OCHIJE, HENRY IKECHUKWU, Ph.D. Differential Effect of Alkali Metal Ions on the Structure and Stability of DNA G-Quadruplexes with Potential Role in the Regulation of Gene Expression. (2018).
Directed by Dr. Ethan Will Taylor. 156pp

G-quadruplexes (G-QPX) have been the subject of great interest in research due to their exclusive structural configuration, functionality and potential roles in biological applications particularly gene regulation and Nano-switch applications. Initial research has identified the presence of G-Quadruplex in the promoter regions of the human genome as well as untranslated regions of RNA. Promoter regions of human DNA have been found to be rich in Guanine which is one of the essential nucleotides required for the formation of G-Quadruplex structures with the potential to regulate gene activities, thereby acting as a biological switch by either allowing or disrupting the binding of transcription factors during gene expression. Ions such as potassium and sodium have been known to play key role in the formation and stability of G-Quadruplex leading to its potential role in up regulation and down regulation of transcription factors by interfering with their ability to bind on their specific binding sites on DNA. Though a variety of research has shown the potential role of alkali metal ions in the stability of quadruplex structure, not much has been done on the effect of concentration of theses ions on the structure of G-Quadruplex. The focus of this research is to understand the differential effect of alkali metal ions on the structure and stability of DNA G-Quadruplexes with potential role in the regulation of gene expression. Three alkali ions have been employed in this research, sodium, potassium ion because of their potential role in cell activities and Lithium ion because of its potential use in therapeutics.

We also worked with four different modified crystal structures of the complex human telomeric repeat G-quadruplex (4fxm,1xav,217v and 2hri) In silico. The study was done under different ion concentration (50,100,150 and 160 mM).

Different computational analysis was done on the structural change of G-Quadruplex DNA with change in ion concentration using Gromacs computational software and algorithms. The obtained result suggests different response in structural change of G-Quadruplex DNA to changes in concentration of different ions. The effect of lithium ion on the structure of G-quadruplex has been an area of debate in the literature. Our results show that different G-Quadruplex and structures respond to variations in ion concentration differently, with some alkali ions favoring stability at lower concentration and acting differently at higher concentration.

We observed deviations or structural changes in 4fxm G-Quadruplex with the three ions used in the simulation, sodium ion caused the most structural change, followed by lithium and potassium, making potassium more favorable in terms of structural change and stability. The same analysis was carried out on 1XAV G-Quadruplex, it also suggests lithium caused the most structural change followed by sodium and potassium, which further confirms potassium favors structural stability more than the other two ions (Li and Na^{+}). A different result was observed on the structure of another G-Quadruplex 217v, the analysis shows a different response compared to the initial results, lithium caused the most structural change followed by potassium and sodium respectively. This could potentially tell that different G-Quadruplex structures may have a different response to variations in ions and ion concentration. Overall, potassium ion seems to favor G-

Quadruplex structural stability more than sodium and lithium. Lithium seem to cause more destabilizing effect at lower concentration and acts as stabilizing agent at higher concentration in some G-Quadruplex structures.

DIFFERENTIAL EFFECT OF ALKALI METAL IONS ON THE STRUCTURE AND STABILITY OF DNA G-QUADRUPLEXES WITH POTENTIAL ROLE IN THE REGULATION OF GENE EXPRESSION

by

Henry Ikechukwu Ochije

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Greensboro
2018

Approved by

_____
Committee Chair

DEDICATION

To God Almighty who has given me the strength to go this far, the completion of this dissertation was by his grace; I was able to make it through this program with limited time on my side.

APPROVAL PAGE

This dissertation written by Henry Ikechukwu Ochije has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____
Ethan Will Taylor, Ph.D.

Committee Members _____
Hemali Rathnayake, Ph.D.

_____
Dennis R. LaJeunesse, Ph.D.

_____
Liam Duffy, Ph.D.

6/27/2018
Date of Acceptance by Committee

6/27/201
Date of Final Oral Examination

ACKNOWLEDGMENTS

Special thank you goes to my advisor Dr. Ethan Will Taylor for believing in me and giving me the opportunity to live my dream. I am thankful for his guidance and suggestions and support that have helped me throughout my stay in school.

I also extend my thanks to Dr. James G. Ryan, Dean of Joint School of Nanoscience and Nanoengineering. As one of my advisory committee members, thank you for your guidance.

I would like to thank Dr. Dennis R. LaJeunesse, Dr. Hemali Rathnayake, Dr. Liam Duffy and Dr. Jianjun Wei for being part of my work and serving in my advisory committee.

I am thankful to my parents, siblings, my wife and children, who have encouraged and supported me during my studies.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION TO G-QUADRUPLEX STRUCTURES

This dissertation is focused on understanding the differential effect of alkali metal ions on the structure and stability of G-Quadruplex DNA with potential role in the regulation of gene expression. In this chapter we will discuss the formation of G-QPX motifs, topologies and structural differences.

1.1 G-QPX Motifs, Topologies and Structural Stability and Differences

DNA is known for its ability to store and pass along genetic information, but the discovery of the form of DNA known as G-quadruplex sheds new light on the on the activities of DNA in the regulatory process at the cellular level [1]. G-Quadruplex secondary structures are non-B form nucleic acid DNA/RNA sequences usually rich in guanine residues that are self-associated. They are helical structures with guanine tetrads formed from one, two or four strands which deviates from Watson-Crick hydrogen bonding found in B-form duplex DNA [2-4].G-quadruplex structure can be parallel (fig 1d), antiparallel (fig 1f) or a combination of both (fig 1e), with G-tetrad guanines adopting anti or syn conformations around the glycosidic bonds. G-tetrad guanines from parallel G-strands have the same glycosidic conformation, whereas the antiparallel G-strands shows the opposite glycosidic conformation (Fig. 1b) [5, 6].The unimolecular forms of G-quadruplex are naturally found near the ends of the chromosomes or telomeric regions of human chromosomes.

and in transcriptional regulatory regions of multiple genes and oncogenes. The planar structure of four guanine bases is held together by hoogsteen hydrogen bonding of guanine bases with a central cavity. The stack of guanine tetrads can form a structure known as G-quadruplex (also called G-tetrad or G-quartet). The resulting structure is stabilized by the presence of monovalent cationic charge, which sits in a central channel between each pair of tetrads, especially ions like potassium (K+) or sodium (Na+). K+ is considered to be more relevant to the structural formation due to its intercellular concentration (K+ ~140 mM and Na+ ~5 -15 mM), coordination with eight guanine (O6) atoms and lower dehydration energy [2, 6-9].

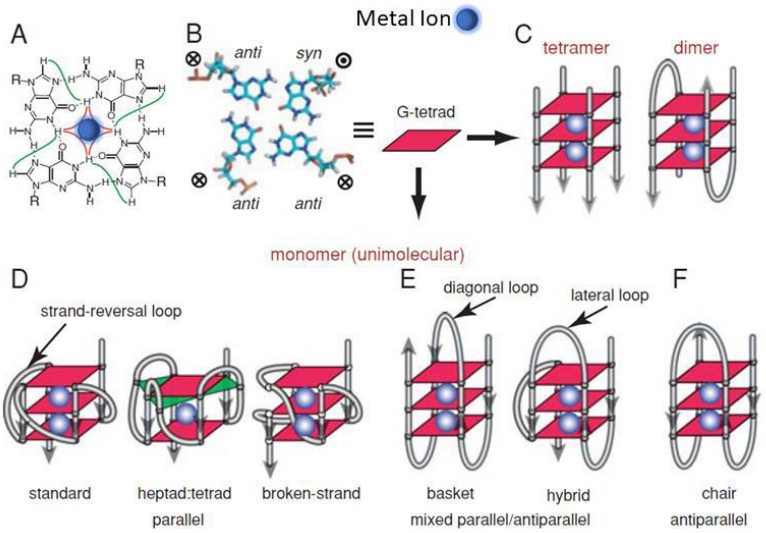

Figure 1. Structure of G-quadruplex and their Folding Topology [10].

1.1.1 Cations in G-Quadruplex Stability

The interaction between cations and quadruplex depends on the ions on the central channel, the central channel is formed due to the quartet arrangement [11]. Ions can be located in the quartet plane or between the planes of the quartet depending on the ion and the structure of the G-quadruplex [12]. The ions play an important role in the formation of G-Quadruplex by contributing to charge screening, and interactions with the loops and the grooves, which supports stability of the formed G-Quadruplex structure. Initial molecular dynamics simulation studies on G-quadruplex suggest that absence of coordinated cation at the center of the quartet destabilizes the G-Quadruplex structure and causes electronic unfavourability to the stability of the structure [13, 14]. Physiologically relevant monovalent ions like Na^+ and K^+ have been extensively studied on their role in the stability and formation of G-quadruplex structure. Other monovalent and divalent ions have also shown influence in the structure and stability of G-quadruplex. Some earlier studies on the role of cations on the stability of G-quadruplex structures suggested that $K^+ > Ca^{2+} > Na^+ > Mg^{2+} > Li^+$ and $K^+ > Rb^+ > Cs^+$ in coordination and stability of G-quadruplex structure [15]. Other studies and analysis suggest interaction of G-Quadruplex with type Ia and IIa cations shows stability in the following order $Sr^{2+} > Ba^{2+} > Ca^{2+} > Mg^{2+}$ and $K^+ > Rb^+ > Na^+ > Li^+ > Cs^+$ [16].

There has been notion and counter notion on the role of Li^+ in the stability of G-Quadruplex structure, some studies have observed that it rather plays a neutral role wherein it neither stabilizes nor destabilizes G-Quadruplex structures, as have been

observed by other research groups [17, 18]. Others are of the opinion of great destabilizing effect of Li^+ on the structure of G-quadruplex [19, 20]. Although K^+ and Na^+ have been studied to be more physiologically relevant to the structure and formation of G-Quadruplex, other ions such as monovalent cations Rb^+, Cs^+, NH+4, and Tl^+ and divalent cations Sr^{2+}, Ba^{2+}, and Pb^{2+} have also been known to promote GQ structures in some specific cases [21].

Some divalent cations in low concentration have been observed to have a disruptive effect on G-Quadruplex structures. Ions like Mn^{2+}, Co^{2+} or Ni^{2+} have been found to disrupt G-Quadruplex structures even in the presence of K^+ ions. It has also been observed that at high or low temperature, divalent cations, for such as Ca^{2+}, Co^{2+}, Mn^{2+}, Zn^{2+}, Ni^{2+}, and Mg^{2+}can induce destabilizing effect on G-Quadruplex [22]. Studies have shown that destabilizing activity by divalent cation in dimeric antiparallel G-Quadruplex $d(G_4T_4G_4)_2$ could be in the following order $Zn^{2+} > Co^{2+} > Mn^{2+} > Mg^{2+} > Ca^{2+}$ and could induce a transition to a parallel G-Quadruplex structure [23]. However, certain divalent metal ions stabilize G-Quadruplex structure more than the others, but such effect depends on the structure of the G-Quadruplex. The studies by Marky et al on metal ions on G-Quadruplex shows that G-Quadruplex formed in the presence of Sr^{2+} and Ba^{2+} are thermodynamically more stable than those formed by Mg^{2+} and Ca^{2+}; this stability has been attributed to their ionic radii [21, 24]. They also characterized two major hydration contributions for the GQ folding: (i) the water uptake upon folding of a single strand into a GQ structure, (ii) dehydration of both cations and guanine (O6) atomic groups. An ESI-

MS study by Vairamain et al. did not observe a smooth correlation between the ionic radii and G-Quadruplex formation, thereby suggesting the size of the metal ion may not be the only factor influencing the stability of the G-Quadruplex structure [25]. Another ESI-MS study on GQ formation by deoxy guanosine with various alkali earth metal ions by Vairamain et al provided evidence that the stability order of GQs formed with divalent cations is $Sr^{2+} > Ba^{2+} > Pb^{2+} > Ca^{2+} >> Mg^{2+}$, suggesting that the metal–oxygen bond is responsible for the stabilization of GQ structures [25].

A computational study on the effect of interaction between alkaline earth metal cations and G-tetrads suggests the cation's ability to stabilize G-Quadruplex structure is in the following order: $Be^{2+} > Mg^{2+} > Ca^{2+}$. Smaller ions are tightly bonded to the G-Quadruplex, signifying dominating effect of electrostatic interaction in the cation–tetraplex systems [21].

A study by Liu et al on the kinetics of G-Quadruplex formation in the presence of divalent cations using an oligonucleotide sequence, $d(GTG_3TAG_3CG_3TTG_2)$ shows the formation of unimolecular G-Quadruplex in the presence of Pb^{2+} [26]. Further studies have also shown that trivalent ions have destabilizing effect on G-Quadruplex structures and potentially promote the stacking of quartets [27, 28]

1.1.2 G-Quadruplex and Metal Ion Coordination

Metal ion interaction sites arise mainly from the negatively charged phosphate backbone of the nucleic acids, causing non-specific charge neutralization. Various studies suggest the central channel metal ions of the G-Quadruplex structure plays a vital role

and contribute to the stability of G-Quadruplex. Structures of G-Quadruplex are formed by the stacking of G-quartets upon each other. The stacking provides potential sites for cation coordination with eight O6 atoms between two stacked G-quartets or four O6 atoms within the plane of a G-quartet. Cation coordination is alleged to contribute more to the stabilization of the G-Quadruplex structure than stacking interactions or the hydrogen bonding [29].

One of the major criteria for the selection of cations for G-Quadruplex is the ionic radius. Large ionic radius is coordinated in-between the plane while smaller atomic radius is coordinated in the plane. For example, K^+ with atomic radius of 1.33 angstrom will fit in-between coordination planes while ion like Na^+ with atomic radius 0.95 angstrom will be coordinated in the plane of a G-quartet because of its size. Coordination interactions between guanine (O6) atoms and cations contribute significantly to the overall stability of a G-Quadruplex. The smaller Na^+ ion allows for in-plane coordination. Multiple Na^+ ions are not restricted to the spacing between G-quartets, and can move further away from each other to reduce electrostatic repulsions [21, 30].

1.1.3 Metal Cations in G-Quadruplex Folding and Stability

A study on the interaction of Na^+ with G-Quadruplex using high resolution crystal structure of DNA hexamer d(TG_4T) with resolution of about 0.95 angstrom suggests the formation of a tetramolecular parallel stranded G-quadruplex structure with Na^+ counterion. Images obtained in this research reveal two G-Quadruplex could stack in 5' to

5' orientation with seven sodium ion coordinated along the axial channel [31]. It was observed that Na^+ at the terminal quartet are coplanar with G-quartets and slightly displaced from quartet plane, due to the electrostatic repulsion between adjacent Na^+. The coordination geometry is also seen to have slightly changed [32]. Characterization of naturally occurring sequences also suggests that *oxytricha nova* telomere sequence $d(G_4T_4G_4)$ forms a bimolecular quadruplex with four quartets and a row of five K^+ ions along the central axis, with an average distance of 3.38 angstrom between any two tandem K^+ [21].

1.1.4 Folding and Topology of G-Quadruplex

G-quadruplex can be formed when four guanine repeats self-associate. It can also be formed in solution of four DNA strands containing short runs of guanine. This can be determined by some major factors such as the nucleotide X, the length of the nucleotide Xn, and the number or length of Guanine Gp.

Xn Gp Xn

Crystallographic and NMR structure of G-Quadruplex studies showed that structures with short guanine runs also known as tetramolecular or intermolecular quadruplex such as $d(TGGGGT)_4$ tend to have parallel structure arrangement [33].

The phosphate backbone is in the same direction and all the bases in *anti* glycosidic direction (Fig. 2c stacked parallel QPX in 5' to 5' head to head interaction). Guanine self-association can occur in four different ways, as shown in Fig. 9a-d although a parallel orientation has been observed in monomeric solutions containing single G runs (Fig. 9a).

The 5' and 3' ends of the runs forming parallel arrangements of G-Quadruplex are capped by non-guanine bases. More complex arrangements are observed in structures that are not capped on both ends by other bases such as d(GGGT). In this complex arrangement, the interlocked quadruplexes are formed from eight associated strands of G rich DNA, which suggests the potential of single stands to form complex structural arrangement in solution while still utilizing guanine tetrads as key building blocks. Strands containing two guanine repeats separated by non-guanine nucleotide form more complex structure and topology. The association of the two strands could form a structure of a four stranded G-quadruplex, also known as dimeric duplexes. The topology formed by these quadruplexes depends on the connectivity of the strands. Strands can associate in four possible ways as can be seen on Fig. 2a and 2c-e. Strands could be linked to the quadruplex in diagonal, lateral (edgewise) or external (strand exchange, propeller or double chain reversal) Fig. 3(a-e). The linking nucleotide is a major factor in the stability of the quadruplex, as well as the length and the sequence formed.

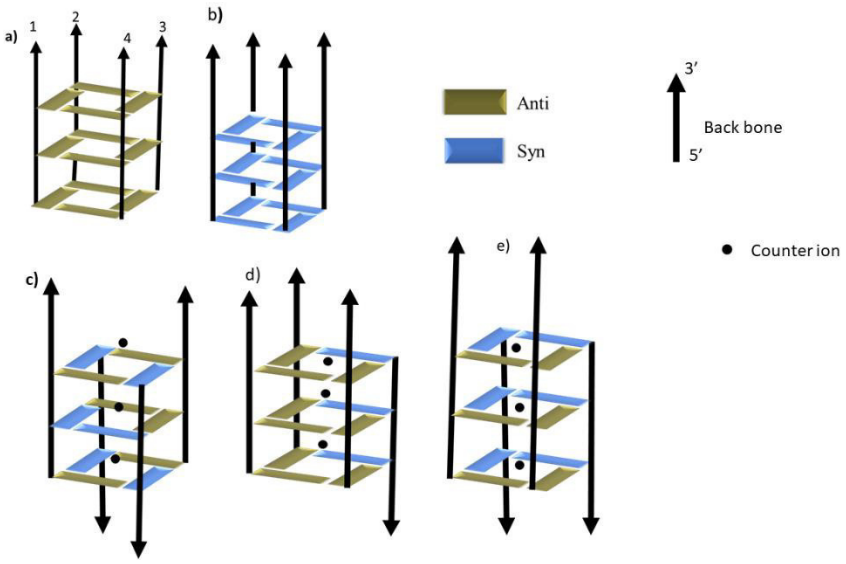

Figure 2. This Shows the Potential Strand Polarity and its Related Glycosidic Torsional Angles for Intermolecular Quadruplexes. In a) the Anti Glycosidic Torsional Angles are Parallel; in b) all Syn Glycosidic Torsional Angles are Parallel; from c) to e) it shows the Alternating Strand arrangements with their mixed Syn/Anti relationships.

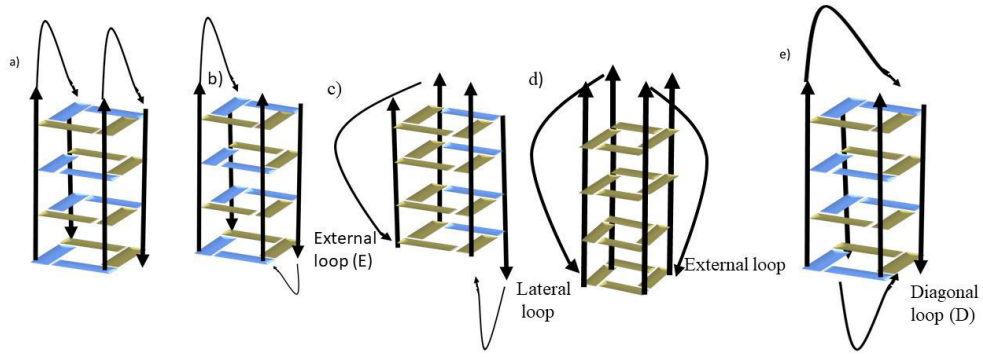

Figure 3. This Shows the Potential Linking Loop Arrangements and Resulting Topologies for Biomolecular Quadruplexes. a) Lateral Loops Linking to Antiparallel Strands arranged on the same face; b) Lateral Loops Linking Antiparallel Strands on either end; c) Diagonal Loops Linking Antiparallel Strands that are opposite; d) External Chain Reversal Loops Linking Parallel Strands together; e) Diagonal Loops Linking Antiparallel Strands that are opposite.

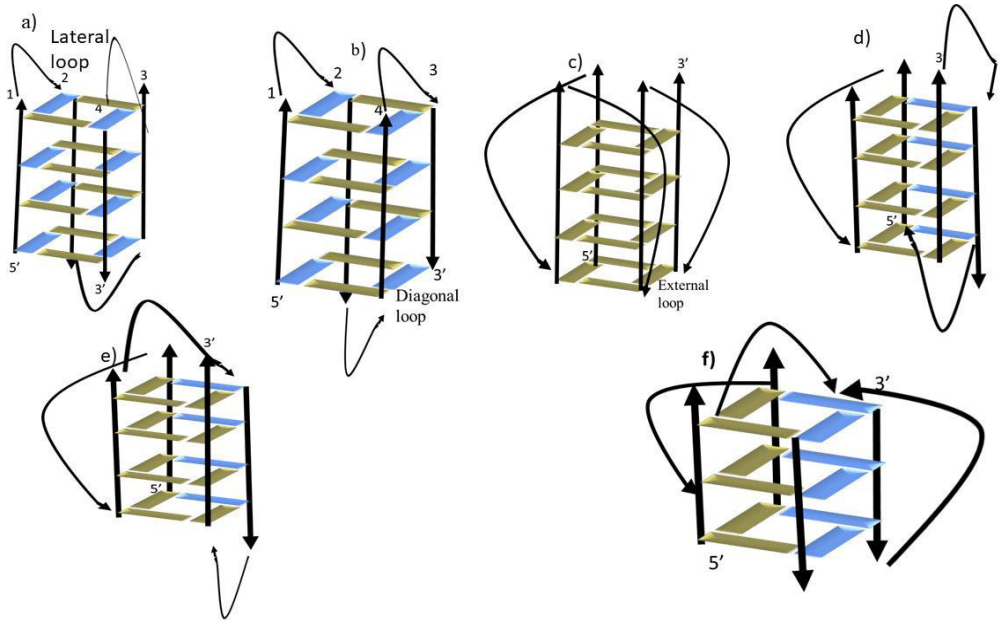

Figure 4. This Shows Potential Linking Loop Arrangements and their Resulting Topologies for Intramolecular Quadruplexes. a) Lateral Loops Linking Antiparallel Strands arranged on the same face; b) Mixed Lateral and Diagonal Loops Linking Antiparallel Strands; c) External Chain Reversal Loops Linking together Parallel Strands; d) Mixed External and Lateral Loops Linking Strands together; e) Mixed External, Diagonal and Lateral Loops Linking Strands together; f) Mixed External and Diagonal Loops Linking Strands together.

1.1.5 Loop Length, Sequence and Stability

One of the constraints to the type of topology or structure formed by intermolecular or bimolecular G-Quadruplex is the number of linking nucleotides or the length of the nucleotide**.** Short lengths or sequences of two or less can accept both lateral and external loop but will prevent diagonal loops from forming due to the distance to be covered across the G tetrad. A stack of up to three guanine tetrads connected by an external loop can be formed by a single connecting nucleotide, while a stack of four guanine tetrads requires at least two nucleotides in the linker. Short linkers can cause

certain topology constraints on the folded quadruplex, causing parallel structure formation. A CD gel electrophoresis experiment using a sequence $d(GGGGTGGGG)_2$ included as nucleotide linker showed that introduction of two thymine residues into the linker resulted in a mixed parallel and anti-parallel structural formation of G-Quadruplex, while lengthening the sequence by three or more nucleotide linkers resulted in only anti parallel folded quadruplexes. Longer linkers have been observed in lateral, diagonal and external connecting loop, although no relationship has been established linking G-Quadruplex stability to the length of the loop [34].

1.2 Potential G-Quadruplexes in the Human Genome

Computational analyses on Prevalence of quadruplexes in the human genome using $(G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+})$ sequence revealed over 300 000 sequences that could potentially form G-quadruplexes (pG4 or PQS) [35, 36].

The localization of pG4s were found to be non-random in the functional regions of the genome and are highly conserved between different species [37]. This indicates that there is a selection compression in retaining such sequences at specific genomic sites. It has been found that the this conservation is highest among mammalian species compared to non-mammalian species and other lower organisms [38]. The mammalian telomere has been known to contain the highest amount of pG4s. In humans, pG4 consist of 5 to 10 000 bp of the tandemly arranged TTAGGG repeat. About 90% of human DNA replication origin contains pG4 motif and have been found to be rich with gene promoters, at the border between introns and exons and target immunoglobulin gene class

switch recombination [39, 40] (Figure5A–C). It has been observed that pG4s present in G-rich RNA transcripts of telomeric DNA, TERRA and the 5′-UTR region of the encoded mRNAs may indicate translation and suggests possible regulatory role due to their ability to for G-Quadruplex structure (Figure5D) [41-43].

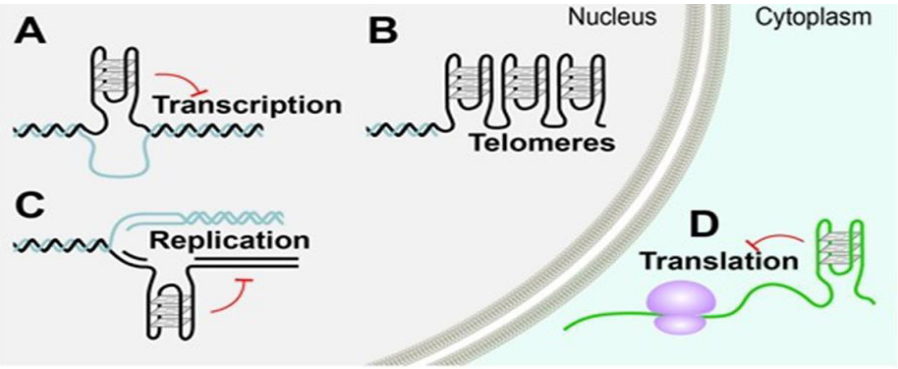

Figure 5. Locations of G-quadruplex Structures in Cells. In the Nucleus, G-quadruplex formation can occur in double Stranded G-rich regions when DNA becomes transiently single stranded, during (A) Transcription and (C) Replication and (B) at the single stranded Telomeric G-rich overhangs. Outside the Nucleus, there is formation of the G-quadruplexes in the mRNA and (D) involves Translational control. Red T-bars indicate disorders affecting Transcription, Replication and Translation [36].

1.2.1 Possible Locations of G-Quadruplex Structures in Cells

Genome wide searches have revealed the location of G-rich regions with G-quadruplex forming potential (pG4). pG4s are non-randomly distributed in the genome and promoters and telomeres are particularly enriched in these sequences. In the nucleus, G-quadruplex formation can occur in double stranded G-rich regions. When DNA

becomes transiently single stranded, during (**A**) transcription and (**C**) replication and (**B**) at the single stranded telomeric G-rich overhangs. Outside the nucleus, G-quadruplexes can also form in mRNA and (**D**) are involved in translational control (fig 5). Red T-bars indicate impediments to transcription, replication and translation Potential involvement of some helicases in unwinding of DNA or RNA G-quadruplexes has become one of the key research interest area. Helicases are part of ATP dependent nucleic acid unwinding enzyme. Genetic disorders and mutations causing functional loss in DNA helicases have been linked to pG4 sequences and genome instability. Studies have shown that G-quadruplex helicases play important roles in DNA replication and telomere maintenance [44]. Telomere maintenance involves human helicases which aids in G-quadruplex-unwinding activity with conserved RQC which is a DNA-binding domain found only in RecQ family enzymes. RecQ family helicases can unwind pG4 DNA and play vital role at G-rich areas of the genome, this includes the telomeres, immunoglobulin switch regions and rDNA. This domain has a helix-turn-helix structure and acts as a high affinity G4 DNA binding domain which has affinity for G-quadruplexes [45, 46].

The kinetic abilities of G-Quadruplex in formation and resolution has been an interesting area of study with regards to its regulatory role in biological system. Studies indicates the possible role of protein chaperons and helicases in the formation and resolution of G-Quadruplexes respectively. The folding kinetics of G-quadruplexes forming sequences could be relatively fast or slow and may range from microseconds to minutes. For sequences such as human telomeric repeats, the folding time scale can be

compared to that of DNA replication [47]. Studies have suggested the possible role protein chaperones in increasing the kinetic of G-Quadruplex formation [48-50]. Literatures have shown that human telomeric binding protein TRF2 contributes in both DNA and RNA (TERRA) G-quadruplex binding [51-53].

Recent studies show that human DNA mismatch recognizing factor MutSα enhances synapsis of transcriptionally triggered immunoglobulin switch regions and binds to G-quadruplexes, this also suggests that nucleophosmin (NPM1) interacts with some G-quadruplex regions in ribosomal DNA *in vitro* and *in vivo*. NPM1 is a nucleolar protein have been found to show up in ribosome maturation, export and have been found to be one of the genes mutated in acute myeloid leukemia [54, 55]. An important nucleolar phosphoprotein nucleolin (NCL) which binds DNA G-quadruplexes and its sequestration was found to be based on the conformation of G-quadruplexes present in terminated RNA transcripts arising from the expansion of the hexanucleotide repeat (GGGGCC)n (Figure 6B). This has been implicated in a kind of nucleolar stress and perturbations in RNA processing, that leads neurodegenerative disorder [56, 57].

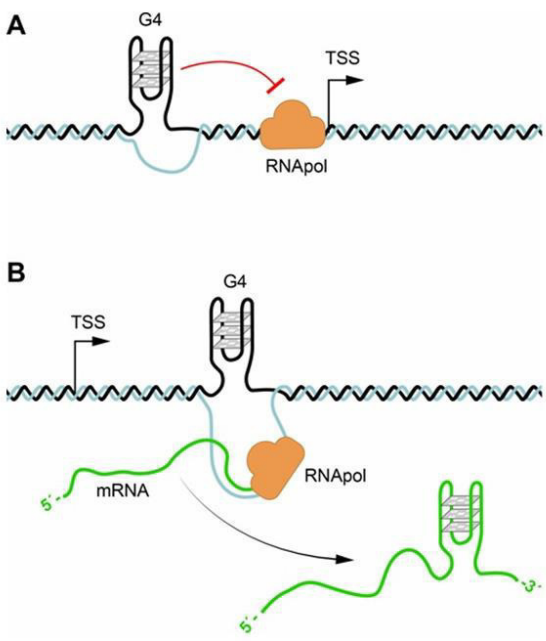

Figure 6. G-quadruplexes in Transcription and Translation. (A) pG4 sequences are present in about 50% of Human Genes Promoters. G-quadruplex formation could impair initiation of Transcription by the RNA Polymerase, or if present in the antisense Strand inhibit Transcription. (B) the presence of G-quadruplexes formed in the 5′ UTR of mRNAs can regulate Translation as well as lead to aborted RNA Transcripts in Hexanucleotide repeat expansion diseases[36].

1.2.2 Quadruplexes Formed in Human Telomeres

G-quadruplex formation in human DNA was first observed in telomeric region of human DNA which consists of tandem repeat of nucleotides d(TTAGGG)n 5–25 kb in length and terminates in a single terminate in a single-stranded 3′-overhang of 35–600 bases [58]. Telomeric DNA has been extensively studied and linked to some telosome proteins. The structure and stability of telomeres have been associated with aging and genetic stability [59, 60]. Human telomeric DNA sequence or repeats are highly conserved, literatures suggests this feature to be related to its ability to form the DNA G-

quadruplex [61]. Studies have also shown that the telomere end-binding proteins TEBPα and TEBPβ are essential in controlling G-quadruplex formation and are resolved by phosphorylation of TEBPβ during replication [58, 62]. G-quadruplex formation was detected at the ends of human chromosome using a radiolabeled G-quadruplex ligand ^{3}H-360A (2,6-*N*,*N*′-methyl-quinolinio-3-yl)-pyridine dicarboxamide [methyl-^{3}H] on a cultured human cell [63]. In a related research, in vivo existence of G-quadruplexes in human chromosomes telomeres and promoter regions was observed by using a structure-specific fluorescent antibody in human cells. Intramolecular G-quadruplex in human telomeric DNA induced by K+ ion has been shown to inhibit the telomerase activity. (fig 4B)Further research showed that the major G-quadruplexes formed in human telomeres in K+ solution are the hybrid-type intramolecular structures (Fig.2A).

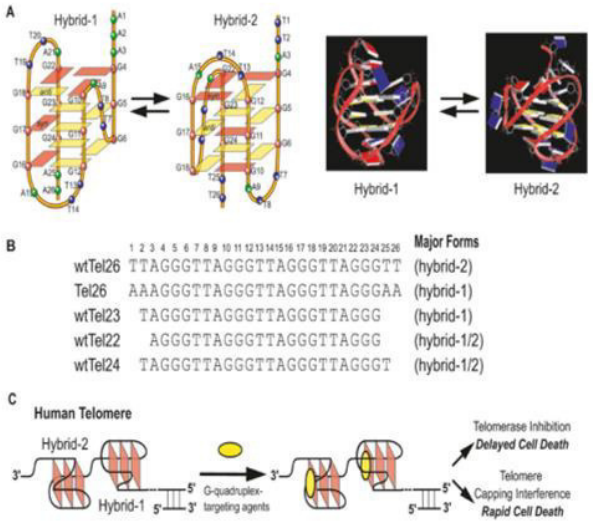

Figure 7. Hybrid G-quadruplex Formation and Human Telomere [64]

During cell replication in somatic cells, Telomeres lose 50- to 200-base in each cell replication, this base loss leads to shortening of the telomeric DNA and the cell undergoes apoptosis at critical loss stage. However, telomeres of cancer cells do not shorten on replication, due to the activation of telomerase, a reverse transcriptase that extends the telomeric sequence at the chromosome ends. Literature has shown that Telomerase is activated in 80–85% of human cancer cells suggesting its key role in maintaining the malignant phenotype. Intramolecular G-quadruplex in human telomeric DNA induced by K+ ion has also shown inhibition of telomerase activity [58, 65-67].

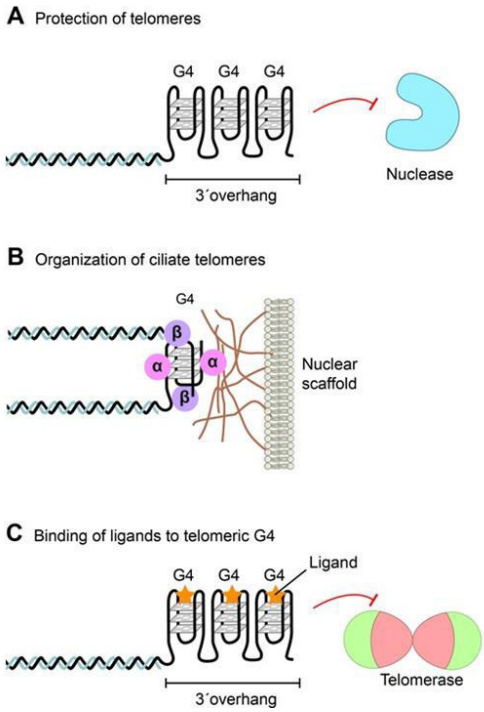

Figure 8. G-quadruplexes at Telomeres. The G-rich overhang of Telomeres can form G-quadruplex Structures involved in Telomere end protection and Telomeric DNA Metabolism. (**A**) The long Human G-rich overhang can form Strings Intramolecularly folded G-quadruplexes that may offer end protection against Nucleases or Regulate Telomerase activity. (**B**) Ciliate Telomeres form Intermolecular G-quadruplex Structures

involving two telomeres promoted by the telomere-end binding protein TEBPβ. Telomeres are attached to a sub-nuclear structure (the nuclear matrix or scaffold) via an interaction of the telomere-end binding protein TEBPα. (**C**) Stabilizing of G-quadruplexes by G-quadruplex binding ligands (yellow stars) impairs telomere repeat synthesis by the telomerase enzyme and lead to telomere shortening [36].

Available research data has shown that many G-quadruplex stabilizing ligands do not target the telomerase enzyme, but the telomere itself [68, 69]. Some of the methods used in identifying G-quadruplexes at human telomeres includes, the use of a fluorescent cyanine dye [70, 71], the binding of a radiolabeled G-quadruplex ligand to metaphase chromosomes [63] and, the use of specific antibodies that binds to human G-quadruplexes [72-74]. Signals were detected at the ends of some of the chromosomes using the above methods of detecting G-Quadruplex. Though G-Quadruplex human telomeres remains to be established because resolution of light microscopy is inadequately high to read whether binding occurred at the very end of the chromosome or at sub telomeric regions which is believed to adopt G-quadruplex structures *in vitro* [75].

However, the evidence of a number of helicases that are known to unwind G-Quadruplex *in vitro* are found in telomeres and are essential for telomere integrity *in vivo*, this finding provides strong inferred evidence for the existence of G-Quadruplex in mammalian telomeres [36, 45, 76-78].Figure8 above shows the potential function of WRN and BLM helicases which seem to ensure proper replication of telomeric DNA by supporting the unwinding G-quadruplex impediments. To prevent accumulation of chromosome aberrations such as chromosome fusions and much loss of telomere during lagging strand replication of the G-rich strand the action of WRN helicase is required[79,

80]. WRN helicase is found around and interacts with telomere binding proteins TRF2 and POT1[81]. The high affinity of WRN and BLM for POT1, suggests that DNA binding telomeric proteins may function in the presence of helicase. Cells that are deficient in RETL1 tend to exhibit telomere fragility. This fragility is enhanced TMPyP4 which is a G-Quadruplex stabilizing agent. However, the role of RETL1 seem more of supporting replication in the genome due to its direct interaction with the replication clamp known as PCNA[76, 82, 83].

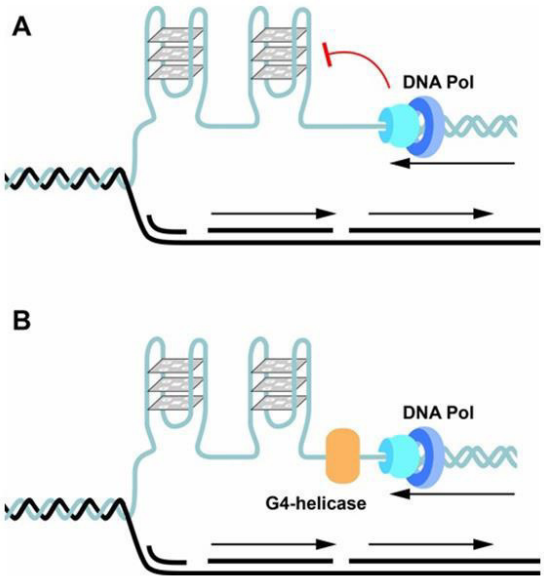

Figure 9. G-quadruplexes and Replication [36].

1.2.3 G-Quadruplexes in Transcription and Translation

The experimental discovery of G-Quadruplex around the promoter region of human gene suggests its potential in regulating gene expression Fig 9. It has been found that guanine rich region of human DNA are more in the oncogenes or regulatory genes

than in tumor suppressor genes[18, 84]. One of the studies on oncogene c-MYC shows that mutation of pG4s at promoters or the addition of a G-quadruplex stabilizing ligand affected transcription *in vivo* and could potentially play a role in gene expression or regulation Fig8A[85, 86].Alterations in gene expression of genes containing pG4s has been observed in a study where G-quadruplex specific single chain antibody was used, the result was seen to affect both promoters and pG4s at the ends of genes, indicating possible role of G-quadruplexes in both transcriptional initiation and termination [36, 87, 88]. Reports form computational studies has shown the presence of RNA pG4s in the 5' untranslated region(UTR) of many genes, this report also shows the connection between G-quadruplexes and regulation of translation both *in vitro* and *in vivo* studies [41]. Studies on fragile X mental retardation, which is traceable to (CGG)n repeat expansion suggests the involvement of G-quadruplexes in mRNA in neuronal function [89] and possible target of the mental retardation protein FMRP. FMRP through G-quadruplex present in its 5′ UTR binds to its mRNA, this binding hinders translation resulting to altered brain mRNA translation [90]. A research on the role of the RNA helicase eIF4A, using ribosome signals to provide snapshots of translation across the transcriptome, has shown the presence of pG4 $(CGG)_4$ which is eIF4A dependent transcripts of 12-nucleotide that can form RNA G-quadruplex structures [91, 92]. eIF4A has been known to promote T-cell acute lymphoblastic leukemia by favoring the translation of mRNAs complex UTRs. Indicating possible role of G-quadruplexes in gene regulation [91, 92].

CHAPTER II

THEORY OF MOLECULAR DYNAMICS

2.1 Molecular Dynamics (MD)

Molecular dynamic simulation has become one of the most important tools in theoretical studies of liquids and biomolecular systems such as deoxyribonucleic acid (DNA) and Proteins. Molecular dynamics give insight into the understanding of properties and assemblies of molecules with regards to the molecular structure and the microscopic interactions between the molecules. it is a many body system consisting of three or more bodies with no general analytical solution. it is a way of simulating material at atomic level with inputs of the atoms position, mass and velocity. The motions of atoms are divided into time steps usually measured in femtoseconds. All the forces between the atoms are integrated for each time step to give the new position and velocity which is iterated through the period of simulation. The analysis of this time steps can be used to evaluate the properties of the given material. This helps in predicting the outcome of conventional experiment using the two main simulation techniques: the molecular dynamics (MD) technique and the Monte Carlo (MC). One of the advantages of MD over MC is that it allows insight into the dynamic properties of the system with regards to the time dependent response to perturbation, transport coefficients, spectra and atomic properties.

Molecular dynamics simulations calculate the motion of the atoms in a molecular assembly using Newtonian dynamics to determine the net force and acceleration experienced by each atom. Each atom i at position r_i , is treated as a point with a mas m_i and a fixed charge q_i .MD simulations solve Newton's equations of motion for a system of N interacting atoms using Newton's second law:

$$F = ma \quad \text{..... (1)}$$

where F is force on an atom, m is mass of the atom, and a is the atom's acceleration. The equation can be written as

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i, i = 1.....N \quad \text{.....(2)}$$

The force acting on the molecule is given as the negative derivative of the potential

$$V(r_i, r_2, ...r_N) \quad \text{..... (3)}$$

The force can be expressed as

$$F_i = -\frac{\partial V}{\partial r_i} \quad \text{..... (4)}$$

where r represents coordinates of all atoms, and V is the potential energy function. Velocity v is the derivative of position, and acceleration is the derivative of velocity. We can thus write the equations of motion as

$$\frac{dr}{dt} = v$$
$$\frac{dr}{dt} = \frac{F(r)}{m} \quad \text{..... (5)}$$

This is a system of ordinary differential equations for N atoms, N position and N coordinate with N velocity coordinates "Analytical" (algebraic) solution is impossible.

Numerical solution is could be given as

$$r_{i+1} = r_i + \delta_t v_t$$
$$v_{i+1} = v_i + \frac{\delta_t F(r_i)}{m} \quad \text{..... (6)}$$

Where time step is represented as δ_t

For simulation purpose we use time symmetric integration method such as Leapfrog Verlet method

$$r_{i+1} = r_i + \delta_t v_{i+1/2} \quad \text{..... (7)}$$

$$v_{i+1/2} = v_{i-1/2} + \delta_t F(r)/m \quad \text{..... (8)}$$

Table 1. Example of Vibrational Frequencies(wavenumbers) in Molecules and Hydrogen-bonded Liquids. Comparing $kT/h = 200\,\mathrm{cm}^{-1}$ at 300K

Type of bond	Type of vibration	Wave number(cm^{-1})
C-H, O-H, N-H	Stretch	3000-3500
C=C, C=O	Stretch	1700-2000
HOH	Bending	1600
C-C	Stretch	1400-1600
H_2CX	Sciss, rock	1000-1500
CCC	Bending	80-1000
O- H···O	Libration	400-700
O- H···O	stretch	50-200

2.1.1 Classical and Quantum Mechanics in MD

Application of Newton's equation of motion may imply the use of classical mechanics in describing the motion of atoms in the system. This may hold to a certain extent for atoms at normal temperatures, with certain exceptions.

Considering atoms like hydrogen atoms which are quite light and the motion of protons which displays some characteristics of quantum mechanics, such as proton tunneling through a potential barrier during a transfer over a hydrogen bond.

Classical mechanics may not also apply with helium liquid at low temperature due to high frequency vibrations of covalent bonds. The statistical mechanics of a quantum oscillator differs from that classical harmonic when the resonance frequency ν approximates or exceeds $\frac{k_B T}{h}$ at room temperature, the wavenumber

$$\sigma = 1/\lambda \qquad 1/\lambda = \nu/c \text{ .... (9)}$$

At this frequency

$$h\nu = k_B T \text{ .... (10)}$$

is approximated at 200 cm^{-1} thus, all frequencies close to or higher than 200 cm^{-1} may misbehave in classical simulations. This could mean that practically all bond and bond-angle vibrations are contributive having hydrogen-bonded motions as translational or liberational making H-bond vibrations beyond the classical limit. Based on the above, corrections may be necessary when MD simulation is performed using harmonic oscillators for bonds. Correction could be made to the total internal energy $U = E_{kinetic} + E_{potential}$ and specific heat Cv, entropy S and free energy G [93].

The corrections to the energy and specific heat of one-dimensional oscillator with frequency

$$U^{QM} = U^{cl} + kT\left(\frac{1}{2}x - 1 + \frac{x}{e^x - 1}\right) \text{ .... (11)}$$

$$C_V^{QM} = C_V^d + k\left(\frac{x^2 e^x}{\left(e^x - 1\right)^2} - 1\right) \ldots\ldots\ldots\ldots\ldots (12)$$

Where

$$x = hv \diagup kT \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (13)$$

The classical oscillator absorbs much energy (kT) when the high frequency quantum oscillator at zero-point energy level of $\frac{1}{2}hv$ which is its ground state.

Another correction that can be done is to treat the bonds and the bond angles as constraints in the equation of motion. This is because quantum oscillator in its ground state resembles a constrained bond more closely than a classical oscillator. Another reason for this is that the algorithm can use larger time steps when the highest frequencies are removed, and time steps can be made four times as large when bonds are constrained than when they are oscillators [94]. The flexibility of the bond angle correction is essential to allow for the realistic motion and coverage of configurational space. [95]

2.1.2 Periodic Boundary Conditions

One of the ways to minimize edge effects in a finite system when simulating a crystalline system is to apply periodic boundary conditions. A space filling box which contains the atoms to be simulated is surrounded by translated copies of itself. This creates an effect no boundaries in the system; the issue caused by unwanted boundaries in an isolated cluster is replaced by periodic conditions.

Although periodicity conditions could cause error if used in a none periodic system such as liquids or solution. Errors generated in this condition can be evaluated by comparing various system sizes; the errors are expected to be less severe than the errors resulting from an unnatural boundary with vacuum. Unit cells could be filled with several possible shapes such as the rhombic dodecahedron and the truncated octahedron these shapes are more towards spherical than cubic [96].

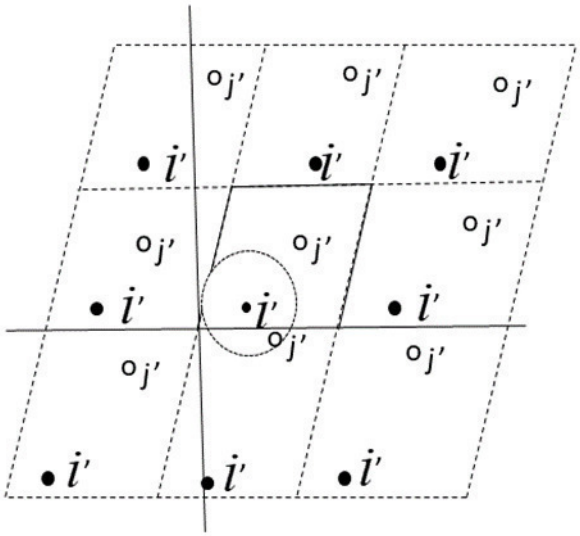

Figure 10. Periodic Boundary Conditions in two Dimensions

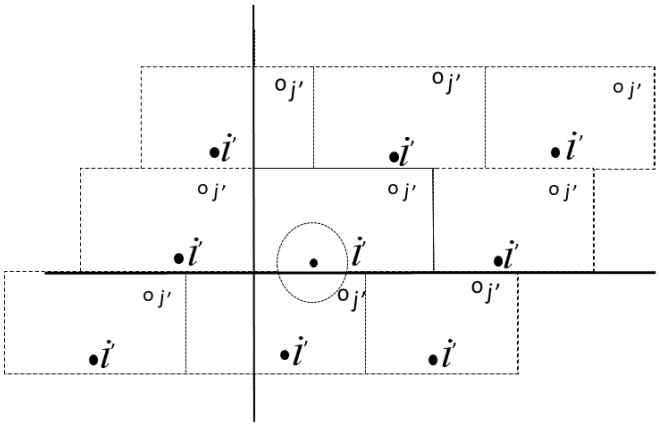

Figure 11. Periodic Boundary Conditions in two Dimensions

This characteristic makes it better for the study of approximately spherical macromolecule in solution, which require fewer solvent molecules to fill the box with minimum distance between the images of the macromolecule. The rhombic dodecahedra and truncated octahedra belongs to a special class of triclinic unit cells which generally fills most unit cells space that comprise all possible space-filling shapes [97]. This is the reason GROMACS is based on triclinic unit cell. GROMACS utilizes periodic boundary conditions along with the minimum image convention allowing only one nearest – image of each particle to be considered for short-range non-bonded interaction terms. For long-range electrostatic interactions this is not always hold in accuracy. To correct this, GROMACS incorporated lattice sum methods such as Ewald Sum, PME and PPPM. This allows GROMACS to supports triclinic boxes of any shape. The simulation box is defined by the 3 box vectors a, b and c. The vectors should satisfy the following conditions:

$$a_y = a_z = b_z = 0 \quad \text{..... (14)}$$

$$a_x > 0 \ , \ b_y > 0, \ c_z > 0 \text{,..... (15)}$$

$$|b_x| \le \frac{1}{2} a_x, |c_x| \le \frac{1}{2} a_x, |c_y| \le \frac{1}{2} b_x \text{..... (16)}$$

Equation 15 can be satisfied by rotating the box and the Inequalities by addition or subtraction of the box vector (15) and (16). Gromacs is designed to keep the particles in a brick shaped and volume for efficiency by using triclinic box as illustrated in figure 11 in a two-dimensional system. This make the output trajectory look like the simulation was done in a triangular box. The trajectory convergence part of the program(trajconv) can be used in converting trajectories for different unit cell representation.

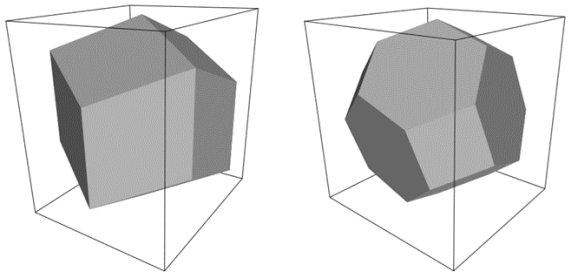

Figure 12. A Rhombic Dodecahedron and Truncated Octahedron (arbitrary orientations)

2.2 Forces

The Potential energy term is one of the terms computed when interaction term is computed. The potential energy for each contribution is summed such as coulomb, Lennard-Jones and bonded terms.

2.2.1 Kinetic Energy and Temperature

The temperature of the system is given by the total kinetic energy of the Nparticle in thy system:

$$E_{kinetic} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (17)$$

Using equation 17 above the absolute temperature T can be calculated from

$$\frac{1}{2} N_{df} kT = E_{kinetic} \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (18)$$

Where k is Boltzmann's constant and N_{df} is the number of degrees of freedom which can be computed from

$$N_{df} = 3N - N_C - N_{com} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (19)$$

N_c is the number of constrains on the system. During simulation $N_{com} = 3$ additional degree of freedom must be deducted because the velocities for the three centers of mass are constant of motions which are set to zero. The rotation around the center of mass can be removed if the simulation is done in a vacuum and this $N_{com} = 6$. When the group of temperature is more than one, the degree of freedom for group i is given as

$$N_{df}^{i} = \left(3N^i - N_c^i\right)\frac{3N - N_c - N_{com}}{3N - N_c} \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (20)$$

The kinetic energy could be represented in the form of a tensor. This is utilized in the calculation of the system pressure where shear forces are imposed or in a triclinic system

$$E_{kinetic} = \frac{1}{2}\sum_{i}^{N} m_i v_i \otimes v_i \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (21)$$

2.2.2 Pressure

The pressure tensor is calculated by taking the difference between kinetic energy $E_{kinetic}$ and Virial Ξ

$$P = \frac{2}{V}\left(E_{kinetic} - \Xi\right) \ldots\ldots\ldots\ldots\ldots\ldots \quad (22)$$

Where V is the volume of the box housing the molecule, In the case of Isotropic the scalar pressure P can be used for pressure coupling and can be computed as

$$P = trace\left(\mathrm{P}\right)/3 \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (23)$$

The virial Ξ tensor is given as

$$\Xi = \frac{1}{2}\sum_{i<j} r_{ij} \otimes F_{ij} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (24)$$

2.3 Integration Methods

2.3.1 The Leap-Frog Integration Method

This is one of the integration methods used in molecular dynamics, it is also the default MD integrator in GROMACS for the integration of the equations of motion. There are also other integrators such as Verlet integrators that may be used if extremely

accurate integration with temperature and/or pressure coupling is required. The leap-frog algorithm utilizes positions r at time t and velocity v at time $t-\frac{1}{2}\Delta t$: this is used to update velocity and position with the force as a function of time $F(t)$

$$v\left(t+\frac{1}{2}\Delta t\right)=v\left(t-\frac{1}{2}\Delta t\right)+\frac{\Delta t}{m}F(t) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (25)$$

$$r(t+\Delta t)=r(t)+\Delta t v\left(t+\frac{1}{2}\Delta t\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (26)$$

The trajectories of leap frog algorithms is identical to the Verlet algorithm with position relation as Verlet [98]

$$r(t+\Delta t)=2r(t)-r(t-\Delta t)+\frac{1}{m}F(t)\Delta t^2+o\left(\Delta t^4\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (27)$$

This algorithm compared with other algorithms is of third order in r and is time-reversible. The equations of motion are modified for temperature coupling and pressure coupling, and extended to include the conservation of constraints, all of which are described below [99]. The equations of motion are modified for temperature coupling and pressure coupling and extended to include the conservation of constraints as shown below.

2.3.2 The Velocity Verlet Integrator

GROMACS algorithm has been built to also implement the velocity Verlet algorithm , though it is not yet fully integrated with all sets of options[100]. To integrate the equation of motion in velocity verlet algorithm, the velocity v , the position r at time t is utilized without the requirement of the velocities at the previous half steps.

$$v\left(t+\frac{1}{2}\Delta t\right)=v(t)+\frac{\Delta t}{2m}F(t) \quad \text{..... (28)}$$

$$r(t+\Delta t)=r(t)+\Delta t v\left(t+\frac{1}{2}\Delta t\right) \quad \text{..... (29)}$$

$$v(t+\Delta t)=v\left(t+\frac{1}{2}\Delta t\right)+\frac{\Delta t}{2m}F(t+\Delta t) \quad \text{..... (30)}$$

The above equation can also be written as

$$r(t+\Delta t)=r(t)+\Delta t v+\frac{\Delta t^2}{2m}F(t) \quad \text{..... (31)}$$

$$v(t+\Delta t)=v(t)+\frac{\Delta t}{2m}F(t)+F(t+\Delta t) \quad \text{..... (32)}$$

Leap-frog and velocity verlet will generate identical trajectories when there is no temperature or pressure coupling as can be seen in the above equation. But if a single starting file with the same starting point $x(0)$ and $v(0)$ is used , leap-frog and velocity Verlet will not give identical trajectories, as leap-frog will interpret the velocities as

corresponding to $t = -\frac{1}{2}\Delta t$, while velocity Verlet will interpret them as corresponding to the timepoint $t = 0$.

2.3.3 Temperature Coupling

Although in Molecular dynamics we can calculate certain ensembles such as constant number, constant volume, constant energy ensemble, temperature is a major factor to changes that happen within the system, so we mostly calculate quantities that from a constant temperature (NVT) ensemble, also called the canonical ensemble. Gromacs utilizes the weak-coupling scheme of Berendsen, stochastic randomization through the Andersen thermostat[101, 102], the extended ensemble Nos´e-Hoover scheme [103, 104] or velocity-rescaling scheme to simulate constant temperature Berendsen temperature coupling. The Berendsen algorithm imitates weak coupling with first-order kinetics to an external heat bath with given temperature T_0 .one of the main difference between Nos e-Hoover scheme is that the deviation of the system temperature from the initial temperature T_0 is slowly corrected according to:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (33)$$

This means exponential decay of the deviation with time at constant τ .This advantage gives room for the strength of the coupling to be varied to the user condition. During equilibration the coupling time could be short for the system, but the run time needs to be longer which rarely affects the conservative dynamics.

Berendsen thermostat works to reduce the fluctuations of the kinetic energy with an error scale within $\frac{1}{N}$. For very large systems the ensemble averages are not affected significantly, except for kinetic energy distribution and the fluctuation properties such as the heat capacity. Another thermostat that does correct ensemble is the velocity rescaling thermostat[104]. The scaling velocity of each particle affects the heat flow in and out of the system every step or every nTC steps, with a time-dependent factor λ, given by:

$$\lambda = \left[1 + \frac{nTC\Delta t}{\tau T} \left\{ \frac{T_0}{T\left(t - \frac{1}{2}\Delta t \right)} \right\} \right]^{1/2} \quad \text{..... (34)}$$

Where τT is approximately equal but not equal to the time constant τ of the temperature coupling(temperature parameter τT), this is because the kinetic energy change due to the scaling of the velocity is partly redistributed between kinetic and potential energy causing change in temperature which is less than the scaling energy.

$$\tau = {2Cv\tau T}/{N_{df}k} \quad \text{..... (35)}$$

Where Cv is the total heat capacity of the system, k is the Boltzmann's constant, and N_{df} is the total number of degree of freedom.

2.3.4 Andersen Thermostat

One of the ways ensemble with thermostat is maintained by taking the *NVE* integrator and periodically resecting the velocities of particles from Maxwell-Boltzmann's distribution [102].

The resecting can be done by randomizing all the velocities simultaneously creating massive collision at every $\tau T/\Delta t$ steps (Andersen-massive),it can also be attained by randomizing all particle and time steps with some small probability every timestep (Andersen), equal to $\Delta t/\tau$ .in both situations Δt is the timestep and τT is a characteristic coupling time scale [102].

2.3.5 Berendsen Barostat

Like temperature coupling, in some cases it is also important to simulate a system at constant pressure which is also a common practice in Monte Carlo simulations. One of the algorithm supported by GROMACS is the Berendsen algorithm [101]. Berendsen algorithm is coupled weakly to an external pressure bath using the principle of least local perturbation, rescaling the coordinates and box vector every step. The pressure coupling expression is derived by adding extra term to the equation of motion which effects pressure change

$$\left(\frac{dp}{dt}\right)_{bath} = \frac{p_{0-}p}{\tau_p} \quad \text{..... (36)}$$

τ_p is the time constant for coupling. The local disturbance is minimized by proportional coordinate scaling associate with volume. The equation of motion is modified as

$$\dot{x} = v + \alpha x \quad (37)$$

The volume will correspondingly change as

$$\dot{V} = 3\alpha V \quad (38)$$

The relationship between pressure change and isothermal compressibility β is given by

$$\frac{dP}{dt} = -\frac{1}{\beta V}\frac{dV}{dt} = -\frac{3\alpha}{\beta} \quad (39)$$

$$\alpha = -\frac{\beta\left(P_0 - P\right)}{3\tau_P} \quad (40)$$

Modifying the equation, we have the equation of motion as

$$\dot{x} = v - \frac{\beta\left(P_0 - P\right)}{3\tau_P} X \quad (41)$$

The equation above represents proportional scaling of coordinates. The compressibility may not be known accurately but occurring in the expression for scaling leaving only the $\frac{\beta}{3\tau_P}$ term occurring in the equation of motion [105].

2.4 Energy Minimization

Energy minimization is the process of finding an arrangement of the constituent atoms in 3D space where, the net inter-atomic force on each atom is close to zero and the position on the potential energy surface is a stationary point.

2.4.1 Steepest Descent

This is a first order iterative algorithm for finding the minimum function. It utilizes gradient descent to find the local minimum, using steps proportional to the negative of the gradient of the function at the current point. This minimization method may not be the most efficient method of minimization, but it is robust and easy to implement. The vector r could be defined as the vector of all $3N$ coordinates with initial displacement h_0 .Given the initial force F and potential energy, the new positions can be calculated by

$$r_{n+1} = r_n + \frac{F_n}{\max\left(\left|F_n\right|\right)} h_n \quad \text{..... (42)}$$

Where h_n is the maximum displacement and F_n is the force and can also be written as the negative gradient V with the notation $\max\left(\left|F_n\right|\right)$ as the largest absolute value of the force components. The number of evaluation can be specified or when the absolute values of the gradient of force is smaller than a specified value $\in$.To avoid endless iterations the stopping criterion is not made too high, a reasonable value of $\in$can be estimated from root mean square force f a harmonic oscillator will exhibit at temperature T .

$$f = 2\pi v\sqrt{2mkT} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (43)$$

Where m is the reduced mass, v is the oscillator frequency and k is the Boltzmann's constant.

2.4.2 Conjugate Gradient

In conjugate gradient algorithm, the first portion of the search takes place opposite the direction of the largest gradient as also observed in the Steepest Descent method. To avoid oscillations that often affects steepest descent method toward the minimum, the conjugate gradient algorithm blends part of the previous direction in the next search. This allows the algorithm to move quickly to the minimum.

2.5 Interaction Function and Force Fields

The force field of a system describes the potential energy associated with various molecular motions involved in the molecular dynamics analysis.

The intra and inter molecular interaction can be presented as the total potential energy which could be written as:

$$E_{total} = E_{vdw} + E_{elect} + E_{bond} + E_{angle} + E_{torsion} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (44)$$

Where E_{total} is the total energy, E_{vdw} is the Vanda Waals energy component, E_{elect} is the electrostatic energy, E_{bond} is the energy due to bonded interaction, E_{angle} is the angle bending energy component and $E_{torsion}$ is the torsion energy component. The Hamiltonian is defined as a classical system, specifying the instantaneous positions and momenta of all the particles constituting the system can specify the

microstate at any time t. Sum of potential and kinetic energy of the system in a classical molecular mechanics system [106]. It is dependent on the velocity (v) and the position (r) of the particles in the system. The momentum p is a function on the velocity v of the particle and the Hamiltonian which can be written as a function of momentum and position.

$$\mathrm{H}(q, p) = \frac{p^2}{2m} + v(r)_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (45)$$

The kinetic energy is defined as a function of the mass and momentum particles while the potential V defines the intermolecular interaction which is dependent on the positions of the particles. The interaction between particles or molecules is a function of the force field of the system. For this research we applied the OPLS force field. The motion and position of particles/atoms in a molecular system can be estimated using the force field and the initial configuration of the system. The total potential energy of the system is the sum of both the intermolecular interaction and intra-molecular interaction. The interaction of the atoms of same molecule is known as the intra-molecular interaction while the interaction of atoms of different molecules is known as intermolecular interaction. In the following section, we describe these two interactions in details.

2.5.1 Non-Bonded Interaction

This includes Lennard-Jones or Buckingham potential, coulomb or modified coulomb interaction. The non-bonded interaction is experienced between atoms in the same molecule and those in other molecules and are computed on the bases of neighbors.

In GROMACS they are pair additive and Centro symmetric.

$$V\left(r_1,....r_N\right)=\sum_{i<j}V_{ij}(r_{ij}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (46)$$

$$F_i=-\sum_j \frac{dV_{ij}(r_{ij})}{dr_{ij}}\frac{r_{ij}}{r_{ij}}=-F_j \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (47)$$

The non-bonded interaction comprises of the dispersion term, the coulomb term and the repulsion term. The dispersion and repulsion terms are combined in the Lennard-Jones or Buckingham potential with partially charged atoms on the coulomb term

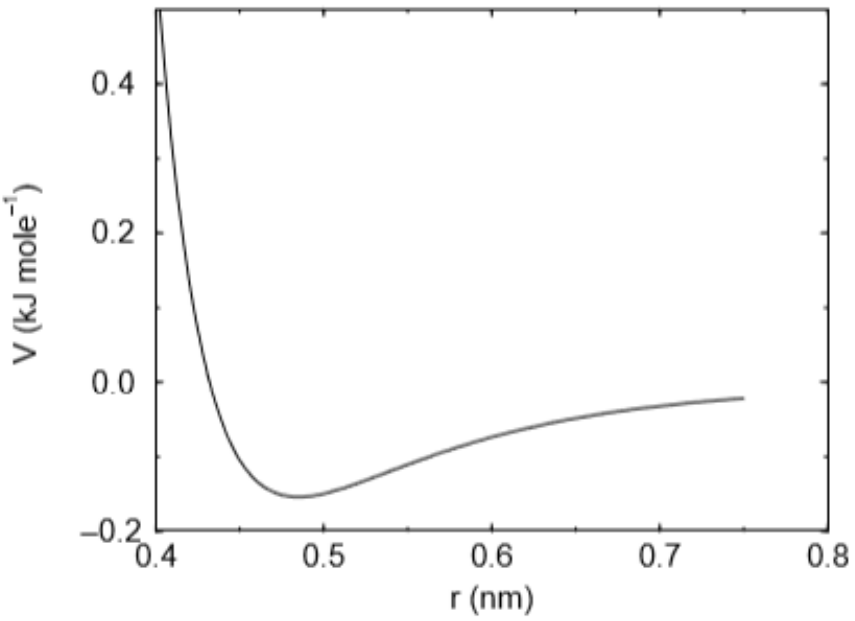

Figure 13. Lennard-Jones Interaction

The Lennard-Jones potential V_{LJ} between to atoms can be represented with the equation below

$$V_{LJ}(r_{ij}) = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}} \quad \text{..... (48)}$$

$$F_i\left(r_{ij}\right) = \left(12\frac{C_{ij}^{(12)}}{r_{ij}^{13}} - 6\frac{C_{ij}^{(6)}}{r_{ij}^{7}}\right)\frac{r_{ij}}{r_{ij}} \quad \text{..... (49)}$$

The Lennard-Jones potential can also be written as

$$V_{LJ}\left(r_{ij}\right) = 4\in_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) \quad \text{..... (50)}$$

2.5.2 Coulomb Interaction

For ionic materials or charged particles, the Coulomb interaction term is dominant and can be represented as

$$V_c\left(r_{ij}\right) = f\frac{q_i q_j}{\varepsilon_r r_{ij}} \quad \text{..... (51)}$$

Where $$f = \frac{1}{4\pi\varepsilon_0} = 138.935485 \quad \text{..... (52)}$$

The force derived from the above potential is given as

$$F_i\left(r_{ij}\right) = f\frac{q_i q_j}{\varepsilon_r r_{ij}^{\,2}}\frac{r_{ij}}{r_{ij}} \quad \text{..... (53)}$$

The interaction of the non- bonded molecules can be modeled using the Lennard Jones potential and the columbic potential[107] as described in the following equation

$$v_{nonbonded(i,j)} = v_{lernard(i,j)} + v_{coulomb(i,j)} \quad \text{..... (54)}$$

where *v* is the potential for non-bonded interaction between atoms of different molecules *i* and *j* with the total potential given as the sum of the coulomb potential and the Lennard Jones potential.

2.5.3 Bonded Interaction

Another form of interaction in a molecular dynamic system is bonded interaction. This interaction occurs between atoms of the same molecules with shared bonds.

This interaction could be in the form of stretching (2 body), torsion or angle bending (3 body) and dihedral angle(4-body) interaction. The bond and angle vibration are harmonic, bond vibrations have very high frequency which will require slow time step. This is taken care of by introducing a rigid constraint bond to replace the bond potential. This replacement will allow the use of bigger time steps [108]. The bond stretching and bond bending energy equation is based on Hooke's law [109].

2.5.4 Bond Stretching

The bond stretching between two covalently bonded atoms i and j can be represented by a harmonic potential

$$V_b\left(r_{ij}\right)=\frac{1}{2}k_{ij}^b\left(r_{ij}-b_{ij}\right)^2 \quad \text{..... (55)}$$

And the force can be represented as

$$F_i\left(r_{ij}\right)=k_{ij}^b\left(r_{ij}-b_{ij}\right)\frac{r_{ij}}{r_{ij}} \quad \text{..... (56)}$$

For bond angle bending interaction harmonic potential can be used

$$E_{bend}\left(\theta\right)=\frac{1}{2}k_0\left(\theta-\theta_0\right)^2 \quad \text{..... (57)}$$

where k_0 is the angle bending force constant, θ is the actual bond angle and θ_0 is the reference bond angle. The energy needed to distort an angle away from equilibrium is much lower than that needed to distort a bond, therefore bond angle bending force constants is proportionally smaller than those for bond stretching.

2.6 Modeling

Human DNA has become one of the key areas of study in science today. its regulatory activities with potential focus on G Quadruplex as a potential bio switch during gene regulation [86]. Computational modeling could help in identifying potential structural changes that happen during regulations at cell level. Modeling as a research

tool could help identify high probability target with a reduced level of error.
The current wet lab activity does not provide full detail of the interaction that happen around G- Quadruplex and its counter ions during gene regulation and expression. Today's computational power available in the field of modeling has made it possible to test, analyze and visualize the activities between these molecules at the cell level in real time. This computing ability has the potential to reduce the amount of time material and error experienced most time in wet lab activity. This could be a tool to fundamental understanding of the physiological activities within the regulatory regions of the human DNA.

Computational modeling could provide a visual understanding of the interactions that happens around G Quadruplex and its counter ions during gene regulation and expression. Computational modeling could help identify binding site of certain molecules involved in gene regulation.

2.6.1 Advantages of Molecular Simulation

Molecular dynamics is powerful technique used to study the macroscopic properties of many body problems at the atomic or molecular level using classical mechanics theories and Newton laws. This concept was introduced 1950's by Alder and wainwright to aid the understanding of interactions between hard spheres [110]. The concept of molecular dynamic simulation was introduced to help solve and understand the problems below.

- To understand the properties, structure of molecule and understand the interaction between molecules.
- To generate the trajectory of molecules to understand the dynamics of molecules in a given time frame.
- To give qualitative understanding physiological changes in molecules due to interactions at the molecular level.
- To help compare experimental results between simulations and wet lab experiments.
- To help reduce guess work to the minimum in research
- To serve as a bridge between the experimental analysis and theory [111].
- To create dynamic time integration method in computing intermolecular motion and forces between molecules
- To help solve the Newton's equation at the atomic/molecular level

Molecular Dynamics can be used to predict the future position and velocity of atoms in a dynamic system if the initial position, force and time interval is known [110]. In recent years the advancement in computing power and software has made molecular dynamic simulations an important tool in research involving biomolecules. This advancement has enabled the investigation of bigger systems with more realistic boundary condition with better sampling time [112]. The ab initio molecular dynamics was aimed at reducing the amount of guess work and fitting to the minimum.

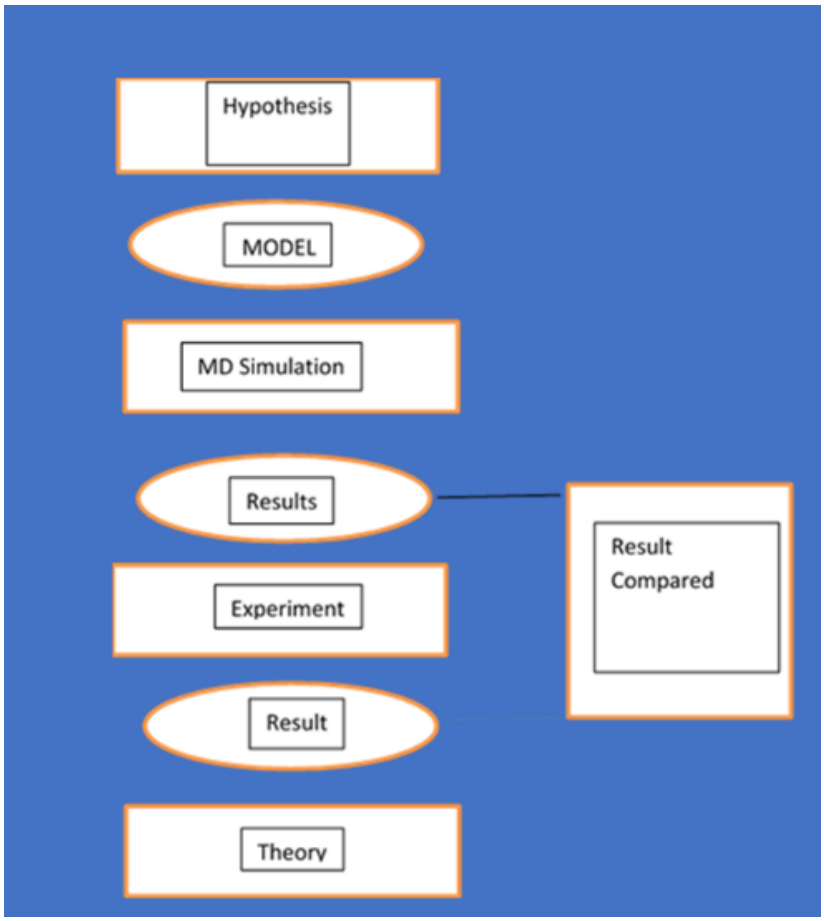

Figure 14. Modelling Flow Chart

For this research we will be using a tool known as GROMACS MD

2.7 Groningen Machine for Chemical Simulation (GROMACS)

This is an open source molecular dynamics package designed for the simulation of proteins, nucleic acids and lipids. The origin of GROMACS can be traced back to the work done by Berendsen-Van Gunsteren group at the University of Groningen, Netherlands in the late 1980s. The core developers of GROMACS, Erik Lindahl, Berk Hess and David van der Spoel adopted the idea of combined parallel computing with parallel simulation algorithms [113-115]. The initial package was written in C language,

but there were performance issues in the early stage of GROMACS which has led to the programming to evolve over time. In mid-1990s, some of the performance gaps were addressed by rewriting the inner loops of the code were in Fortran language [113, 116, 117]. The code currently operates on various operating platforms including UNIX and LINUX operating systems and could run on a multiple CPU core. GROMACS is a multipurpose package optimized for molecular simulation with capacity to run at high performance in parallel computing and work station. GROMACS was designed to extract virial and periodic boundary conditions from pairwise and loop interaction. GROMACS has a simple and consistent user interface void of scripting language. Programs running on GROMACS are controlled by command line arguments in a UNIX environment. GROMACS is extensively compatible with a range of algorithms from molecular dynamics with features that supports the generalized reaction field (GRF) and liquid systems such as Reaction field potentials [118, 119]. GROMACS has the potential to apply analytic corrections to energy or both energy and pressure for dispersion interactions beyond the cut-off points [120]. GROMACS minimizes Energy using either conjugate gradients or steepest descent algorithms and can be used to determine the potential of a mean force acting on a molecule [116], this allows the simulation of atomic force microscope experiments, averaged distance of ensemble restraints, model Nuclear magnetic resonance(NMR) and X-RAY effects on molecules [121, 122]. GROMACS undergoes constant development and it is one of the most efficient, friendly MD software packages with the potential to convert protein data bank (pdb) structure files to GROMACS compatible structure files understood by GROMACS MD code.

GROMACS has no force field of its own, but it is compatible with generalized force fields such as OPLS, GROMOS, AMBER and ENCAD [123].Next we will discuss molecular dynamics as pertains to this research.

CHAPTER III

MOLECULAR DYNAMIC SIMULATION OF 4FXM G-QUADRUPLEX IN DIFFERENT SALT SOLUTIONS (POTASSIUM CHLORIDE, SODIUM CHLORIDE AND LITHIUM CHLORIDE)

3.1 Introduction

G-quadruplex has been studied both experimentally and computationally but not much has been known about its potential role as a molecular switch at the telomeric region of human DNA, its potential regulatory role in gene expression and the effect of ion concentration in its regulatory role. Our intention is to understand the differential effect of alkali metal ions on the structure and stability of DNA g-quadruplexes with potential role in the regulation of gene expression. The result obtained will help in validating initial experimental work on Ionic Modulation of QPX Stability as a Nano Switch Regulating Gene Expression in Neurons which was carried out in the lab earlier. We begin studying the effect of ion concentration on G-Quadruplex, we will also replace each cation at the center of the structures with these different ions (Na^+, K^+ and Li^+) in a sodium, potassium and lithium chloride environment respectively. For this research, we will analyze various thermodynamic quantities in addition to G-quadruplex structural parameters such as radius of gyration and root mean square deviation (RMSD)for molecular systems. This study helps us verify the consistency of our present simulation method by comparing with the results from the existing literature and the earlier result from the lab and the computational result.

It also serves as a reference to compare the G-quadruplex structure and its dynamics under different ion concentration.

As stated earlier, we will use GROMACSMD code for this purpose. In the following section below, we will describe the simulation details of G-Quadruplex under different ion concentration in water.

3.1.1 Simulation Details

Simulation of G-quadruplex DNA performed using GROMACS MD code with various G-Quadruplex structures is obtained from protein databank. In this chapter we will be looking at the pdb code '4fxm' which is the crystal structures of G-Quadruplex obtained from the protein data bank. The pdb structures of G-Quadruplex DNA were modified by removing attached drug ligand and converting to a GROMACS compatible structure. A simulation directory was created for minimization file minim.mdp, the equilibration files nvt.mdp, npt.mdp, md.mdp and the pdb file. The counter ion in the G-Quadruplex was replaced with three different ions namely Na^+, K^+ and Li^+. The counter ions where chosen because of its abundance and biological role in the body. Lithium was chosen because of its possible role therapeutic in depression. TIP3P water was used in solvating the system [124], while computational visualization tool VMD [125] was used to visualize progress in simulation and visually inspect the G-Quadruplex DNA structure. All computations were performed on an in-house computation cluster, at the Joint School of Nano Science and Nano engineering Greensboro North Carolina. The following steps were employed in the simulation to arrive at the final structures for simulation.

- We modified the structure be removing the drug ligand and placing the desired counterion at the center of the structure
- We converted the G-quadruplex structure obtained from the protein data bank to a GROMACS compatible structure
- By using the right configuration and force field, we generated the system that was viewed using the VMD software.
- We generated output files for the minimization which will help keep the system stable at the minimum energy state. By using conjugate gradient method, the system was minimized.
- The G-Quadruplex structure was solvated with TIP3P water molecules.
- The solvated G-Quadruplex structure was equilibrated using NVT and NPT ensembles to ensure the stability of the system.
- The NPT and NVT equilibration were carried out for 200ps with a 1fs time step at a temperature of 310K and the pressure of 1 bar. The final configuration from NPT equilibration was used to initiate the MD.
- Production was run for half microseconds(500ns). All the simulation details are presented in the Appendix A

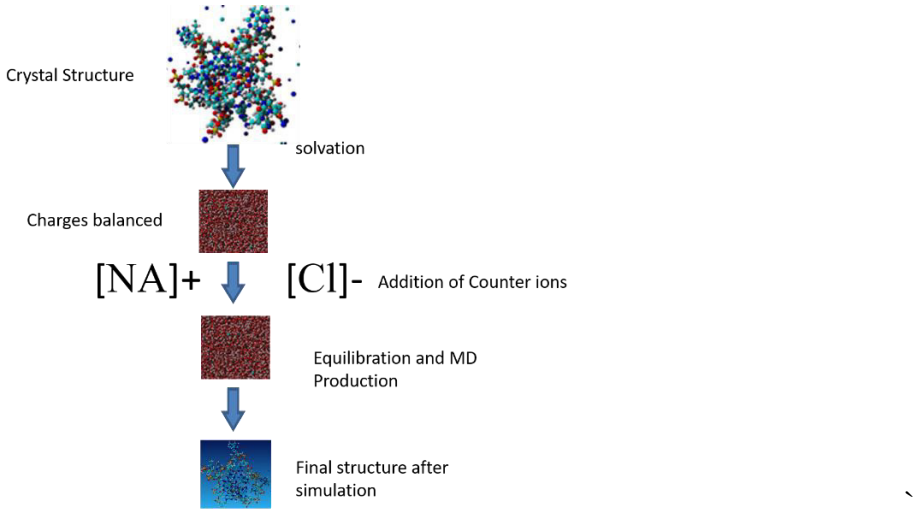

Figure 15. Illustration of Simulation Steps from the Crystal Structure to the Final Structure

We will run our first simulation on a crystal structure with code 4fxm from RCSB website. 4fxm is a complex of a human telomeric repeat G-quadruplex DNA and N-methyl mesoporphyrin(p21212). We modified the structure by removing the attached drug leaving only the G-Quadruplex structure. The simulation was run for half a microsecond(500ns) with other simulation details in Appendix A.

Table 2. Description of 1xav G-quadruplex DNA Molecule

Organism	G-Quadruplex DNA	Ion	Length	Resolution (Angstrom)
Homo sapiens	(5'D (*AP*GP*GP*GP*TP*TP*AP*GP*GP *GP* TP*TP*AP*GP*GP*GP*TP*TP*AP*G P*GP*G)-3')	K^+	22	1.651

Further in this chapter we will evaluate some energy parameter and the analysis of the md production

3.2 System Run, Stability and Conformational Changes

During the production dynamics simulation of solvated 4fxm, we recorded the trajectory by collecting data at every 100ps for post-processing.

Figure 20 displays the initial structure and final structure of 4fxm G-quadruplex in 50,100,150 and 160 millimolar of different ion concentration. (The structure above is a structure without the water molecules). The generated data includes basic thermodynamic quantities such as total energy, density, temperature, RMSD, RDF, of the system. We describe these quantities along with the quantities used in analyzing the simulation data. Next, we will look at energy and temperature terms.

3.2.1 Energy and Temperature

The stability of the system was analyzed in relations to the energy of the system and the temperature of the system. The simulation trajectories generated were used in evaluating the energy of the system which includes total energy, potential energy and density. The result of the analysis provided information about the stability of the simulation system. Figure 17 shows the energy of the system remained constant indicating the thermostat is working as expected. The potential energy stability also indicates the systems equilibrium. The total energy which is the sum of kinetic and potential energy was also evaluated and it showed energy was conserved throughout the simulation period. The temperature of the system remained steady at 310 K also indicating the correct working of the thermostat as can be seen in figure 19.

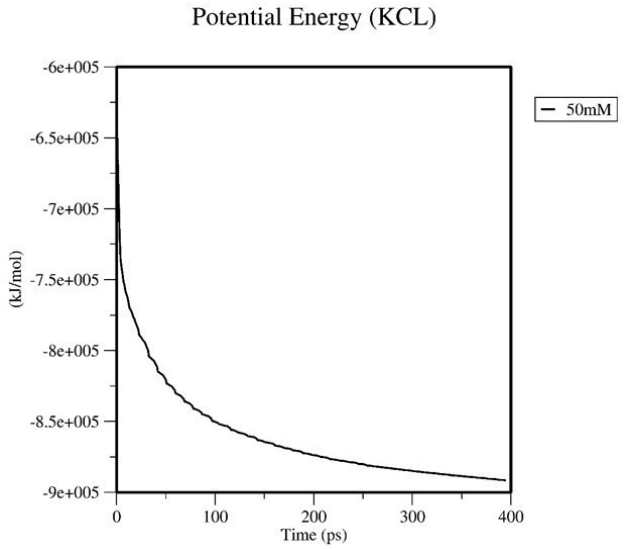

Figure 16. Potential of the System 50 mM KCl

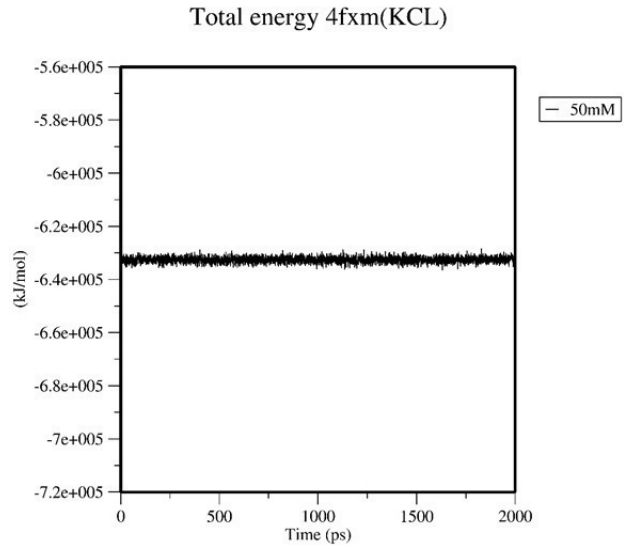

Figure 17. Total Energy of the System 50mM KCl

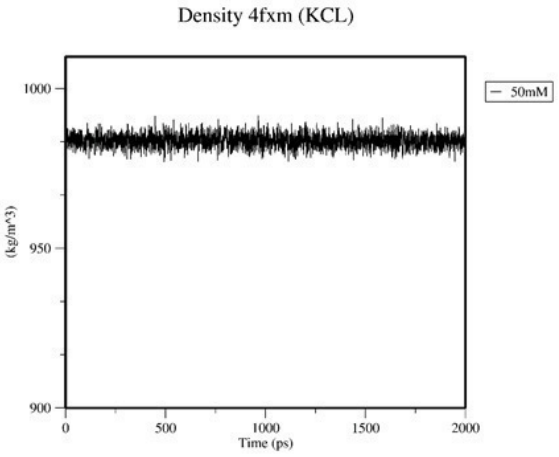

Figure 18. Density of Atoms in System for 50 mM KCl

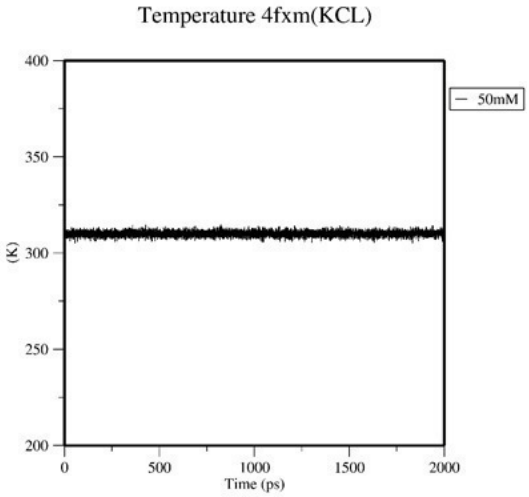

Figure 19. Temperature of the System at 310 K

3.2.2 Structural Change and Conformity

At the end of the simulation we analyzed and compared the thermodynamic quantities. We examined the G-quardruplex DNA structure visually with VMD for any apparent visible structural changes in the in the DNA structure before and after the simulation. The initial and final structures are shown in (figure 21-32) for 4fxm G-quadruplex DNA in different salt solutions and concentration as stated earlier.

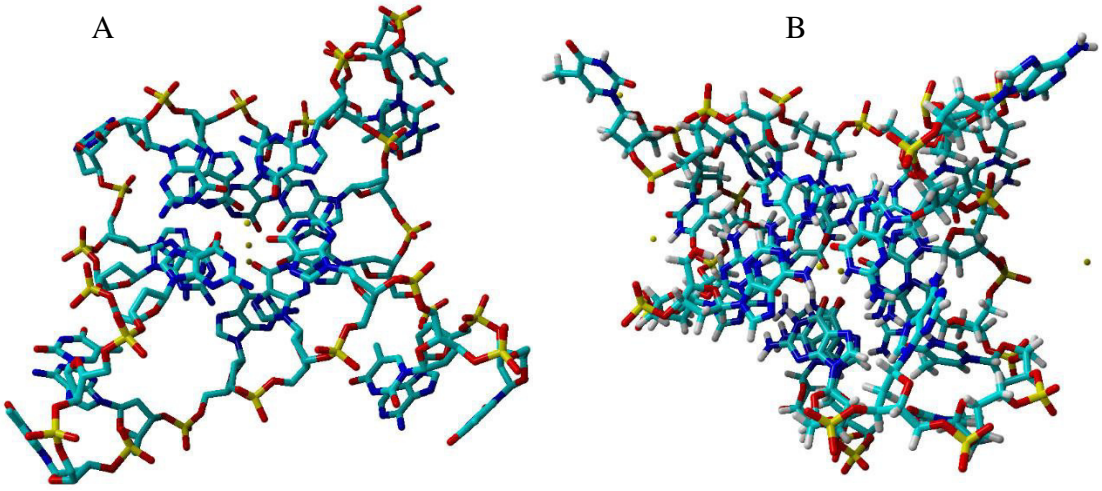

Figure 20. (A) Initial Structure 4fxm G-quadruplex (B) Final Structure of 4fxm G-quadruplex after Simulation in 50 mM KCl.

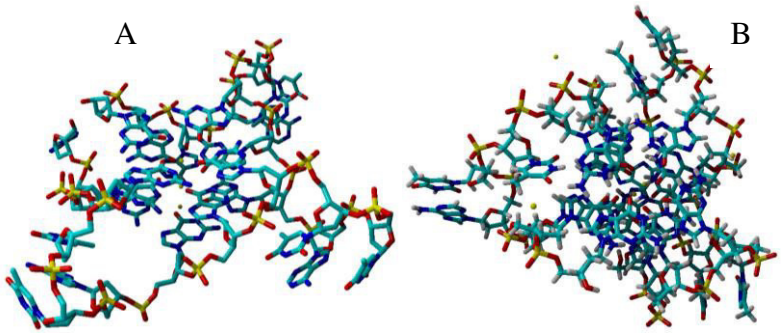

Figure 21. (A) Initial Structure 4fxm G-quadruplex (B) Final structure of 4fxm G-quadruplex after Simulation in 100 mM KCl

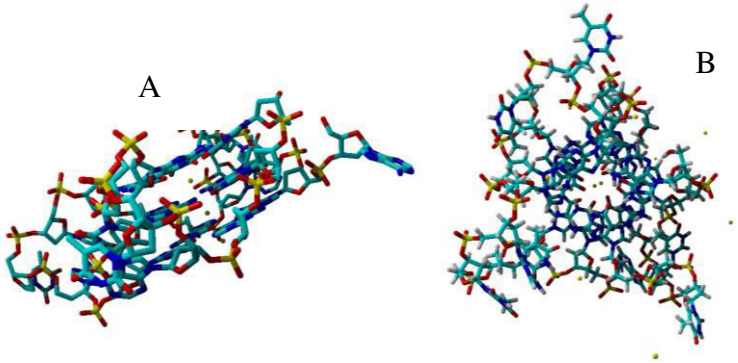

Figure 22. (A) Initial structure 4fxm G-quadruplex (B) final structure of 4fxm G-quadruplex after simulation in 150 mM KCl

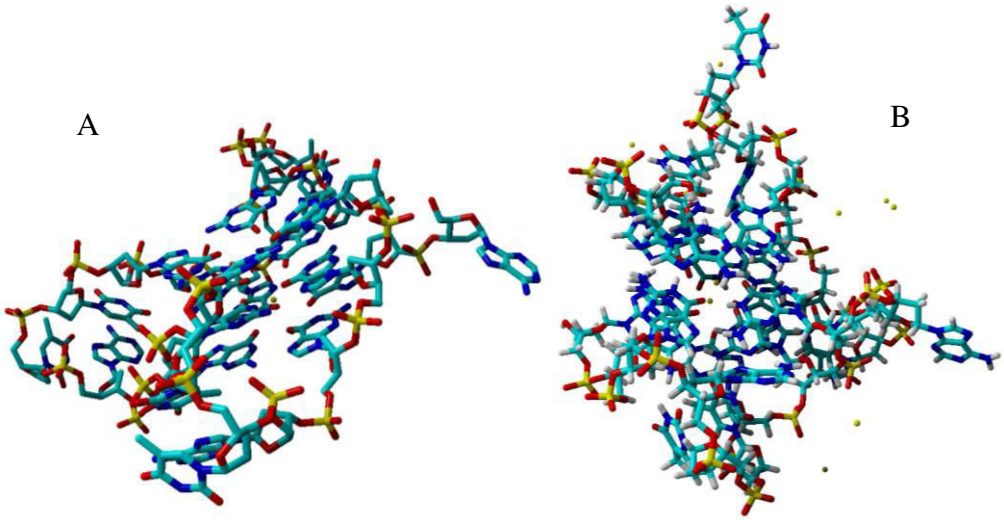

Figure 23. (A) Initial Structure 4fxm G-quadruplex (B) Final structure of 4fxm G-quadruplex after Simulation in 160 mM KCl

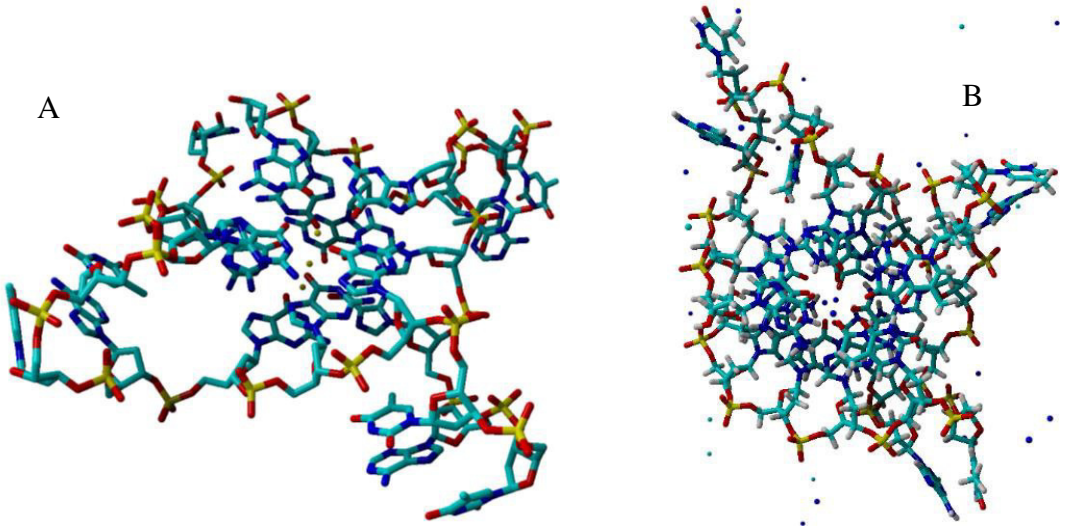

Figure 24. (A) Initial Structure 4fxm G-quadruplex (B) Final structure of 4fxm G-quadruplex after Simulation in 50 mM NaCl

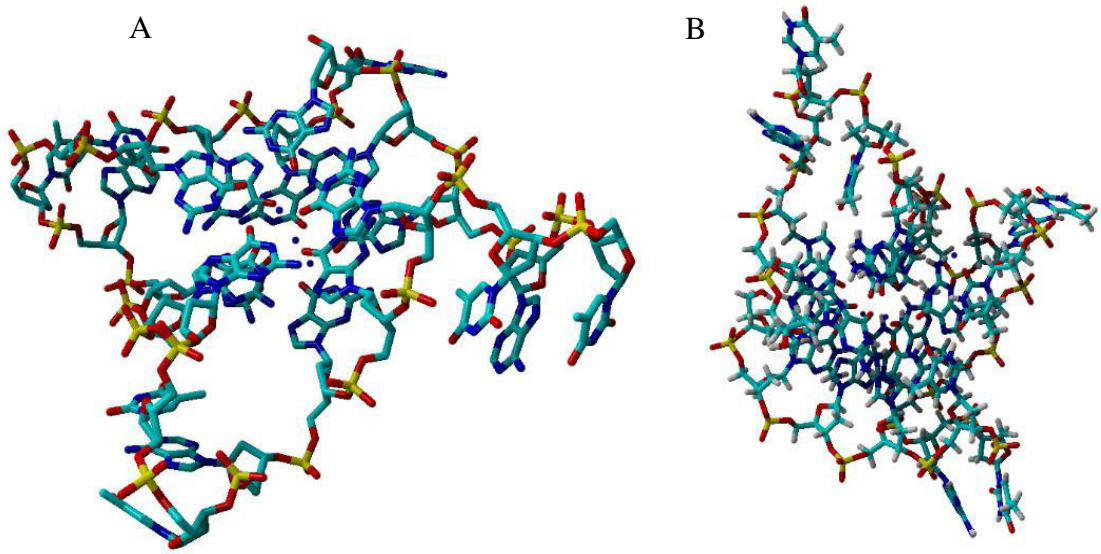

Figure 25. (A) Initial Structure 4fxm G-quadruplex (B) Final Structure of 4fxm G-quadruplex after Simulation in 100 mM NaCl

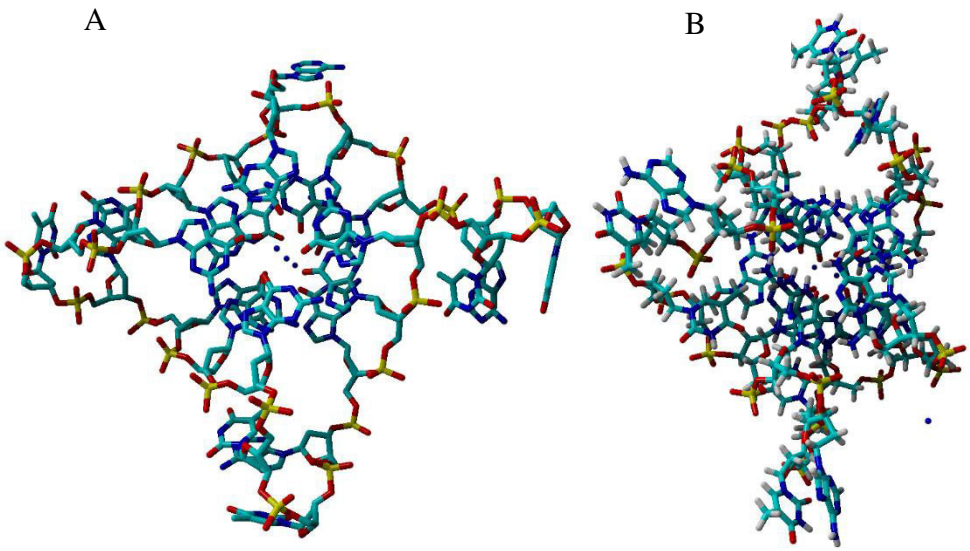

Figure 26. (A) Initial Structure 4fxm G-quadruplex (B) Final structure of 4fxm G-quadruplex after Simulation in 150 mM NaCl

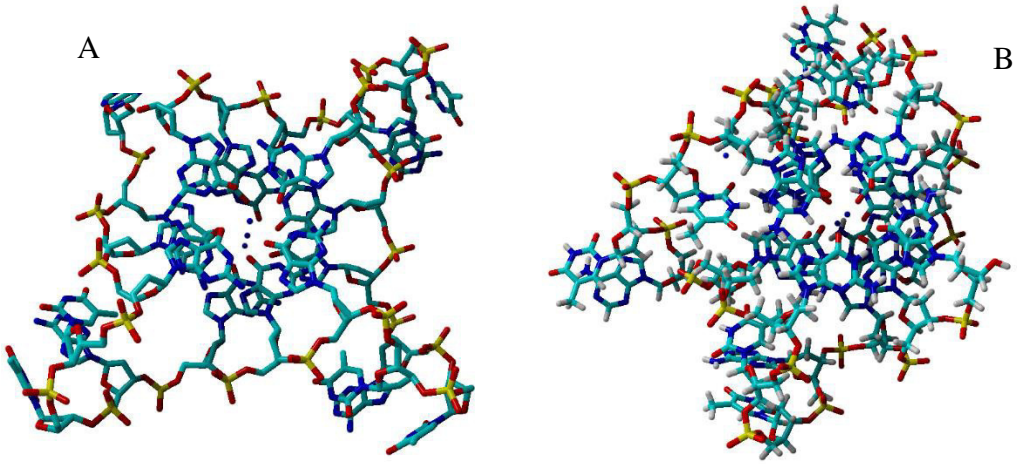

Figure 27. (A) Initial Structure 4fxm G-quadruplex (B) Final Structure of 4fxm G-quadruplex after Simulation in 160 mM NaCl

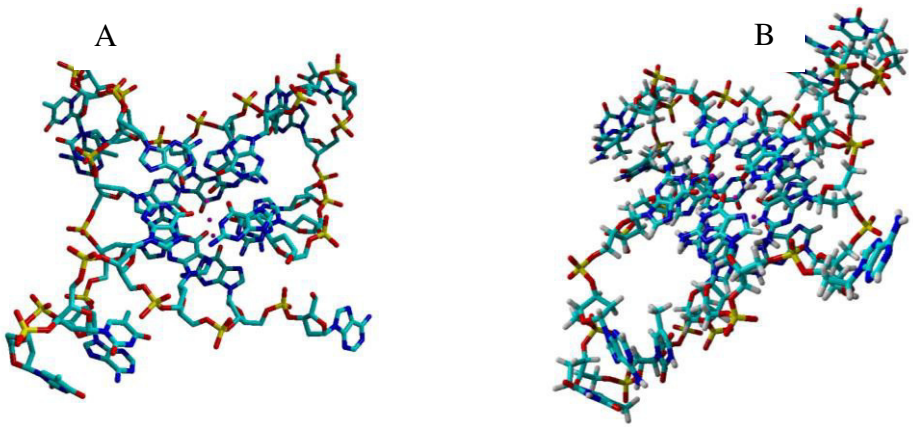

Figure 28. (A) Initial Structure 4fxm G-quadruplex (B) Final Structure of 4fxm G-quadruplex after Simulation in 50 mM LiCl

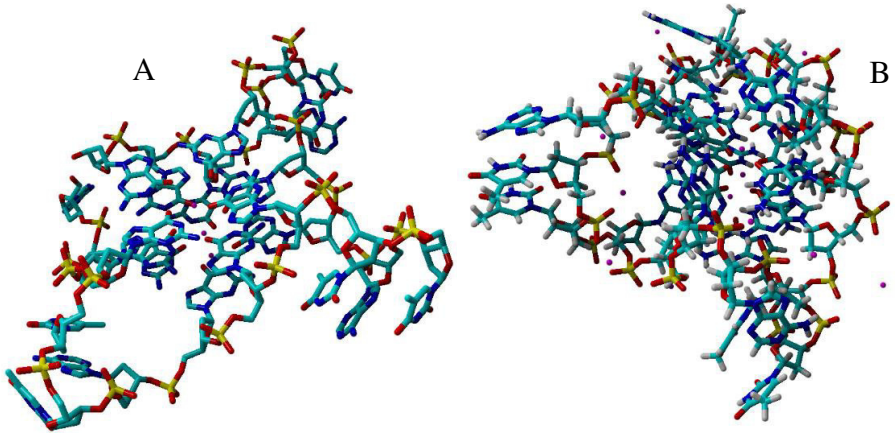

Figure 29. (A) Initial Structure 4fxm G-quadruplex (B) Final Structure of 4fxm G-quadruplex after Simulation in 100 mM LiCl

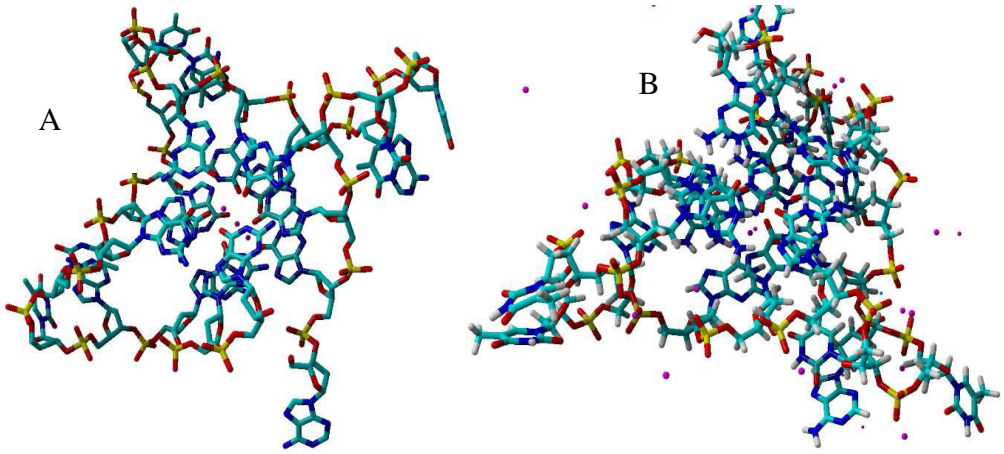

Figure 30. (A) Initial Structure 4fxm G-quadruplex (B) Final Structure of 4fxm G-quadruplex after Simulation in 150 mM LiCl

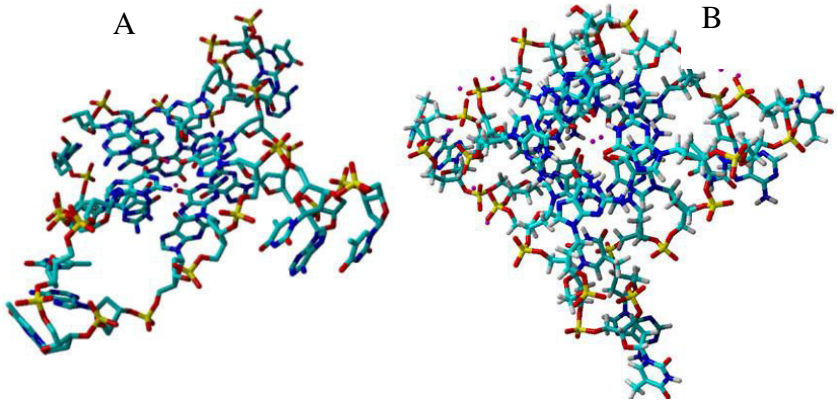

Figure 31. (A) Initial Structure 4fxm G-quadruplex (B) Final Structure of 4fxm G-quadruplex after Simulation in 160 mM LiCl

In analyzing the structural changes observed after simulation visually, the structure of 4fxm changed with different concentration and salt solution. We observed that potassium and sodium have great destabilizing effect on the 4fxm G-quadruplex with structural change. From figures 19-26, the third counter ion seems to be displaced with as concentration increases. Looking at figures 27-30 we observed structural change with lithium counter ion in lithium chloride solution without displacement of the counterion. The lithium counter ion remained intact even with the G-quadruplex structure destabilized. To further understand the observations mentioned above we will analyze the structure using another thermodynamic quantity known as radius of gyration

3.3 Radius of Gyration

This measure helps in understanding the compactness of the G-Quadruplex DNA structure. For example, how much the structure has spread or contracted with regards to its degree of freedom. A relatively compact structure or molecule will maintain a level of stability in its radius of gyration R g. To understand this with respect to our molecule, we plotted the graph of radius of gyration of our molecule against time first with all the concentration potassium to understand the effect of potassium concentration on our structure. The same analysis was done for sodium and lithium. To further understand the effect of this ions, we plotted and compared the three different ions with the same concentration as shown in figures (33-39) below.

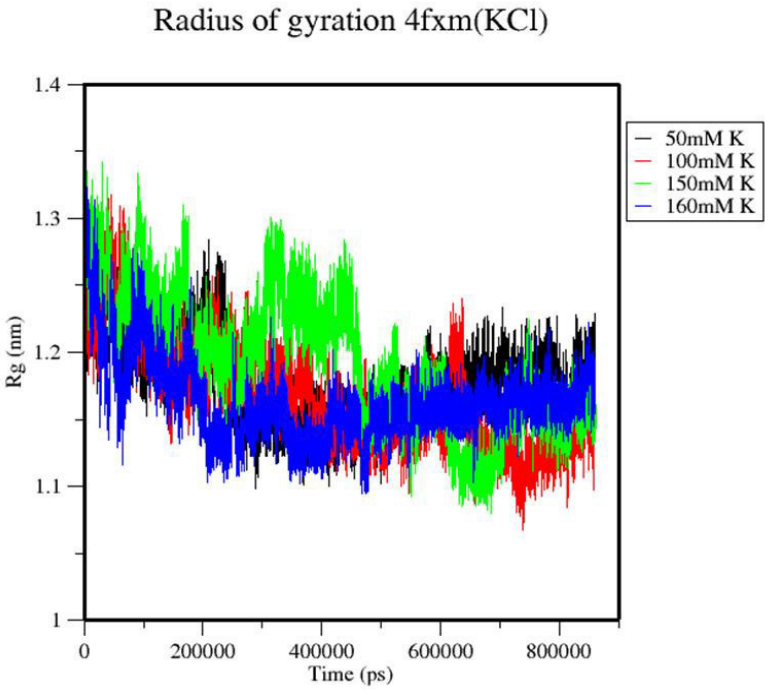

Figure 32. Radius of Gyration of 4fxm in 50,100,150 and 160 mM of 4fxm in KCL

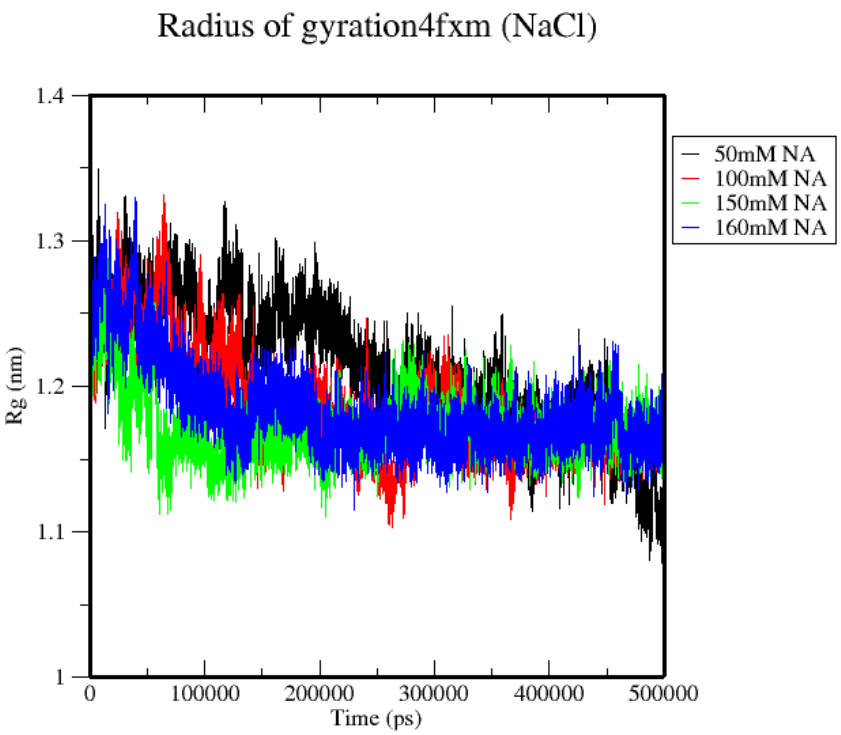

Figure 1. Radius of Gyration of 4fxm in 50,100,150 and 160 mM of NaCl

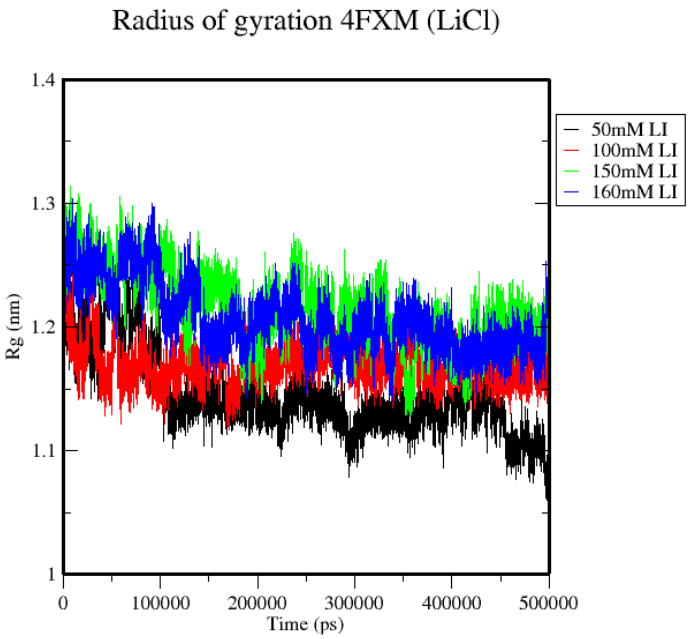

Figure 34. Radius of Gyration of 4fxm in 50,100,150 and 160 mM of LiCl

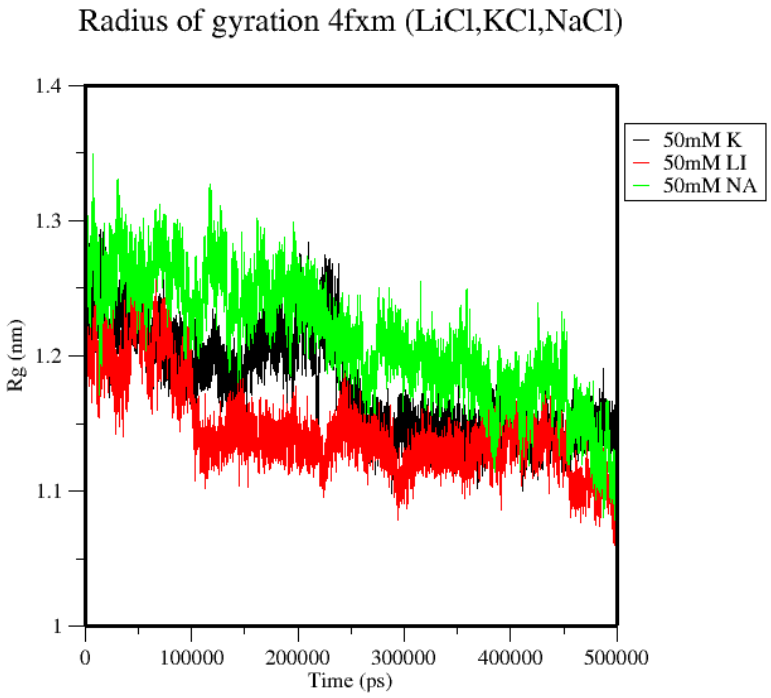

Figure 35. Radius of Gyration of 4fxm in 50 mM of KCl, NaCl and LiCl

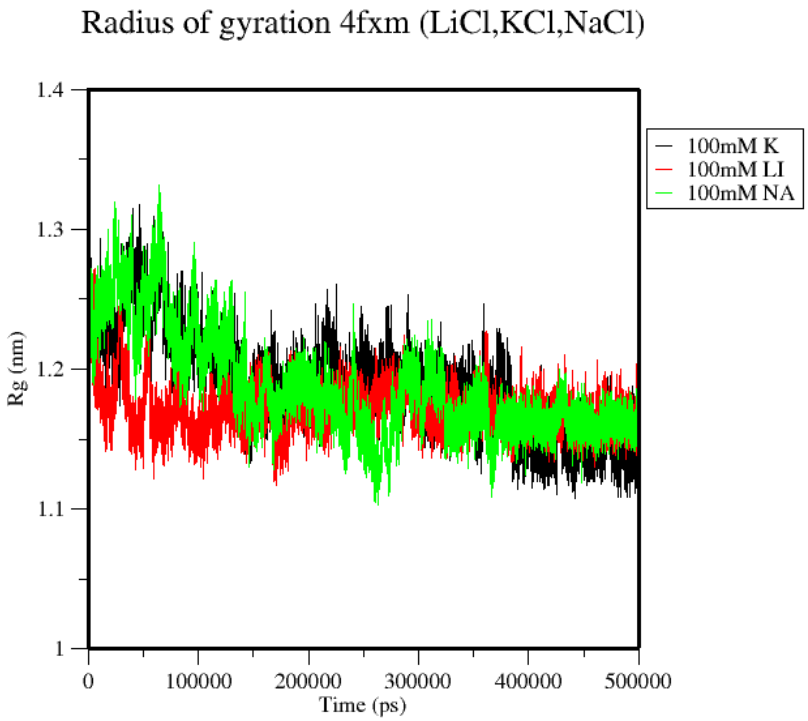

Figure 36. Radius of Gyration of 4fxm in 100 mM of KCl, NaCl and LiCl

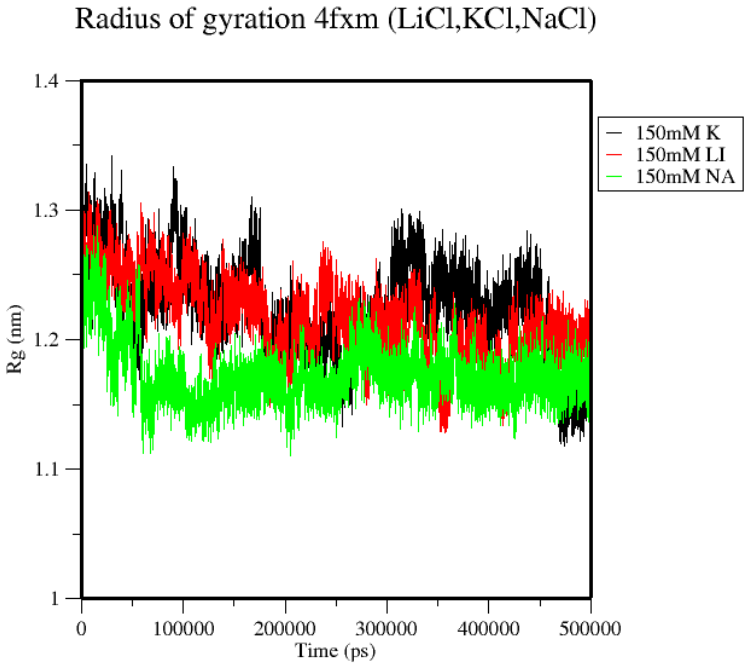

Figure 37. Radius of Gyration of 4fxm in 150 mM of KCl, NaCl and LiCl

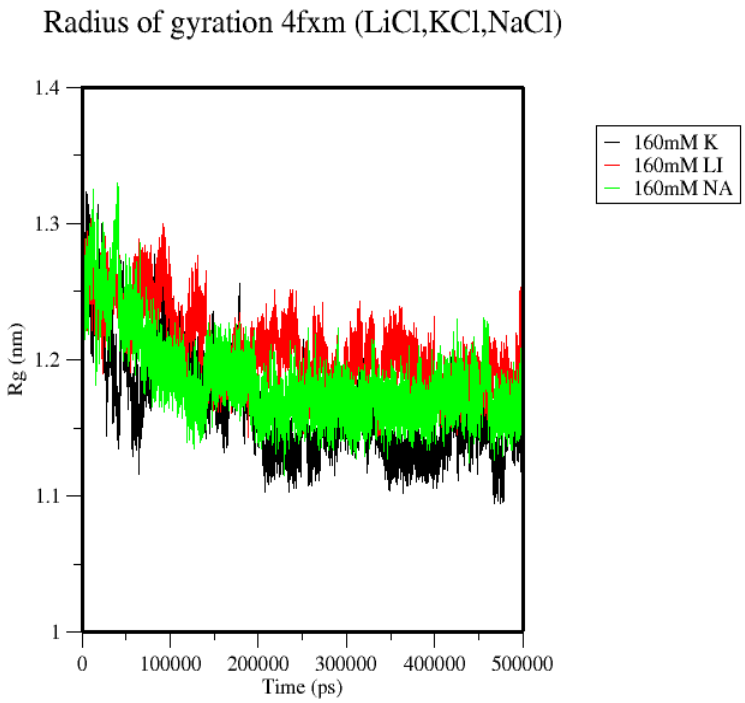

Figure 38. Radius of Gyration of 4fxm in 160 mM of KCl, NaCl and LiCl

From (figures 33-35) it seen that at lower salt concentration of 150 mM concentration potassium had the highest radius of gyration of about 1.25(nm) compared to 50 and100 mM which were stable at an average of 1.15 to 1.2 nm. 160 mM concentration indicates a lower radius of gyration of about 1.13 which may suggest a stabilization mechanism at higher concentration above 150. Figure 34 indicates sodium at 50 mM concentration had the highest radius of gyration of about 1.15nm compared 100,150 1nd 160 mM having an average of 1.2nm indicating the systems compactness was better maintained at concentration above 50 mM. Lithium ion in figure35 had the opposite at lower concentration of 50 mM it maintained the compactness at an average

1.13nm and increased to about 1.23 at 160 mM concentration. We also analyzed the radius of gyration of the three ions compared under different concentration. At a lower concentration of about 50 mM and 100 mM sodium had the highest Rg compared to potassium and lithium but at 160 mM lithium seem to have higher Rg compared to sodium and potassium indicating the compactness of the system may be affected at higher concentration of lithium. To further verify the structural analysis above we measured the rmsd of the structure.

3.3.1 Root Mean Square Deviation (RMSD)

We further evaluated the stability of the structure by calculating the root mean square deviation which is a measure of deviation of the atoms of final structure form that of the original structure. This was done by plotting the root mean square deviation against time for all the concentration for the respective counter ions and the compared concentration between the three counter ions.

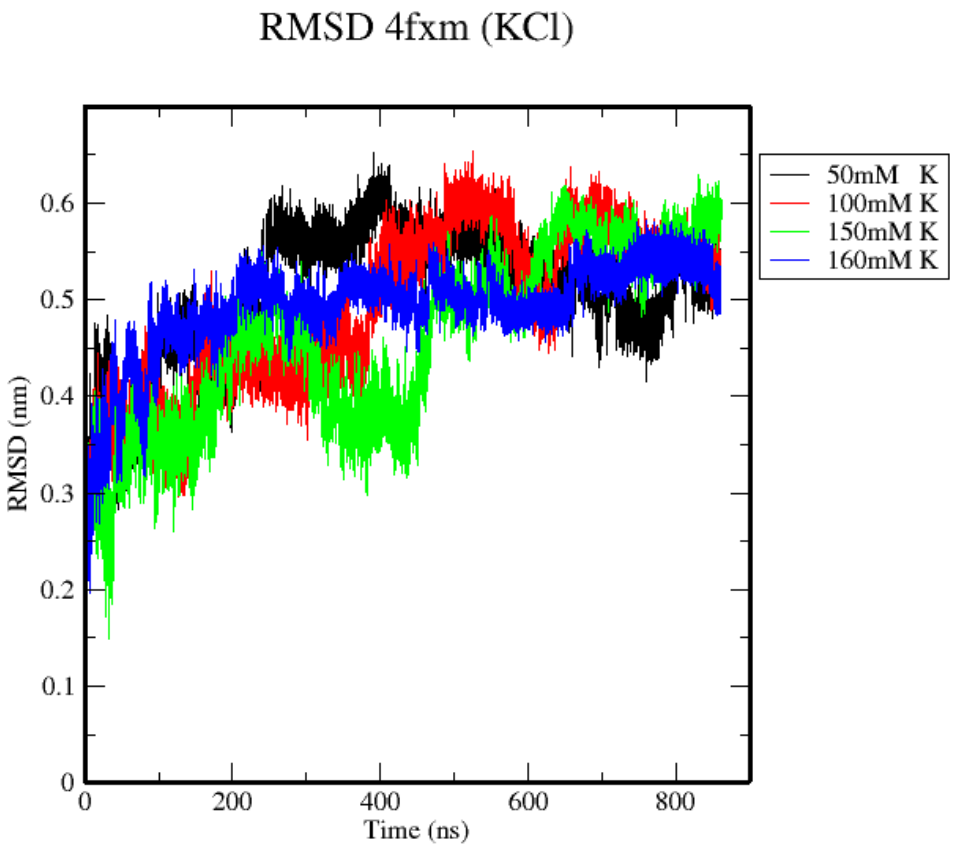

Figure 39. RMSD of 4fxm in 50,100,150 and 160 mM of KCl

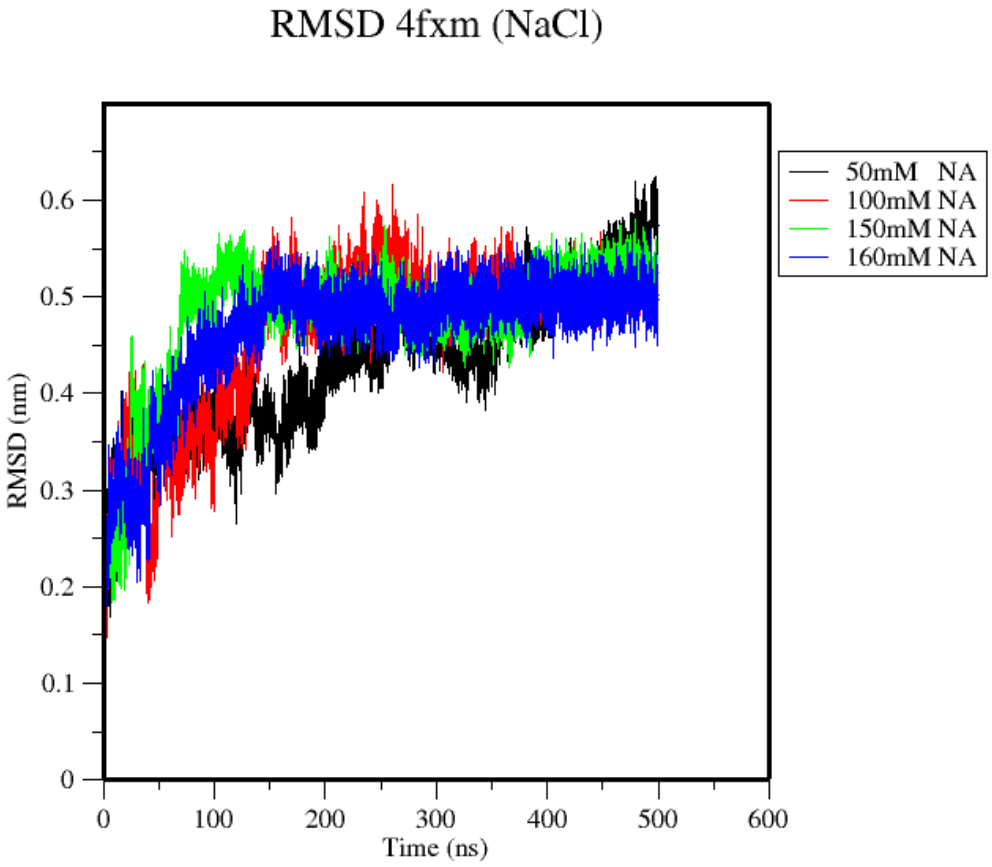

Figure 40. RMSD of 4fxm in 50,100,150 and 160 mM of NaCl

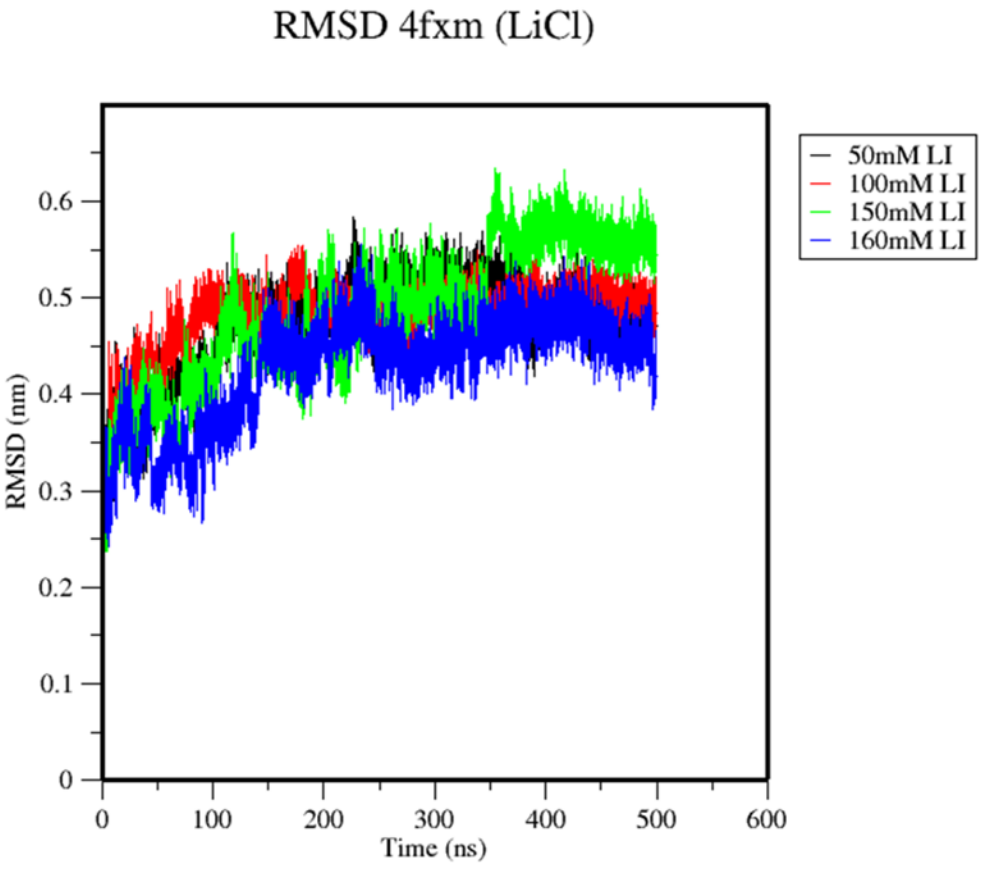

Figure 41. RMSD of 4fxm in 50,100,150 and 160 mM of LiCl

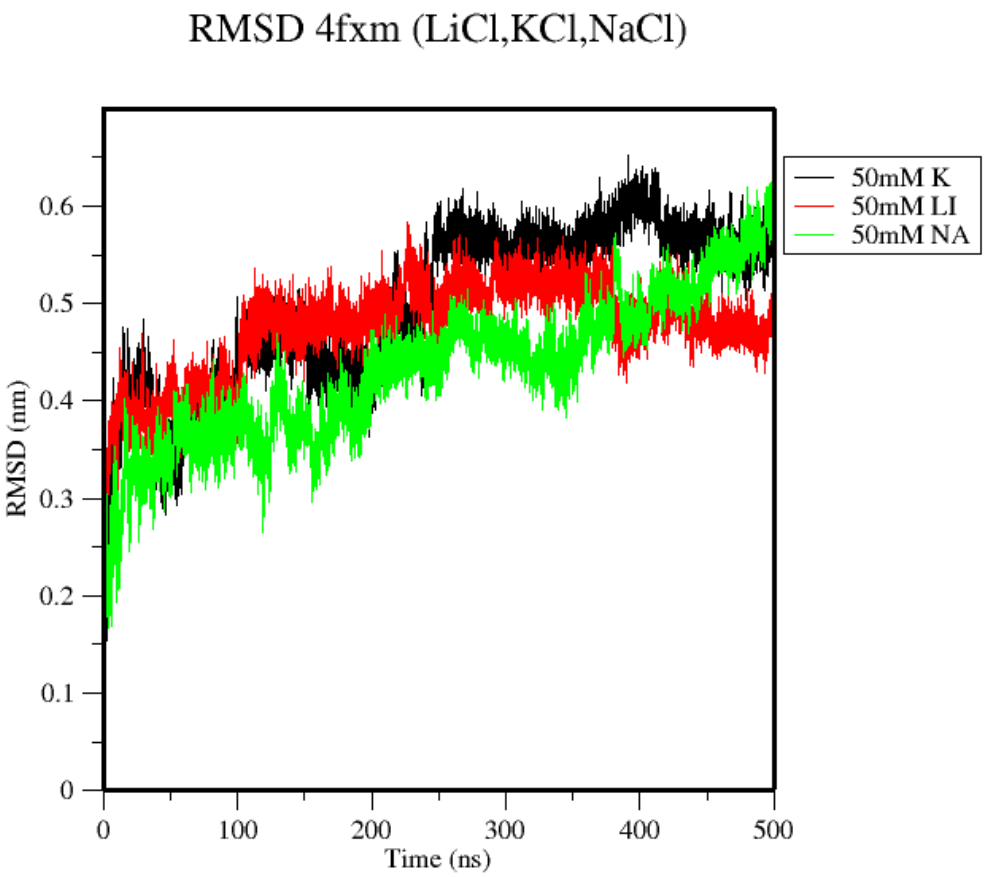

Figure 42. RMSD of 4fxm in 50 mM of KCl, NaCl and LiCl

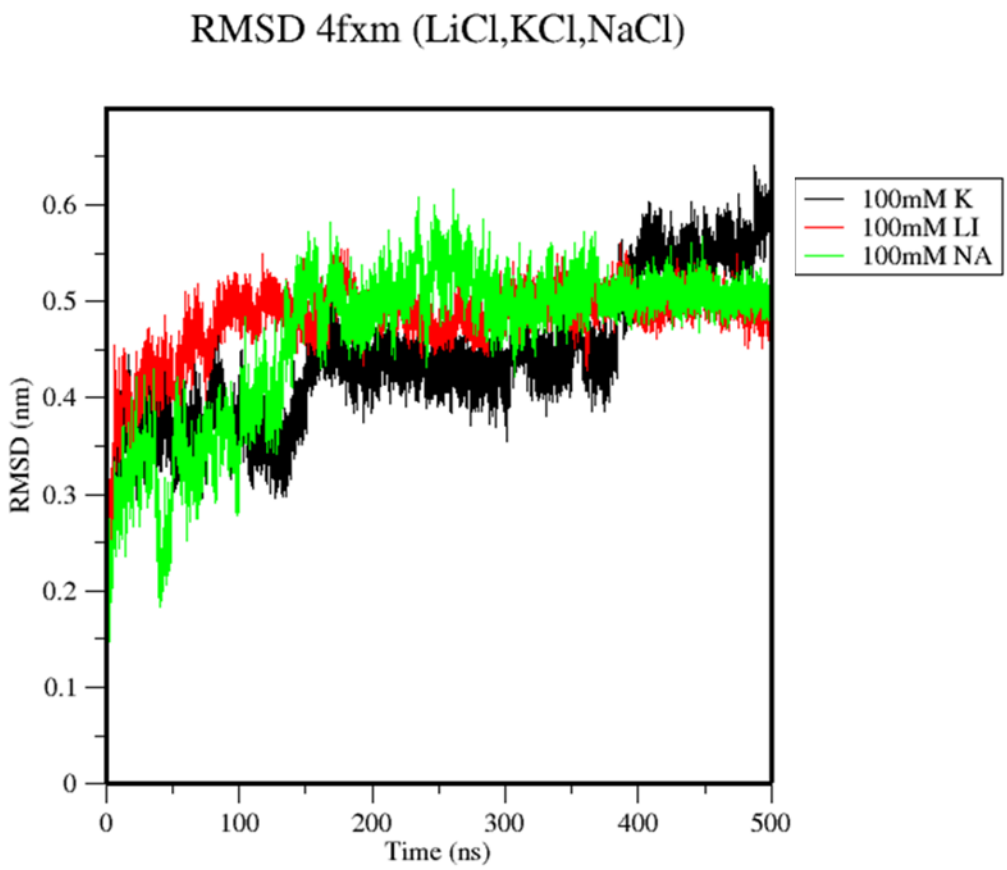

Figure 43. RMSD of 4fxm in 100 mM of KCl, NaCl and LiCl

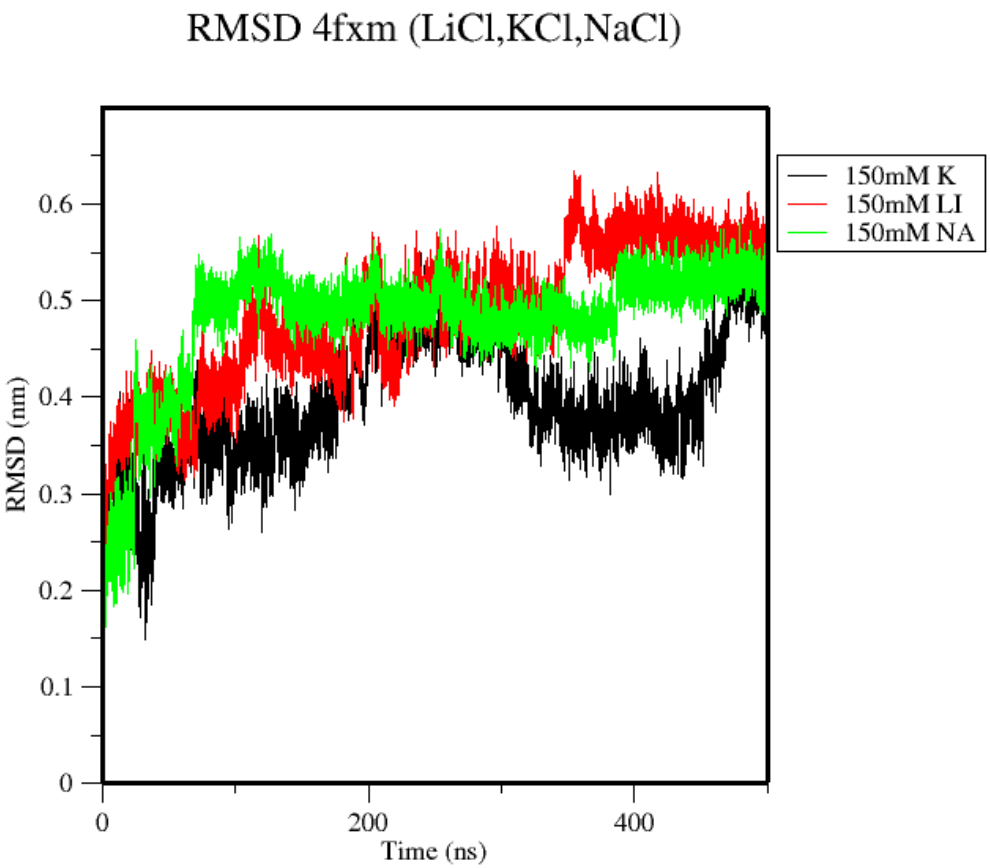

Figure 44. RMSD of 4fxm in 150 mM of KCl, NaCl and LiCl

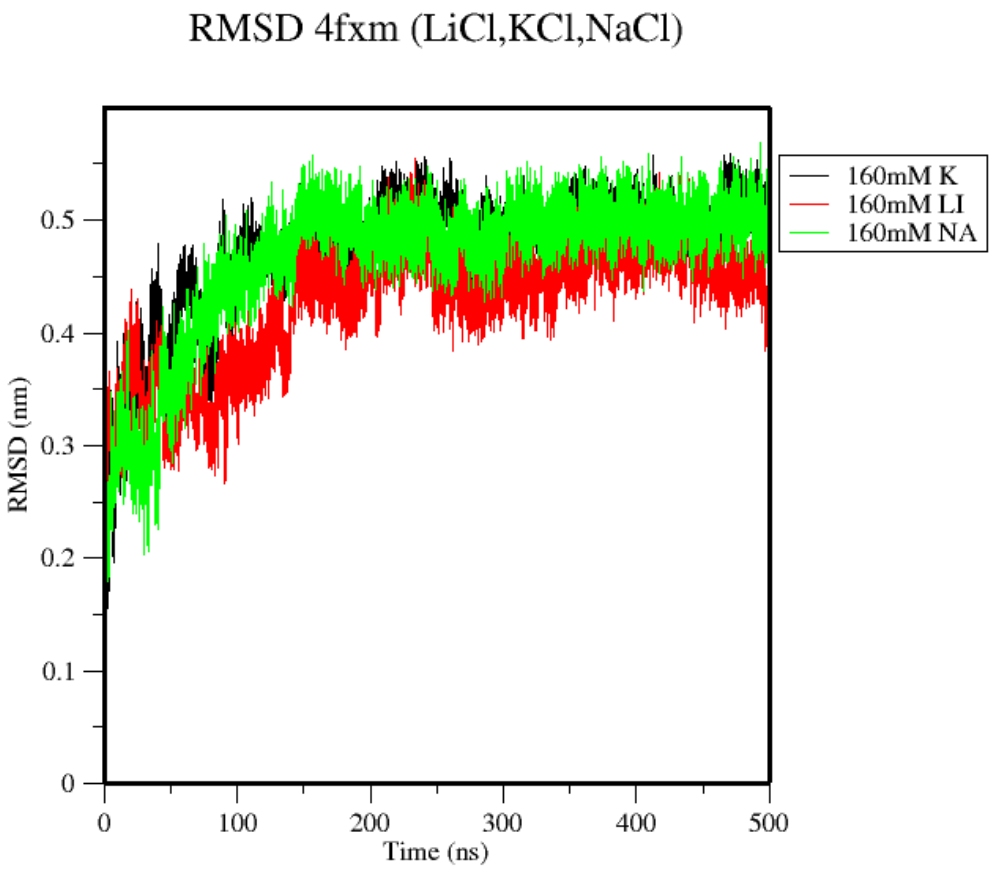

Figure 45. RMSD of 4fxm in 160 mM of KCl, NaCl and LiCl

Comparing the rmsd of potassium chloride solution in figure39, 150 mM concentration of potassium chloride seem to have the lowest initial raise of about 0.45 nm compared to 50, 100 and 150 mM of the salt solution having average rmsd of 0.55 to 0.60 nm. This may suggest the favorability of structure at higher concentration of KCl solution. Figures 40 show the rmsd of our structure in different concentration of sodium chloride solution. 150 mM concentration of our solution showed the highest rmsd initial raise of about 0.55 nm before convergence compared to other concentration which were

within the average of 0.5nm. Figure 41 shows the rmsd of our structure in lithium solution with average initial raise of 0.42 to 0.45 nm before convergence. Figure 42 shows the graphs of the rmsd for the three counter ions compared. The result suggests indicating on the average sodium had higher destabilizing effect on the structure compared to potassium and lithium ion. To further investigate this, we plotted the graph of rmsd against concentration to see the effect on the structural change.

3.3.2 Structural Change

We further compared the average structure of the trajectory after calculating the RMSD of the system and compared it with the initial structure used in the simulation. The conformational change where seen in the structure as shown in table 3-5 and figure 46 below.

Table 3. RMSD of 4fxm for Different Concentration of Potassium Chloride Solution

Concentration (mM)	RMSD (nm)
50	0.450465
100	0.645096
150	0.484542
160	0.535709

Table 4. RMSD of 4fxm for Different Concentration of Sodium Chloride Solution

Concentration (mM)	RMSD (nm)
50	0.571999
100	0.502591
150	0.526222
160	0.499074

Table 5. RMSD of 4fxm for Different Concentration of Lithium Chloride Solution

Concentration (mM)	RMSD (nm)
50	0.467933
100	0.483872
150	0.547501
160	0.419753

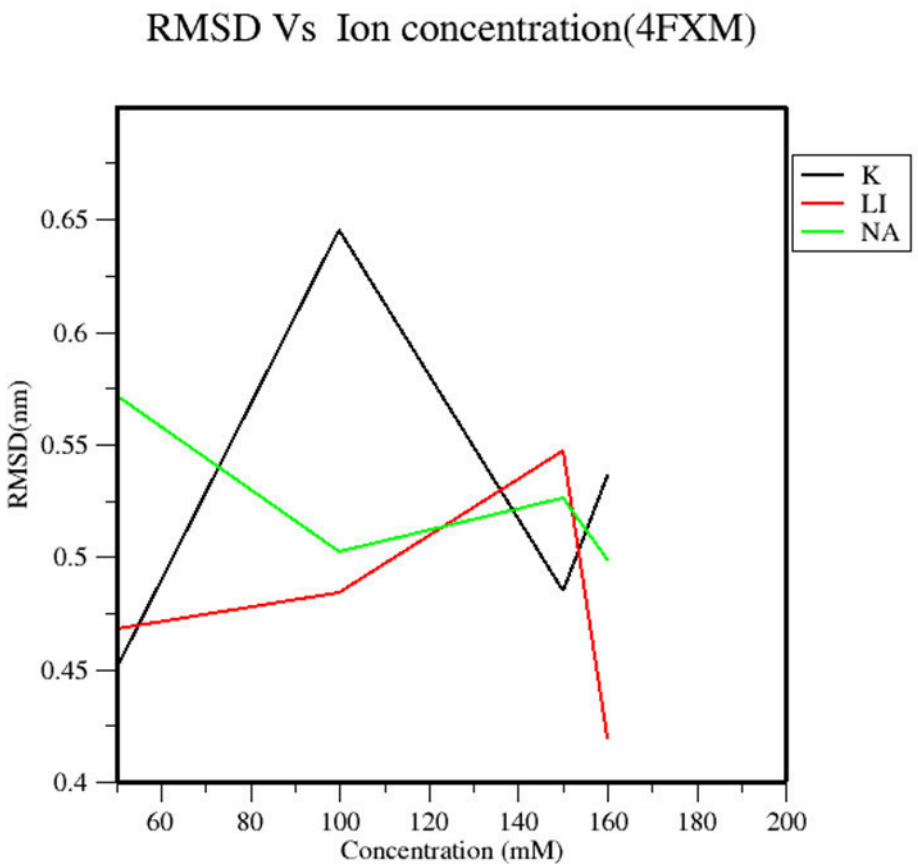

Figure 46. RMSD against Concentration of Ions for 4fxm in 50 mM,100 mM,150 mM and160 mM of KCl, NaCl and LiCl

Figure 46 above shows the graph of RMSD against concentration which shows sodium ion to have the highest structural change of about 0.57nm followed by lithium which was 0.47 nm and potassium which was 0.45nm.

3.4 Summary

In summary, this chapter presented a detailed study of 4fxm G-quadruplex DNA structure and dynamics in different ion concentration employing GROMACS molecular dynamics simulations. MD analysis studies involved the solvation of the system by using appropriate number of TIP3P water molecules. The solvated structure was equilibrated in NVT and NPT ensembles for 200ps in each case. The final equilibrated structure of G-quadruplex DNA was used as the starting structure for the production, dynamics simulation. The dynamics analysis was conducted for 500ns with a 1fs time step at 1atm pressure and 310K temperature with the simulation data collected at every 100ps for post-production analysis.

Average values for both the energy terms and other thermodynamic analysis were evaluated and confirmed the validity of our simulation setup. The average structural change for 4fxm in solution of potassium, sodium and lithium ion concentration were 0.57, 0.47 and 0.45 nm respectively most structural deviation was observed in sodium ion solution followed by potassium and lithium ion. Sodium also maintained the highest Rg in most cases followed by potassium and lithium. The next chapters discuss the effect of ion concentration on 1xav G-quadruplex DNA structure in different concentration of potassium, sodium and lithium ion.

CHAPTER IV

MOLECULAR DYNAMIC SIMULATION OF 1XAV G-QUADRUPLEX IN DIFFERENT SALT SOLUTIONS (POTASSIUM CHLORIDE, SODIUM CHLORIDE AND LITHIUM CHLORIDE)

4.1 Introduction

In the previous chapter we examined the effect of ion concentration on 4fxm G-quadruplex. We also carried out some computational analysis on the simulation data generated. To further understand the effect of ion concentration on G-Quadruplex we carried out the same simulation process and steps on a different molecule known as 1xav. 1xav RCSB code for a G-quadruplex structure formed in human c-MYC oncogene in the promoter region of human DNA. It is a monomeric parallel stranded DNA that has been studied both experimentally and computationally but not much has been known about its potential role as a molecular switch at the telomeric region of human DNA, its potential regulatory role in gene expression and the effect of ion concentration in its regulatory role. Our intention is to understand the differential effect of alkali metal ions on the structure and stability of 1xav g-quadruplexes. For this set up we will replace the cation at the of the structure with Na^+, K^+ and Li^+ ions respectively. Next, we will run simulation of different concentrations of 50, 100, 150 and 160 millimolar of NaCl, KCl and LiCl. We will analyze various thermodynamic quantities in addition to G-quadruplex. Structural parameters such as radius of gyration and root mean square deviation (RMSD) and root mean square fluctuation.

4.1.1 Simulation Details

Simulation of 1xav G-quadruplex DNA performed using GROMACS MD code with various. The pdb structures of 1xav G-Quadruplex DNA was downloaded from RCSB website. A simulation directory was created for minimization file minim.mdp, the equilibration files nvt.mdp, npt.mdp, md.mdp and the pdb file. The counter ion in the G-Quadruplex was replaced with three different ions namely Na^+, K^+ and Li^+.TIP3P water was used in solvating the system [121], while computational visualization tool VMD [122] was used to visualize progress in simulation and visually inspect the G-Quadruplex DNA structure. All other simulations parameters remain the same as in the previous chapter. Additional simulation detail can be seen on Appendix B

Table 6. Description of 1xav G-quadruplex DNA Molecule

Organism	G-Quadruplex DNA	Ion	Length
Human	5'-D(*TP*GP*AP*GP*GP*GP*TP*GP*GP*GP* TP*AP*GP*GP*GP*TP*GP*GP*GP*TP*AP*A)-3'	K^+	22

4.1.2 System Run, Stability and Conformational Changes

During the production of dynamics simulation of solvated 1xav, we recorded the trajectory by collecting data at every 100ps for post-processing. (Figures 47-57) displays the initial structure and final structure of 1xav G-quadruplex in 50,100,150 and 160 millimolar of different ion concentration. (The structure above is a structure without the water molecules). The generated data was used to compute basic thermodynamic quantities such as RMSD, RDF, RG and RMSF of the system. Next, we analyze the results of simulation starting from visual analysis of the initial and final structures of 1xav G-quadruplex for different ions and concentration.

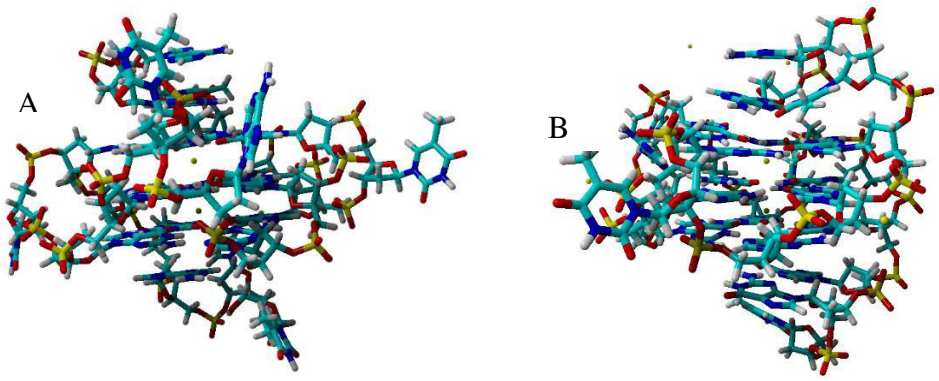

Figure 47. Initial Structure 1xav G-quadruplex (B) Final Structure of 1xav G-quadruplex after Simulation in 50 mM KCl

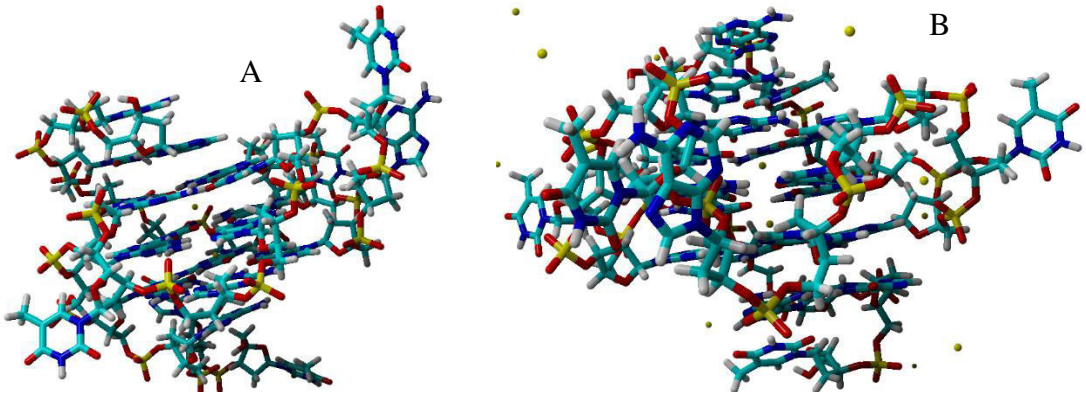

Figure 48. Initial Structure 1xav G-quadruplex (B) Final Structure of 1xav G-quadruplex after Simulation in 100 mM KCl

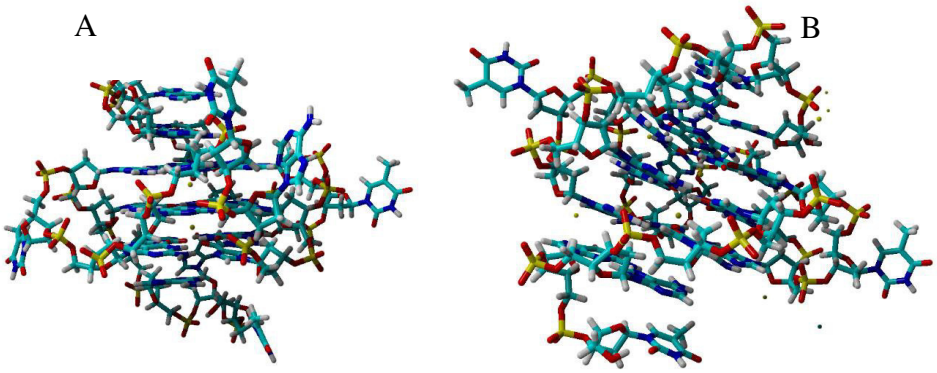

Figure 49. Initial Structure 1xav G-quadruplex (B) Final Structure of 1xav G-quadruplex after Simulation in 150 mM KCl

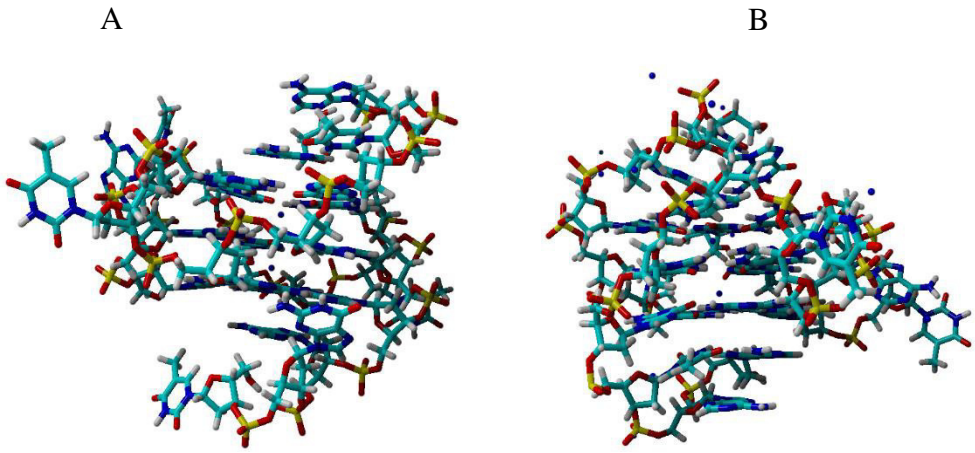

Figure 50. Initial Structure 1xav G-quadruplex (B) Final Structure of 1xav G-quadruplex after Simulation in 50 mM NaCl

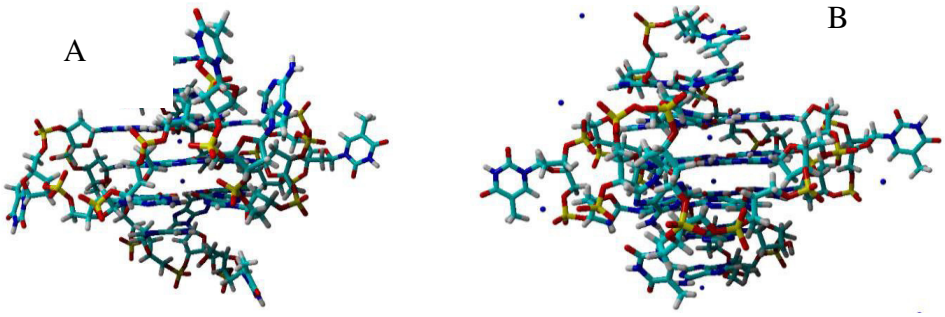

Figure 51. Initial Structure 1xav G-quadruplex (B) Final Structure of 1xav G-quadruplex after Simulation in 100 mM NaCl

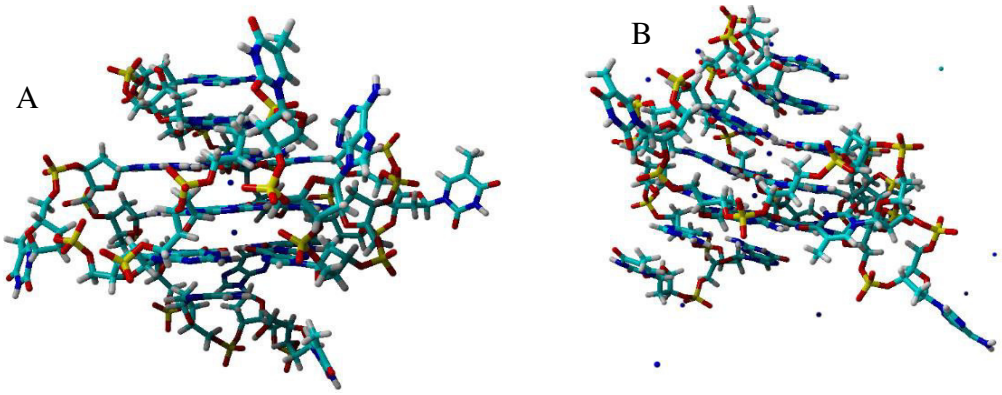

Figure 52. Initial Structure 1xav G-quadruplex (B) Final Structure of 1xav G-quadruplex after Simulation in 150 mM NaCl

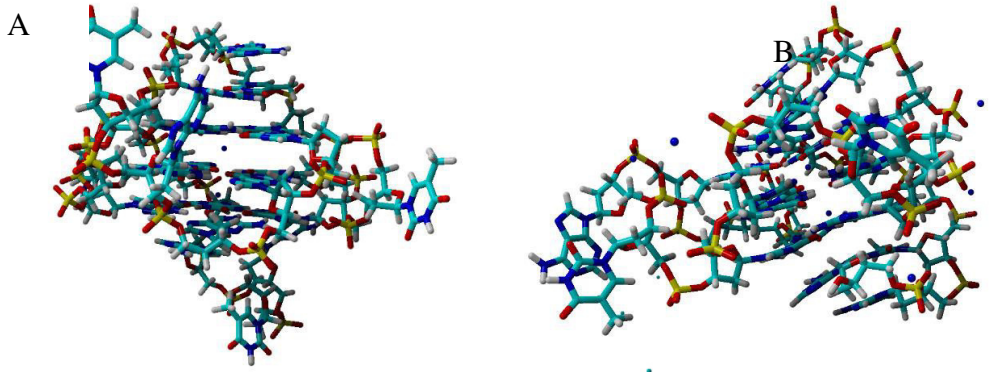

Figure 53. Initial Structure 1xav G-quadruplex (B) Final Structure of 1xav G-quadruplex after Simulation in 160 mM NaCl

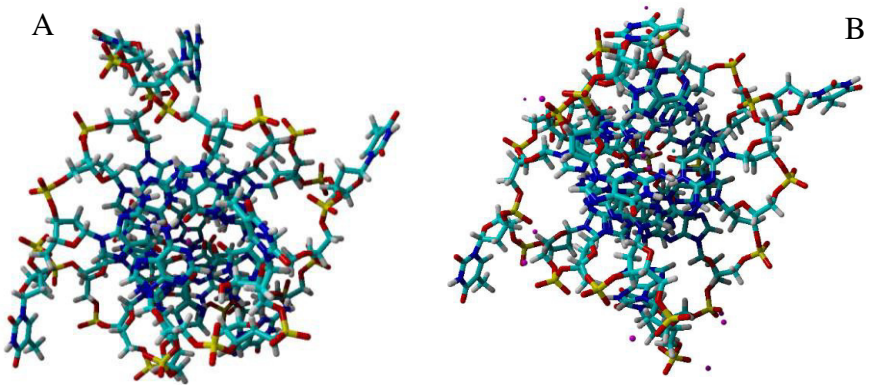

Figure 54. Initial Structure 1xav G-quadruplex (B) Final Structure of 1xav G-quadruplex after Simulation in 50 mM LiCl

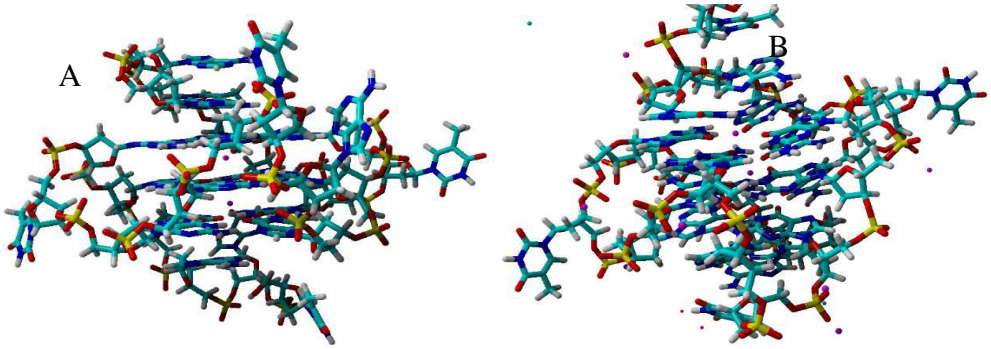

Figure 55. Initial Structure 1xav G-quadruplex (B) Final Structure of 1xav G-quadruplex after Simulation in 100 mM LiCl

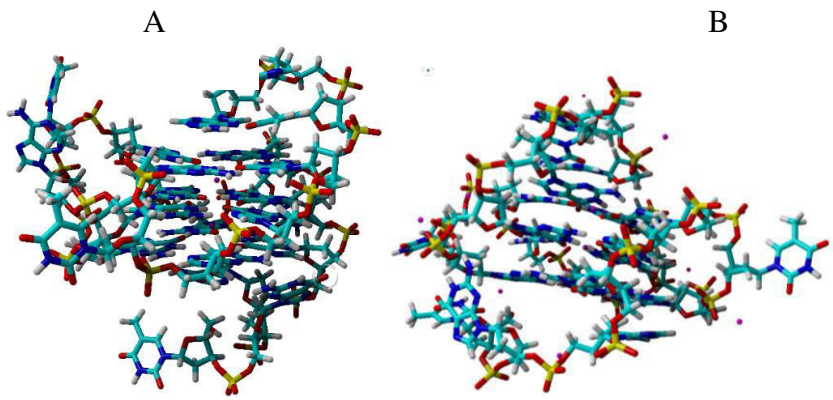

Figure 56. Initial Structure 1xav G-quadruplex (B) Final Structure of 1xav G-quadruplex after Simulation in 150 mM LiCl

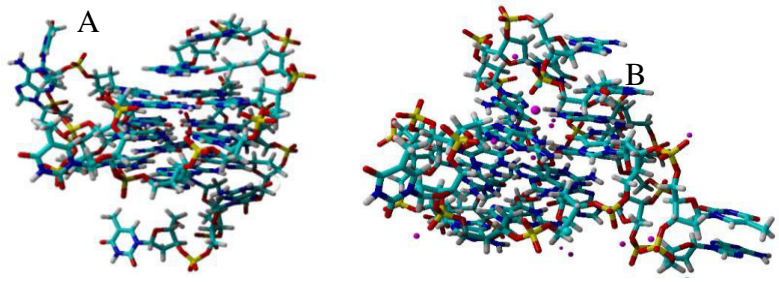

Figure 57. Initial Structure 1xav G-quadruplex (B) Final Structure of 1xav G-quadruplex after Simulation in 160 mM LiCl

4.1.3 Summary

As shown in (figures 47-57), we observed that 1xav G-quadruplex DNA structure showed significant changes in the size of the G-quadruplex molecule. The final structure appears destabilized compared to the initial structure. A closer examination reveals significant difference in the 4fxm G-quadruplex molecule and 1xav G-quadruplex molecule as the cation on the center of 1xav seem to remain intact irrespective of the structural changes. This observation compared to 4fxm shows there could be potential

different response or interaction between the cations and different molecules. We also observed that more structural change seems to happen around the molecule in sodium ion concentration compared to potassium and lithium ion concentration.

To further confirm our findings, we calculated and compared thermodynamic quantities such as radius of gyration (Rg), root mean square deviation (RMSD) and root mean square fluctuation (RMSF).

4.2 Radius of Gyration of 1xav G-Quadruplex in Different Ion Concentration

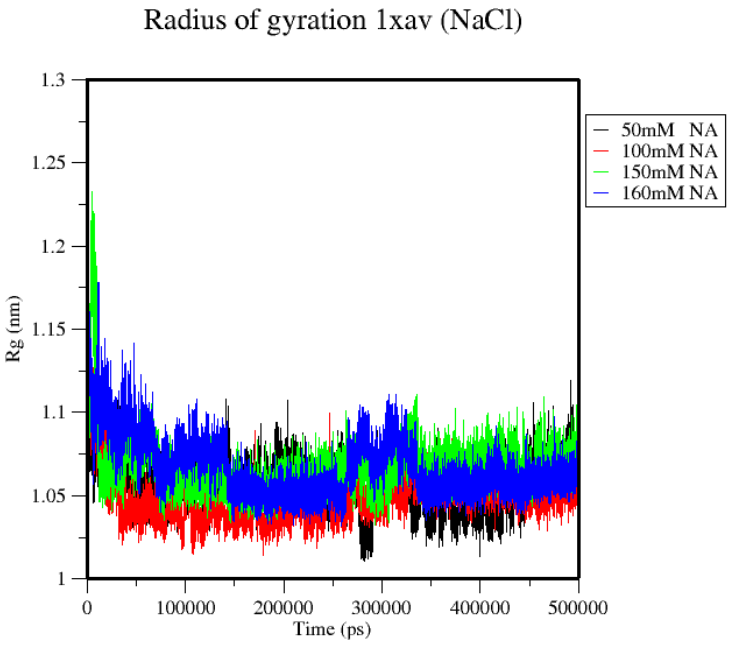

Figure 58. Radius of Gyration of 1xav in 50,100,150 and 160 mM of NaCl

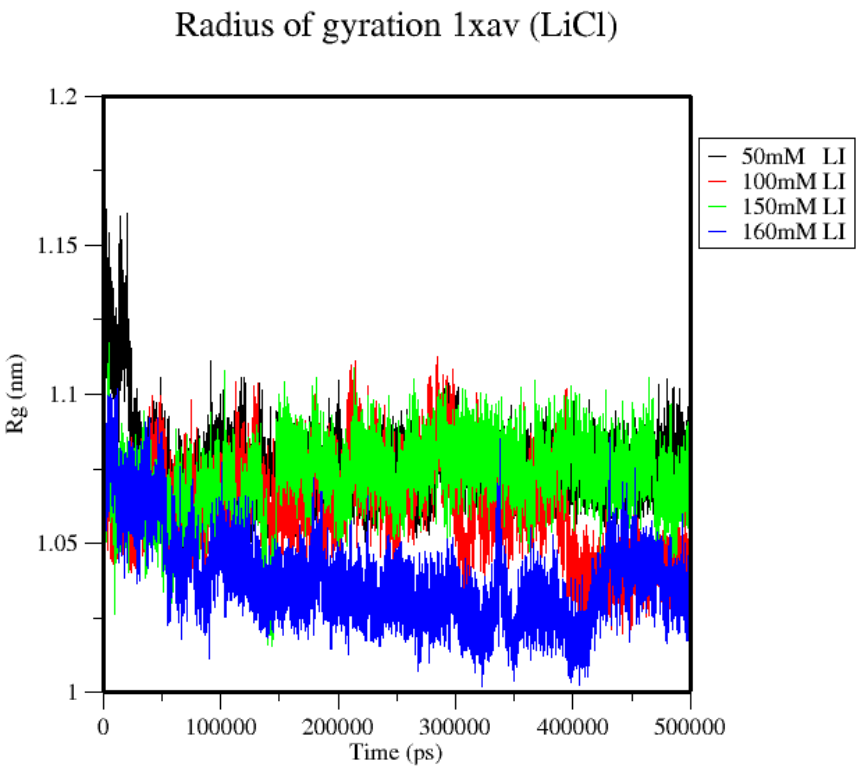

Figure 59. Radius of Gyration of 1xav in 50,100,150 and 160 mM of LiCl

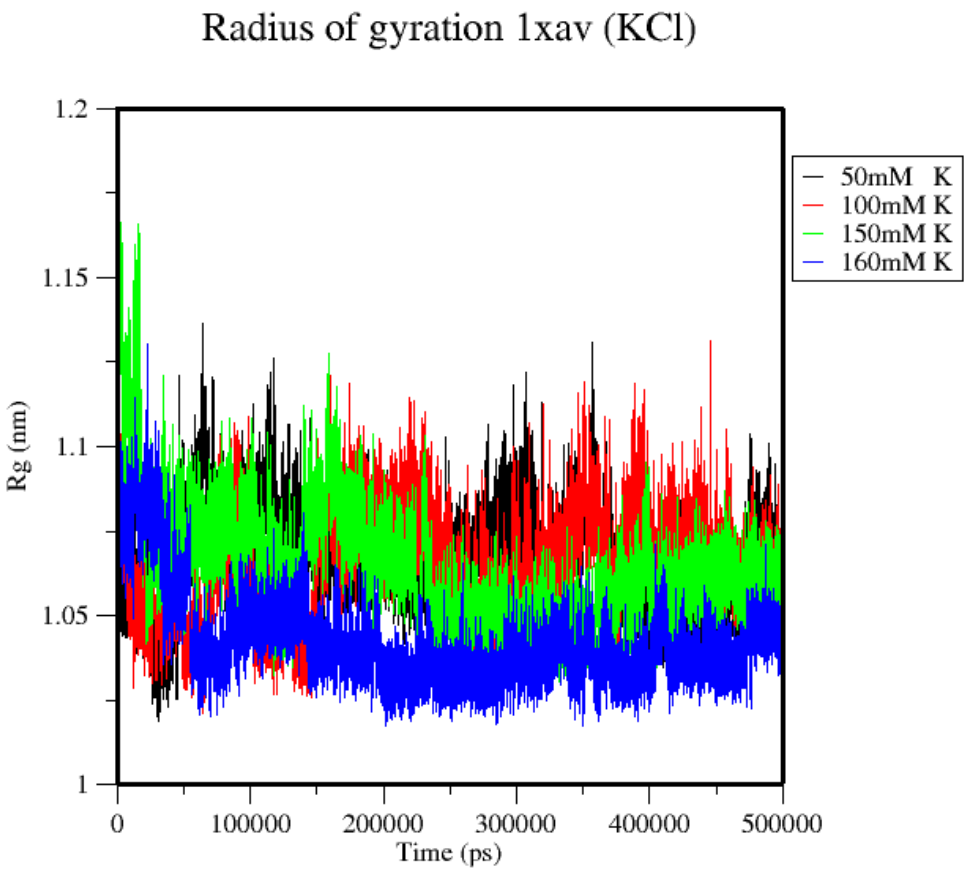

Figure 60. Radius of Gyration of 1xav in 50,100,150 and 160 mM of KCl

We analyzed the radius of gyration to understand the compactness of 1xav G-quadruplex DNA molecule.First, we looked at all the concentration of sodium chloride in figure 58,the radius of gyration of the system had the highest initial raise at 150 and 160mM concentration at about 1.23nm and converges at about 1.05nm indicating our structures compactness. In figure showing the radius of gyration of potassium chloride, contrary to what would have been expected the average radius of gyration was lowest at 160 mM though the initial raise was higher for 150 and 160mM concentration. On the average the system had an average radius of gyration of about 1.08 nm. Figure59 for

lithium chloride showed an average of 1.07 for 50 mM to 150 mM but at 160 mM the reduced to average of 1.03 nm which could indicate the structure stabilizing at higher concentration. Next, we carried out further analysis on the structure to determine the deviation of the atoms from its original structure.

4.3 Root Mean Square Deviation Analysis

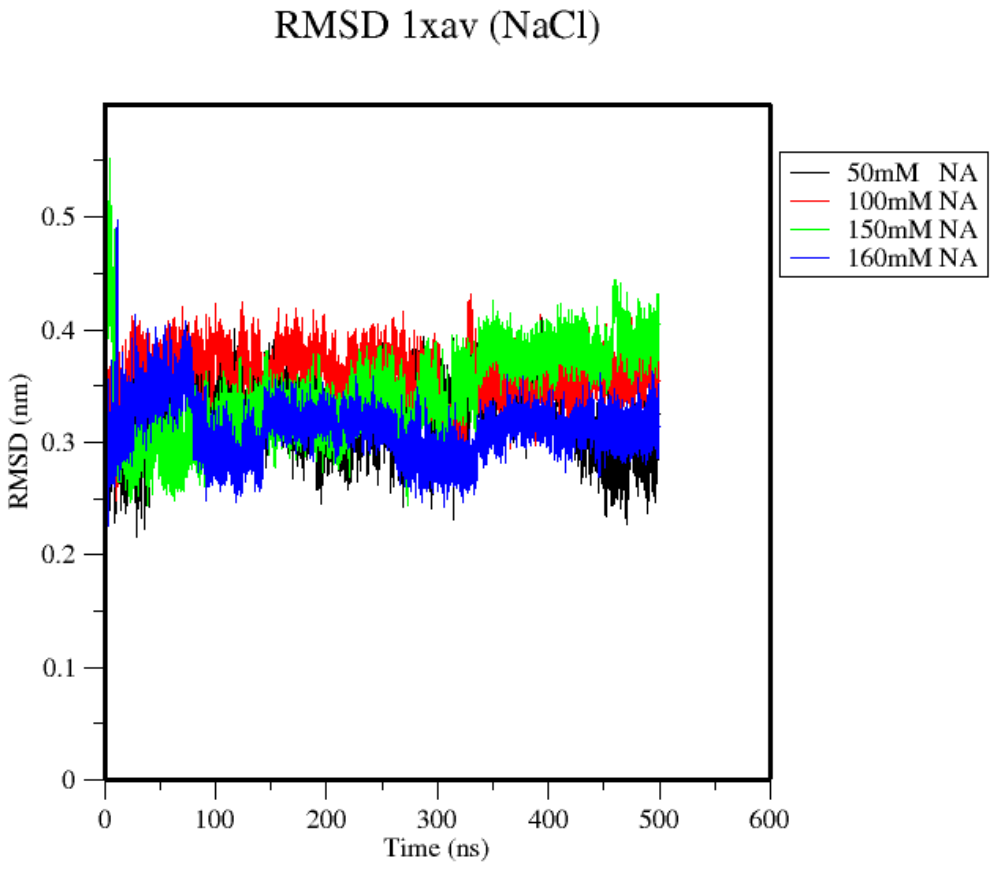

Figure 61. RMSD of 1xav in 50,100,150 and 160 mM of NaCl

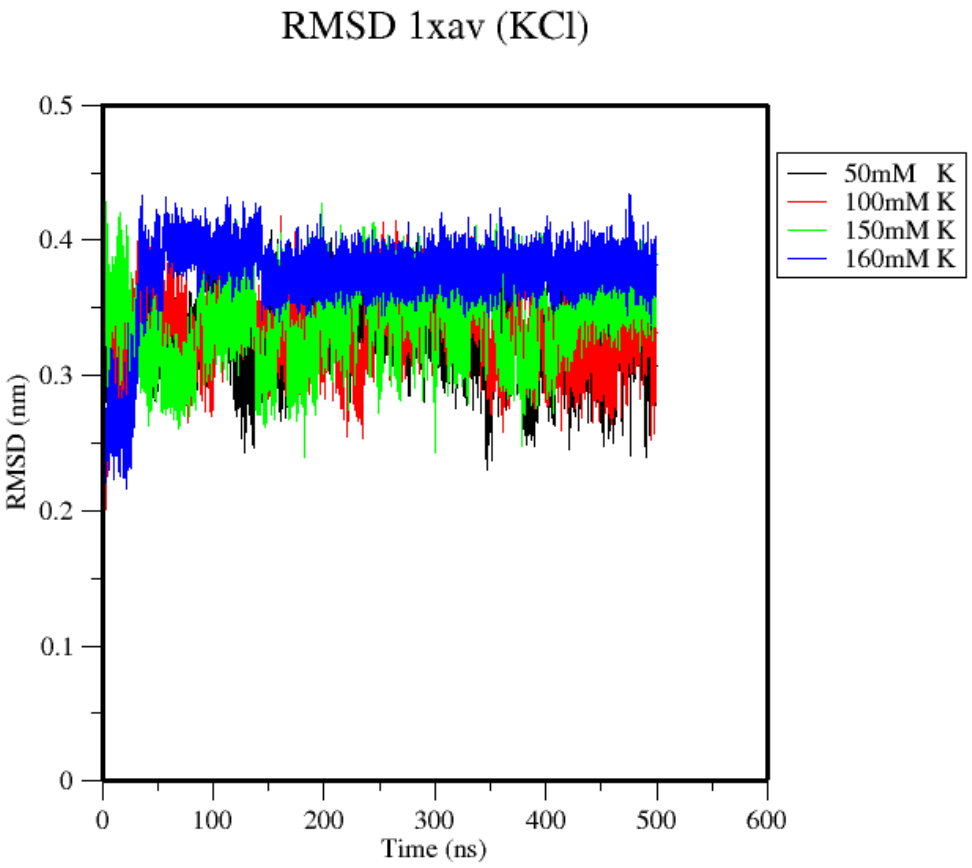

Figure 62. RMSD of 1xav in 50,100,150 and 160 mM of KCl

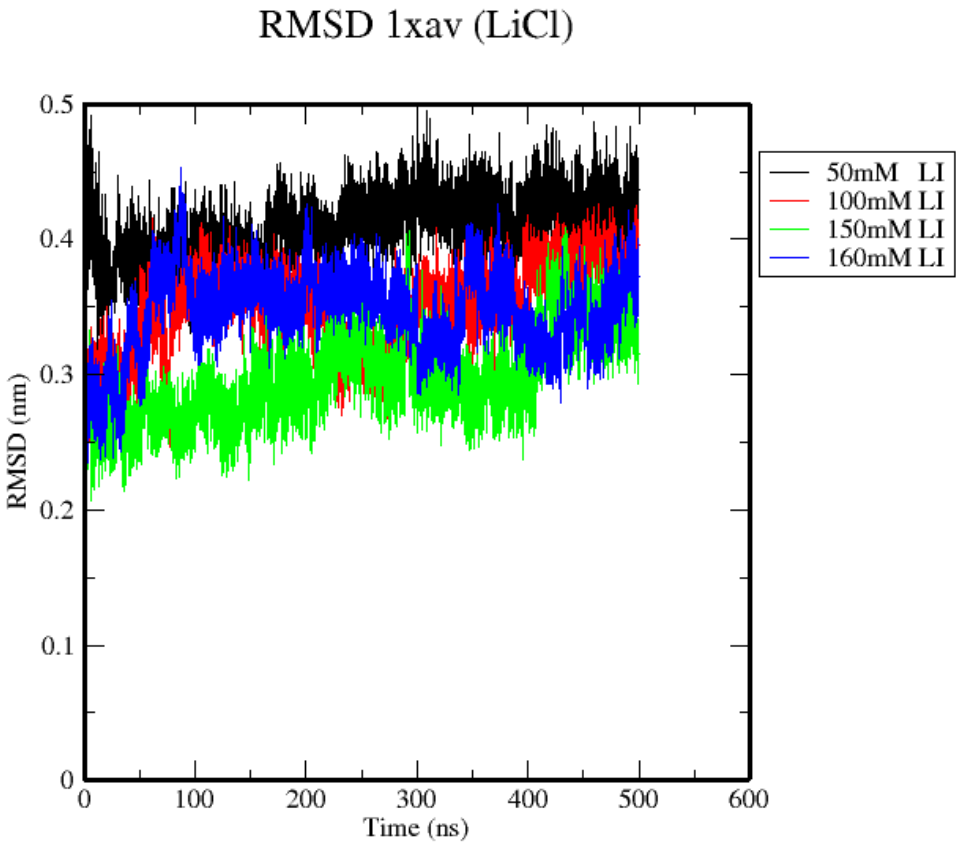

Figure 63. RMSD of 1xav in 50,100,150 and 160 mM of LiCl

4.3.1 Root Mean Square Deviation (RMSD)

The stability of the structure was evaluated by analyzing the root mean square deviation of the atoms of the simulated structure from the original structure and this was done by plotting the deviation against time. The rmsd shows the initial rise of RMSD and its relaxing or convergence point in the solution indicating equilibrated dynamical state of the system. We recorded and plotted the RMSD for the different concentration as shown in (figures 61-63) the rmsd of sodium in figure 61 converged below 0.4 with highest initial raise at 0.5 nm for the 150 mM concentration converging at 0.35nm.

The structure's rmsd converged below 0.4 for 50 to 150 mM of Potassium ion concentration. But at 160 mM ion concentration it converged at about 0.39 indicating possible structural change. Figure 63 shows lithium converging at 0.39 nm for 50 mM ion concentration with 100 and 160mM ion concentration converging at 0.35 nm. The 150 mM concentration converged at 0.28 nm indicating a potential deviation from the original structure. Next, we further verified the possible structural change by looking at the fluctuation of the molecule in the system.

4.3.2 Root Mean Square Fluctuation

This quantity describes the effect of thermal fluctuation of residues on the DNA structure and its function by local conformation and flexibility. We plotted fluctuation against residue to show the conformational change and will discourse our findings below.

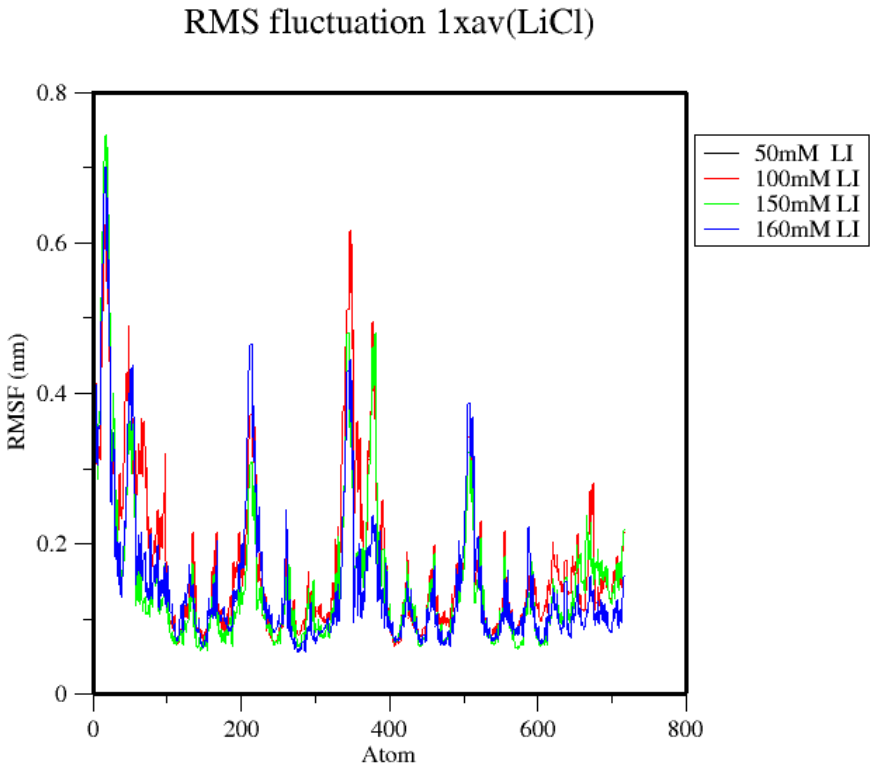

Figure 64. RMSF of 1xav in 50,100,150 and 160 mM of LiCl

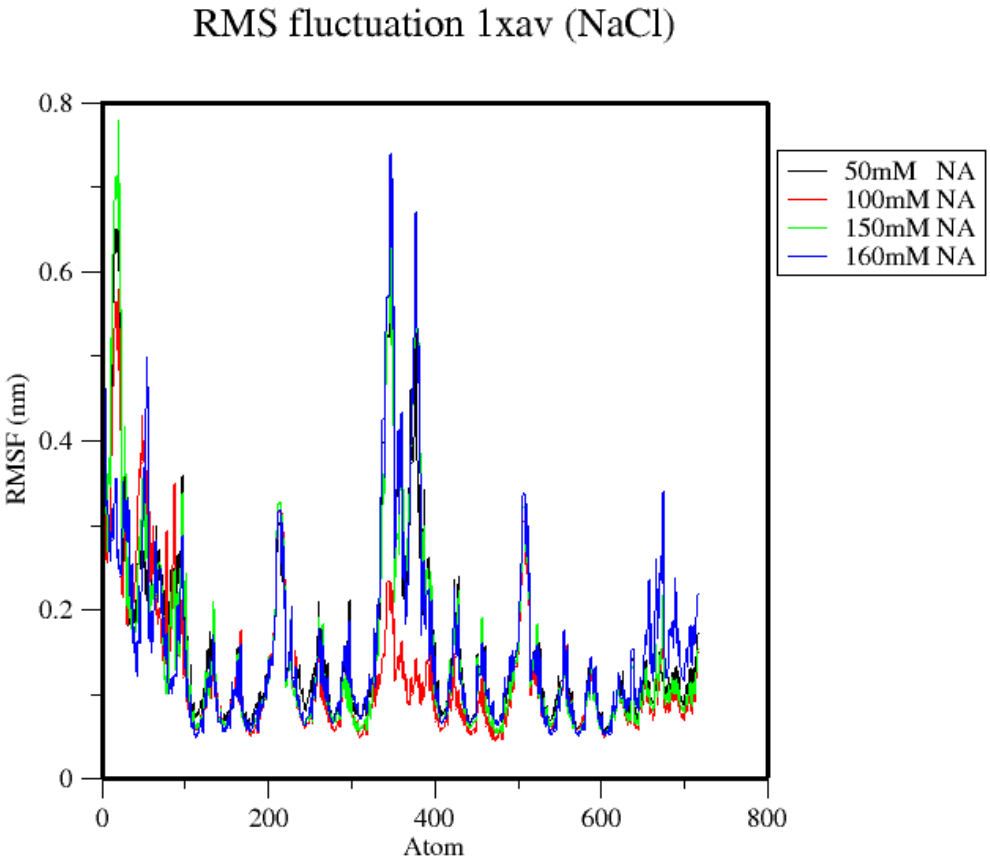

Figure 65. RMSF of 1xav in 50,100,150 and 160 mM of NaCl

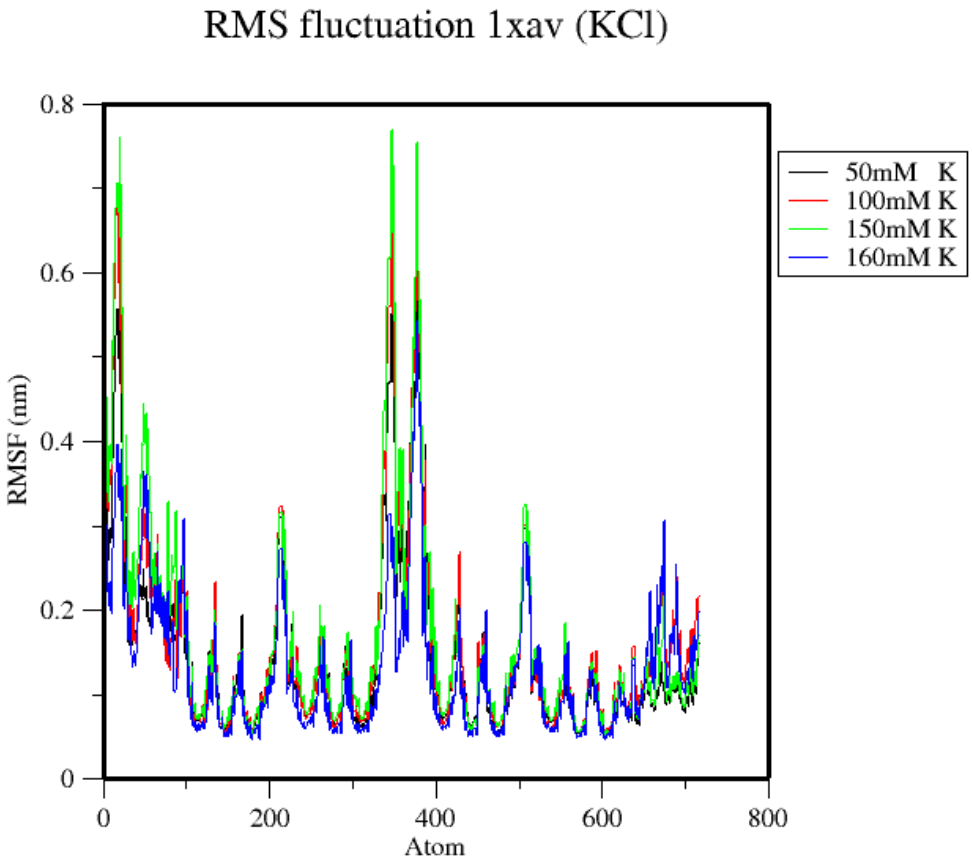

Figure 66. RMSF of 1xav in 50,100,150 and 160 mM of KCl

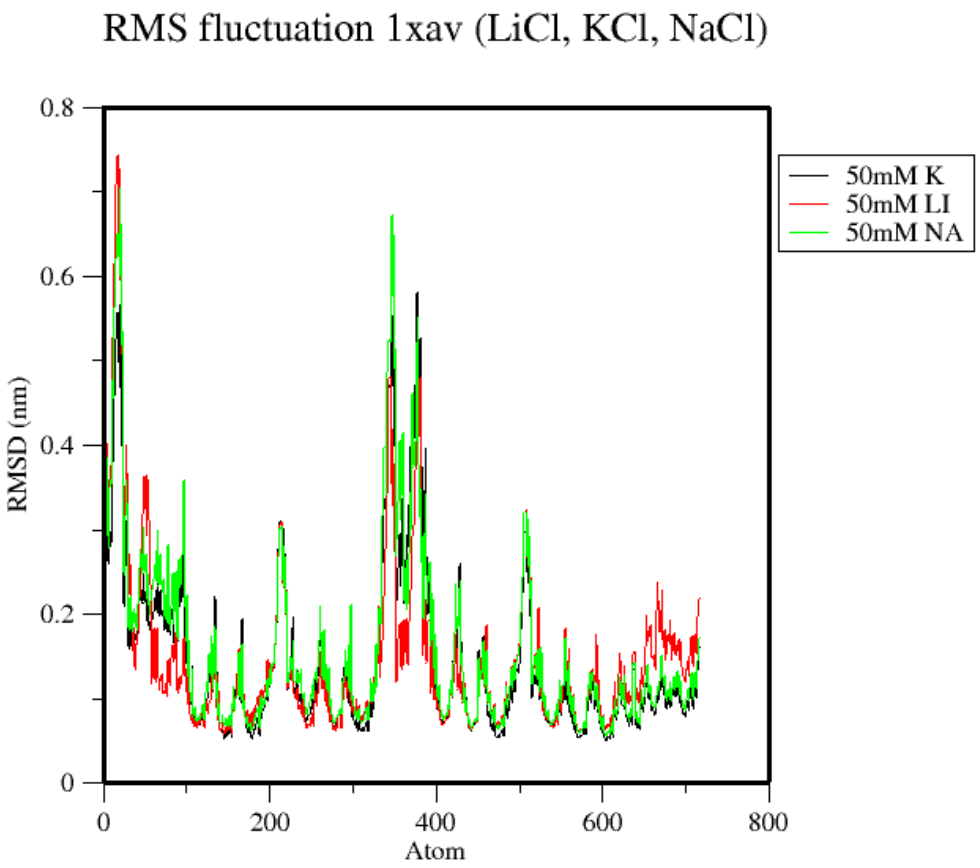

Figure 67. RMSF of 1xav in 50 mM of NaCl, KCl and LiCl

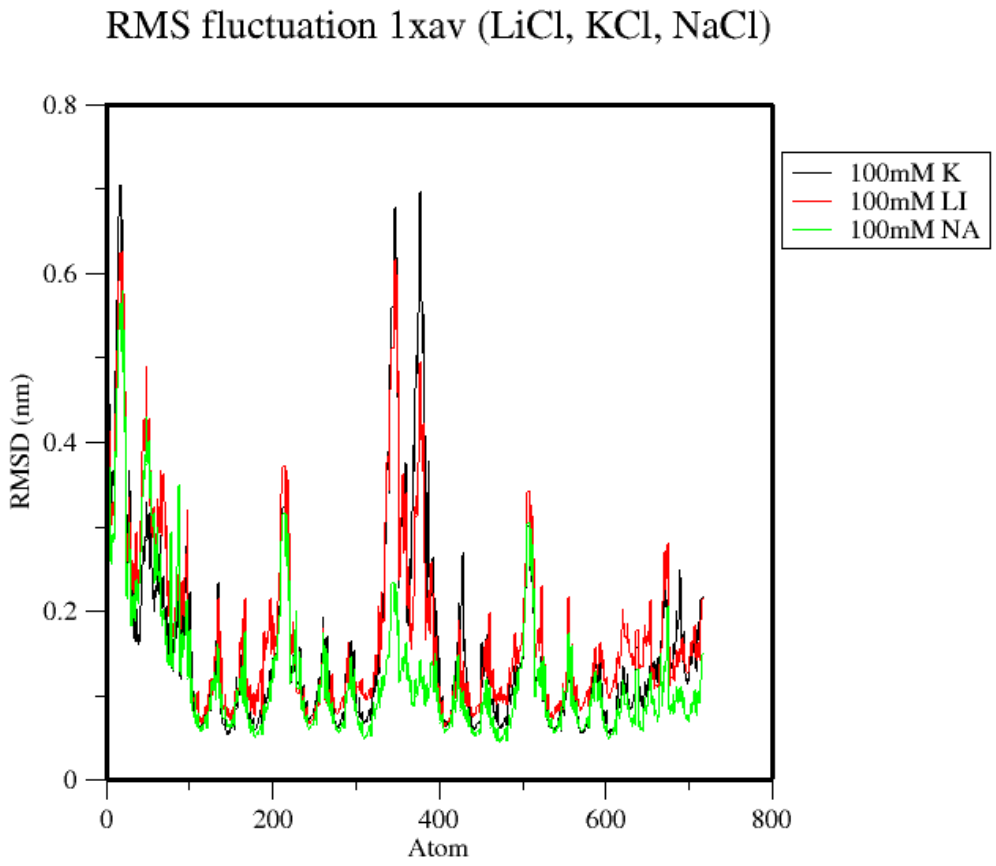

Figure 68. RMSF of 1xav in 100 mM of NaCl, KCl and LiCl

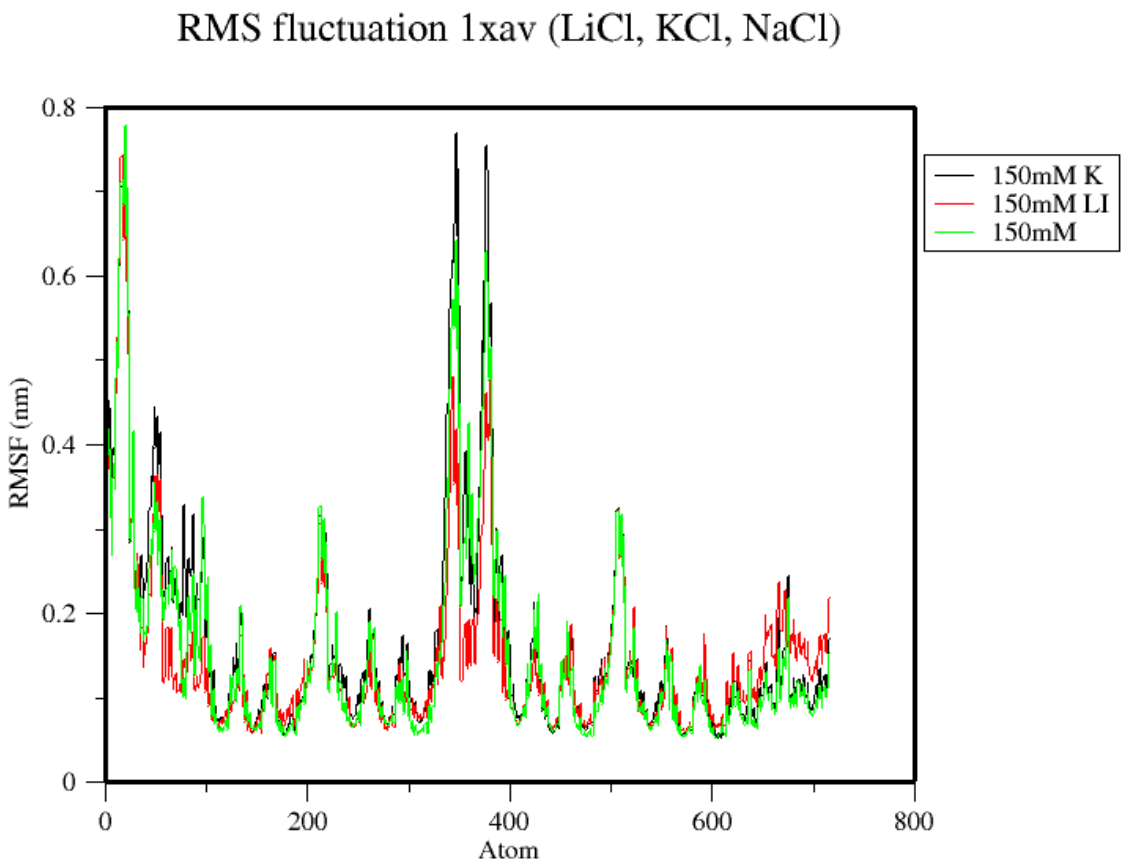

Figure 69. RMSF of 1xav in 150 mM of NaCl, KCl and LiCl

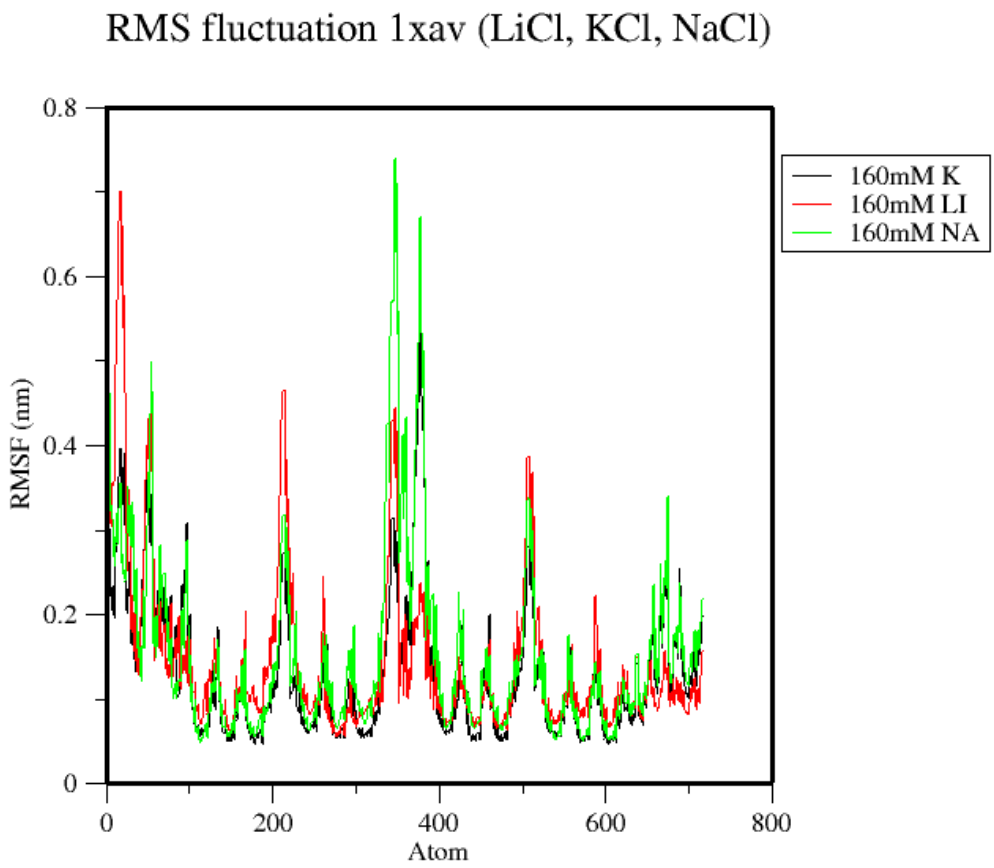

Figure 70. RMSF of 1xav in 160 mM of NaCl, KCl and LiCl

We observed a common symmetry between the 300 and the 400 atoms in all the ions and concentration for all the figures. In figure 68 and 70 the fluctuation of lithium concentration ions occurred around 100 to 160 mM concentration with most fluctuation by the 100 mM concentration suggesting 100 mM concentration had the most effect on the structure. For sodium in figure 64 the fluctuation happened most between 150 and 160 mM and sodium in figure 65 had the most fluctuation at 150 mM concentration. Comparing the three ions, at low concentration of 50 mM sodium had the highest

fluctuation followed by potassium at 100 mM and sodium at 160. Showing lithium potentially caused the least fluctuation in the system. To further understand the potential change in structure we plotted RMSD against concentration.

4.3.3 Structural Change

We Further compared the average structure of the trajectory after calculating the RMSD of the system and compared it with the initial structure used in the simulation. The conformational change where seen in the structure as shown in (table 7-9). Changes were observed in the structure after the simulation.

Table 7. RMSD of 1xav for Different Concentration of Lithium Chloride Solution

Concentration (mM)	RMSD (nm)
50	0.374259
100	0.395273
150	0.310927
160	0.374259

Table 8. RMSD of 1xav for Different Concentration of Sodium Chloride Solution

Concentration (mM)	RMSD (nm)
50	0.328227
100	0.352943
150	0.404894
160	0.313959

Table 9. RMSD of 1xav for Different Concentration of Potassium Chloride Solution

Concentration(mM)	RMSD (nm)
50	0.306504
100	0.328687
150	0.384251
160	0.38449

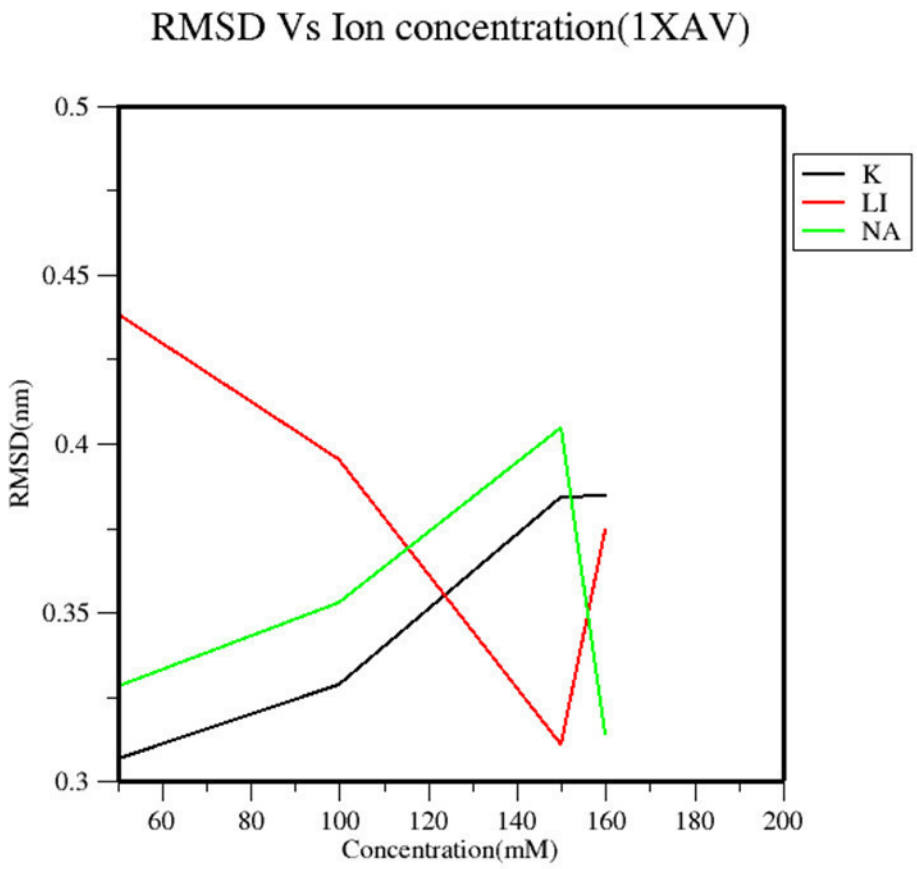

Figure 71. RMSD against Concentration of Ions for 1XAV in 50 mM,100 mM,150 mM and 160 mM of KCl, NaCl and LiCl

From figure 71 above shows the effect of different ion concentration on the structure if 1xav G-quadruplex DNA reveals lithium may have caused the most structural change of about 0.43 nm, followed by sodium with structural change of 0.33 nm. Potassium had the least structural change of about 0.31 nm. Compared to result in chapter, lithium displayed the least favorability to the stability of 1xav G-quadruplex DNA.

4.4 Summary

In this chapter we studied 1xav G-quadruplex DNA structure and analyzed the effect of different ion concentration on the structure of our molecule. The simulation results showed different response from the results of the previous chapter. Lithium seem to show a higher RMSD fluctuation followed by sodium and potassium. For lower concentration potassium and sodium showed a higher RMSF. Lithium showed higher RMSF compared to potassium and sodium at higher concentration which affirms potential changes in structure. The plot of RMSD against concentration indicates lithium had the most structural change on the structure of 1xav G-quadruplex DNA. To further confirm the observed changes, we proceed to study the effect of these ion concentrations on another G-quadruplex molecule known as 2l7v. For this purpose, we use similar simulation setup as discussed in the previous chapters to create and study the effect of different ion concentration on our new molecule. The effect of ion concentration on 2l7v is presented in the next chapter.

CHAPTER V

MOLECULAR DYNAMIC SIMULATION OF 2L7V G-QUADRUPLEX IN DIFFERENT SALT SOLUTIONS (POTASSIUM CHLORIDE, SODIUM CHLORIDE AND LITHIUM CHLORIDE)

5.1 Introduction

The differential effect of ion concentration on 4fxm and 1xav G-quadruplex DNA have been studied in the previous chapter. We also performed some analysis on the simulation data. For more insight into the conformational changes observed in the analysis performed, we will perform another computational simulation on another G-quadruplex molecule known as 2l7v form RCSB protein data bank. 2L7V is a G-quadruplex complex formed in human c-MYC oncogene promoter region of human DNA. (DNA/Inhibitor) We performed the same simulation process and steps as in the previous chapters. It is another parallel stranded DNA that has been studied both experimentally and computationally but not much has been known about differential effect of ion concentration on its structure and it's as a potential regulatory role at the telomeric region of human DNA. Our intention is to understand the differential effect of alkali metal ions on the structure and stability of 2l7V G-quadruplexes. For this set up we will modify the molecule by detaching the attached drug ligand, replace the cation at the of the structure with Na^+, K^+ and Li^+ ions respectively. Next, we will run simulation of different concentrations of 50, 100, 150 and 160 millimolar of NaCl, KCl and LiCl.

We will analyze various thermodynamic quantities in addition to G-quadruplex structural parameters such as radius of gyration and root mean square deviation (RMSD) and root mean square fluctuation.

5.1.1 Simulation Details

Simulation of 2l7V G-quadruplex DNA was performed using GROMACS MD code. The pdb structures of 2L7V G-Quadruplex DNA was downloaded from RCSB website. A simulation directory was created for minimization file minim.mdp, the equilibration files nvt.mdp, npt.mdp, md.mdp and the pdb file. The counter ion in the G-Quadruplex was replaced with three different ions namely Na^+, K^+ and Li^+.TIP3P water was used in solvating the system [121], while computational visualization tool VMD [122] was used to visualize progress in simulation and visually inspect the G-Quadruplex DNA structure. All other steps were as in previous chapters. Other simulation details are presented in appendix C.

We ran the simulation for 500 ns and at the end of the simulation we carried out some structural analysis.

Table 10 . Description of 2L7V G-quadruplex DNA Molecule

Chain	G-Quadruplex DNA	Ion	Length
A	DNA (5'D(*TP*GP*AP*GP*GP*GP*TP*GP*GP* GP*TP*AP*GP*GP*GP*TP*GP*GP*GP*TP*AP*A)-3')	K^+	22

5.1.2 System Run, Stability and Conformational Changes

Trajectories of solvated 2L7V G-quadruplex were recorded during MD production run by collecting data at every 100ps for post-processing. (Figures 72-83) displays the initial structure and final structure of 2L7V G-quadruplex in 50,100,150 and 160 millimolar of different ion concentration. (The structure above is a structure without the water molecules). The generated data was used to compute basic thermodynamic quantities such as RMSD, RG and RMSF of the system. Next, we analyzed the results of simulation.

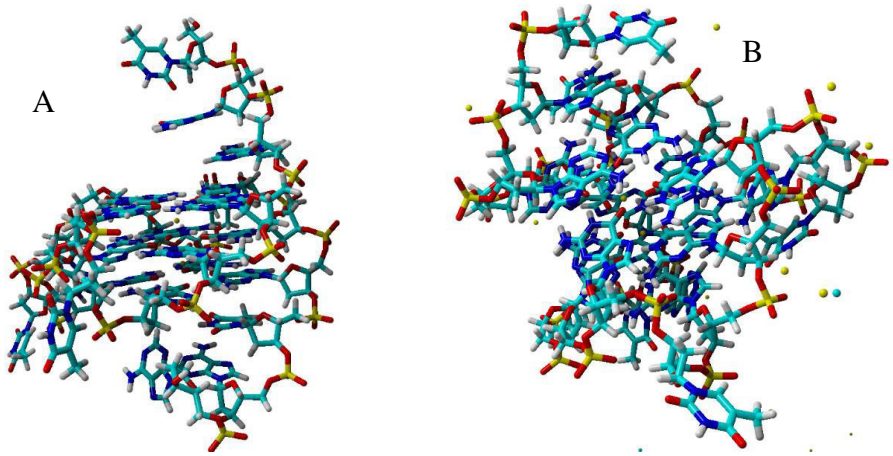

Figure 72. (A) Initial Structure 2l7v G-quadruplex (B) Final Structure of 2l7v G-quadruplex after Simulation in 50 mM KCl

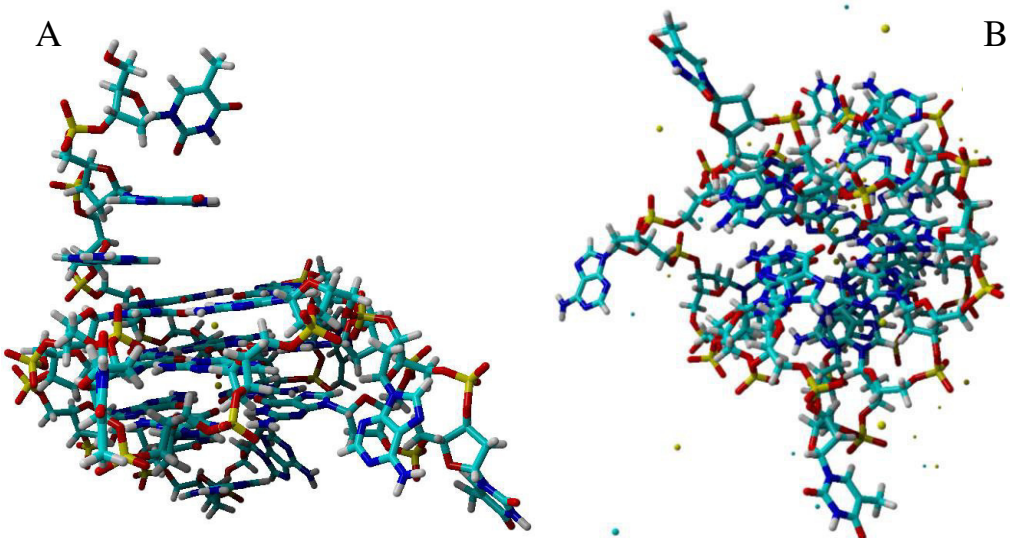

Figure 73. (A) Initial Structure 2l7v G-quadruplex (B) Final Structure of 2l7v G-quadruplex after Simulation in 100 mM KCl

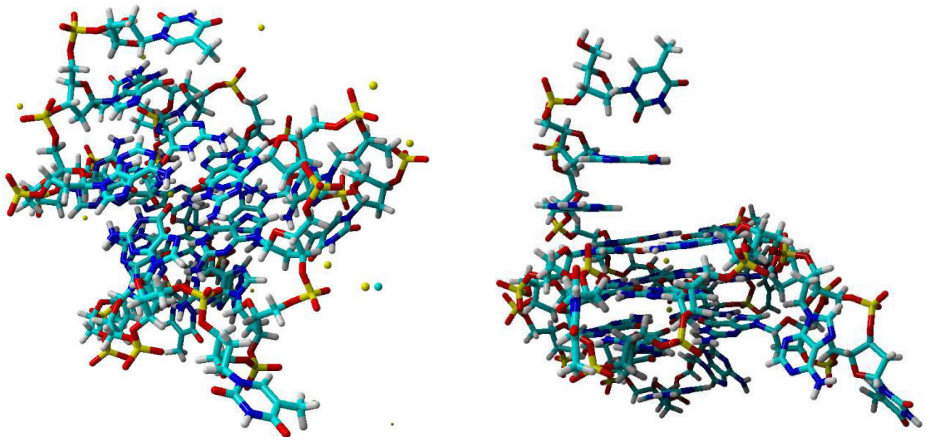

Figure 74. (A) Initial Structure 2l7v G-quadruplex (B) Final Structure of 2l7v G-quadruplex after Simulation in 150mM KCl

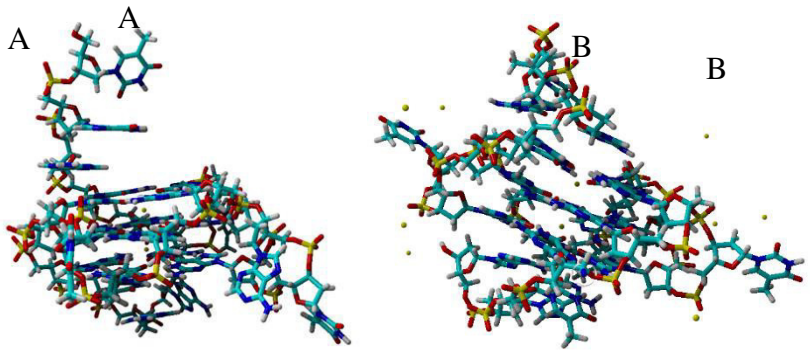

Figure 75. (A) Initial Structure 2l7v G-quadruplex (B) Final Structure of 2l7v G-quadruplex after Simulation in 160 mM KCl

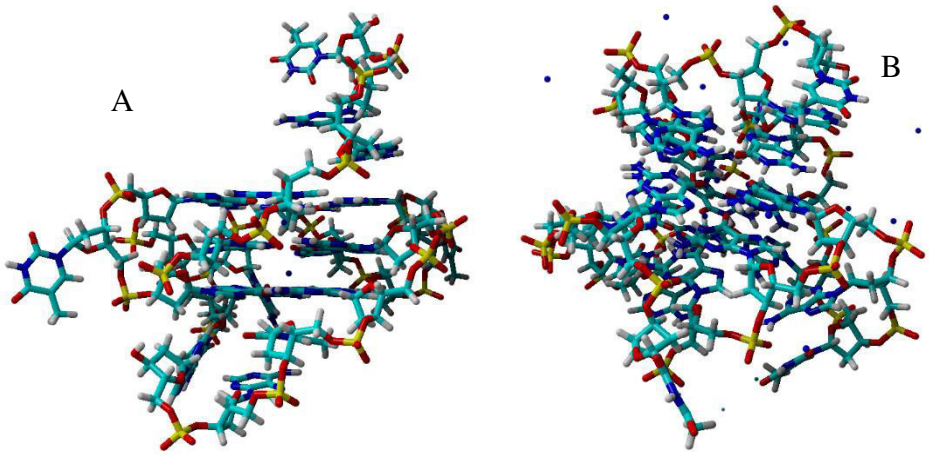

Figure 76. (A) Initial Structure 2l7v G-quadruplex (B) Final Structure of 2l7v G-quadruplex after Simulation in 50 mM NaCl

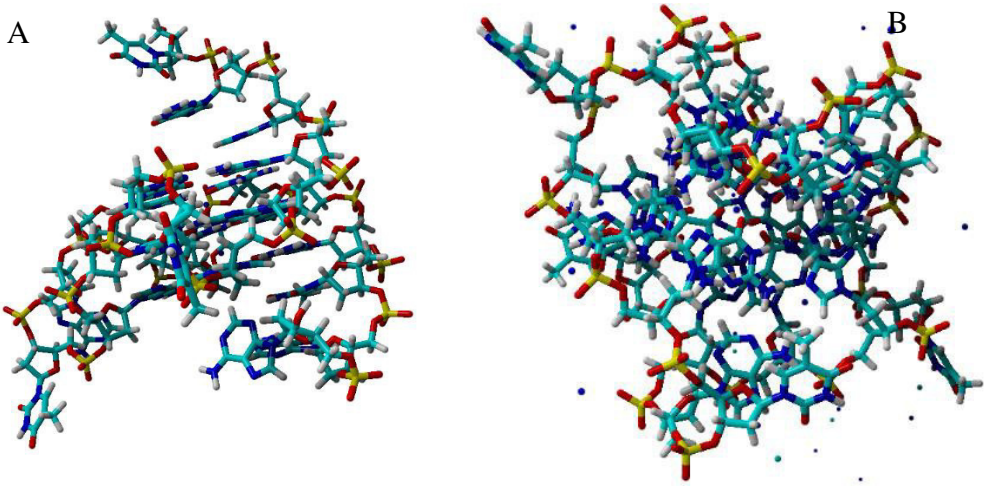

Figure 77. (A) Initial Structure 2l7v G-quadruplex (B) Final Structure of 2l7v G-quadruplex after Simulation in 100 mM NaCl

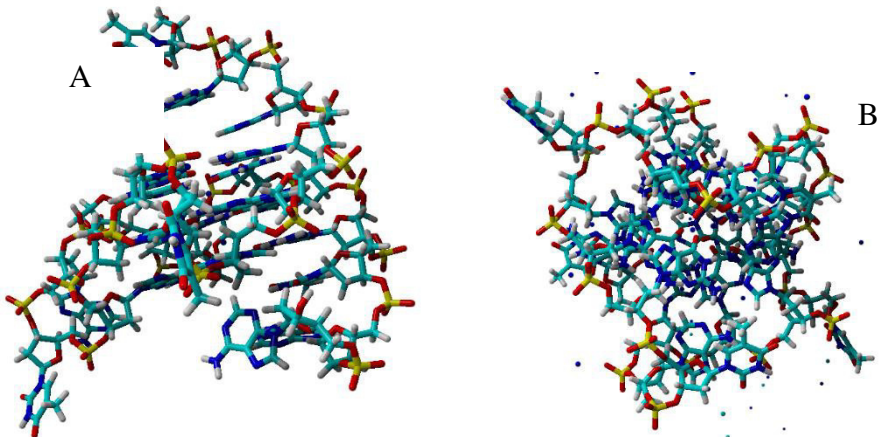

Figure 78. (A) Initial Structure 2l7v G-quadruplex (B) Final Structure of 2l7v G-quadruplex after Simulation in 150 mM NaCl

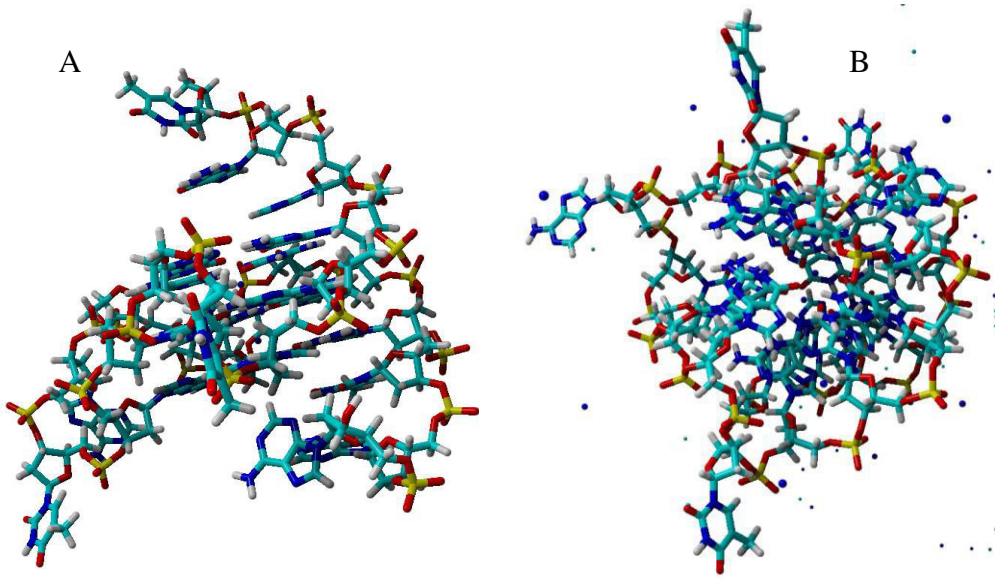

Figure 79. (A) Initial Structure 2l7v G-quadruplex (B) Final Structure of 2l7v G-quadruplex after Simulation in 160 mM NaCl

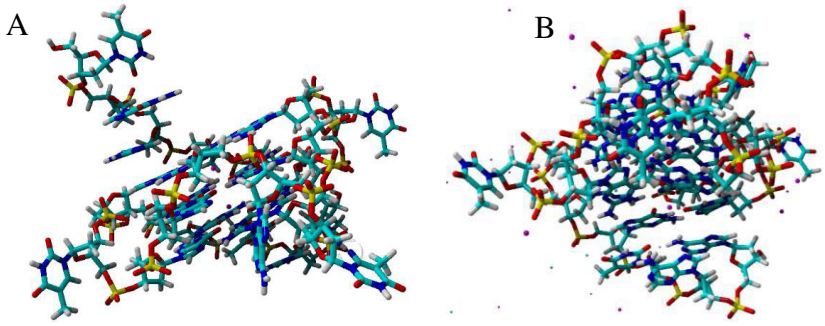

Figure 80. (A) Initial Structure 2l7v G-quadruplex (B) Final Structure of 2l7v G-quadruplex after Simulation in 50 mM LiCl

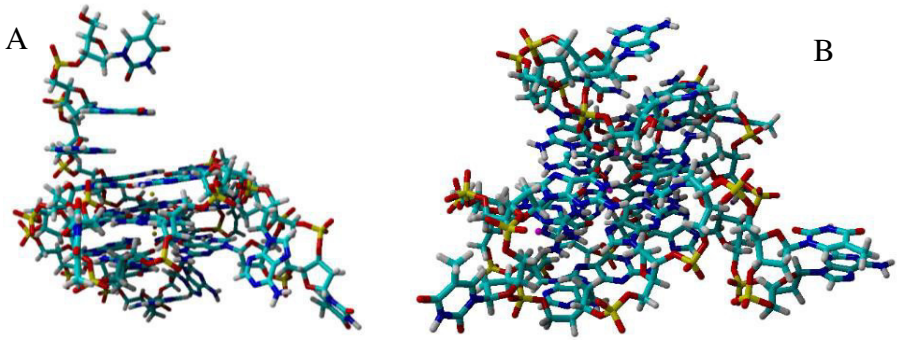

Figure 81. (A) Initial Structure 2l7v G-quadruplex (B) Final Structure of 2l7v G-quadruplex after Simulation in 100 mM LiCl

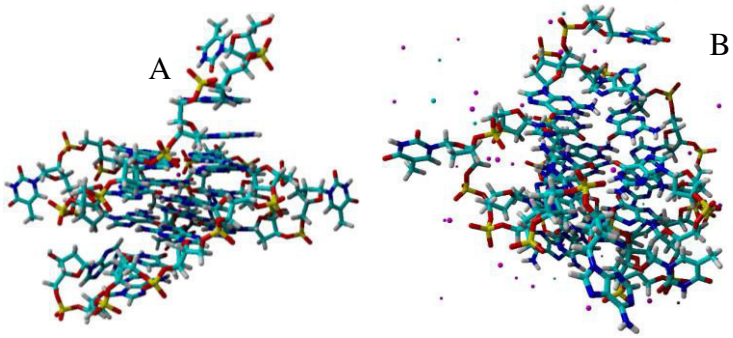

Figure 82. (A) Initial Structure 2l7v G-quadruplex (B) Final Structure of 2l7v G-quadruplex after Simulation in 150 mM LiCl

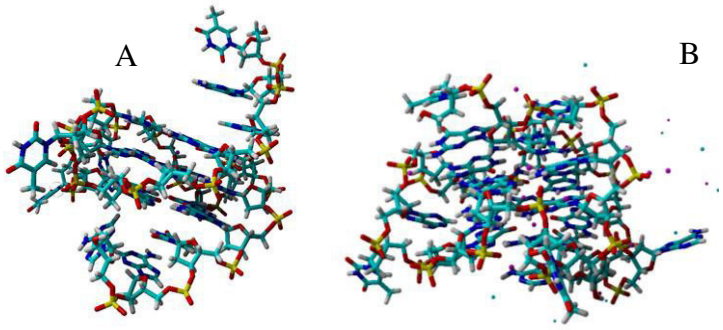

Figure 83. (A) Initial Structure 2l7v G-quadruplex (B) Final Structure of 2l7v G-quadruplex after Simulation in 160 mM LiCl

We carefully examined the initial and final structure of 2L7V G-quadruplex visually using VMD. The structure showed significant changes in the size of the G-quadruplex molecule and position of atoms. The final structure seems destabilized compared to the initial structure. A closer examination showed the final structure of 2L7V maintained its counter ion at the center but structurally changed the shape of the and atom positions of the molecule. This change showed potential destabilization of the molecule. Compared to the two earlier molecules (4fxm and 1xav), the structural change is closer to 1xav molecule as they both retained the ions in the quadruplex channel while 4fxm molecule lost some of the channel ions. Concentrations

We also observed that more structural changes happen around the molecule in lithium ion concentration compared followed by potassium and sodium.

To further confirm our findings, we calculated and compared thermodynamic quantities such as radius of gyration (Rg), root mean square deviation (RMSD) and root mean square fluctuation (RMSF).

5.2 Radius of Gyration

This measure helps in understanding the compactness of the G-Quadruplex DNA structure. For example, how much the structure has spread or contracted with regards to its degree of freedom. A relatively compact structure or molecule will maintain a level of stability in its radius of gyration Rg.

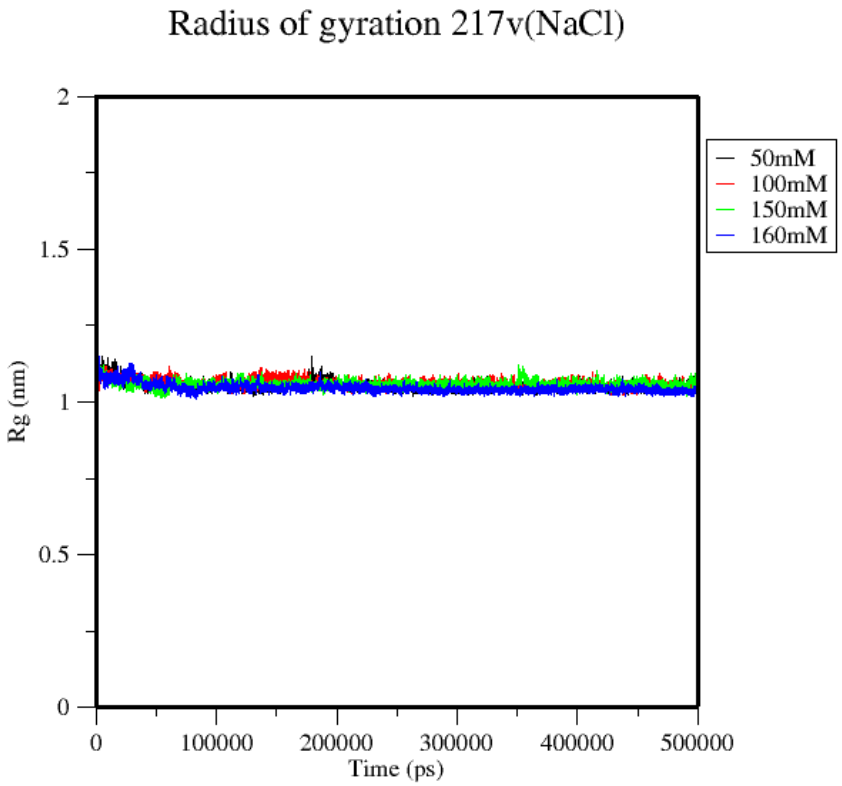

Figure 84. Radius of Gyration of 2L7V in 50,100,150 and 160 mM NaCl

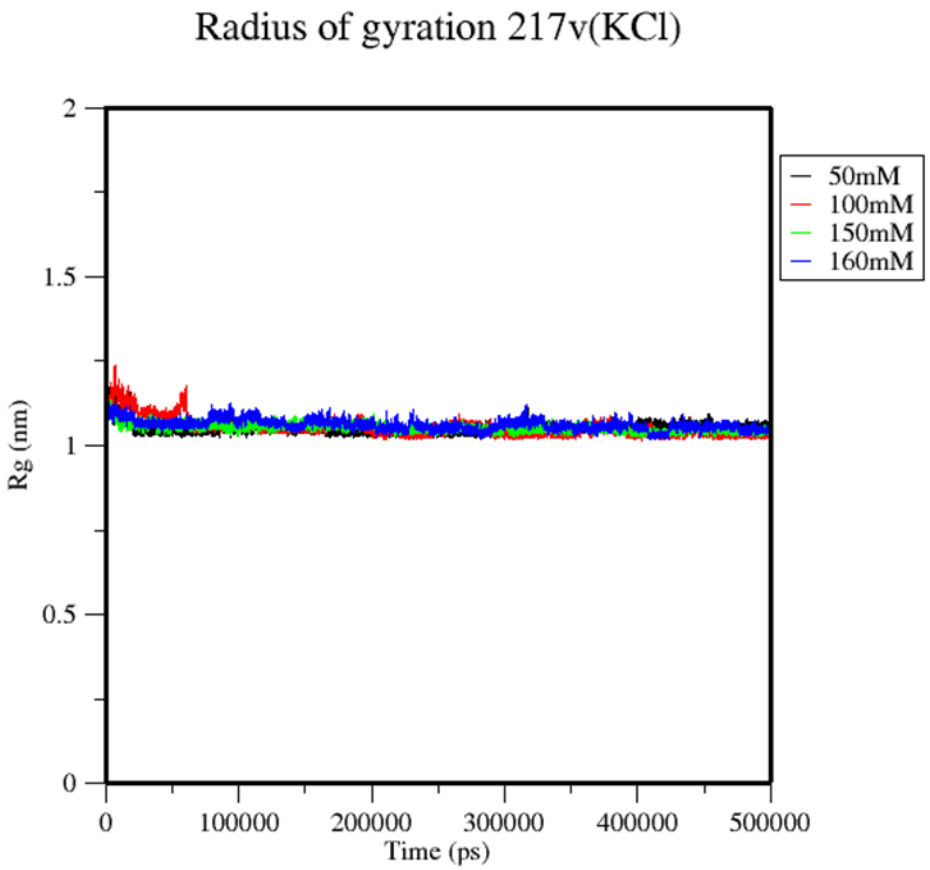

Figure 85. Radius of Gyration of 2L7V in 50,100,150 and 160 mM of KCl

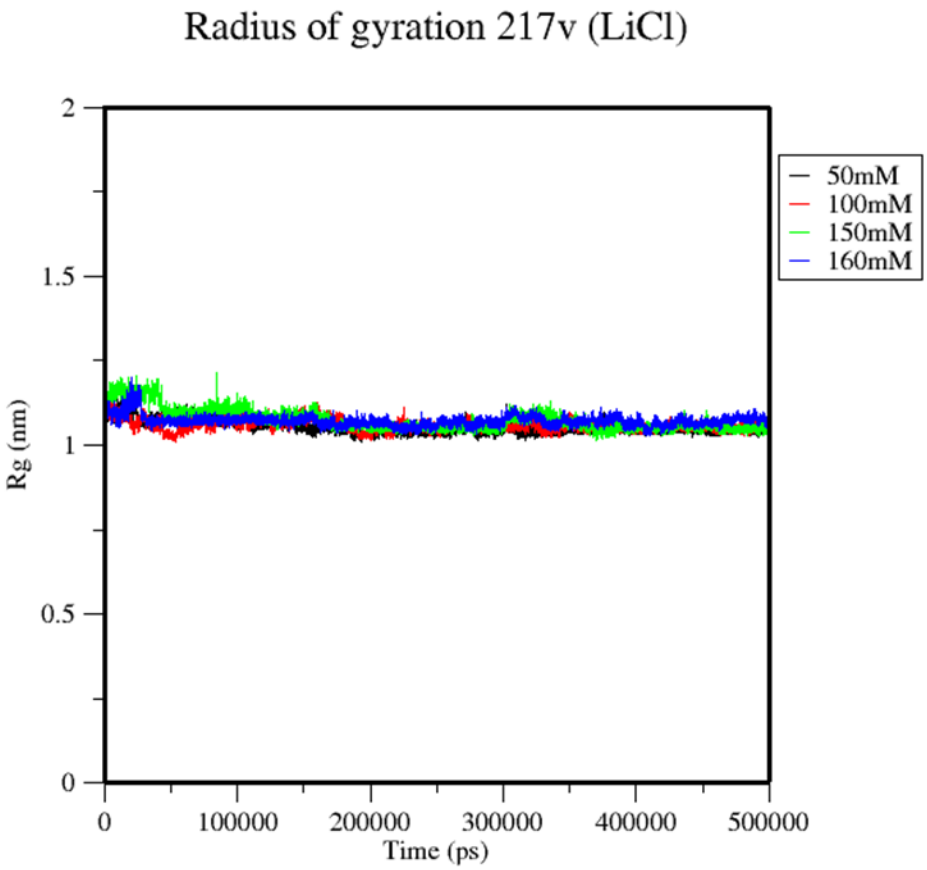

Figure 86. Radius of Gyration of 2L7V in 50,100,150 and 160 mM of LiCl

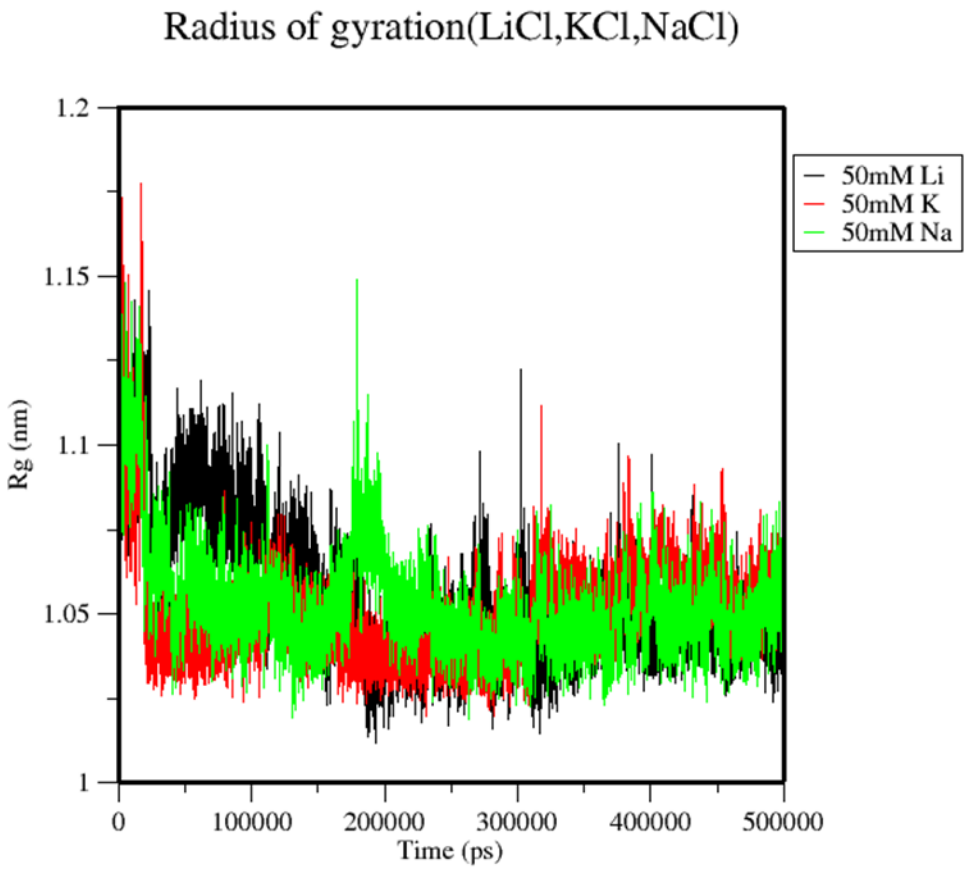

Figure 87. Radius of Gyration of 2L7V in 50 mM of NaCl, KCl and LiCl

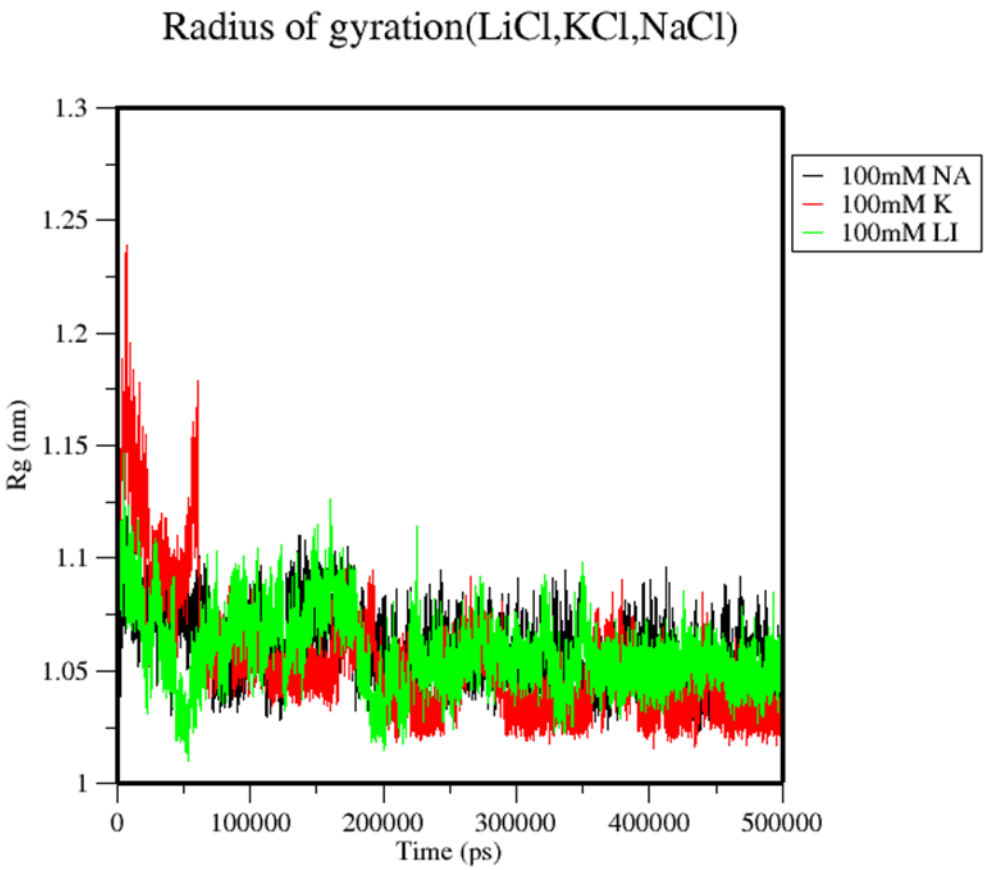

Figure 88. Radius of Gyration of 2L7V in 100 mM of NaCl, KCl and LiCl

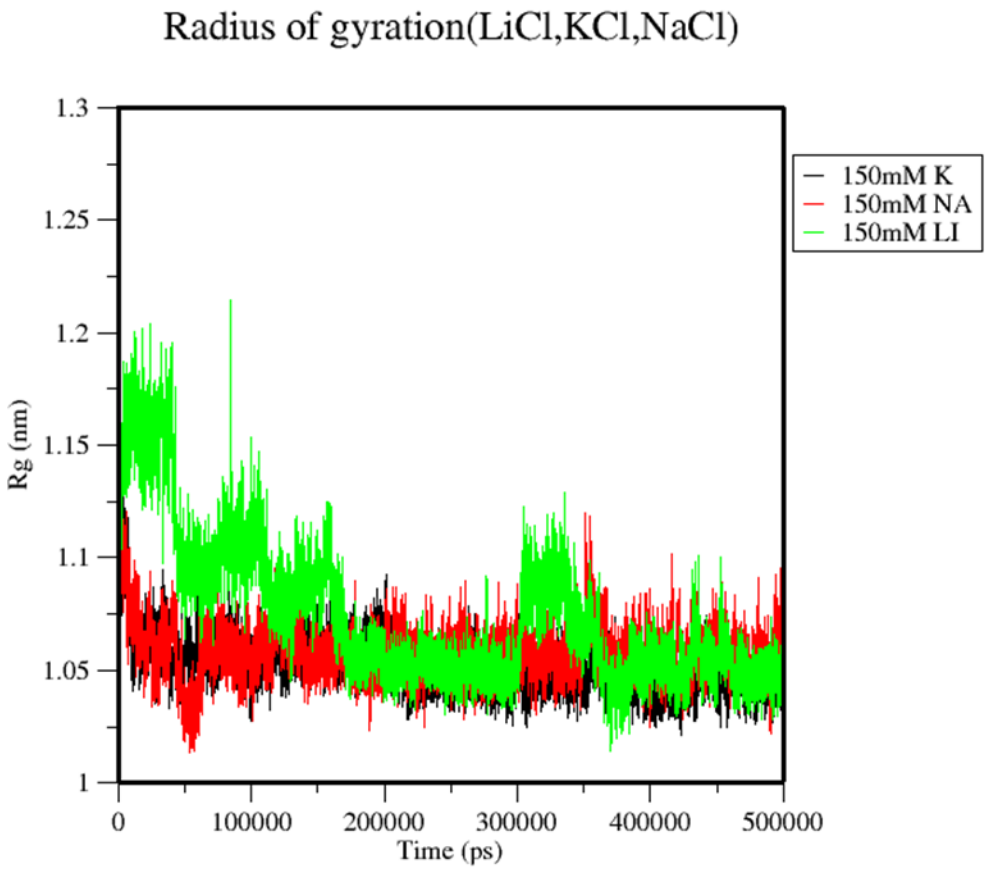

Figure 89. Radius of Gyration of 2L7V in 150 mM of NaCl, KCl and LiCl

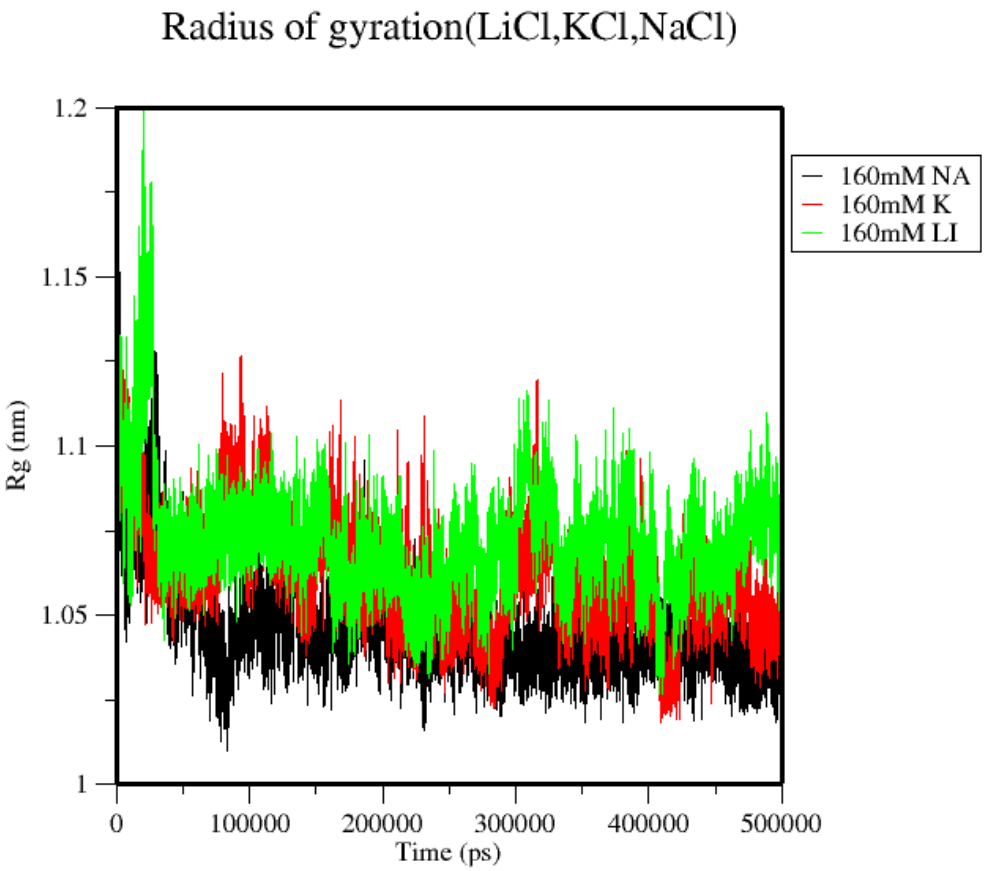

Figure 90. Radius of Gyration of 2L7V in 160 mM of NaCl, KCl and LiCl

We examined the radius of gyration to understand the compactness of 2L7V G-quadruplex DNA molecule. First, we looked at all the concentration of sodium chloride in figure 84.The radius of gyration of the system had the highest initial raise at 50 mM with 100 and 150 mM having the highest radius of gyration of about 1.2 nm indicating our structures compactness. In figure 85 shows the radius of gyration for lithium showing higher Rg for concentration of 150 and 160mM indication of possible structural change at higher concentration. Figure 86 is the radius of gyration for the molecule with potassium

counter ion in KCl solution. With 100 mM concentration had the highest initial raise of about 1.25 nm followed by 150 and 160mM concentration indicating the most structural change may have happened at 100 mM concentration. The lower radius of gyration at 160 mM and maybe indication of no further structure change. Figure 87 compares the effect of the three counter-ion at 50mM concentration, at lower concentration sodium showed the potential to cause more structural change compared to lithium and potassium. Figure 88 indicate no further structural change at 100 mM concentration. Figure 89 shows a potential structural change by lithium ion further increase in concentration in figure 88 indicates potassium and lithium may have caused a structural change. Next, we carried out further analysis on the structure to determine the deviation of the atoms from its original structure.

5.2.1 Root Mean Square Deviation (RMSD)

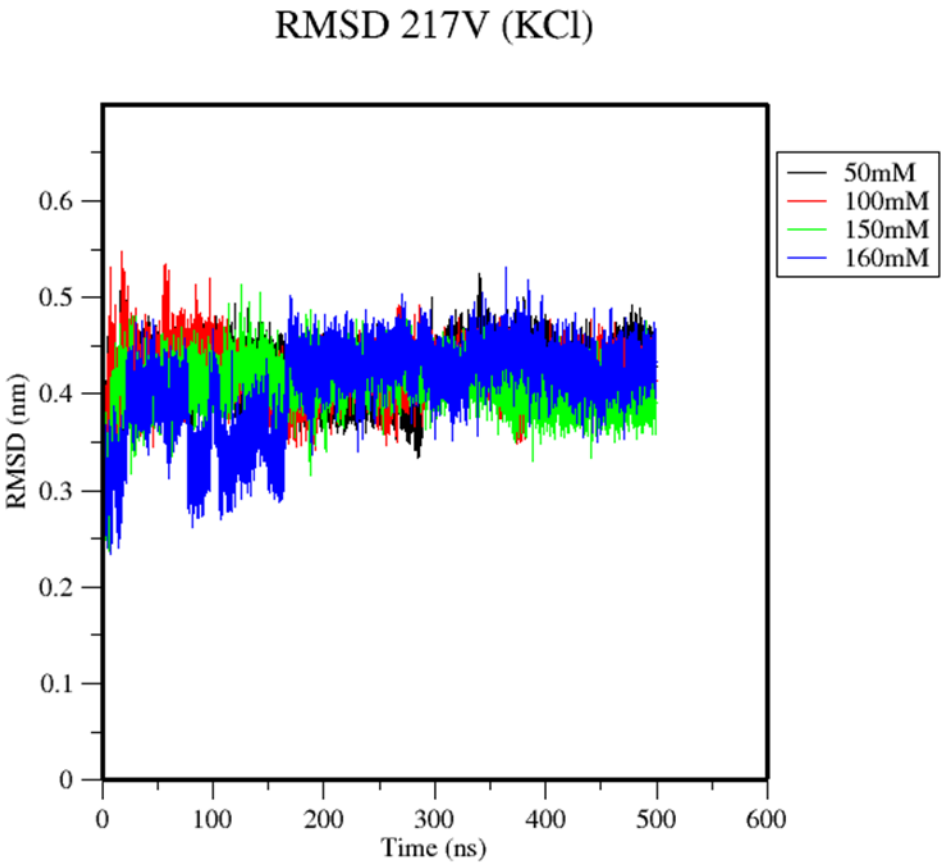

Figure 91. RMSD of 2L7V in 50,100,150 and 160 mM of KCl

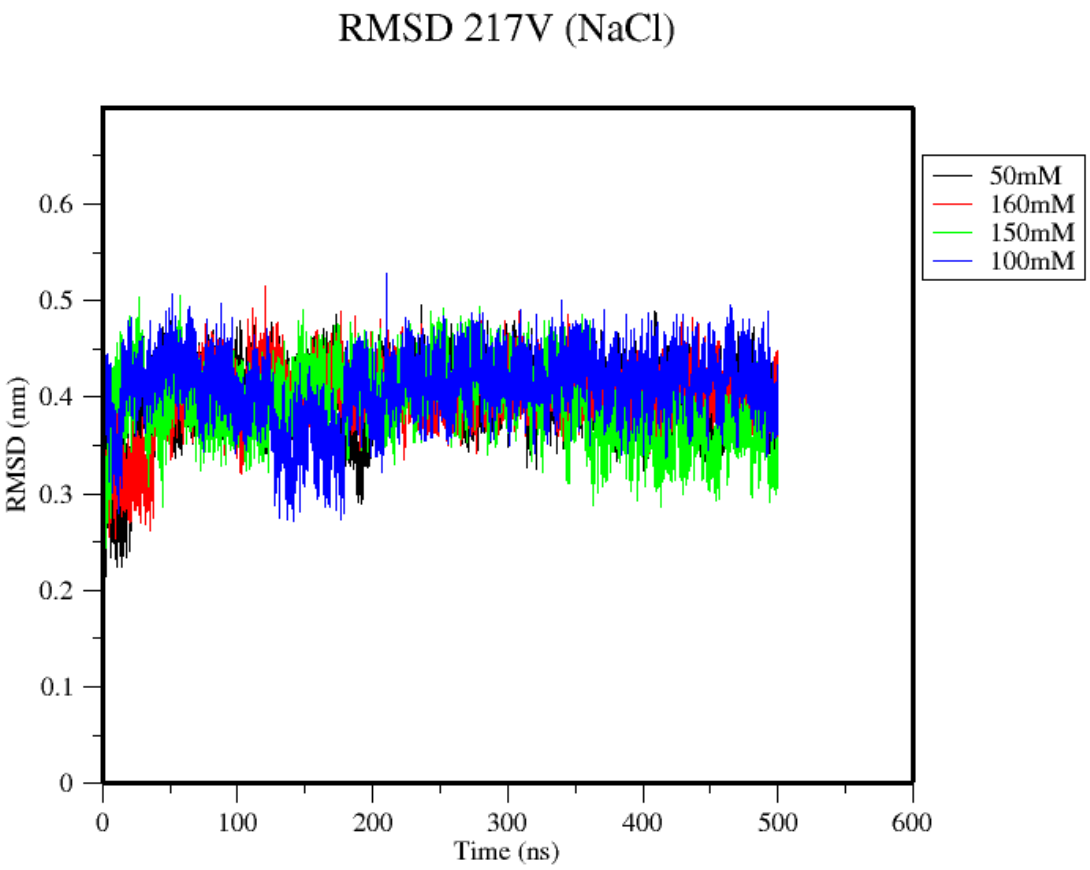

Figure 92. RMSD of 2L7V in 50,100,150 and 160 mM of NaCl

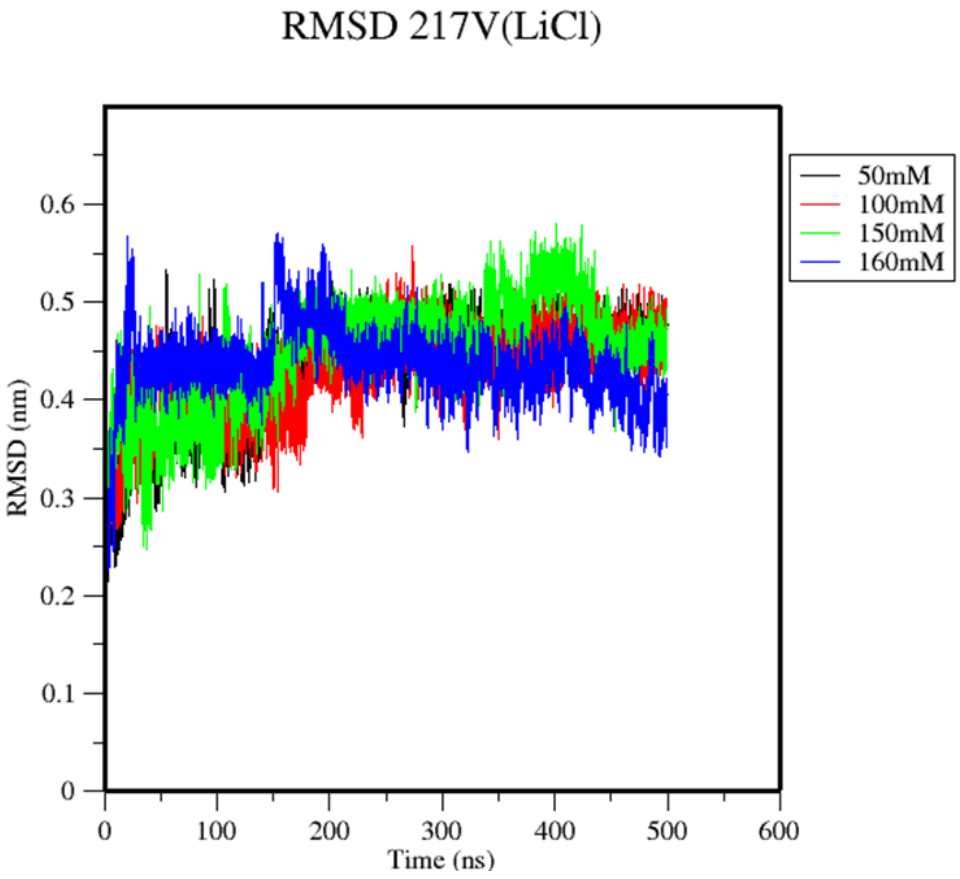

Figure 93. RMSD of 2L7V in 50,100,150 and 160 mM of LiCl

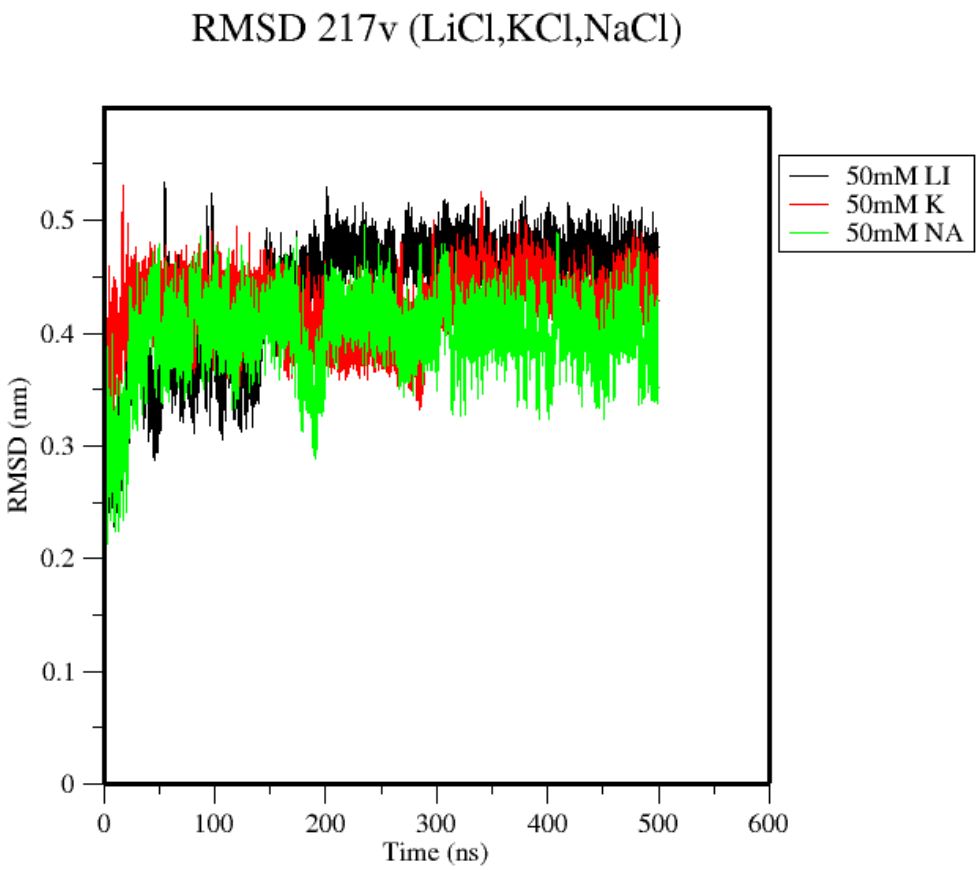

Figure 94. RMSD of 2L7V in 50 mM of KCl, NaCl and LiCl

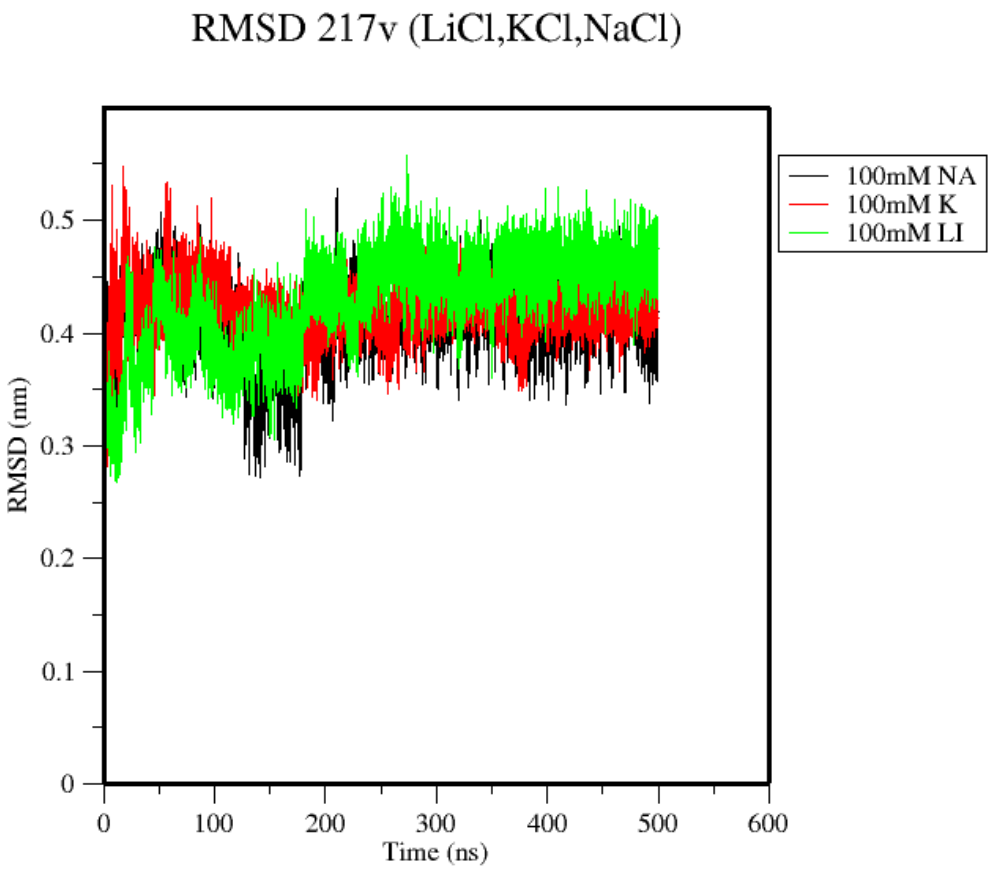

Figure 95. RMSD of 2L7V in 100 mM of KCl, NaCl and LiCl

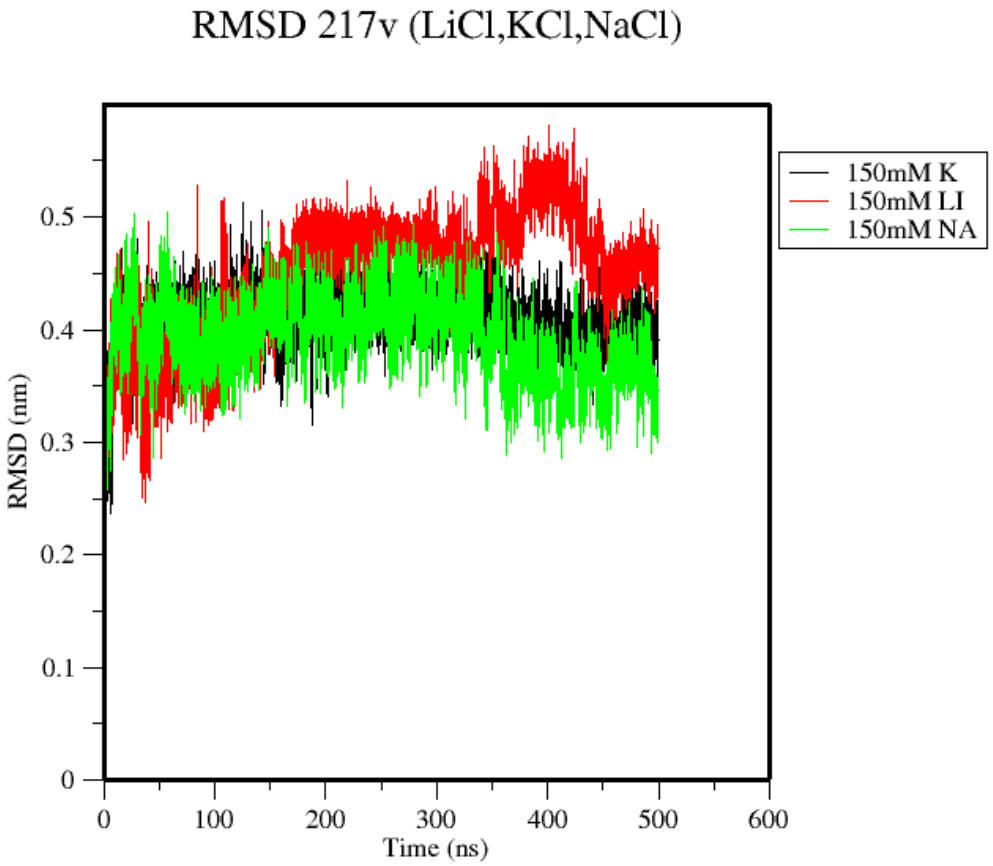

Figure 96. RMSD of 2L7V in 150 mM of KCl, NaCl and LiCl

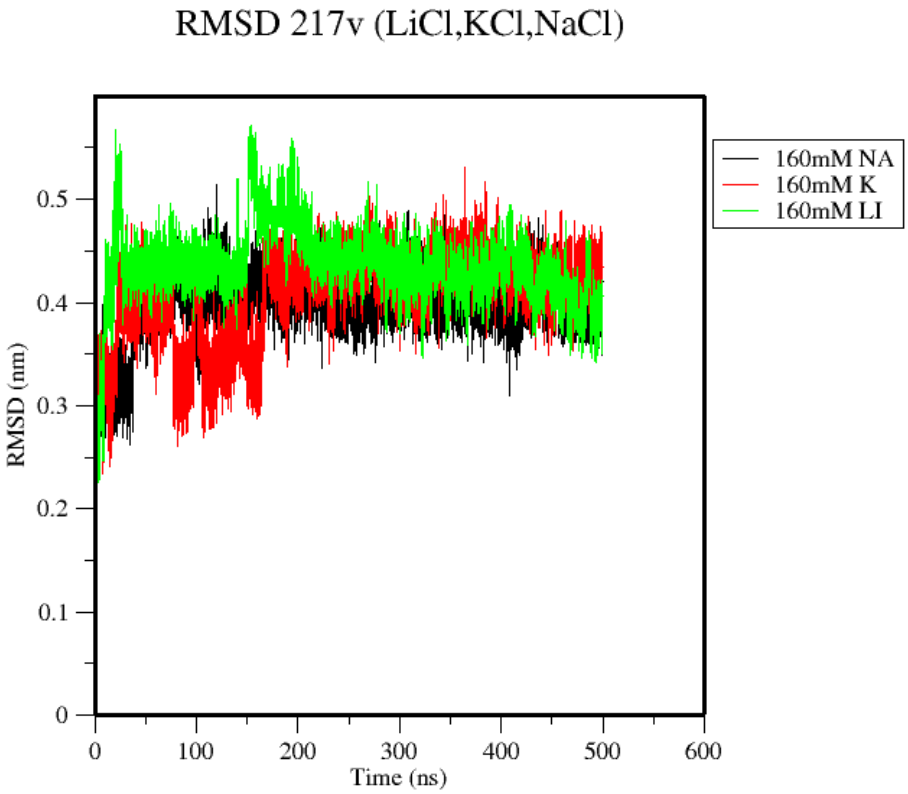

Figure 97. RMSD of 2L7V in 160 mM of KCl, NaCl and LiCl

We evaluated the stability of the structure by analyzing the root mean square deviation of the atoms of the simulated structure from the original structure. We plotted RMSD against time. The RMSD of KCl concentration remained relatively steady at 50mM with average RMSD of 0.423nm Figure 91, structure maintained average of 0.41nm in NaCl solution Figure 92. The highest RMSD was observed in figure 94 with lithium having RMSD of 0.426 nm. We further compared the effect of the three ions by plotting all equal concentration against time. For shows the initial rise of RMSD and its relaxing or convergence point lithium was observed to maintain the highest RMSD followed by potassium and the least RMSD was observed in NaCl ion concentration.

To further validate this, we analyzed the fluctuation of the structure by analyzing the RMSF.

5.2.2 Root Mean Square Fluctuation Analysis

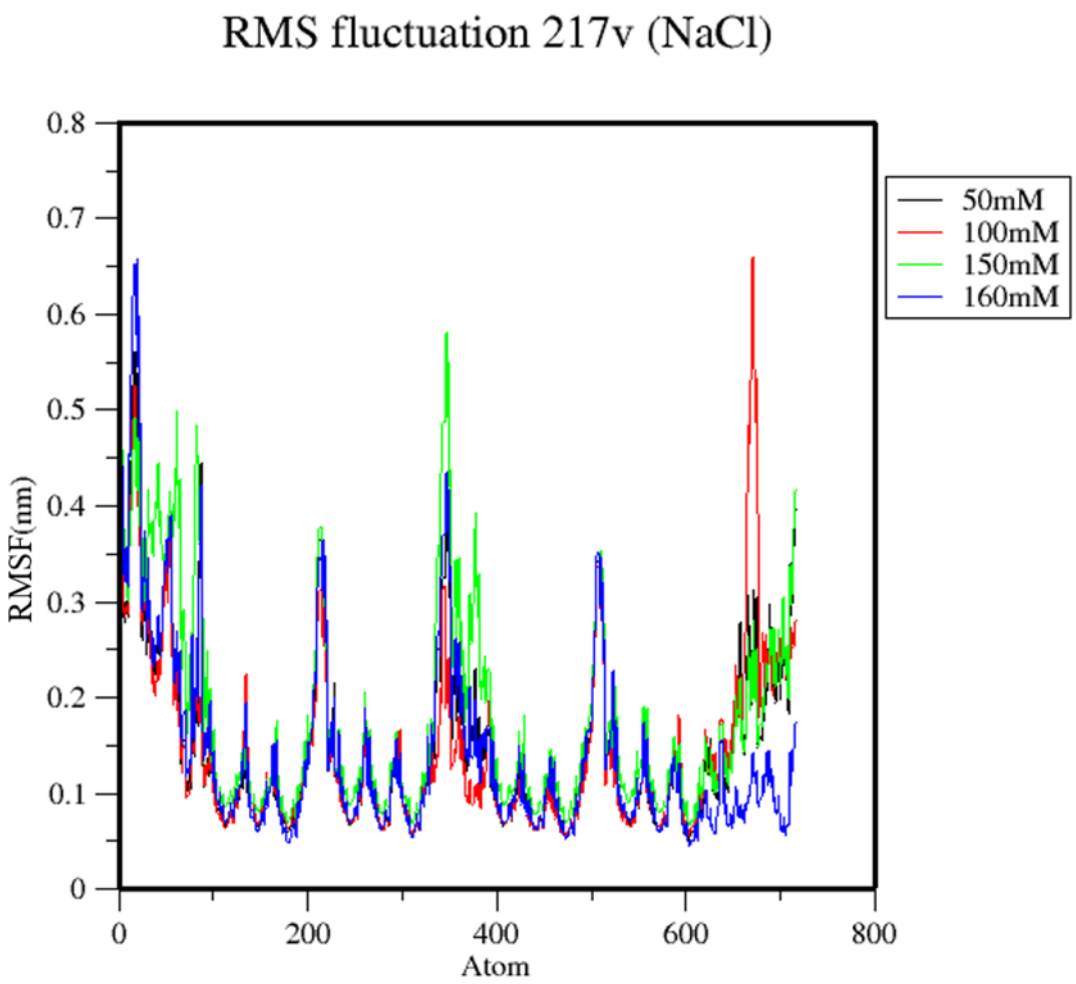

Figure 98. RMSF of 2L7V in 50,100,150 and 160 mM of NaCl

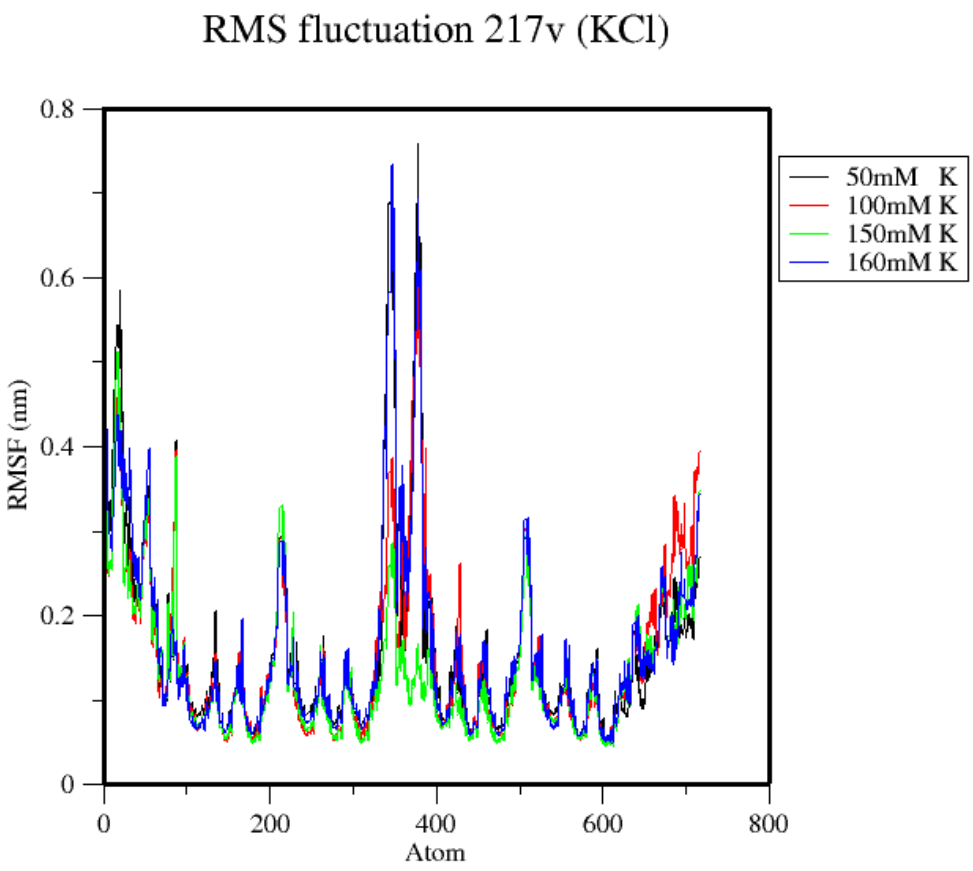

Figure 99. RMSF of 2L7V in 50,100,150 and 160 mM of KCl

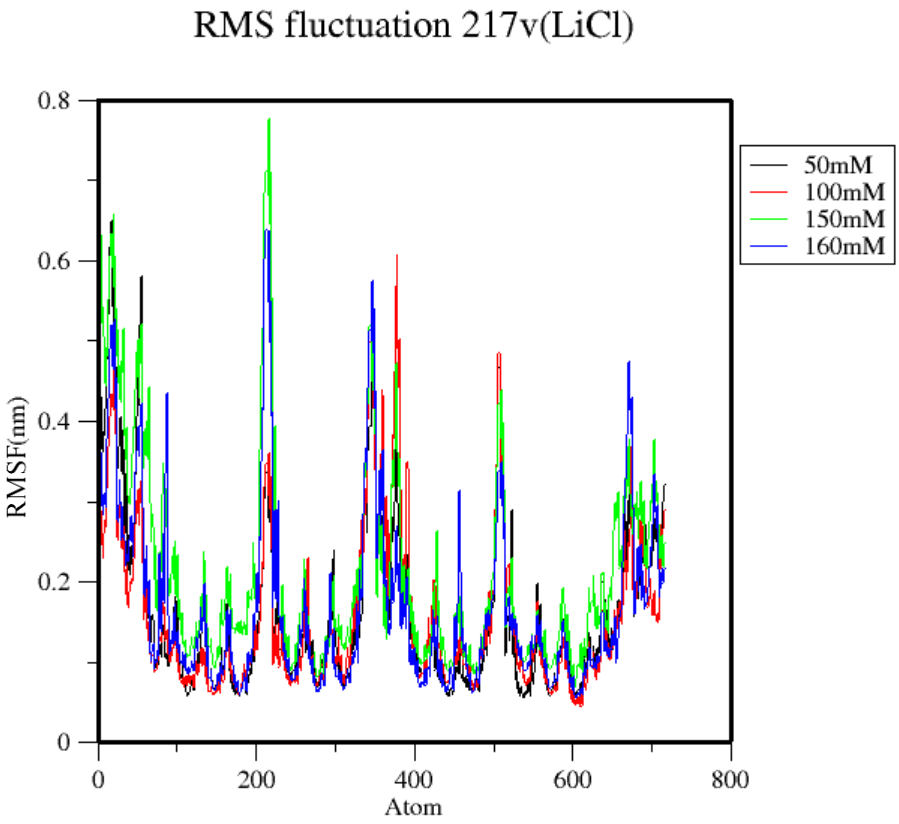

Figure 100. RMSF of 2L7V in 50,100,150 and 160 mM of LiCl

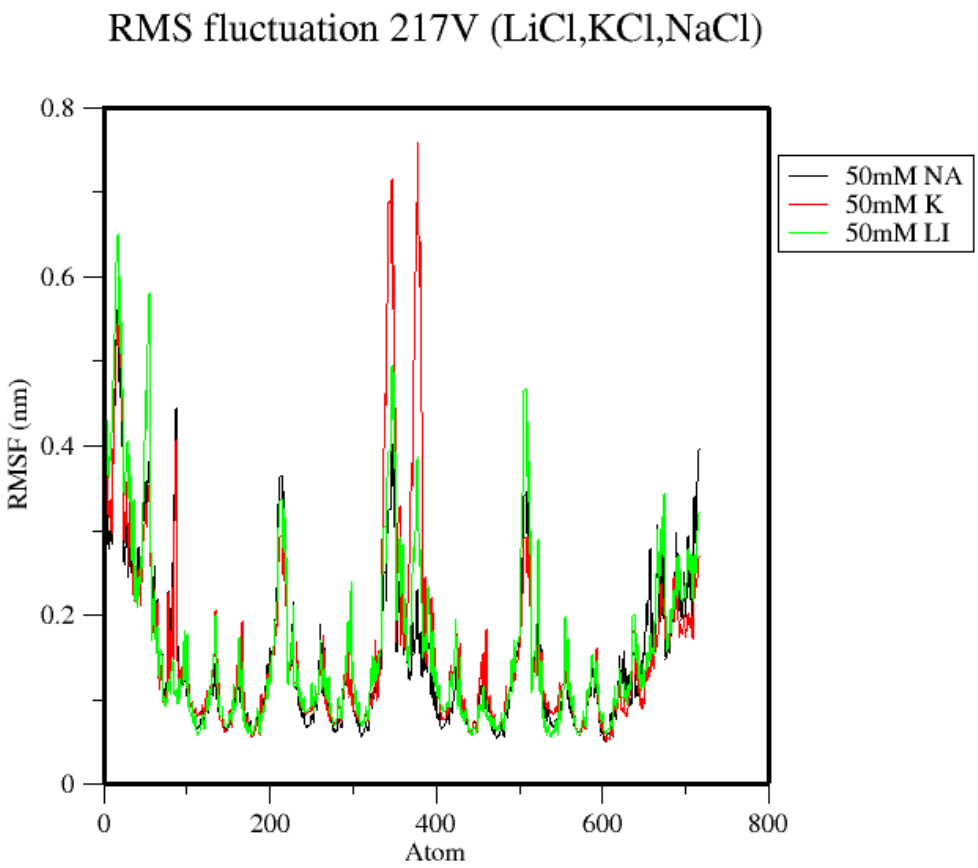

Figure 101. RMSF of 2L7V in 50 mM of NaCl, KCl, and LiCl

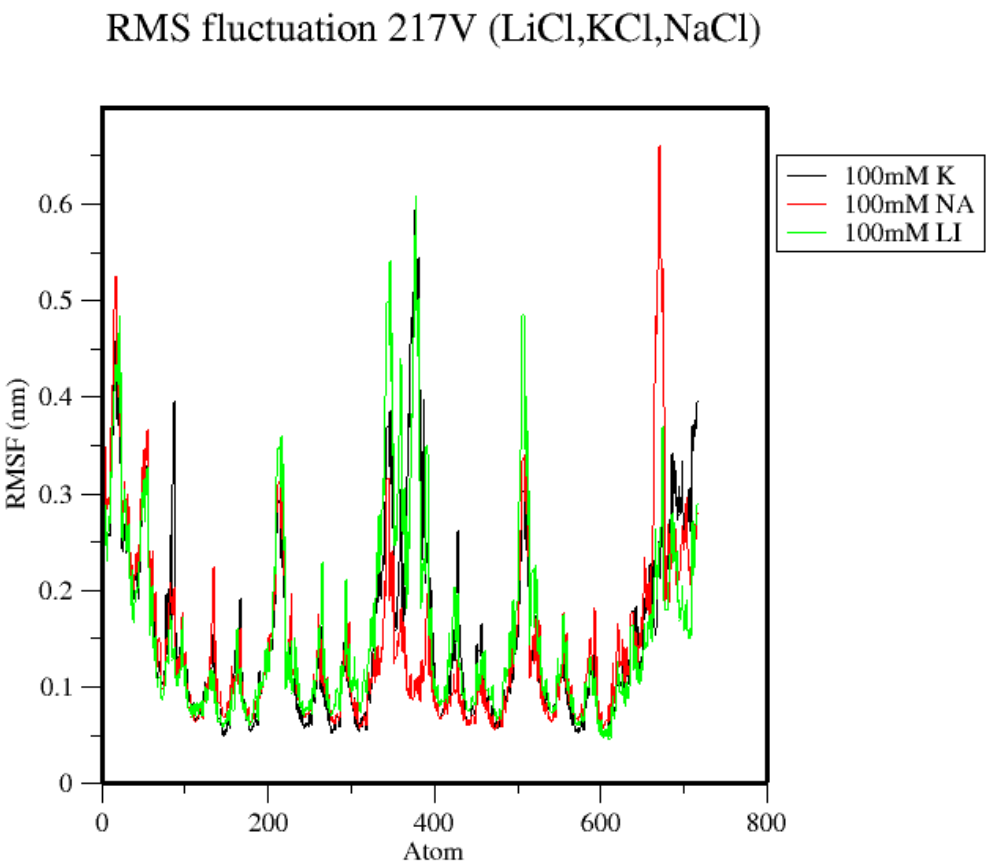

Figure 102. RMSF of 2L7V in 100 mM of NaCl, KCl, and LiCl

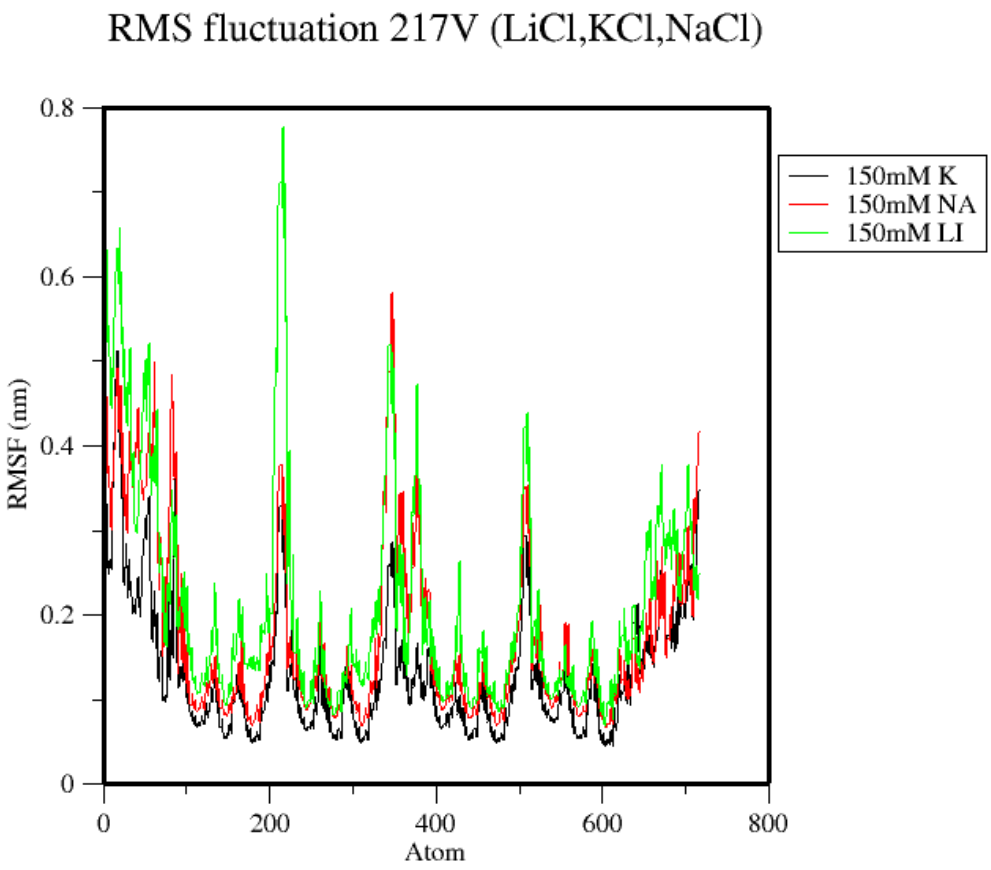

Figure 103. RMSF of 2L7V in 150 mM of NaCl, KCl, and LiCl

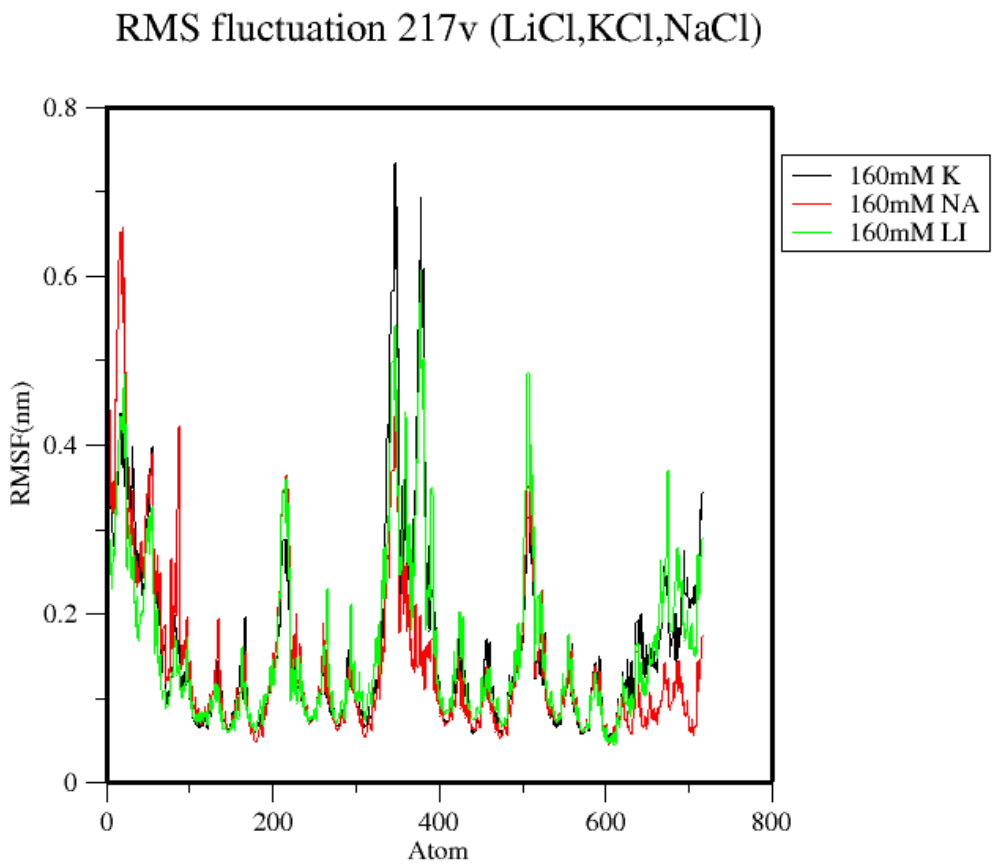

Figure 104. RMSF of 2L7V in 160 mM of NaCl, KCl, and LiCl

Quantities in (Figures98-104) above describes the effect of thermal fluctuation of residues on the DNA structure, local conformation and flexibility. We plotted fluctuation against residue to show the conformational change and will discourse our findings bellow. We observed a common symmetry between the 300 and the 400 atoms in all the ions and concentration for all the figures. In figure 98 the fluctuation of our structure in NaCl ion concentration was highest between100 and 150 mM. We observed increase in fluctuation with increase in concentration indicating potential structural change by

lithium ion on the structure figure100. We compared the three different ions and concentration (Figures101-103). The analysis of the three ions and concentration indicates that at lower concentration of 50 mM KCl ions caused the highest fluctuation followed by Li and Na ions. For concentration 100 mM Na ions seems to cause the most fluctuation followed by Li and K ions respectively. For 150 and 160mM concentration Li ion was seen to cause the most fluctuation followed K ion and Na ions respectively. This indicates lithium may cause the most structural change in the system. Subsequently we will look at the plot of RMSD after fit against concentration to see the potential structural change.

5.2.3 Structural Change

We Further compared the average structure of the trajectory after calculating the RMSD of the system with the initial structure used in the simulation. The conformational change where seen in the structure as shown in (Tables 10-12). Changes were observed in the structure after the simulation.

Table 11. RMSD of 2l7V for Different Concentration of Potassium Chloride Solution

Concentration(mM)	RMSD (nm)
50	0.430101
100	0.41768
150	0.393872
160	0.437788

Table 12. RMSD of 2L7V for Different Concentration of Sodium Chloride Solution

Concentration(mM)	RMSD (nm)
50	0.350951
100	0.415585
150	0.348849
160	0.423943

Table 13. RMSD of 2L7V for Different Concentration of Lithium Chloride Solution

Concentration(mM)	RMSD (nm)
50	0.478296
100	0.475279
150	0.471309
160	0.407768

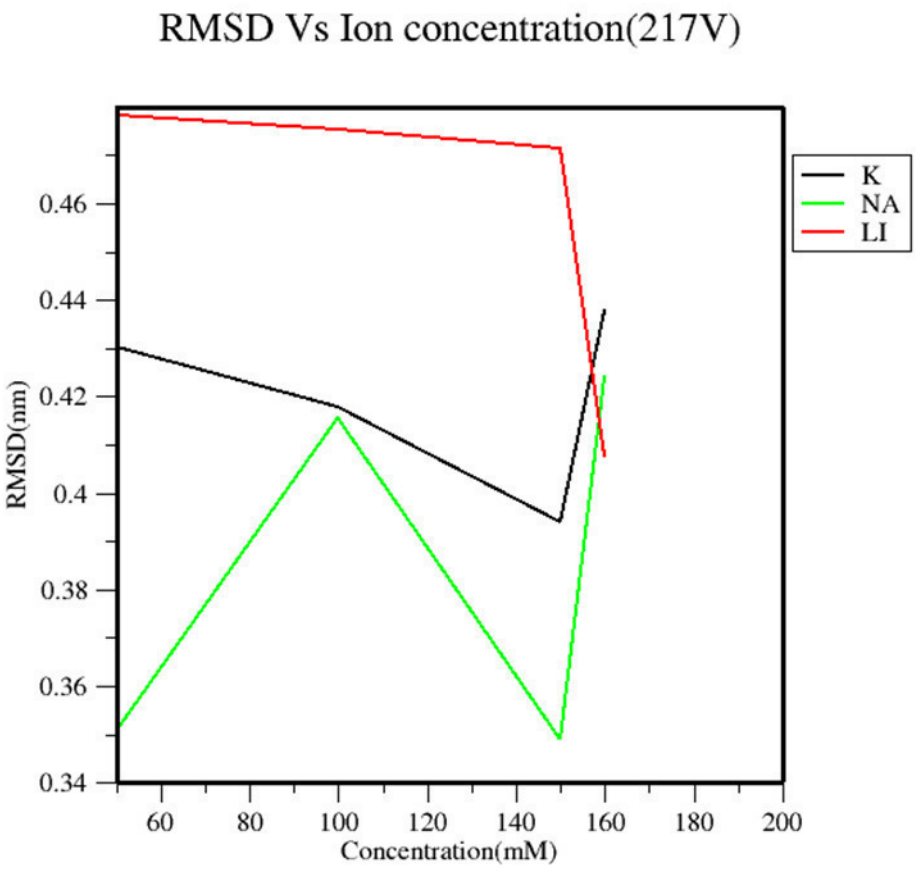

Figure 105. RMSD against Concentration of Ions for 2L7V in 50 mM,100 mM,150 mM and 160 mM of KCl, NaCl and LiCl

From figure 105 above shows the effect of different ion concentration on the structure if 1xav G-quadruplex DNA reveals lithium caused the most structural change of about 0.49 nm, followed by potassium with structural change of 0.43 nm. sodium had the least structural change of about 0.35 nm.

5.3 Summary

In summary, this chapter presented a detailed study of 2L7V G-quadruplex DNA structure and dynamics in different alkali ion alkali ion concentration employing GROMACS molecular dynamics simulations. MD analysis studies involved the solvation of 2L7V G-quadruplex DNA the by using appropriate number of TIP3P water molecules. The solvated structure was equilibrated in NVT and NPT ensembles for 200ps in each case. The final equilibrated structure was used as the starting structure for the production, dynamics simulation. The dynamics analysis was conducted for 500ns with simulation data collected at every 100ps for post-production analysis. Sodium chloride ion concentration was high between 100mM and 50mM with 50mM concentration causing initial structural change. lithium chloride, potential structural change was observed with increase in concentration both at high and low concentration.

Potassium chloride ion was found to have higher structural change compared to sodium. The structure of 2L7V seem to be more favored by sodium chloride solution with a structural change of about 0.35 nm from the original structure followed by potassium ion which had a structural change of about 0.425nm from the original structure. The most structural change was observed with lithium ions with structural change of 0.469 nm from the original structure

CHAPTER VI

CONCLUSION

6.1 Summary of Work

Differential effect of alkali metal ions on the structure and stability of DNA G-quadruplexes with potential role in the regulation of gene expression have been investigated with computational techniques. G-quadruplexes are noncanonical secondary structures formed both in DNA and RNA sequences containing consecutive runs of guanines. Literatures have shown that Quadruplexes (QPX), can form as an alternate structure in competition with the famous double helix and cells use fluxes of ions, such as sodium and potassium currents in the electrical signals of neurons, to stabilize or destabilize QPX in the nuclei of cells and can play potential role in turning on and off the expression of genes. Previous studies from our lab was conducted using a G-QPX containing region of the promoter of the choline acetyltransferase (ChAT) gene, in driving expression of Green Fluorescent Protein as a "reporter gene". The transcription factors used to activate the ChAT gene promoter significantly up-regulate the GFP expression Aconitine, a Na+ -channel opening drug. The presence of active transcription factor, led to significant increase in GFP, supporting the idea that monovalent cations have impact on stability / instability of G-QPX (in this case, destabilizing the G-QPX) TMPYP4. TMPYP4 a drug which stabilizes G-QPX formation, was found to significantly decrease GFP expression, suggesting the G-QPX formation is an OFF state

for gene expression in this system. Our study further validated earlier results in the lab by investigating the differential effect of alkali metal ions on the structure and stability of DNA G-quadruplexes with potential role in the regulation of gene expression.
In this dissertation, we explored structure and dynamics of G-quadruplex in different alkali ion (Li, K+ and Na+) concentration. Our interest on Li ion was because of its therapeutic application in the treatment of bipolar disorder. K+ and Na+ ions was selected because of their role in human cell activity.

We conducted an extensive computational molecular dynamics simulation study using G-Quadruplex element in the human c-MYC and C-MIK promoter region of the human telomere. The QPX molecules were downloaded from the RCSB website (4fxm, 1xav,2l7v and 2hri). In each case, we performed detailed molecular dynamic simulation and analysis on the thermodynamic and structural quantities of above structures. Simulation analysis studies were performed using GROMACS molecular dynamic simulation code. We performed production simulation analysis based on time step of 1fs, pressure of one bar, temperature of 310K for 500 ns of total simulation time. We performed the simulation process for the G-quadruplex structures in the chloride solution of each ion. We also carried out some thermodynamic, visual and structural analysis by calculating the total energy, root mean square deviation (RMSD), root mean square fluctuation (RMSF) and radius of gyration. Visual examination of the structures raveled structural changes with lithium remaining intact in the ion channel at the center of the structure, K+ and Na+ lost one ion each from the ion channel in the center of the structure. Results from thermodynamic analysis also indicate that in 4fxm qpx molecule

Na+ ion has the potential to destabilize the structure up to 0.56nm from the original structure at lower concentration of 50mM further increase in concentration tends to stabilize the structure which remained constant between 100 and 150 mM. Li+ ion had a destabilization of about 0.47 nm from the original structure for lower concentration of 50 mM and remained constant to 100 mM before further destabilization to 150 mM. We noticed a sharp drop toward stabilization at 160mM concentration. K+ had the least change in structure of 0.45 nm at lower concentration of 50mM but increased with increase in concentration with a drop towards stabilization at concentration above 100mM. The result in the simulation of 4fxm indicate K+ ion favored the structural stability of 4fxm qpx molecule more than Li+ and Na+ respectively but lithium ion has the potential to keep ions in the structure of the channel more than the Li+ and Na+ ion. 1XAV structure was also analyzed both visually and thermodynamically quantities calculated. Visual analysis indicates structural changes with all the channel's counter ions remaining intact.

At lower ion concentration of 50 mM Li+ ion caused the most structural change from the original structure of about 0.44 nm. Further increase in concentration indicated stabilization with a reversal at concentration above 150 mM. Na+ ion caused a change of 0.33 nm from the original structure and gradually increased with increase in concentration and a reversal at concentration above 150 mM. K+ ion produced a 0.31 nm change in structure from the original structure also with a gradual increase with increase in concentration. The structure stabilized with no further change at concentration above 150 mM. The result indicates K+ favored 1xav qpx structurally compared to Na+ and Li+

respectively. 2L7V qpx structure was also studied, the visual examination of the structure after simulation revealed structural changes with the center channel ions remaining intact. Li+ ion produced the most structural change of about 0.49 nm and remained constant with increase in concentration. A sharp move towards stabilization was observed at 150 mM concentrations. K+ ion showed a structural change of about 0.43nm at lower concentration of 50 mM further increase in concentration showed stabilization with reversal at 150 mM. Na+ destabilized the structure with about 0.35 nm at lower concentration and increased with increase in concentration, a reversal at 100 mM and another reversal at 150 mM. The result indicates Na+ ion favored the structural stability 2L7V qpx more than Na+ and Li+ ion respectively. Inconclusion we observed that lithium has the most destabilizing effect on the used qpx molecules and has the potential to act as stabilizer to 4fxm qpx: as observed in its stability neutrality when Na+ and K+ counter ions was lost form the center counter ion of 4fxm. The qpx used was favored most by K+ which is also consistent with literatures. We also observed different qpx structure respond to different ion concentration differently and this could be attributed to different structural conformations. This could also be an answer to the divide in literatures on the role of lithium in stability of qpx structures. Some are of destabilizing effect while others are of neutral and stabilizing effect. The results also go to validate the initial result from the lab on ChAT Gene.

6.2 Future Work

In the present work we have simulated and analyzed the dynamics differential effect of alkali metal ions on the structure and stability of DNA G-quadruplexes with potential role in the regulation of gene expression. With the knowledge acquired from this work, future studies can explore its potential use as molecular switch for therapeutic purpose employing MD analysis methodology. Along these lines, one can study the response mechanisms of qpx structural response to ion concentration. Despite the use of lithium in treating bipolar effect with a potential link to qpx, understanding its mechanism as a potential biomolecular switch could lead to potential cure of some auto immune disease.

REFRENCES

1. Ou, T.-M., et al., *Stabilization of G-quadruplex DNA and down-regulation of oncogene c-myc by quindoline derivatives.* Journal of medicinal chemistry, 2007. **50**(7): p. 1465-1474.
2. Qin, Y. and L.H. Hurley, *Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions.* Biochimie, 2008. **90**(8): p. 1149-1171.
3. Vorlíčková, M., et al., *Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes.* Chirality, 2012. **24**(9): p. 691-698.
4. Bacolla, A. and R.D. Wells, *Non-B DNA conformations as determinants of mutagenesis and human disease.* Molecular carcinogenesis, 2009. **48**(4): p. 273-285.
5. Sen, D. and W. Gilbert, *A sodium-potassium switch in the formation of four-stranded G4-DNA.* Nature, 1990. **344**(6265): p. 410.
6. Hud, N.V. and J. Plavec, *The role of cations in determining quadruplex structure and stability.* Quadruplex nucleic acids, 2006: p. 100-130.
7. Gabelica, V., et al., *Multiple and cooperative binding of fluorescence light-up probe thioflavin T with human telomere DNA G-quadruplex.* Biochemistry, 2013. **52**(33): p. 5620-5628.
8. Sen, D. and W. Gilbert, *Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis.* Nature, 1988. **334**(6180): p. 364.
9. Weisman-Shomer, P. and M. Fry, *QUAD, a protein from hepatocyte chromatin that binds selectively to guanine-rich quadruplex DNA.* Journal of Biological Chemistry, 1993. **268**(5): p. 3306-3312.
10. Chen, Y. and D. Yang, *Sequence, stability, and structure of G-Quadruplexes and their interactions with drugs.* Current protocols in nucleic acid chemistry, 2012: p. 17.5. 1-17.5. 17.
11. Bidzinska, J., et al., *G-quadruplex structures in the human genome as novel therapeutic targets.* Molecules, 2013. **18**(10): p. 12368-12395.
12. Hu, D., J. Ren, and X. Qu, *Metal-mediated fabrication of new functional G-quartet-based supramolecular nanostructure and potential application as controlled drug release system.* Chemical Science, 2011. **2**(7): p. 1356-1361.
13. Fadrná, E., et al., *Molecular dynamics simulations of guanine quadruplex loops: advances and force field limitations.* Biophysical journal, 2004. **87**(1): p. 227-242.
14. Štefl, R., et al., *Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of the substates.* Biophysical journal, 2003. **85**(3): p. 1787-1804.

15. Hardin, C.C., et al., *Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d (CGCG3GCG).* Biochemistry, 1992. **31**(3): p. 833-841.
16. Venczel, E.A. and D. Sen, *Parallel and antiparallel G-DNA structures from a complex telomeric sequence.* Biochemistry, 1993. **32**(24): p. 6220-6228.
17. Burge, S., et al., *Quadruplex DNA: sequence, topology and structure.* Nucleic acids research, 2006. **34**(19): p. 5402-5415.
18. Huppert, J.L. and S. Balasubramanian, *G-quadruplexes in promoters throughout the human genome.* Nucleic acids research, 2006. **35**(2): p. 406-413.
19. Wanrooij, P.H., et al., *G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation.* Proceedings of the National Academy of Sciences, 2010. **107**(37): p. 16072-16077.
20. Rachwal, P.A., T. Brown, and K.R. Fox, *Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes.* Biochemistry, 2007. **46**(11): p. 3036-3044.
21. Bhattacharyya, D., G. Mirihana Arachchilage, and S. Basu, *Metal cations in G-quadruplex folding and stability.* Frontiers in chemistry, 2016. **4**: p. 38.
22. Miyoshi, D., et al., *Effect of divalent cations on antiparallel G-quartet structure of d (G4T4G4).* FEBS letters, 2001. **496**(2-3): p. 128-133.
23. Miyoshi, D., A. Nakao, and N. Sugimoto, *Structural transition from antiparallel to parallel G-quadruplex of d (G 4 T 4 G 4) induced by Ca 2+.* Nucleic acids research, 2003. **31**(4): p. 1156-1163.
24. Kankia, B.I. and L.A. Marky, *Folding of the thrombin aptamer into a G-quadruplex with Sr2+: stability, heat, and hydration.* Journal of the American Chemical Society, 2001. **123**(44): p. 10799-10804.
25. Vairamani, M. and M.L. Gross, *G-quadruplex formation of thrombin-binding aptamer detected by electrospray ionization mass spectrometry.* Journal of the American Chemical Society, 2003. **125**(1): p. 42-43.
26. Guo, L., et al., *A G-quadruplex based label-free fluorescent biosensor for lead ion.* Biosensors and Bioelectronics, 2012. **35**(1): p. 123-127.
27. Sannohe, Y. and H. Sugiyama, *Overview of formation of G-quadruplex structures.* Current protocols in nucleic acid chemistry, 2010: p. 17.2. 1-17.2. 17.
28. Neidle, S., *Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer.* The FEBS journal, 2010. **277**(5): p. 1118-1125.
29. Tahara, H., et al., *G-Quadruplex stabilization by telomestatin induces TRF2 protein dissociation from telomeres and anaphase bridge formation accompanied by loss of the 3′ telomeric overhang in cancer cells.* Oncogene, 2006. **25**(13): p. 1955.
30. Hud, N.V., et al., *The selectivity for K+ versus Na+ in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by 1H NMR.* Biochemistry, 1996. **35**(48): p. 15383-15390.

31. Kettani, A., et al., *Solution structure of a Na cation stabilized DNA quadruplex containing G· G· G· G and G· C· G· C tetrads formed by GGGC repeats observed in adeno-associated viral DNA1.* Journal of molecular biology, 1998. **282**(3): p. 619-636.
32. Hud, N.V., et al., *Binding sites and dynamics of ammonium ions in a telomere repeat DNA quadruplex1.* Journal of molecular biology, 1999. **285**(1): p. 233-243.
33. Martino, L., et al., *Structural and thermodynamic studies of the interaction of distamycin A with the parallel quadruplex structure [d (TGGGGT)] 4.* Journal of the American Chemical Society, 2007. **129**(51): p. 16048-16056.
34. Müller, S., *Quadruplex Nucleic Acids. Edited by Stephen Neidle and Shankar Balasubramanian*. 2008, Wiley Online Library.
35. Huppert, J.L. and S. Balasubramanian, *Prevalence of quadruplexes in the human genome.* Nucleic acids research, 2005. **33**(9): p. 2908-2916.
36. Rhodes, D. and H.J. Lipps, *G-quadruplexes and their regulatory roles in biology.* Nucleic acids research, 2015. **43**(18): p. 8627-8637.
37. König, S.L., A.C. Evans, and J.L. Huppert, *Seven essential questions on G-quadruplexes.* Biomolecular concepts, 2010. **1**(2): p. 197-213.
38. Frees, S., et al., *QGRS-Conserve: a computational method for discovering evolutionarily conserved G-quadruplex motifs.* Human genomics, 2014. **8**(1): p. 8.
39. Maizels, N. and L.T. Gray, *The G4 genome.* PLoS genetics, 2013. **9**(4): p. e1003468.
40. Cayrou, C., et al., *Genome-scale identification of active DNA replication origins.* Methods, 2012. **57**(2): p. 158-164.
41. Bugaut, A. and S. Balasubramanian, *5′-UTR RNA G-quadruplexes: translation regulation and targeting.* Nucleic acids research, 2012. **40**(11): p. 4727-4741.
42. Beaudoin, J.-D. and J.-P. Perreault, *Exploring mRNA 3′-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening.* Nucleic acids research, 2013. **41**(11): p. 5898-5911.
43. Xu, Y., et al., *Telomeric repeat-containing RNA structure in living cells.* Proceedings of the National Academy of Sciences, 2010. **107**(33): p. 14579-14584.
44. Brosh Jr, R.M., *DNA helicases involved in DNA repair and their roles in cancer.* Nature Reviews Cancer, 2013. **13**(8): p. 542.
45. Lipps, H.J. and D. Rhodes, *G-quadruplex structures: in vivo evidence and function.* Trends in cell biology, 2009. **19**(8): p. 414-422.
46. Huber, M.D., et al., *A conserved G4 DNA binding domain in RecQ family helicases.* Journal of molecular biology, 2006. **358**(4): p. 1071-1080.
47. Zhang, A.Y. and S. Balasubramanian, *The kinetics and folding pathways of intramolecular G-quadruplex nucleic acids.* Journal of the American Chemical Society, 2012. **134**(46): p. 19297-19308.
48. Chang, C.-C., et al., *Detection of quadruplex DNA structures in human telomeres by a fluorescent carbazole derivative.* Analytical chemistry, 2004. **76**(15): p. 4490-4494.

49. Fletcher, T.M., et al., *Effect of DNA secondary structure on human telomerase activity.* Biochemistry, 1998. **37**(16): p. 5536-5541.
50. Li, Q.-J., et al., *Characterization of the intramolecular G-quadruplex promoting activity of Est1.* FEBS letters, 2013. **587**(6): p. 659-665.
51. Biffi, G., D. Tannahill, and S. Balasubramanian, *An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2.* Journal of the American Chemical Society, 2012. **134**(29): p. 11974-11976.
52. Han, H., R.J. Bennett, and L.H. Hurley, *Inhibition of unwinding of G-quadruplex structures by Sgs1 helicase in the presence of N, N '-bis [2-(1-piperidino) ethyl]-3, 4, 9, 10-perylenetetracarboxylic diimide, a G-quadruplex-interactive ligand.* Biochemistry, 2000. **39**(31): p. 9311-9316.
53. Paeschke, K., et al., *Cell cycle-dependent regulation of telomere tethering in the nucleus.* Chromosome research, 2008. **16**(5): p. 721-728.
54. Larson, E.D., et al., *MutSα binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions.* Current biology, 2005. **15**(5): p. 470-474.
55. Chiarella, S., et al., *Nucleophosmin mutations alter its nucleolar localization by impairing G-quadruplex binding at ribosomal DNA.* Nucleic acids research, 2013. **41**(5): p. 3228-3239.
56. Dempsey, L.A., et al., *G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, A role for GG pairing in immunoglobulin switch recombination.* Journal of Biological Chemistry, 1999. **274**(2): p. 1066-1071.
57. Haeusler, A.R., et al., *C9orf72 nucleotide repeat structures initiate molecular cascades of disease.* Nature, 2014. **507**(7491): p. 195.
58. Lin, C. and D. Yang, *Human Telomeric G-Quadruplex Structures and G-Quadruplex-Interactive Compounds.* Telomeres and Telomerase: Methods and Protocols, 2017: p. 171-196.
59. Bodnar, A.G., et al., *Extension of life-span by introduction of telomerase into normal human cells.* science, 1998. **279**(5349): p. 349-352.
60. Sun, D., et al. *Modulation of the catalytic activity and processivity of human telomerase from HeLa cell*. in *Proc Am Assoc Cancer Res*. 1998.
61. Neidle, S. and G. Parkinson, *Telomere maintenance as a target for anticancer drug discovery.* Nature Reviews Drug Discovery, 2002. **1**(5): p. 383.
62. Paeschke, K., et al., *Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo.* Nature Structural and Molecular Biology, 2005. **12**(10): p. 847.
63. Granotier, C., et al., *Preferential binding of a G-quadruplex ligand to human chromosome ends.* Nucleic acids research, 2005. **33**(13): p. 4182-4190.
64. Mathad, R.I. and D. Yang, *G-quadruplex structures and G-quadruplex-interactive compounds*, in *Telomeres and Telomerase*. 2011, Springer. p. 77-96.
65. Zahler, A.M., et al., *Inhibition of telomerase by G-quartet DMA structures.* nature, 1991. **350**(6320): p. 718.

66. Dai, J., et al., *Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence.* Nucleic acids research, 2007. **35**(15): p. 4927-4940.
67. Ambrus, A., et al., *Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution.* Nucleic acids research, 2006. **34**(9): p. 2723-2735.
68. Di Antonio, M., R. Rodriguez, and S. Balasubramanian, *Experimental approaches to identify cellular G-quadruplex structures and functions.* Methods, 2012. **57**(1): p. 84-92.
69. Rizzo, A., E. Salvati, and A. Biroccio, *Methods of studying telomere damage induced by quadruplex-ligand complexes.* Methods, 2012. **57**(1): p. 93-99.
70. Yang, Q., et al., *Verification of specific G-quadruplex structure by using a novel cyanine dye supramolecular assembly: II. The binding characterization with specific intramolecular G-quadruplex and the recognizing mechanism.* Nucleic acids research, 2009. **38**(3): p. 1022-1033.
71. Yang, Q., et al., *Verification of specific G-quadruplex structure by using a novel cyanine dye supramolecular assembly: I. recognizing mixed G-quadruplex in human telomeres.* Chemical Communications, 2009(9): p. 1103-1105.
72. Biffi, G., et al., *Quantitative visualization of DNA G-quadruplex structures in human cells.* Nature chemistry, 2013. **5**(3): p. 182.
73. Henderson, A., et al., *Detection of G-quadruplex DNA in mammalian cells.* Nucleic acids research, 2013. **42**(2): p. 860-869.
74. Henderson, A., et al., *Detection of G-quadruplex DNA in mammalian cells.* Nucleic acids research, 2017. **45**(10): p. 6252-6252.
75. Piazza, A., et al., *Stimulation of gross chromosomal rearrangements by the human CEB1 and CEB25 minisatellites in Saccharomyces cerevisiae depends on G-quadruplexes or Cdc13.* PLoS genetics, 2012. **8**(11): p. e1003033.
76. Ding, H., et al., *Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein.* Cell, 2004. **117**(7): p. 873-886.
77. Lin, W., et al., *Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity.* The EMBO journal, 2013. **32**(10): p. 1425-1439.
78. Bochman, M.L., K. Paeschke, and V.A. Zakian, *DNA secondary structures: stability and function of G-quadruplex structures.* Nature Reviews Genetics, 2012. **13**(11): p. 770.
79. Crabbe, L., et al., *Defective telomere lagging strand synthesis in cells lacking WRN helicase activity.* Science, 2004. **306**(5703): p. 1951-1953.
80. Karlseder, J., et al., *Telomere Lagging Strand Synthesis Dysfunction in Cells Lacking WRN Helicase Activity.* Klinische Neurophysiologie, 2004. **35**(03): p. 126.
81. Opresko, P.L., et al., *Telomere-binding protein TRF2 binds to and stimulates the Werner and Bloom syndrome helicases.* Journal of Biological Chemistry, 2002. **277**(43): p. 41110-41119.

82. Opresko, P.L., et al., *POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates.* Journal of Biological Chemistry, 2005. **280**(37): p. 32069-32080.
83. Vannier, J.-B., et al., *RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity.* Cell, 2012. **149**(4): p. 795-806.
84. Eddy, J. and N. Maizels, *Gene function correlates with potential for G4 DNA formation in the human genome.* Nucleic acids research, 2006. **34**(14): p. 3887-3896.
85. Simonsson, T., M. Kubista, and P. Pecinka, *DNA tetraplex formation in the control region of c-myc.* Nucleic acids research, 1998. **26**(5): p. 1167-1172.
86. Siddiqui-Jain, A., et al., *Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription.* Proceedings of the National Academy of Sciences, 2002. **99**(18): p. 11593-11598.
87. Huppert, J.L., et al., *G-quadruplexes: the beginning and end of UTRs.* Nucleic acids research, 2008. **36**(19): p. 6260-6268.
88. Fernando, H., et al., *Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression.* Nucleic acids research, 2009. **37**(20): p. 6716-6722.
89. Darnell, J.C., et al., *Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function.* Cell, 2001. **107**(4): p. 489-499.
90. Brown, V., et al., *Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome.* Cell, 2001. **107**(4): p. 477-487.
91. Wolfe, A.L., et al., *RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer.* Nature, 2014. **513**(7516): p. 65.
92. Wendel, H., et al., *558 RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer.* European Journal of Cancer, 2014. **50**: p. 181.
93. Kumar, S., et al., *The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method.* Journal of computational chemistry, 1992. **13**(8): p. 1011-1021.
94. Van Gunsteren, W. and H. Berendsen, *Algorithms for macromolecular dynamics and constraint dynamics.* Molecular Physics, 1977. **34**(5): p. 1311-1327.
95. Liberatore, E., R. Meli, and U. Rothlisberger, *A Versatile Multiple Time Step Scheme for Efficient ab Initio Molecular Dynamics Simulations.* Journal of chemical theory and computation, 2018.
96. Adams, D.J., E.M. Adams, and G.J. Hills, *The computer simulation of polar liquids.* Molecular Physics, 1979. **38**(2): p. 387-400.
97. Bekker, H., et al., *An efficient, box shape independent non-bonded force and virial algorithm for molecular dynamics.* Molecular Simulation, 1995. **14**(3): p. 137-151.
98. Verlet, L., *Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules.* Physical review, 1967. **159**(1): p. 98.

99. Berendsen, H.J. and W.F. Van Gunsteren, *Practical algorithms for dynamic simulations.* Molecular-dynamics simulation of statistical-mechanical systems, 1986: p. 43-65.
100. Swope, W.C., et al., *A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters.* The Journal of Chemical Physics, 1982. **76**(1): p. 637-649.
101. Berendsen, H.J., et al., *Molecular dynamics with coupling to an external bath.* The Journal of chemical physics, 1984. **81**(8): p. 3684-3690.
102. Andersen, H.C., *Molecular dynamics simulations at constant pressure and/or temperature.* The Journal of chemical physics, 1980. **72**(4): p. 2384-2393.
103. Hoover, W.G., *Canonical dynamics: equilibrium phase-space distributions.* Physical review A, 1985. **31**(3): p. 1695.
104. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling.* The Journal of chemical physics, 2007. **126**(1): p. 014101.
105. Rühle, V., *Pressure coupling/barostats.* Journal Club, 2008.
106. Dirac, P.A.M., *Generalized hamiltonian dynamics.* Can. J. Math, 1950. **2**(2): p. 129-148.
107. Müller-Plathe, F., *YASP: A molecular simulation package.* Computer physics communications, 1993. **78**(1-2): p. 77-94.
108. Comba, P. and T.W. Hambley, *Frontmatter*. 2007: Wiley Online Library.
109. Wang, Y. and D.J. Patel, *Solution structure of a parallel-stranded G-quadruplex DNA.* Journal of molecular biology, 1993. **234**(4): p. 1171-1183.
110. Astuti, A. and A. Mutiara, *Performance Analysis on Molecular Dynamics Simulation of Protein Using GROMACS.* arXiv preprint arXiv:0912.0893, 2009.
111. Frenkel, D., et al., *Understanding molecular simulation.* Computers in Physics, 1997. **11**(4): p. 351-354.
112. Hansson, T., C. Oostenbrink, and W. van Gunsteren, *Molecular dynamics simulations.* Current opinion in structural biology, 2002. **12**(2): p. 190-196.
113. Hess, B., et al., *GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation.* Journal of chemical theory and computation, 2008. **4**(3): p. 435-447.
114. Hess, B., *P-LINCS: A parallel linear constraint solver for molecular simulation.* Journal of Chemical Theory and Computation, 2008. **4**(1): p. 116-122.
115. Bjelkmar, P., et al., *Implementation of the CHARMM force field in GROMACS: analysis of protein stability effects from correction maps, virtual interaction sites, and water models.* Journal of Chemical Theory and Computation, 2010. **6**(2): p. 459-466.
116. Lindahl, E., B. Hess, and D. Van Der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis.* Molecular modeling annual, 2001. **7**(8): p. 306-317.

117. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.* SoftwareX, 2015. **1**: p. 19-25.
118. Neumann, M. and O. Steinhauser, *The influence of boundary conditions used in machine simulations on the structure of polar systems.* Molecular Physics, 1980. **39**(2): p. 437-454.
119. Tironi, I.G., et al., *A generalized reaction field method for molecular dynamics simulations.* The Journal of chemical physics, 1995. **102**(13): p. 5451-5459.
120. Allen, M.P. and D.J. Tildesley, *Computer simulation in chemical physics*. Vol. 397. 2012: Springer Science & Business Media.
121. Scheek, R., et al., *Structure determination by NMR: The modeling of NMR parameters as ensemble averages*, in *Computational aspects of the study of biological macromolecules by nuclear magnetic resonance spectroscopy*. 1991, Springer. p. 209-217.
122. Neutze, R., et al., *Potential for biomolecular imaging with femtosecond X-ray pulses.* Nature, 2000. **406**(6797): p. 752.
123. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* Journal of computational chemistry, 2005. **26**(16): p. 1701-1718.
124. Jorgensen, W.L. and C. Jenson, *Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density.* Journal of Computational Chemistry, 1998. **19**(10): p. 1179-1186.
125. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* Journal of molecular graphics, 1996. **14**(1): p. 33-38.