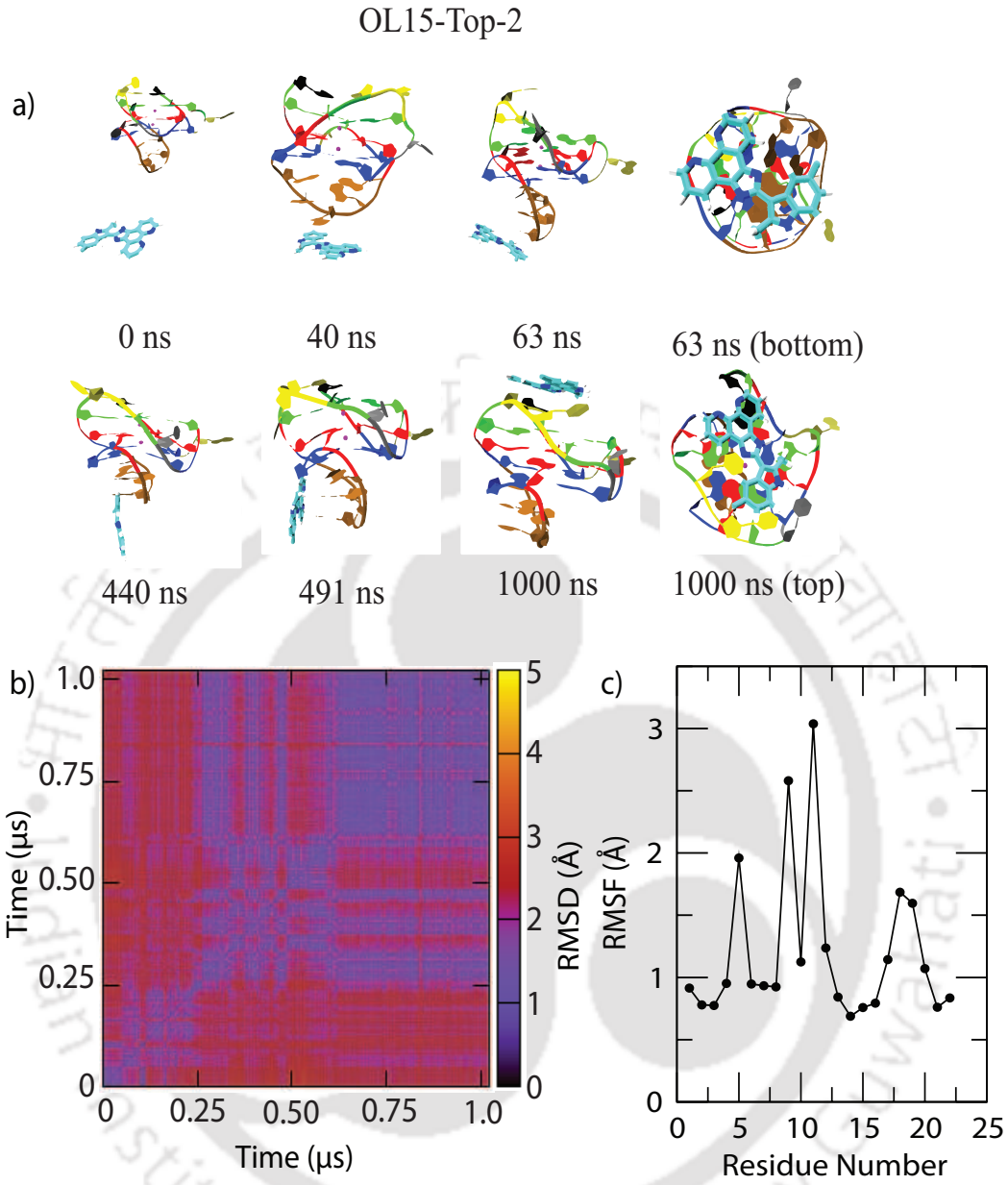


Figure 6.19: *(a) Snapshots representations of the complex formation of the c-KIT quadruplex DNA with time progress, (b) taking into account all heavy atoms of c-KIT G-quadruplex DNA, pairwise 2D-RMSDs of different systems, and (c) Root-mean-square fluctuations (RMSFs) of all heavy atoms of c-KIT G-quadruplex DNA for different systems.*

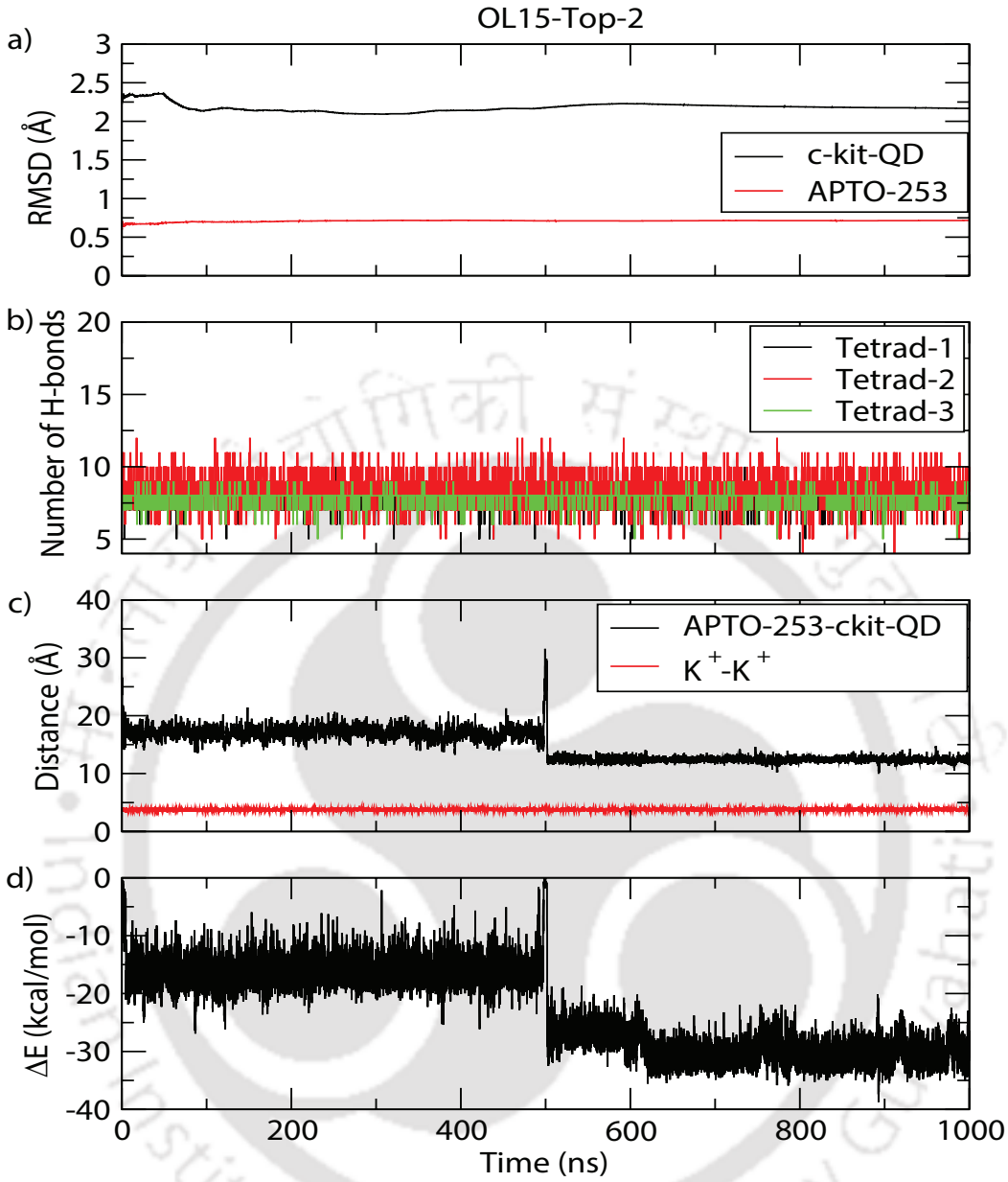


Figure 6.20: *(a) Time progression of the root-mean-square deviations (RMSDs) of all heavy atoms of c-KIT G-quadruplex DNA, (b) number of hydrogen bonds for tetrads, (c) the distance between center of masses of c-KIT G-quadruplex DNA and APTO-253, and the distance between the two K^+ central cations, and (d) the binding free energy of complex formation of APTO-253 ligand and c-KIT G-quadruplex DNA with time progression.*

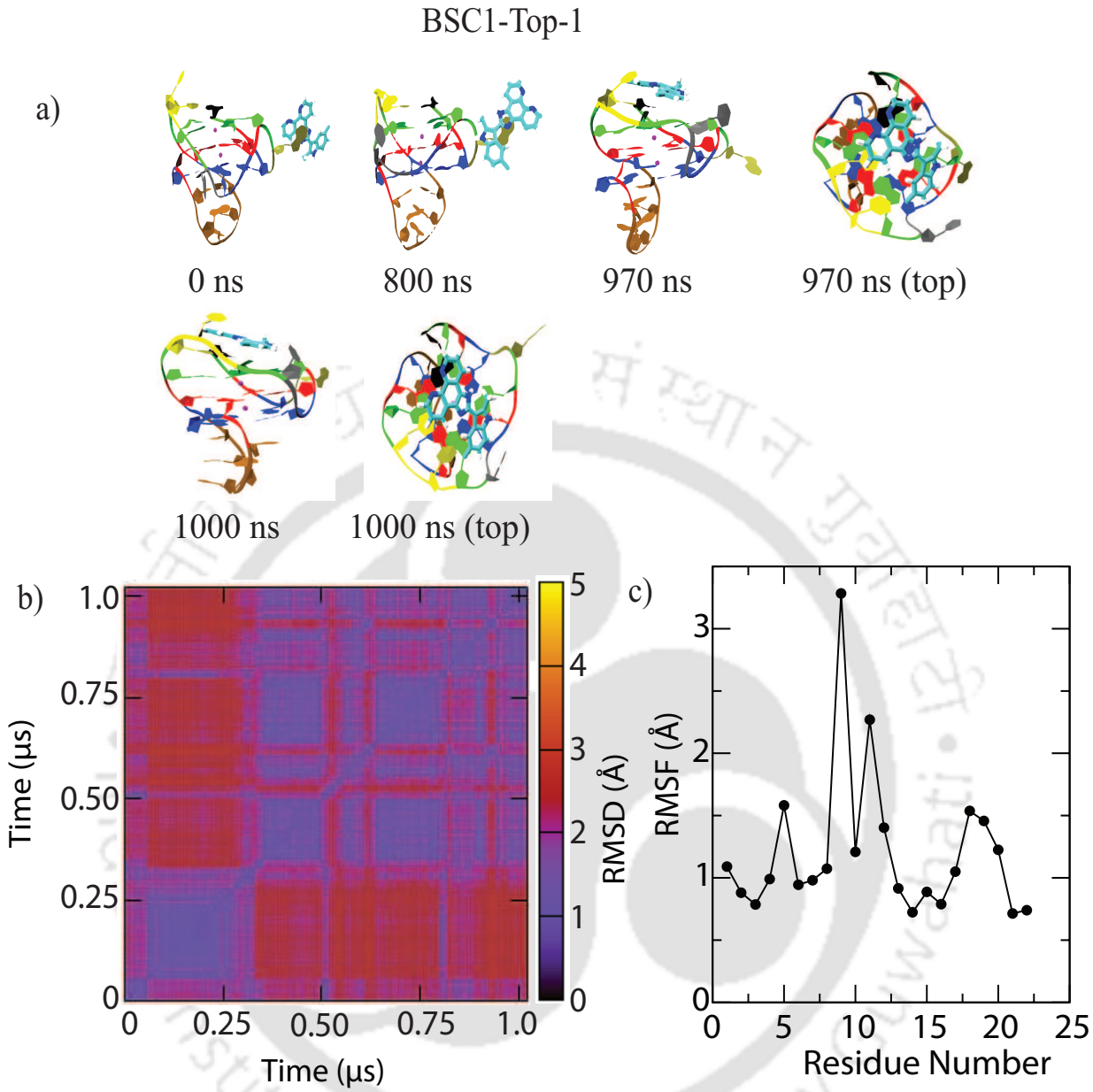


Figure 6.21: *(a) Snapshots representations of the complex formation of the c-KIT quadruplex DNA with time progress, (b) taking into account all heavy atoms of c-KIT G-quadruplex DNA, pairwise 2D-RMSDs of different systems, and (c) Root-mean-square fluctuations (RMSFs) of all heavy atoms of c-KIT G-quadruplex DNA for different systems.*

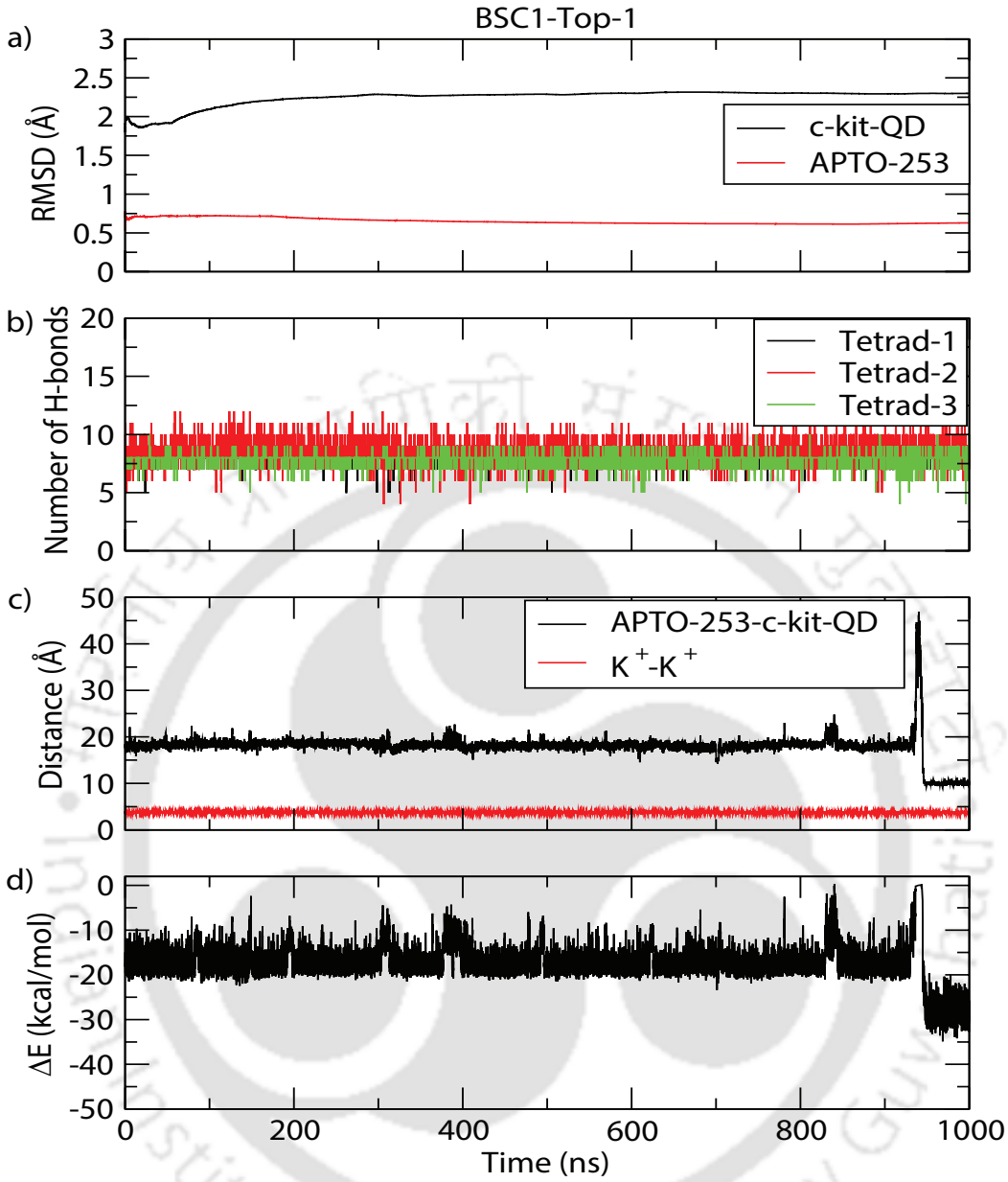


Figure 6.22: *(a) Time progression of the root-mean-square deviations (RMSDs) of all heavy atoms of c-KIT G-quadruplex DNA, (b) number of hydrogen bonds for tetrads, (c) the distance between center of masses of c-KIT G-quadruplex DNA and APTO-253, and the distance between the two K^+ central cations, and (d) the binding free energy of complex formation of APTO-253 ligand and c-KIT G-quadruplex DNA with time progression.*

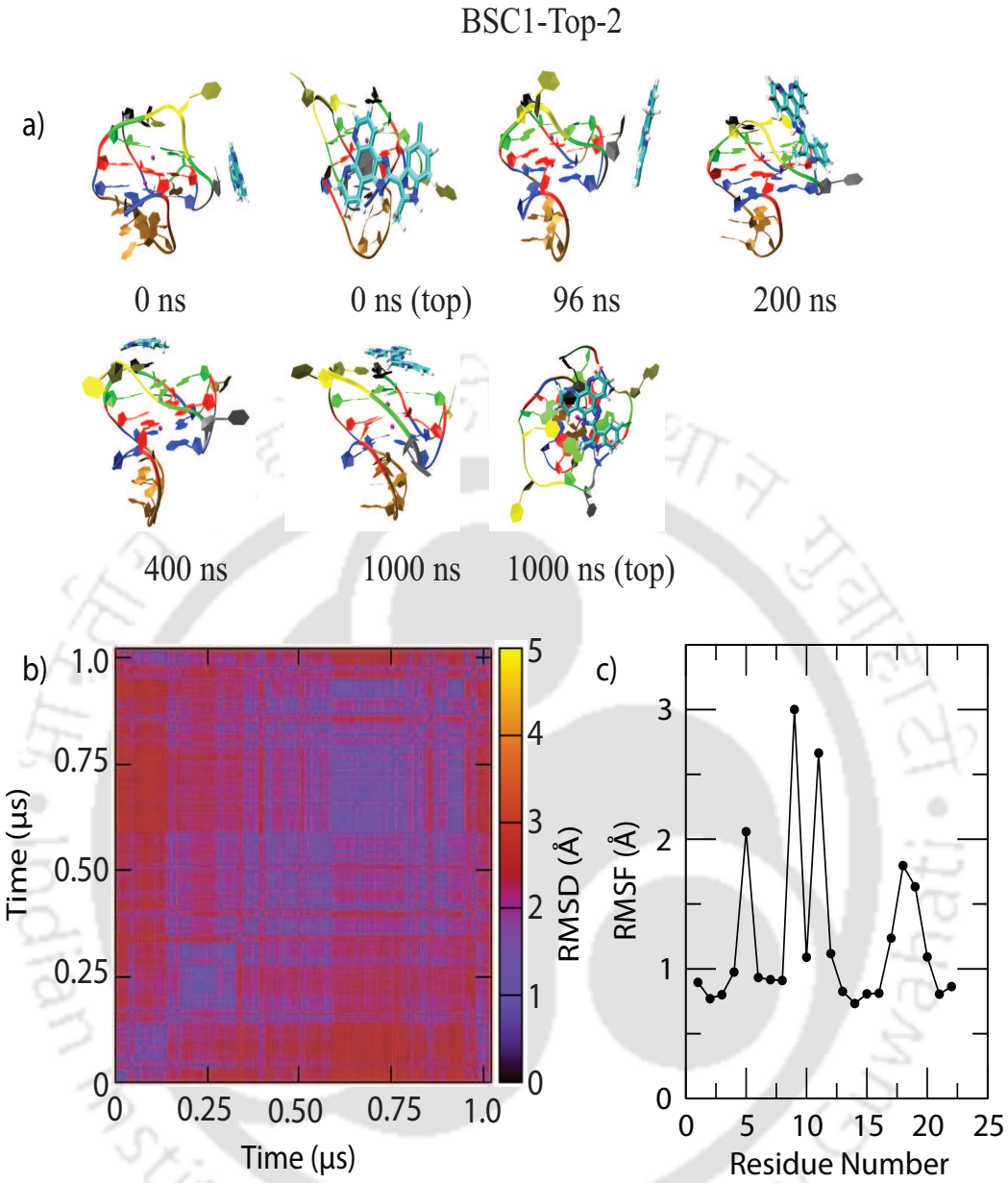


Figure 6.23: *(a) Snapshots representations of the complex formation of the c-KIT quadruplex DNA with time progress, (b) taking into account all heavy atoms of c-KIT G-quadruplex DNA, pairwise 2D-RMSDs of different systems, and (c) Root-mean-square fluctuations (RMSFs) of all heavy atoms of c-KIT G-quadruplex DNA for different systems.*

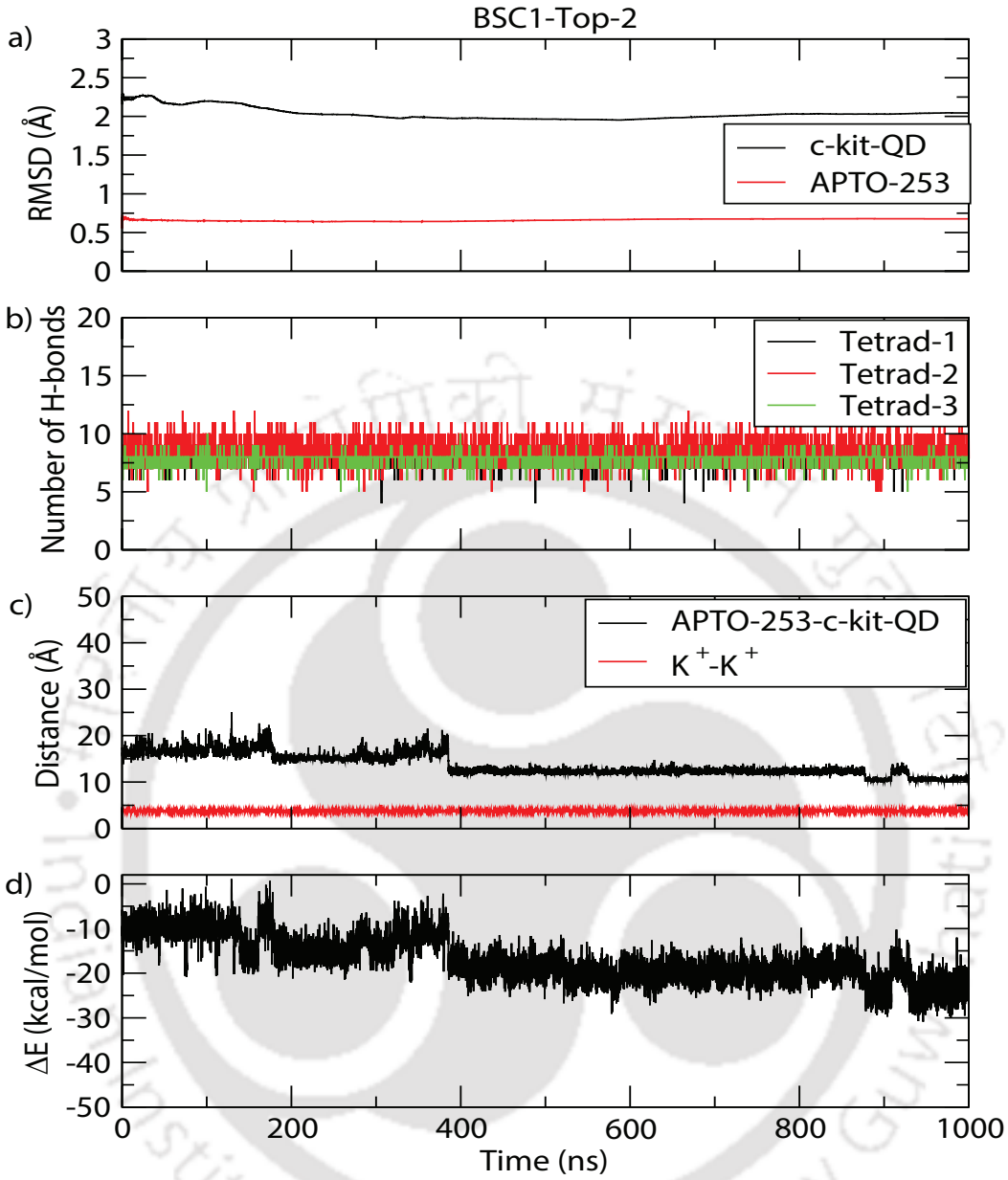


Figure 6.24: *(a) Time progression of the root-mean-square deviations (RMSDs) of all heavy atoms of c-KIT G-quadruplex DNA, (b) number of hydrogen bonds for tetrads, (c) the distance between center of masses of c-KIT G-quadruplex DNA and APTO-253, and the distance between the two K^+ central cations, and (d) the binding free energy of complex formation of APTO-253 ligand and c-KIT G-quadruplex DNA with time progression.*

6.3.5 Driving Force of the Complex Formation

In this section, we have investigated the driving force of the complex formation of the APTO-253 ligand and the c-KIT quadruplex DNA. In **Table 6.3**, we have presented the average MM-GBSA binding free energy, considering the last 4 ns trajectory for each simulation. Here, we notice that among all the considered interactions, van der Waals (vdW) interactions followed by electrostatic (EEL) interactions play a major role in the stabilization of the complex formation for all the binding modes. On the other hand, the other interaction terms i.e., polar solvation energy (EGB) and the change of the conformation energy are unfavorable to the binding formation. From **Table 6.3**, we see that for OL15 force field the binding free energy for top, bottom and side loop mode are −18.18 kcal/mol, −34.24 kcal/mol, −9.46 kcal/mol respectively. Hence, among the three binding modes, the bottom binding mode of APTO-253 on the c-KIT quadruplex DNA is most favorable. On the other hand, for BSC1 force field, the binding free energy for top, bottom and side loop mode are −34.34 kcal/mol, −30.91 kcal/mol and −10.32 kcal/mol respectively. Here, we notice that the top and bottom binding mode are very much comparable to each other. However, the side loop binding mode has the least binding free energy for both the force fields.

Table 6.3: *MM-GBSA binding free Energy (kcal/mol) of APTO-253 to c-KIT quadruplex DNA. Change of van der Waals energy (ΔE_{vdW}) on complex formation, change of electrostatic energy (ΔE_{EEL}) on complex formation, change of polar solvation energy (ΔE_{EGB}) on complex formation, change of non polar solvation energy (ΔE_{ESURF}) on complex formation, change of conformational energy (ΔE_{Conf}) on complex formation, and change of total potential energy (ΔE_{TOT}) on complex formation.*

Force field	Pose	ΔE_{vdW}	ΔE_{EEL}	ΔE_{EGB}	ΔE_{ESURF}	ΔE_{Conf}	ΔE_{TOT}
OL15	Top	-26.47	-8.10	17.29	-1.40	0.50	-18.18
		(± 1.74)	(± 5.04)	(± 5.12)	(± 0.08)	(± 0.14)	(± 1.68)
	Bottom	-48.20	-12.06	28.87	-2.85	0.002	-34.24
		(± 4.04)	(± 7.20)	(± 5.82)	(± 0.19)	(± 0.02)	(± 3.20)
	Side	-16.09	0.43	6.64	-0.93	0.47	-9.46
		(± 2.49)	(± 14.25)	(± 13.68)	(± 0.13)	(± 0.14)	(± 2.15)
BSC1	Top	-48.10	-1.89	17.80	-2.65	0.50	-34.34
		(± 1.85)	(± 4.81)	(± 4.79)	(± 0.13)	(± 0.15)	(± 2.37)
	Bottom	-44.36	-15.11	30.75	-2.72	0.52	-30.91
		(± 2.51)	(± 6.06)	(± 4.83)	(± 0.15)	(± 0.14)	(± 2.16)
	Side	-16.67	-9.99	16.77	-0.94	0.51	-10.32
		(± 1.90)	(± 7.55)	(± 7.23)	(± 0.12)	(± 0.15)	(± 1.54)

Now, it is important to find the orientation or arrangement of APTO-253 ligand and the respective nucleoside residues. Hence, we have estimated the probability of the distance (r) between the COM of the APTO-253 ligand and the COM of the considered nucleoside base and angle (θ) between the respective aromatic planes (**Fig. 6.25**). In **Figures**

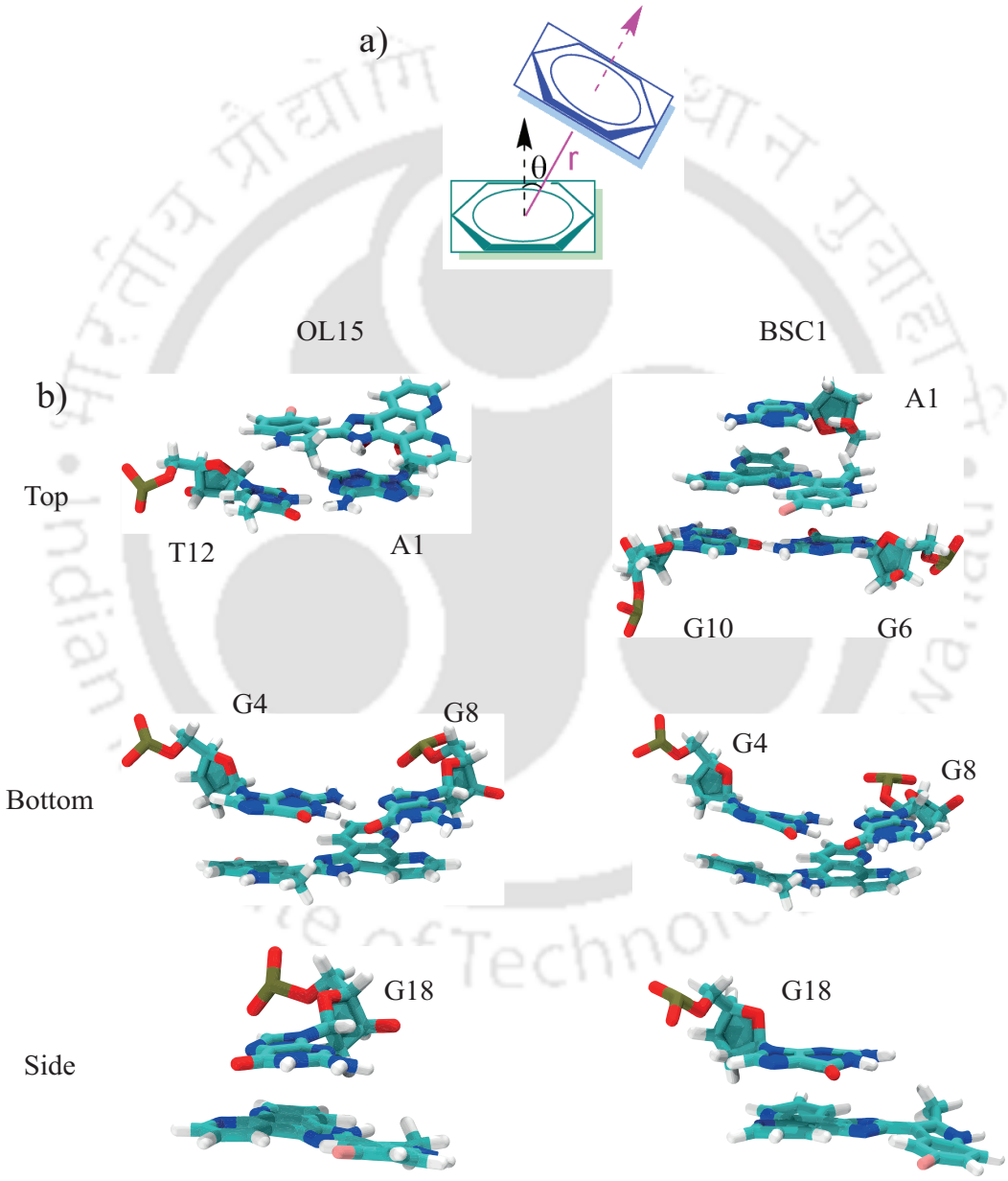


Figure 6.25: *(a) Schematic representations of π-π stacking interaction with respect to distance (r) and angle (θ) between the corresponding planes and (b) Snapshots of π-π stacking interactions of APTO-253 with different nucleoside bases.*

6.26-6.28, we have shown the above mentioned distance and angle probability for both the OL15 and BSC1 force fields. It is found that, in most of the cases, the π-π stacking

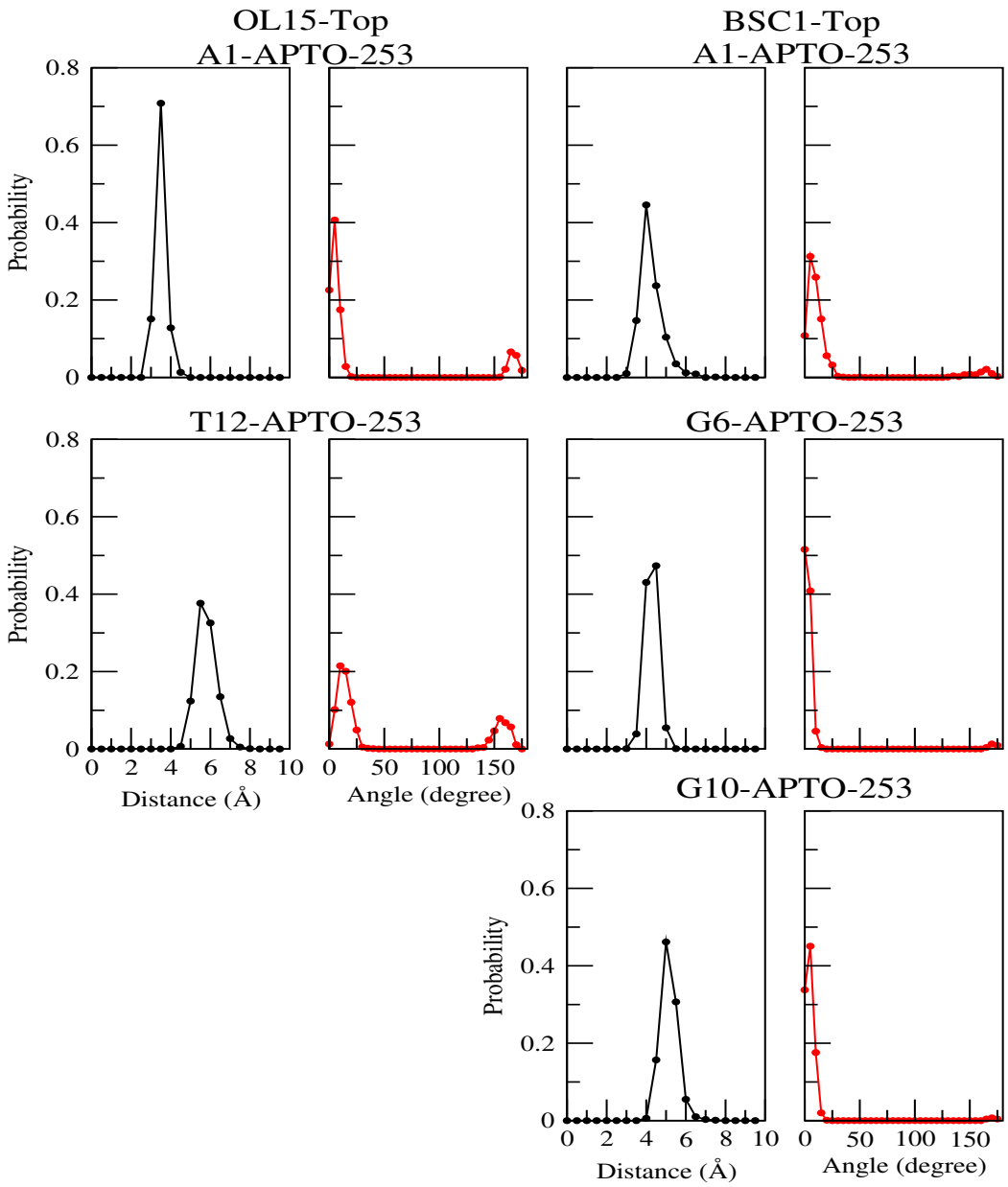


Figure 6.26: *Stacking probability with respect to distance and angle between the corresponding planes.*

interactions are present. For example, in top stacking mode, APTO-253 ligand binds with the A1 and T12 nucleoside residues in a π-π stacking manner (for OL15 force field). In case of BSC1, among the tetrad-1 (G2-G6-G10-G13) residues, APTO-253 forms π-π stacking arrangement with only G6 and G10 nucleoside residues. Moreover, the π-π

stacking formation is also observed between A1 and APTO-253 molecules. This indicates that APTO-253 ligand mainly binds with the A1, G6 and G10 residues of the c-KIT

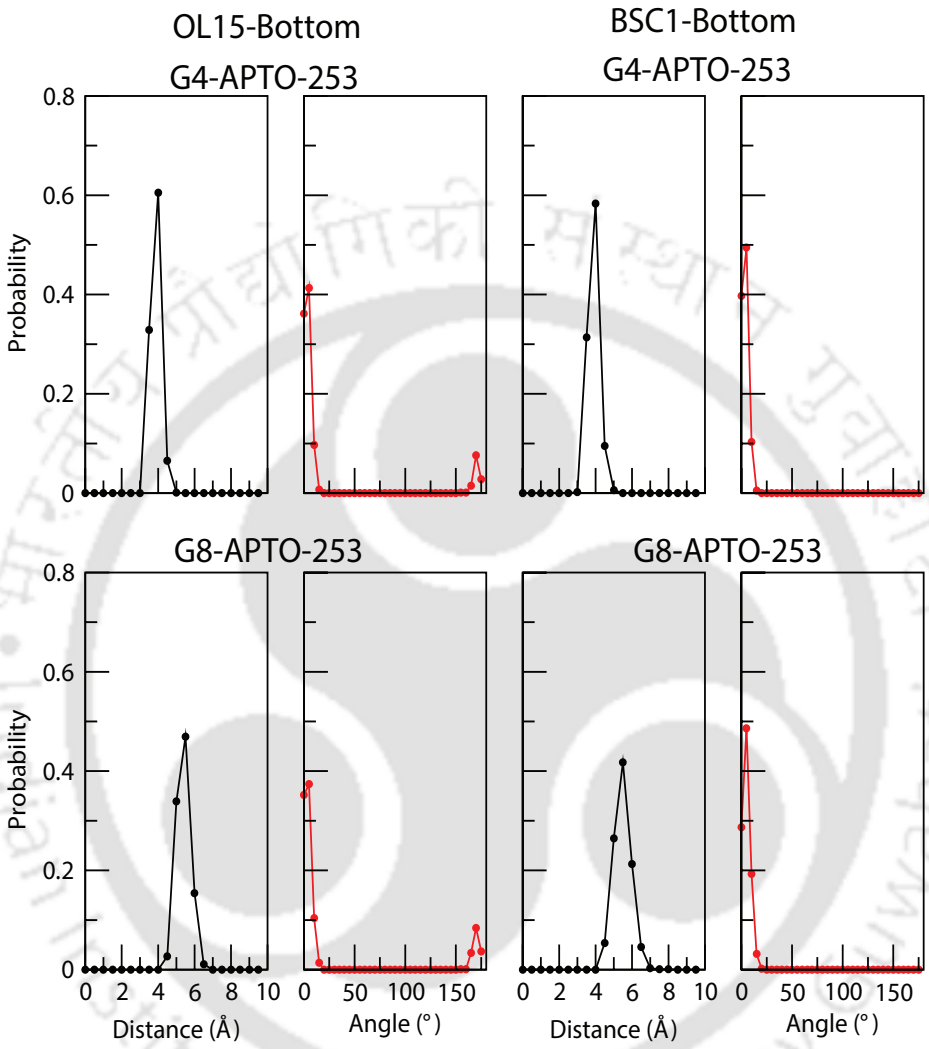


Figure 6.27: *Stacking probability with respect to distance and angle between the corresponding planes.*

quadruplex DNA in a 'sandwich' manner. Furthermore, for both the OL15 and BSC1 force fields, in the bottom binding mode, π-π stacking formation is noticed between APTO-253 ligand and the G4 and G8 residues of the tetrad-3 regions of the c-KIT quadruplex DNA. In side loop binding mode, the G18 residue of loop-4 , takes part in π-π stacking for all the force fields. Thus, it can be proposed that due to the presence of π-π stacking interaction between APTO-253 and the respective nucleoside bases of the c-KIT quadruplex DNA, the above mentioned binding poses are very stable.

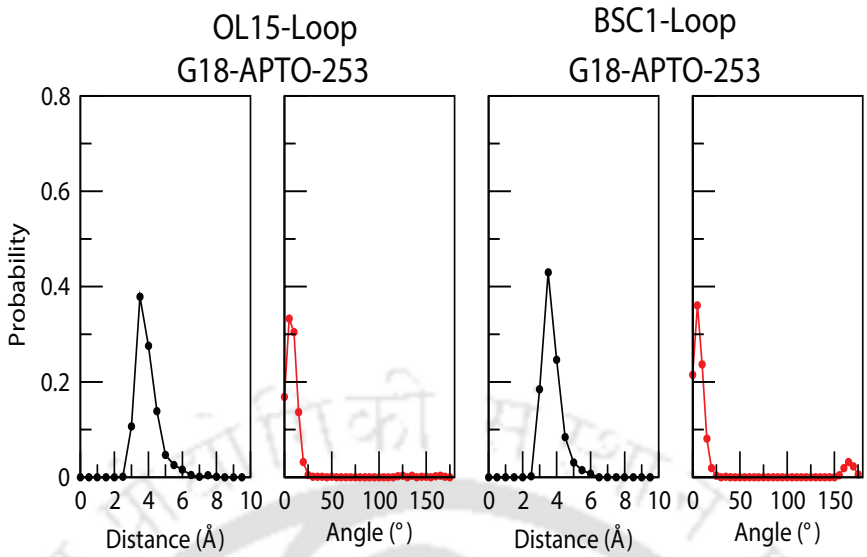


Figure 6.28: *Stacking probability with respect to distance and angle between the corresponding planes.*

6.3.6 Binding Pathway Studies Using Umbrella Sampling Method

In order to understand the pathway of complex formation, we have estimated the potential of mean force (PMF) for every system. In **Fig. 6.29**, for both OL15 and BSC1 DNA force fields, we have presented the free energy profiles of APTO-253 ligand and the c-KIT quadruplex DNA complex formation for all the three binding modes. The negative values of the minimum in PMF curves indicate that all the three binding modes are energetically favorable. Considering the top binding mode for all the force fields (**Fig. 6.29 (a)**), a minimum PMF value of −2.97 kcal/mol is reached at 4.35 Å for COM of APTO-253 and COM of nucleoside bases, A1 and T12 residues, of the c-KIT quadruplex DNA. On the other hand, for COM of APTO-253 and COM of nucleoside bases, A1, G2 and G6 residues of the considered DNA, a minimum PMF value of −5.52 kcal/mol is noticed at 5.81 Å. In case of the bottom binding mode, we have calculated the PMF between the COM of APTO-253 and COM of G4 and G6 nucleoside bases for both OL15 and BSC1 force fields (**Fig. 6.29 (b)**). Comparable minima in the PMF values for OL15 (−4.51 kcal/mol) at 4.61 Å and for BSC1 (−3.98 kcal/mol) at 5.04 Å are noticed. From **Fig. 6.29 (c)**, for the side loop binding mode, we have observed that the minimum PMF values of COM of APTO-253 and COM of G18 nucleoside base are −0.69 kcal/mol at 4.6 Å and −0.93 kcal/mol at 5.57 Å for OL15 and BSC1 force fields respectively. From the above observations, it is clear

that the side loop binding mode is the least favorable in both the force fields. Moreover, for OL15 force field, the bottom binding mode is energetically most favorable among all the three binding modes. However, in case of BSC1 force field, both the top and bottom binding modes of APTO-253 on the c-KIT quadruplex DNA, are energetically favorable. These minimum PMF values also support the binding free energy calculations discussed previously.

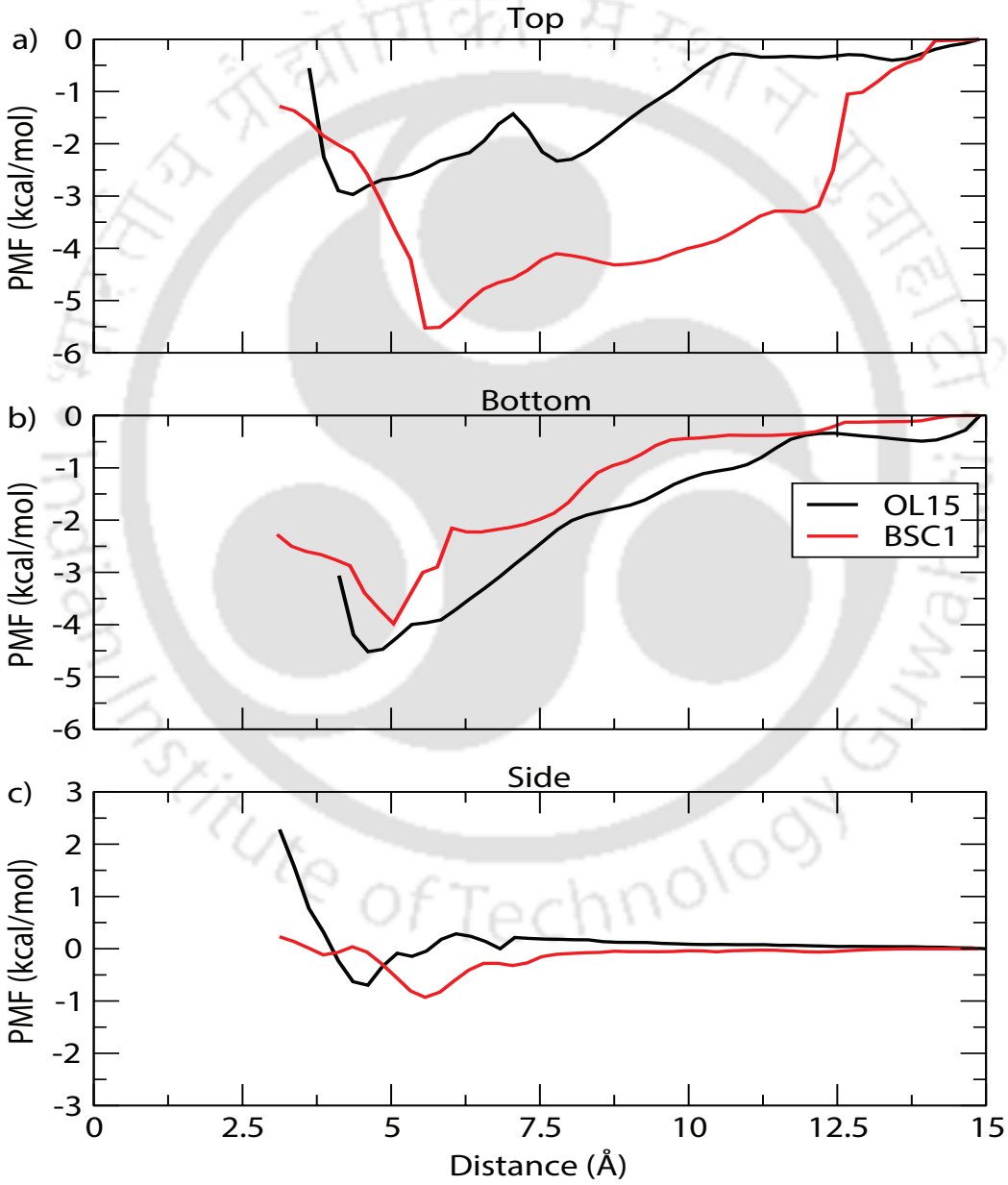


Figure 6.29: *Potential of mean forces (PMFs) between APTO-253 and respective nucleoside bases of the c-KIT quadruplex DNA.*

6.4 Discussions

As we all know that the force field parameters of different atomic sites play very important roles in molecular dynamics simulation. Thus, in this work, we have used two independent latest DNA force fields i.e., OL15 [75, 76] and BSC1 [78] for a better clarity to find the pathway and binding site of APTO-253 on c-KIT G-quadruplex DNA. From MM-GBSA binding energy calculations and PMF analysis, it is very clear that all the three binding modes (namely top, bottom and side loop) of APTO-253 on c-KIT G-quadruplex DNA are energetically favorable for both OL15 and BSC1 force fields of DNA. However, the side loop stacking mode is the least favorable, where the APTO-253 binds with only one nucleoside residue i.e., G18 residue of loop-4 regions for both OL15 and BSC1 force fields. Here, we have noticed that the binding free energy and PMF values are similar for both the force field cases. In case of bottom binding mode of APTO-253 on c-KIT G-quadruplex DNA, APTO-253 molecule binds with two nucleoside residues (G4 and G8) of tetrad-3 regions and also the values of binding free energy and PMF seems to be close to each other. However, major difference of binding free energy and PMF values are noticed for top binding mode for OL15 and BSC1 force fields. In case of BSC1 force field, APTO-253 binds with three nucleoside residues i.e., A1 (flanking base), G6 and G10 (of tetrad-1) in a 'sandwich' manner, while for OL15 force field, APTO-253 ligand binds with two nucleoside bases only i.e, A1 and T12, but mostly with A1. It seems that, when APTO-253 ligand binds with the similar nucleoside bases, the binding free energy and PMF values are very similar for both the OL15 and BSC1 force fields. However, these values mismatch when the participating nucleoside bases are different. This observation also indicate that, when APTO-253 ligand binds with nucleoside bases of tetrad regions, the complex formation is energetically most favorable. Here, it is worth mentioning that we have performed a total of 10 independent extended simulation run using different initial configurations using OL15 and BSC1 force fields and the previously mentioned binding modes are observed. It is possible when the force field is changed the binding modes can be different.

6.5 Summary and Conclusions

To enhance the selectivity and efficiency of the drug or ligand molecules on c-KIT quadruplex DNA, a detailed study of complex formation between them is required. The computer aided drug design method provide very useful techniques in the growing field of drug design. In this regard, we have performed all atom molecular dynamics simulation to investigate

the binding behavior of APTO-253 ligand on c-KIT quadruplex DNA using OL15 and BSC1 force fields (both the force fields have been recently developed for DNA structures). At first, we have characterized the dynamics of the c-KIT quadruplex DNA and APTO-253 molecules by different analyses such as, RMSD, RMSF, 2D-RMSD etc.. The low RMSD and RMSF values indicate that both c-KIT quadruplex DNA and APTO-253 hardly deviate from their initial structures through out the simulation runs. Moreover, the number of hydrogen bonds of the tetrad regions (namely tetrad-1, tetrad-2, and tetrad-3) and the distance between the two central K^+ cations suggest that the c-KIT quadruplex DNA maintains its structure in the complex formation with the APTO-253 ligand. Now, it is very crucial to identify the binding mode and binding site of the APTO-253 ligand in c-KIT quadruplex DNA. Thus, we have carried out the cluster structure analysis, where mainly three types of binding modes are identified. These are termed as top, bottom and side loop binding modes, according to the position of the ligand on the c-KIT quadruplex DNA. Negative values of binding free energy confirm that all the binding modes are energetically favorable. Moreover, in case of OL15, from the values of binding free energy, it is clear that the bottom mode is the most favorable among all the three poses. However, for BSC1 force fields, both the top and bottom modes are energetically favorable owing to the similar values of the binding free energy values. Next, we have investigated the driving force of binding of the APTO-253 ligand with the c-KIT quadruplex DNA, where we have noticed that mainly van der Waals (vdW) and π-π stacking interactions are responsible for the complex formation. Moreover, to probe the binding pathway of the APTO-253 ligand, we have estimated the PMFs between the APTO-253 ligand and the respective nucleoside bases which are involved in stacking formation. The minimum PMF values also support that all the binding modes are energetically favorable while the bottom binding mode complex formation is most favorable for OL15 force field. However, for BSC1 force field, both the top and bottom binding modes are energetically favorable. Our detailed study provide valuable information about the binding modes and binding pathway of the APTO-253 ligand on c-KIT quadruplex DNA. Furthermore, these observations provide useful information which can help to develop novel drugs for G-quadruplex DNA.

References

[1] T. J. Shaw, E. J. Keszthelyi, A. M. Tonary, M. Cada and B. C. Vanderhyden, *Exp. Cell Res.*, 2002, **273**, 95 – 106.

[2] C. E. Edling and B. Hallberg, *Int. J. Biochem. Cell Biol.*, 2007, **39**, 1995 – 1998.

[3] L. K. Ashman, *Int. J. Biochem. Cell Biol.*, 1999, **31**, 1037 – 1051.

[4] M. A. Babaei, B. Kamalidehghan, M. Saleem, H. Z. Huri and F. Ahmadipour, *Drug Des Devel Ther*, 2016, **10**, 2443.

[5] S. Sakurai, T. Fukasawa, J.-M. Chong, A. Tanaka and M. Fukayama, *Jpn. J. Cancer Res.*, 1999, **90**, 1321–1328.

[6] M. Sattler and R. Salgia, *Leuk. Res.*, 2004, **28**, 11–20.

[7] Y.-Y. Wang, G.-B. Zhou, T. Yin, B. Chen, J.-Y. Shi, W.-X. Liang, X.-L. Jin, J.-H. You, G. Yang, Z.-X. Shen, J. Chen, S.-M. Xiong, G.-Q. Chen, F. Xu, Y.-W. Liu, Z. Chen and S.-J. Chen, *Proc. Natl. Acad. Sci. U.S.A.*, 2005, **102**, 1104–1109.

[8] S. Ferrari, A. Grande, P. Zucchini, R. Manfredini, E. Tagliafico, E. Rossi, P. Temperani, G. Torelli, G. Emilia and U. Torelli, *Leuk. Lymphoma*, 1993, **9**, 495–501.

[9] M. Miettinen and J. Lasota, *Appl. Immunohistochem. Mol. Morphol.*, 2005, **13**, 205–220.

[10] R. Simak, P. Capodieci, D. Cohen, W. Fair, H. Scher, J. Melamed, M. Drobnjak, W. Heston, U. Stix, G. Steiner *et al.*, *Histol Histopathol.*, 2000, **15**, 365–374.

[11] P. Micke, J. G. Hengstler, H. Albrecht, A. Faldum, F. Bittinger, K. Becker, R. Wiewrodt, B. Fischer and R. Buhl, *Tumor Biol.*, 2004, **25**, 235–242.

[12] A. Yasuda, H. Sawai, H. Takahashi, N. Ochi, Y. Matsuo, H. Funahashi, M. Sato, Y. Okada, H. Takeyama and T. Manabe, *Mol. Cancer.*, 2006, **5**, 46.

[13] A. McIntyre, B. Summersgill, B. Grygalewicz, A. J. Gillis, J. Stoop, R. J. van Gurp, N. Dennis, C. Fisher, R. Huddart, C. Cooper *et al.*, *Cancer Res.*, 2005, **65**, 8085–8089.

[14] K. S. Smalley, K. L. Nathanson and K. T. Flaherty, *Cancer Res.*, 2009, **69**, 3241–3244.

[15] H. Nagata, A. S. Worobec, C. K. Oh, B. A. Chowdhury, S. Tannenbaum, Y. Suzuki and D. D. Metcalfe, *Proc. Natl. Acad. Sci. U.S.A.*, 1995, **92**, 10560–10564.

[16] D. A. Tuveson, N. A. Willis, T. Jacks, J. D. Griffin, S. Singer, C. D. Fletcher, J. A. Fletcher and G. D. Demetri, *Oncogene*, 2001, **20**, 5054–5058.

[17] A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle and D. J. Patel, *J. Am. Chem. Soc.*, 2007, **129**, 4386–4392.

[18] D. Wei, G. N. Parkinson, A. P. Reszka and S. Neidle, *Nucleic Acids Res.*, 2012, **40**, 4691–4700.

[19] M. Bejugam, S. Sewitz, P. S. Shirude, R. Rodriguez, R. Shahid and S. Balasubramanian, *J. Am. Chem. Soc.*, 2007, **129**, 12926–12927.

[20] M. Gunaratnam, S. Swank, S. M. Haider, K. Galesa, A. P. Reszka, M. Beltran, F. Cuenca, J. A. Fletcher and S. Neidle, *J. Med. Chem.*, 2009, **52**, 3774–3783.

[21] A. Głuszyńska, K. Bajor, I. Czerwińska, D. Kalet and B. Juskowiak, *Tetrahedron Lett.*, 2010, **51**, 5415 – 5418.

[22] A. Głuszyńska, B. Juskowiak, M. Kuta-Siejkowska, M. Hoffmann and S. Haider, *Molecules*, 2018, **23**, 1134.

[23] C. W. Ong, M.-C. Liu, K.-D. Lee, K. W. Chang, Y.-T. Yang, H.-W. Tung and K. R. Fox, *Tetrahedron*, 2012, **68**, 5453 – 5457.

[24] X. Wang, C.-X. Zhou, J.-W. Yan, J.-Q. Hou, S.-B. Chen, T.-M. Ou, L.-Q. Gu, Z.-S. Huang and J.-H. Tan, *ACS Med. Chem. Lett.*, 2013, **4**, 909–914.

[25] T. P. Garner, H. E. L. Williams, K. I. Gluszyk, S. Roe, N. J. Oldham, M. F. G. Stevens, J. E. Moses and M. S. Searle, *Org. Biomol. Chem.*, 2009, **7**, 4194–4200.

[26] M. C. Nielsen, A. F. Larsen, F. H. Abdikadir and T. Ulven, *Eur. J. Med. Chem.*, 2014, **72**, 119 – 126.

[27] X. Cui, S. Lin and G. Yuan, *Int. J. Biol. Macromol.*, 2012, **50**, 996 – 1001.

[28] F.-H. Shen, J. Jin, J. Li, Y. Wang, S.-H. Zhu, Y.-J. Lu, T.-M. Ou, Z.-S. Huang, M. Huang and Z.-Y. Huang, *Pharm. Biol*, 2013, **51**, 447–454.

[29] K. Jantos, R. Rodriguez, S. Ladame, P. S. Shirude and S. Balasubramanian, *J. Am. Chem. Soc.*, 2006, **128**, 13662–13663.

[30] K. M. Rahman, A. P. Reszka, M. Gunaratnam, S. M. Haider, P. W. Howard, K. R. Fox, S. Neidle and D. E. Thurston, *Chem. Commun.*, 2009, 4097–4099.

[31] M. Bejugam, M. Gunaratnam, S. Mller, D. A. Sanders, S. Sewitz, J. A. Fletcher, S. Neidle and S. Balasubramanian, *ACS Med. Chem. Lett.*, 2010, **1**, 306–310.

[32] K. V. Diveshkumar, S. Sakrikar, S. Harikrishna, V. Dhamodharan and P. I. Pradeepkumar, *ChemMedChem*, 2014, **9**, 2754–2765.

[33] K. I. E. McLuckie, Z. A. E. Waller, D. A. Sanders, D. Alves, R. Rodriguez, J. Dash, G. J. McKenzie, A. R. Venkitaraman and S. Balasubramanian, *J. Am. Chem. Soc.*, 2011, **133**, 2658–2663.

[34] K. V. Diveshkumar, S. Sakrikar, F. Rosu, S. Harikrishna, V. Gabelica and P. I. Pradeepkumar, *Biochemistry*, 2016, **55**, 3571–3585.

[35] C. Wei, L. Ren and N. Gao, *Int. J. Biol. Macromol.*, 2013, **57**, 1 – 8.

[36] Z. A. E. Waller, S. A. Sewitz, S.-T. D. Hsu and S. Balasubramanian, *J. Am. Chem. Soc.*, 2009, **131**, 12628–12633.

[37] S. Manaye, R. Eritja, A. Aviñó, J. Jaumot and R. Gargallo, *Biochim. Biophys. Acta, Gen. Subj.*, 2012, **1820**, 1987 – 1996.

[38] M. Petenzi, D. Verga, E. Largy, F. Hamon, F. Doria, M.-P. Teulade-Fichou, A. Guédin, J.-L. Mergny, M. Mella and M. Freccero, *Chem. Eur. J.*, 2012, **18**, 14487–14496.

[39] K.-w. Zheng, D. Zhang, L.-x. Zhang, Y.-h. Hao, X. Zhou and Z. Tan, *J. Am. Chem. Soc.*, 2011, **133**, 1475–1483.

[40] A. K. Sahoo, B. Bagchi and P. K. Maiti, *J. Chem. Phys*, 2019, **151**, 164902.

[41] K. G. Moghaddam, A. H. de Vries, S. J. Marrink and S. Faraji, *Biophys. Chem.*, 2019, **253**, 106220.

[42] N. Deng, L. Wickstrom, P. Cieplak, C. Lin and D. Yang, *J. Phys. Chem. B*, 2017, **121**, 10484–10497.

[43] V. M. Hridya, J. T. Hynes and A. Mukherjee, *J. Phys. Chem. B*, 2019, **123**, 10904–10914.

[44] S. Satpathi, R. K. Singh, A. Mukherjee and P. Hazra, *Phys. Chem. Chem. Phys.*, 2018, **20**, 7808–7818.

[45] B. Jawad, L. Poudel, R. Podgornik, N. F. Steinmetz and W.-Y. Ching, *Phys. Chem. Chem. Phys.*, 2019, **21**, 3877–3893.

[46] R. Galindo-Murillo and T. E. C. III, *J. Biomol. Struct. Dyn.*, 2018, **36**, 3311–3323.

[47] Z. Shen, K. A. Mulholland, Y. Zheng and C. Wu, *J. Mol. Model*, 2017, **23**, 256.

[48] K. Mulholland, F. Siddiquei and C. Wu, *Phys. Chem. Chem. Phys.*, 2017, **19**, 18685–18694.

[49] B. Machireddy, G. Kalra, S. Jonnalagadda, K. Ramanujachary and C. Wu, *J. Chem. Inf. Model.*, 2017, **57**, 2846–2864.

[50] H.-J. Sullivan, C. Readmond, C. Radicella, V. Persad, T. J. Fasano and C. Wu, *ACS Omega*, 2018, **3**, 14788–14806.

[51] M. Huesca, L. S. Lock, A. A. Khine, S. Viau, R. Peralta, I. H. Cukier, H. Jin, R. A. Al-Qawasmeh, Y. Lee, J. Wright and A. Young, *Mol. Cancer Ther.*, 2009, **8**, 2586–2596.

[52] W. G. Rice, A. Vellanki, Y. Lee, J. Lightfoot, R. Peralta, M. Jamerlan, H. Jin, R. Lum and T. Cheng, *Blood*, 2014, **124**, 4813–4813.

[53] A. Local, H. Zhang, K. D. Benbatoul, P. Folger, X. Sheng, C.-Y. Tsai, S. B. Howell and W. G. Rice, *Mol. Cancer Ther.*, 2018, **17**, 1177–1186.

[54] C.-Y. Tsai, S. Sun, H. Zhang, A. Local, Y. Su, L. A. Gross, W. G. Rice and S. B. Howell, *Mol. Cancer Ther.*, 2018, **17**, 1167–1176.

[55] R. Dennington, T. A. Keith and J. M. Millam, *GaussView Version 6*, 2019, Semichem Inc. Shawnee Mission KS.

[56] G. A. Petersson, A. Bennett, T. G. Tensfeldt, M. A. AlLaham, W. A. Shirley and J. Mantzaris, *J. Chem. Phys.*, 1988, **89**, 2193–2218.

[57] G. A. Petersson and M. A. AlLaham, *J. Chem. Phys.*, 1991, **94**, 6081–6090.

[58] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian16 Revision C.01*, 2016, Gaussian Inc. Wallingford CT.

[59] J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model.*, 2006, **25**, 247 – 260.

[60] C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.

[61] D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER 2018*, University of California, San Francisco, 2018.

[62] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, **25**, 1157–1174.

[63] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

[64] I. S. Joung and T. E. Cheatham, *J. Phys. Chem. B*, 2008, **112**, 9020–9041.

[65] B. Islam, P. Stadlbauer, M. Krepl, M. Havrila, S. Haider and J. Sponer, *J. Chem. Theory Comput.*, 2018, **14**, 5011–5026.

[66] J. Šponer, G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R. A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka, N. G. Walter and M. Otyepka, *Chem. Rev.*, 2018, **118**, 4177–4338.

[67] N. G. J. Richards and M. M. Georgiadis, *Acc. Chem. Res.*, 2017, **50**, 1375–1382.

[68] B. Islam, P. Stadlbauer, A. Gil-Ley, G. Pérez-Hernández, S. Haider, S. Neidle, G. Bussi, P. Banas, M. Otyepka and J. Sponer, *J. Chem. Theory Comput.*, 2017, **13**, 2458–2480.

[69] A. K. Sieradzan, P. Krupa and D. J. Wales, *J. Phys. Chem. B*, 2017, **121**, 2207–2219.

[70] P. Stadlbauer, L. Mazzanti, T. Cragnolini, D. J. Wales, P. Derreumaux, S. Pasquali and J. Šponer, *J. Chem. Theory Comput.*, 2016, **12**, 6077–6097.

[71] D. R. Gruber, J. J. Toner, H. L. Miears, A. V. Shernyukov, A. S. Kiryutin, A. A. Lomzov, A. V. Endutkin, I. R. Grin, D. V. Petrova, M. S. Kupryushkin, A. V. Yurkovskaya, E. C. Johnson, M. Okon, E. G. Bagryanskaya, D. O. Zharkov and S. L. Smirnov, *Nucleic Acids Res.*, 2018, **46**, 10827–10839.

[72] S. Haider, *J. Indian Inst. Sci*, 2018, **98**, 325–339.

[73] P. D. Dans, J. Walther, H. Gómez and M. Orozco, *Curr. Opin.*, 2016, **37**, 29 – 45.

[74] P. D. Dans, L. Danilne, I. Ivani, T. Dršata, F. Lankaš, A. Hospital, J. Walther, R. I. Pujagut, F. Battistini, J. L. Gelpí, R. Lavery and M. Orozco, *Nucleic Acids Res.*, 2016, **44**, 4052–4066.

[75] M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo and P. Jurečka, *J. Chem. Theory Comput.*, 2015, **11**, 5723–5736.

[76] R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka and T. E. Cheatham, *J. Chem. Theory Comput.*, 2016, **12**, 4114–4127.

[77] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton and M. Orozco, *Biophysical J.*, 2007, **92**, 3817 – 3829.

[78] I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case and M. Orozco, *Nat. Methods.*, 2016, **13**, 55–58.

[79] L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, *J. Comput. Chem.*, 2009, **30**, 2157–2164.

[80] D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.

[81] S. Pal and S. Paul, *Int. J. Biol. Macromol.*, 2019, **121**, 350 – 363.

[82] S. Pal and S. Paul, *J. Phys. Chem. C*, 2019, **123**, 11686–11698.

[83] S. Pal and S. Paul, *J. Phys. Chem. B*, 2020, **124**, 3123–3136.

[84] R. W. Pastor, B. R. Brooks and A. Szabo, *Mol. Phys.*, 1988, **65**, 1409–1419.

[85] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys*, 1984, **81**, 3684–3690.

[86] B. P. Uberuaga, M. Anghel and A. F. Voter, *J. Chem. Phys.*, 2004, **120**, 6363–6374.

[87] D. J. Sindhikara, S. Kim, A. F. Voter and A. E. Roitberg, *J. Chem. Theory Comput.*, 2009, **5**, 1624–1631.

[88] J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

[89] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

[90] W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.

[91] M. Shapovalov, S. Vucetic and R. L. Dunbrack, Jr., *PLOS Comput. Biol.*, 2019, **15**, 1–32.

[92] F. Heinkel, L. Abraham, M. Ko, J. Chao, H. Bach, L. T. Hui, H. Li, M. Zhu, Y. M. Ling, J. C. Rogalski, J. Scurll, J. M. Bui, T. Mayor, M. R. Gold, K. C. Chou, Y. Av-Gay, L. P. McIntosh and J. Gsponer, *Proc. Natl. Acad. Sci. U.S.A.*, 2019, **116**, 16326–16331.

[93] S. Pal and S. Paul, *J. Phys. Chem. B*, 2020, **124**, 210–223.

[94] M. Ester, H.-P. Kriegel, J. Sander and X. Xu, Proc. of 2nd International Conference on Knowledge Discovery and, 1996, pp. 226–231.

[95] S. Balasubramanian, S. Pal and B. Bagchi, *Phys. Rev. Lett.*, 2002, **89**, 115505.

[96] A. Chandra, *Phys. Rev. Lett.*, 2000, **85**, 768.

[97] A. Chandra, *J. Phys. Chem. B*, 2003, **107**, 3899–3906.

[98] S. Paul and A. Chandra, *J. Chem. Theory Comput.*, 2005, **1**, 1221–1231.

[99] S. Pal, R. Roy and S. Paul, *J. Phys. Chem. B*, 2020, **124**, 7598–7610.

[100] B. Machireddy, H.-J. Sullivan and C. Wu, *Molecules*, 2019, **24**, 1010.

[101] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Acc. Chem. Res.*, 2000, **33**, 889–897.

[102] B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theory Comput.*, 2012, **8**, 3314–3321.

[103] J. Kongsted, P. Söderhjelm and U. Ryde, *J. Comput. Aided Mol. Des.*, 2009, **23**, 395–409.

[104] T. Hou, J. Wang, Y. Li and W. Wang, *J. Comput. Chem.*, 2011, **32**, 866–877.

[105] T. Hou, J. Wang, Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, **51**, 69–82.

[106] H. Sun, Y. Li, S. Tian, L. Xu and T. Hou, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16719–16729.

[107] L. Xu, H. Sun, Y. Li, J. Wang and T. Hou, *J. Phys. Chem. B*, 2013, **117**, 8408–8421.

[108] G. Biffi, D. Tannahill, J. Miller, W. J. Howat and S. Balasubramanian, *PLOS One*, 2014, **9**, 1–9.

[109] J. Šponer, J. E. Šponer, A. Mládek, P. Jurečka, P. Banáš and M. Otyepka, *Biopolymers*, 2013, **99**, 978–988.

[110] P. Banáš, A. Mládek, M. Otyepka, M. Zgarbová, P. Jurečka, D. Svozil, F. Lankaš and J. Šponer, *J. Chem. Theory Comput.*, 2012, **8**, 2448–2460.

[111] J. Šponer, K. E. Riley and P. Hobza, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2595–2610.

[112] H. Kruse, P. Banáš and J. Šponer, *J. Chem. Theory Comput.*, 2019, **15**, 95–115.

[113] R. Roy and S. Paul, *J. Phys. Chem. B*, doi.org/10.1021/acs.jpcb.0c07830.

[114] J. Kästner, *WIREs Comput. Mol. Sci.*, 2011, **1**, 932–942.

[115] A. Grossfield, *WHAM: the weighted histogram analysis method*, `http://membrane.urmc.rochester.edu/?page\_id=126/`.

[116] E. A. Venczel and D. Sen, *Biochemistry*, 1993, **32**, 6220–6228.

[117] S. Chowdhury and M. Bansal, *J. Phys. Chem. B*, 2001, **105**, 7572–7578.

Chapter 7

Summary and Our View on The Drug Binding Pathways and Conformational Deviations of G-quadruplex DNA under Different Chemical and Thermal Conditions

"While the biological properties of deoxypentose nucleic acid suggest a molecular structure containing great complexity, X-ray diffraction studies described here show the basic molecular configuration has great simplicity."

– Maurice Wilkins

This thesis covers the conformational deviations of the G-quadruplex DNA under different chemical and thermal conditions. At first, we have systematically investigated the role of a divalent cation, Sr^{2+}, on the conformation of Thrombin Binding Aptamer (TBA) by classical MD simulation. Here, we have observed that the TBA-G-quadruplex DNA is not only stable throughout the 1 μs simulation run at different $SrCl_2$ salt concentrations, but also its rigidity increases with increasing salt concentration, due to the interaction between backbone atoms (mainly phosphate regions) and Sr^{2+} ions. The Sr^{2+} ion exchange in and out of the channel core of TBA is not possible. We have also noticed that loop-1 and loop-3 regions fluctuate more than the loop-2 residues. Moreover, with increasing $SrCl_2$ salt concentrations, the fluctuations of the atoms of the residues decrease, which lead us to conclude that Sr^{2+} ion provides more rigidity to the TBA-G-quadruplex DNA. Furthermore, π-π stacking interaction between T3-T4 residues of loop-1, which is observed at low salt concentrations i.e., 0 mM and 50 mM, disappears with increasing salt concentrations.

Next, we have studied the conformational deviations of human telomeric parallel G-quadruplex DNA at different temperatures and KCl salt concentrations under polar and non-polar confinement conditions and we observed that, at high temperature, the G-quadruplex DNA deviates and fluctuates from its native state, but not at high salt concentration. Moreover, the exchange of K^+ ion in and out of the channel core does not occur either at high temperature or in high salt concentrations under both polar and non-polar confinement conditions. Here, we also observed that the density of K^+ ions is higher mainly in two regions: (i) the G-quadruplex channel core and (ii) the sugar-phosphate backbone regions. The higher number of K^+ ions is primarily around the G2, G3, G8, G14, G20 and G21 residues of the tetrad regions of the human telomeric G-quadruplex DNA under both the non-polar and polar confinement conditions.

Furthermore, the conformational deviation of G-quadruplex DNA is analyzed in deep eutectic solvents (DESs) such as choline-urea (1:2) eutectic mixture, namely reline, medium. Here, we notice that, with decreasing reline concentration, TBA deviates and fluctuates from its native state, indicating that reline provides rigidity to TBA. The choline and urea molecules of reline interacts and form hydrogen bonds with the TBA due to which the rigidity of the DNA is maintained. Hence, when the concentration of water increases, the deviation from its native structure is enhanced. Moreover, a higher density of reline is observed around the phosphate regions of the TBA molecule. Here, the G8 and T9 nucleoside bases of loop-2 stack to each other at lower reline concentration, which is vanished at high reline concentrations. The enhanced rigidity of the G-quadruplex DNA in

DES as compared to that in water medium leads us to conclude that for long-term nucleic acid storage, reline is a good choice for solvent as a storage medium.

We, further, examine whether DES provides thermal stability to G-quadruplex DNA. For this, we have studied the effect of reline on c-KIT G-quadruplex DNA a regime of temperatures, starting from 300 to 500 K at 50 K temperature interval. The conformational change of c-KIT DNA in water medium at 450 K is also analyzed for comparison. The DNA unfolds in water medium at 450 K, but it maintains its native state in reline medium at the same temperature, even though it becomes unstable at 500 K. Thus, reline provides the ultra stability to c-kit G-quadruplex DNA at elevated temperature in comparison to water medium, which is also validated by previous experimental work. It is found that the hydrogen bonds between reline and c-kit G-quadruplex DNA are responsible for its stability at high temperature (up to 450 K). Howover, the compact distribution of hydrogen-bonded choline and urea molecules around c-kit G-quadruplex DNA decreases with increasing temperature. Interestingly, the position of the central cation K^+ fluctuates between the tetrad-1 and tetrad-2 and tetrad-2 and tetrad-3 at 450 and 500 K in reline medium. The rate of this oscillation is increased when temperature rises from 450 to 500 K temperatures. However, this phenomenon is absent in water medium at 450 K. This distinctive up down phenomenon of the central cation of the ion channel in reline medium has not been reported elsewhere.

Finally, we have studied the binding mode and pathway of APTO-253, a phase I clinical trial drug molecule, on c-KIT G-quadruplex DNA via all atom molecular dynamics simulations using OL15 and BSC1 force fields (both force fields have been recently developed for DNA structures). Starting from an unbound ligand and c-KIT quadruplex DNA, different binding modes are observed for the complexation using different characterized methods. Initially, we have checked the deviations and fluctuations of DNA from its native state, and observed that they are very small on complex formation. Moreover, the c-KIT G-quadruplex DNA maintains its initial structure through out the simulation run on binding with APTO-253. Here, mainly three binding modes are identified i.e., top, bottom and side loop stacking mode (characterized according to the position of the G-quadruplex DNA binding with the drug). The negative binding energy for all systems indicates that all the binding modes are energetically favorable for both OL15 and BSC1 force fields. The bottom mode is the most favorable for OL15 and, for BSC1 force field, both the top and bottom modes are energetically favorable. The van der Waals (vdW) and π-π stacking interactions between the DNA and the ligand are mainly the driving force of the complex

formation. Our detailed study provides valuable information about the binding modes and binding pathway of the APTO-253 ligand on c-KIT quadruplex DNA.

In short, this thesis enlightens valuable insight about the conformational deviation of the G-quadruplex DNA. Moreover, our studies also provide the useful information for the discovery of the new drugs in field of stabilization of the G-quadruplex DNAs.

List of Figures

List of Tables

Publications

1. Pal, S.; Paul, S. Conformational Deviation of Thrombin Binding G-quadruplex Aptamer (TBA) in Presence of Divalent Cation Sr^{2+}: A Classical Molecular Dynamics Simulation Study. *Int. J. Biol. Macromol.* **2019**, *121*, 350-363.

2. Pal, S.; Paul, S. Effect of Hydrated and Nonhydrated Choline Chloride-Urea Deep Eutectic Solvent (Reline) on Thrombin-Binding G-quadruplex Aptamer (TBA): A Classical Molecular Dynamics Simulation Study. *J. Phys. Chem.* C **2019**, *123*, 11686-11698.

3. Pal, S.; Paul, S. ATP Controls the Aggregation of Aβ_{16-22} Peptides. *J. Phys. Chem. B* **2020**, *124*, 210-223.

4.Pal, S.; Paul, S. Understanding The Role of Reline, a Natural DES, on Temperature-Induced Conformational Changes of C-Kit G-Quadruplex DNA: A Molecular Dynamics Study. *J. Phys. Chem. B* **2020**, *124*, 3123-3136.

5. Giri, R. S.; Pal, S.; Roy, S.; Dolai, G.; Manne, S. R.; Paul, S.; Mandal, B. Nanostructures from Protected L/L and D/L Amino Acid Containing Dipeptides. *Pept. Sci.* **2020**, e24176.

6. Pal, S.; Roy, R.; Paul, S. Potential of a Natural Deep Eutectic Solvent, Glyceline, in the Thermal Stability of the Trp-Cage Mini-protein. *J. Phys. Chem. B* **2020**, *124*, 7598-7610.

7. Pal, S.; Paul, S. An in-Silico Investigation of Binding Modes and Pathway of APTO-253 on c-KIT G-Quadruplex DNA. *Phys Chem. Chem. Phys* doi.org/10.1039/D0CP05210H

8. Pal, S.; Paul, S. Theoretical Investigation of Conformational Deviation of the Human Parallel Telomeric G-quadruplex DNA in Presence of Different Salt Concentrations and Temperatures under Confinement.(manuscript under revision)

Conferences and Workshops

1. Participated in 'Intra IIT latex workshop' (2017) held at IIT Guwahati.

2. Presented a poster in the conference 'Emerging Trends in Chemical Sciences' (ETCS-2018) held at Dibrugarh University.

3. Presented a poster in the conference 'Frontiers in Chemical Sciences' (FICS-2018) held at IIT Guwahati.

4. Presented a poster in the conference 'Theoretical Chemistry Symposium' (TCS- 2019) held at BITS-PILANI.

5. Oral presentation in the conference 'Research Conclave' (RC-2019) held at IIT Guwahati.