MOLECULAR DYNAMICS SIMULATION STUDIES OF ION TRANSPORT ALONG G-QUADRUPLEX DNA CHANNELS

By

Parisa Akhshi

A thesis submitted to the Department of Chemistry

In conformity with the requirements for

the degree of Doctor of Philosophy

Queen's University

Kingston, Ontario, Canada

(January, 2013)

Abstract

Guanine-rich DNA and RNA sequences can fold, in the presence of alkali metal ions such as Na^+ and K^+, into G-quadruplex structures. These alkali metal ions are necessary for the stabilization of G-quadruplex structures. However, little is known about the ion dynamics in G-quadruplex structures. In this thesis, we used molecular dynamics (MD) simulations to study the energetics of ion transport in G-quadruplex DNA channels. In particular, we applied, for the first time, adaptive biasing force (ABF) and umbrella sampling (US) methods to obtain potential of mean force (PMF) profiles for Na^+, K^+, and NH_4^+ ion movement along $[d(TG_4T)]_4$ and $[d(G_3T_4G_4)]_2$ channels. We found that the ABF and US methods produce very similar PMF profiles, in qualitative agreement with the very limited experimental data in the literature.

We found that, within a G-quadruplex channel, K^+ and NH_4^+ ions experience significant energy barriers (13-17 kcal/mol) to cross a G-quartet, whereas the Na^+ movement encounters minimal resistance (5-7 kcal/mol). All ions are nearly fully dehydrated inside the channel but quickly become hydrated after exiting the channel. Our simulations suggested that the free energy landscapes for ion movement between the channel exit points and bulk solution are quite flat (ca. 2-4 kcal/mol) regardless of the loop topology in the region. We discovered that the directional symmetry of the ion movement within any G-quadruplex channel depends critically on both the DNA sequence and the folding of the G-quadrupelx structure. While the ion movement inside the $[d(TG_4T)]_4$ channel shows the same free energy barrier in either direction, the $[d(G_3T_4G_4)]_2$ channel exhibits a free energy difference of 3–4 kcal/mol for NH_4^+ ions exiting from the two ends. We hypothesized that the mode of base-stacking is the determining factor for the G-quartet stiffness and this stiffness then contributes to the free energy barrier for any ion to across it. This hypothesis appears to be consistent with all currently available experimental observations. When

a G-quadruplex channel contains multiple ions, we found that the ion-ion repulsion is an important factor that must be considered in order to have a complete understanding of the ion movement within G-quadruplex DNA channels.

Acknowledgements

I gratefully acknowledge my supervisor Prof. Gang Wu for his guidance throughout my graduate work. His insightful suggestions and direct contributions significantly simplified my work. I also thank Drs. Anne Petitjean, Donal H. Macartney, Nicholas Mosey, and Natalie Cann for being on my supervisory committee and many helpful suggestions. I would also like to thank all the group members in Dr. Gang Wu's lab especially Dr.s Jianfeng Zhu, Xianqi Kong, and Irene Kwan and Jennifer Adams for their help and friendship.

I thank my husband and my best friend, Mohammad, for all the love, scientific discussions and support he offered me in the past years. He has always been supporting and believing in me.

I am also grateful to my parents and siblings for their support, love, and encouragement throughout all these years.

Financial support from NSERC and Queen's University, and the use of computer resources on HPCVL and SHARCNET are also gratefully acknowledged.

Statement of Originality

I hereby certify that all of the work described within this thesis is the original work of the author. Any published (or unpublished) ideas and/or techniques from the work of others are fully acknowledged in accordance with the standard referencing practices.

Parisa Akhshi

December, 2012

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

DNA	Deoxynucleic acid
RNA	Ribonucleic acid
C	Cytosine
G	Guanine
A	Adenine
T	Thymine
U	Uracil
3'-GMP	Guanosine 3'-monophosphate
5'-GMP	Guanosine 5'-monophosphate
1D	1-dimensional
2D	2-dimensional
3D	3-dimensional
IR	Infra-red
NMR	Nuclear magnetic resonance
PDB	Protein data bank: http://www.rcsb.org/pdb
ppm	parts per million
MD	Molecular dynamics
ABF	Adaptive biasing force
US	Umbrella sampling
SMD	Steered molecular dynamics
FEP	Free energy perturbation

TI	Thermodynamic integration
PMF	Potential of mean force
WHAM	Weighted histogram analysis method
NAMD	Nanoscale molecular dynamics
VMD	Visual molecular dynamics
CHARMM	Chemistry at Harvard molecular mechanics
RMSD	Root mean square deviation
RC	Reaction coordinate
SHARCNET	Shared hierarchical academic research computing network
TIP3P	Transferable intermolecular potential 3 point
PME	Particle mesh Ewald
GBA	Glycosidic bond angle
FWHM	Full width at the half maximum
GUS	Guided umbrella sampling
MM-PBSA	Molecular mechanics Poisson-Bolzmann surface area

Chapter 1

Introduction

1.1 Nucleic acids

Nucleic acids are essential biological molecules that function in encoding, transmitting and expressing genetic information. There are two types of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA was first isolated in 1869 by Friedrich Miescher[1] from the nuclei of human blood cells. He called this weakly acidic substance "nuclein". Later on, he separated nuclein into protein and nucleic acid. RNA was discovered as a multiple ribose nucleic acid that plays important roles in protein synthesis.[2] Afterward, nucleic acids were identified to be a significant part of chromosomes. The basic unit of nucleic acids is nucleotide that contains a nucleobase (nitrogenous base), a pentose sugar (ribose or deoxyribose) and a phosphate group.[3] Nucleobases are a group of molecules that are nitrogen-based. The main nucleobases in DNA are cytosine (C), guanine (G), adenine (A) and thymine (T). In RNA, however, uracil (U) replaces T. A nucleotide without the phosphate group is called a nucleoside. Figure 1-1 shows the structures of common nucleobases and their corresponding nucleosides. There are two types of nitrogen-containing bases: pyrimidine and purine. Pyrimidine is a six-membered nitrogen-containing ring similar to pyridine but containing two nitrogen atoms at positions 1 and 3. Purine consists of a six-membered and a five-membered nitrogen-containing ring that are fused together. Among the nucleobases mentioned above, A and G belong to the purines, whereas C, T, and U are pyrimidines.

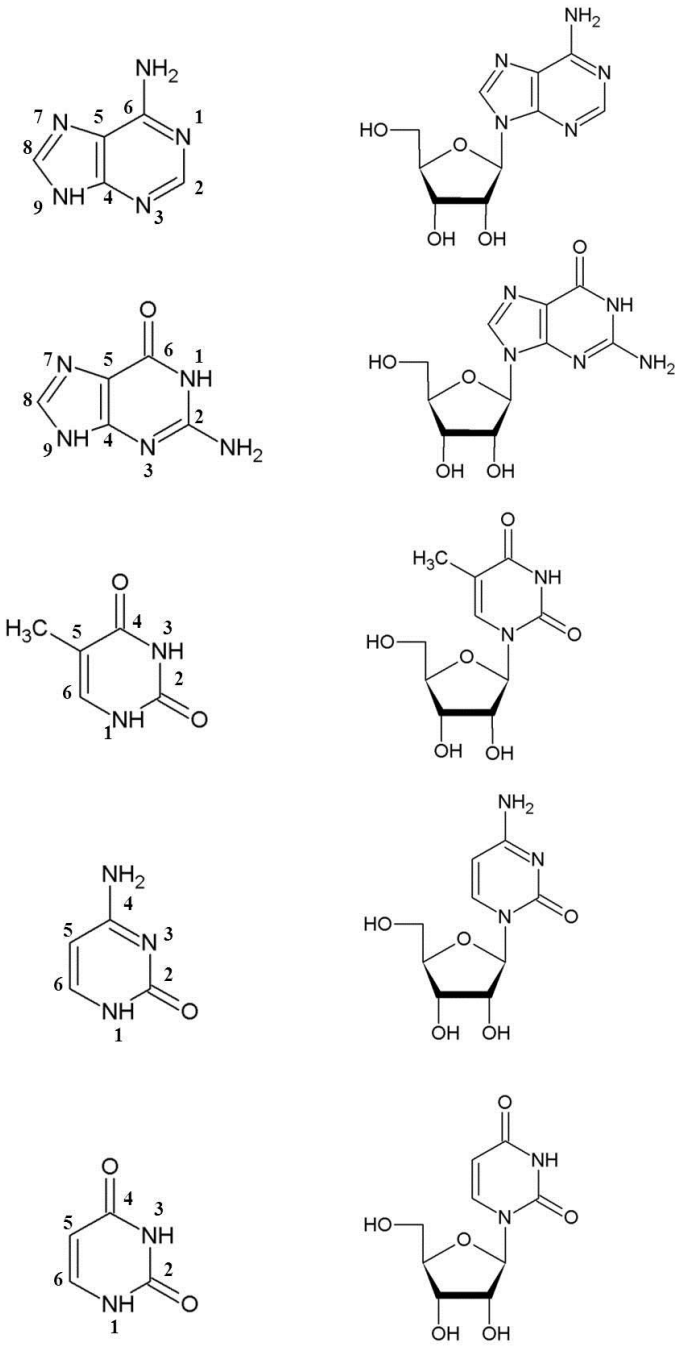

Figure 1-1. Structures of (left) nucleobases and (right) their corresponding nucleosides.

The locations of hydrogen bond donors, (-NH and -NH_2), and hydrogen bond acceptors, (C=O and -N=), in nucleobases suggest that several hydrogen bonds can form between purines and pyrimidines. In fact, these hydrogen bond connections make base pairs. In Watson-Crick base pairing, A and T are linked by two hydrogen bonds, whereas G forms base pair with C through three hydrogen bonds (see Figure 1-2).[3] In RNA, T is replaced by U. The DNA double helix is a twisting polymer of nucleic acids, bound together in a helical fashion by nucleotide base pairs. Other forces are also important to maintain the double-stranded structure such as intrastrand base stacking interactions. These interactions are strongest between G and C stacks. The bases are located inside of double helix and the entire structure is made without covalent bonds between the two strands. The stability of the double helix depends on the length and the GC content of the structure. Usually, long helices with a high GC content form stronger interactions and thus are more stable. The two strands of DNA are aligned in an anti-parallel manner. That is, one has the polarity from top to bottom (5'→3') and the other has a (3'→5') polarity from the same direction.[3]

Depending on the direction of helix twisting, DNA sequence, the level of hydration as well as the ion concentration, the DNA duplex could exist in three major conformations: A-DNA, B-DNA, and Z-DNA.[4] Figure 1-3 shows the structures of A, B, and Z-DNA. While A-DNA and B-DNA are right-handed helices, Z-DNA is found to form a left-handed helix. B-DNA is the most common form of DNA in living organisms. In B-DNA, the mean rotation per base pair is 3.5 nm compared to the 3.4 nm in A-DNA. One turn of A-DNA includes 11 base pairs, whereas B-DNA has around 10 to 10.6 base pairs per turn. B-DNA is longer and thinner than A-DNA. Z-DNA was originally found under high salt conditions. However, several reasons assist the existence of Z-DNA in cellular environments under low salt conditions. For example, in a recent study, Zhou et al.[5] reported that the ruthenium complex $[Ru(dip)_2dppz]^{2+}$ can induce the transition

from B-DNA to Z-DNA of various DNA sequences. In Z-DNA, syn-anti glycosidic angles alternate with variable groove widths. Compared to B-DNA, the A-DNA form is a wider helix, with a shallower and wider minor groove and a narrower and deeper major groove.

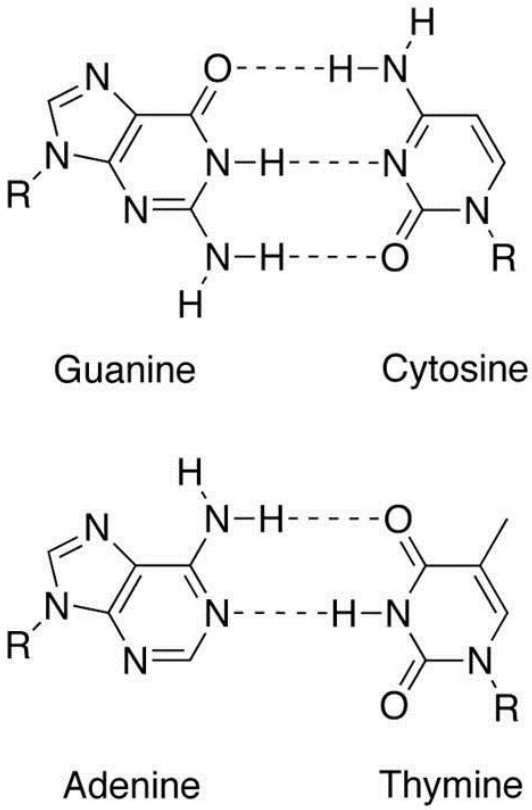

Figure 1-2. Watson-Crick base pairs in DNA. R represents the deoxyribose and phosphate groups. (Figures reproduced from http://chemistry.about.com/od/factsstructures/ig/Chemical-Structures---G/Guanine.-LYw.htm), December 2012.

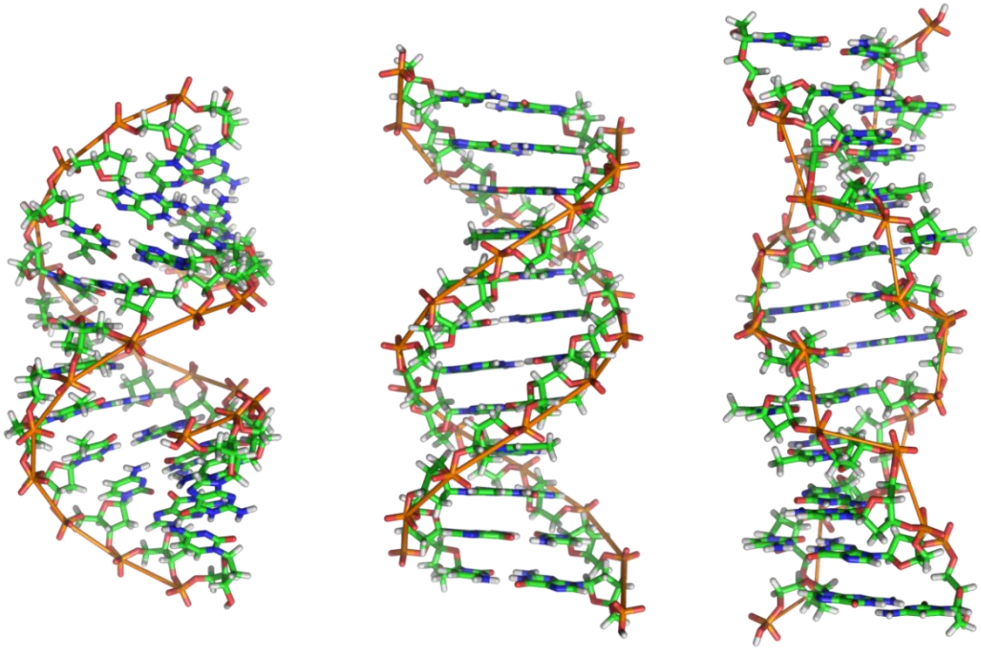

Figure 1-3. (Left to right) A, B, and Z-forms of DNA. (Figures reproduced from http://en.wikipedia.org/wiki/File:A-DNA,_B-DNA_and_Z-DNA.png), December 2012.

1.2 G-quadruplex DNA structures

In 1910, Ivar Christian Bang[6] discovered that guanylic acid forms highly viscous gels in concentrated solutions. After half a century, Khorana et al.[7] obtained evidence on formation of organized structures from deoxyriboguanylic acid oligonucleotides. X-ray fiber diffraction studies of guanosine 3'-monophosphate (3'-GMP) and guanosine 5'-monophosphate (5'-GMP) led Davies et al.[8] to propose that guanylic acid aggregates into two different structures specified as "dimers" and "G-quartets". Each G-quartet is formed via the cyclic arrangement of four guanine residues linked together by Hoogsteen hydrogen bonds.[9] A Hoogsteen base pair refers to any base pairing

between any donor and acceptor sites other than those in Watson-Crick model. The Hoogsteen hydrogen bonds between N1H . . .O6 and N2H . . .N7 atoms in G-quartet are shown in Figure 1-4. Davies et al. further proposed that G-quartets can stack on top of one another by π-π interactions between guanine rings to form a G-quadruplex helical structure.[8,10] Certain metal cations with an appropriate ionic radius, such as Na^+ or K^+, are sandwiched between G-quartets having ion-dipole interactions with the oxygen atoms of guanine-carbonyl groups.

At about the same time, Miles and Frazier[11] discovered that in addition to gel formation under acidic pH, 5'-GMP can also form ordered structures in a neutral solution. Chantot et al.[12] studied the role of ionic strength and the nature of ions in the stability of 5'-GMP aggregates. They found that the K^+ salt forms more stable aggregates than the Na^+ salt. Also, anions play no significant roles in the stability of aggregates. Nucleoside concentration and temperature were the most important factors contributing to the aggregate stability.

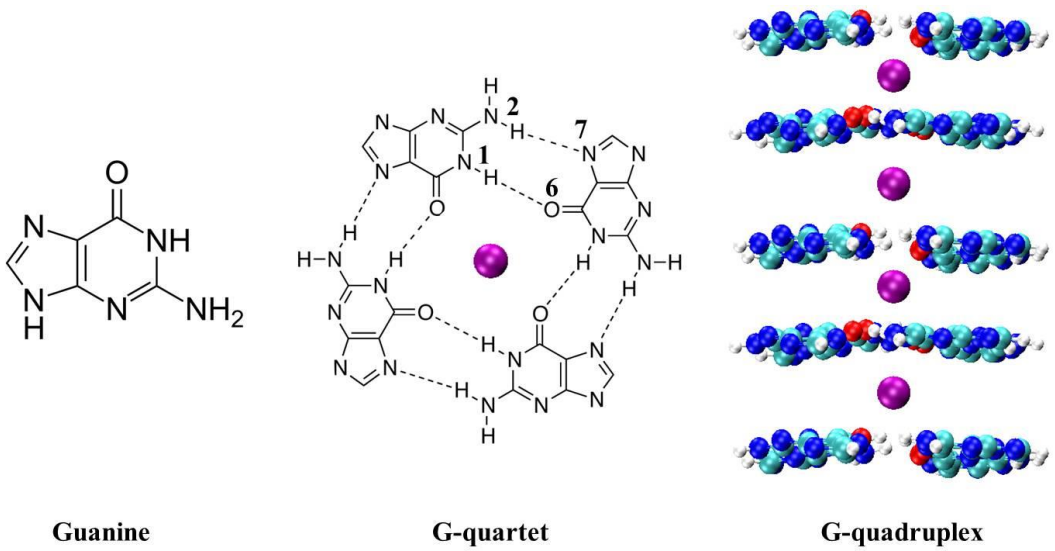

Figure 1-4. (From left to right) Guanine, G-quartet, and G-quadruplex structures. The metal cations are shown in purple.

Several forms of G-quadruplex DNA structures exist depending on the orientation and the number of their chains. The simplest G-quadruplex structures are those formed by four individual DNA strands, i.e., tetramolecular G-quadruplexes. Association of two DNA strands can produce a bimolecular complex with two loops. Loops can orient in two different ways: a "diagonal" loop that crosses over the G-quadruplex helix and connects antiparallel chains and an "edge-type" loop that links adjacent antiparallel chains. Unimolecular G-quadruplexes can be formed from only a single-strand DNA that contains three loops in the structure (see Figure 1-5).[10,13-16]

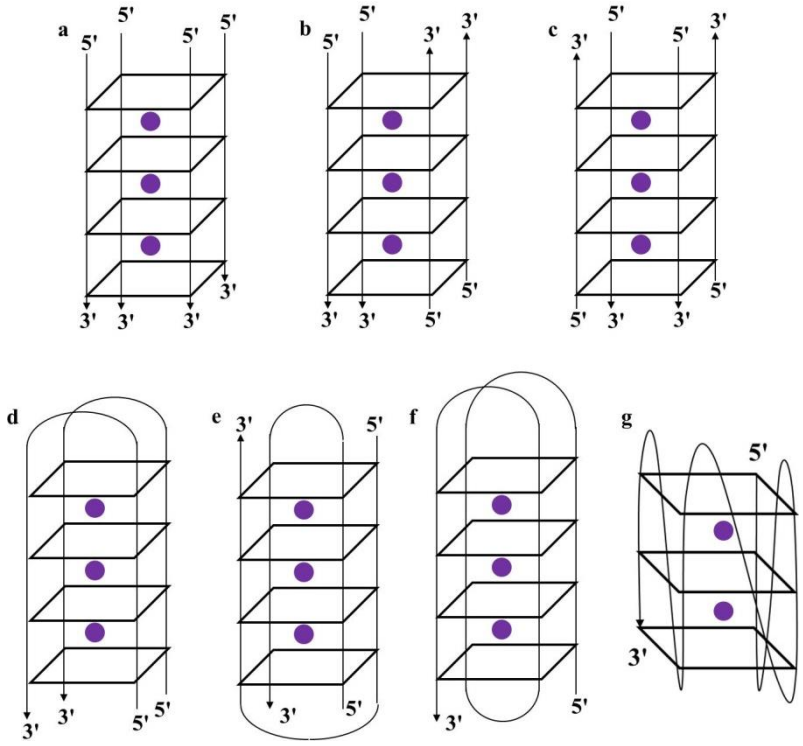

Figure 1-5. (a) A parallel tetramolecular G-quadruplex, (b) an antiparallel tetramolecular G-quadruplex (up-up-down-down), (c) an antiparallel tetramolecular G-quadruplex (up-down-up-down), (d) a bimolecular G-quadruplex with two edge-type loops, (e) a bimolecular G-quadruplex with two diagonal loops, (f) a unimolecular G-quadruplex, and (g) a unimolecular G-quadruplex with propeller loop. The purple circles show the metal cations. The arrows indicate the DNA strand direction from 5' to 3'.

Although G-qudruplex structures were initially studied from a chemistry point of view, several discoveries including the crystal structure of four-stranded *Oxytricha* telomeric DNA[17] and the solution structure of the human telomeric quadruplex[18] revealed their biological importance. Telomeres are specialized structures at the end of chromosomes and they function as protective structures capping both ends of the chromosomes. Human telomeres consist of long, repetitive DNA sequences of TTAGGG. During cell replication, telomeric DNAs cannot be completely replicated. They are gradually shortened until they reach a critical threshold, at which cell replication is arrested, a condition called replicative senescence. An enzyme known as telomerase plays an important role in the elongation of telomeric DNAs (see Figure 1-6). Telomerase activity is almost absent in normal somatic cells, but it is detected in the vast majority of tumor cells. One approach to prevent telomerase from lengthening the telomeric DNAs is to block the enzyme-substrate interaction. Whereas single-stranded DNA is a substrate for telomerase, G-quadruplex DNA is not. Telomerase inhibition depends on shifting the equilibrium between single-stranded and quadruplex DNA structure.[10,19-21]

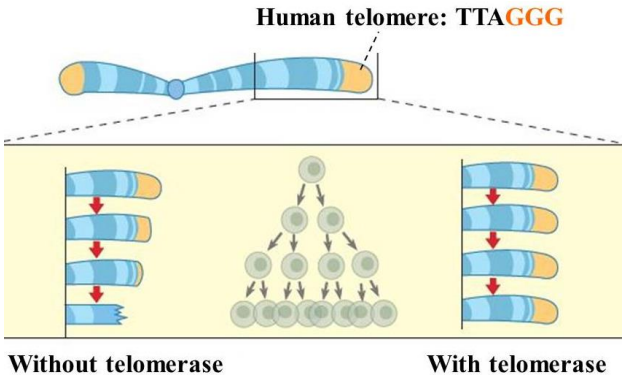

Figure 1-6. A representation of cell division in the absence and the presence of telomerase. (Figure reproduced from http://www.greenpacks.org/2009/10/05/graphics-explain-nobel-prize-work-on-telomeres/), December 2012.

1.3 Experimental NMR studies of G-quadruplex structures

In 1975, Pinnavaia et al.[22] reported the first nuclear magnetic resonance (NMR) study of 5'-GMP self-assembly in neutral solution. They studied the 5'-GMP aggregation under different conditions including high concentration, low temperature, and neutral pH. They observed four different sets of ^{1}H NMR signals for 5'-GMP, indicating the presence of four different 5'-GMP species in solution. A few years later, Pinnavaia et al.[23] showed that the nature of alkali metal ions could influence the self-aggregation of 5'-GMP and that the trend is mainly determined by the cation size. They reported a value of 2.3 Å as the distance between the center of the G-quartet and the four carbonyl oxygen atoms of the same G-quartet. A cation can reside between two G-quartets having ion-dipole interactions with carbonyl-oxygen atoms. Na^+, K^+, and Rb^+ ions promote G-quartet formation. A Li^+ ion, however, is too small to stabilize the aggregate. Eventually, a Cs^+ ion with the largest size is not a good candidate to fit within the G-quadruplex cavity.

In 1982, Pinnavaia et al.[24] recognized that there are at least two major types of bound Na^+ ions in a G-quadruplex structure. The first type is defined by the ions bound to the negatively charged phosphate groups (denoted as phosphate-bound ions). The second type is defined by the ions located inside the G-quadruplex channel (denoted as channel ions). Each channel ion is coordinated to eight guanine oxygen atoms. While phosphate-bound ions undergo fast exchange with free ions in the bulk, the exchange process between the channel and free ions in the bulk is relatively slow (see Figure 1-7). Independent research carried out by Laszlo et al.[25-28] indicated that the aggregate formation of Na_2(5'-GMP) is highly promoted by the presence of K^+ ions. In the presence of Na^+ and K^+ ions, K^+ ions prefer to reside inside the G-quadruplex each being

sandwiched between two G-quartets, whereas Na^+ ions have stronger binding affinity to the peripheral phosphate groups. They also reported the trend of $K^+ > Rb^+$, $Na^+ >> Li^+$, Cs^+ for the G-quadruplex stabilization.

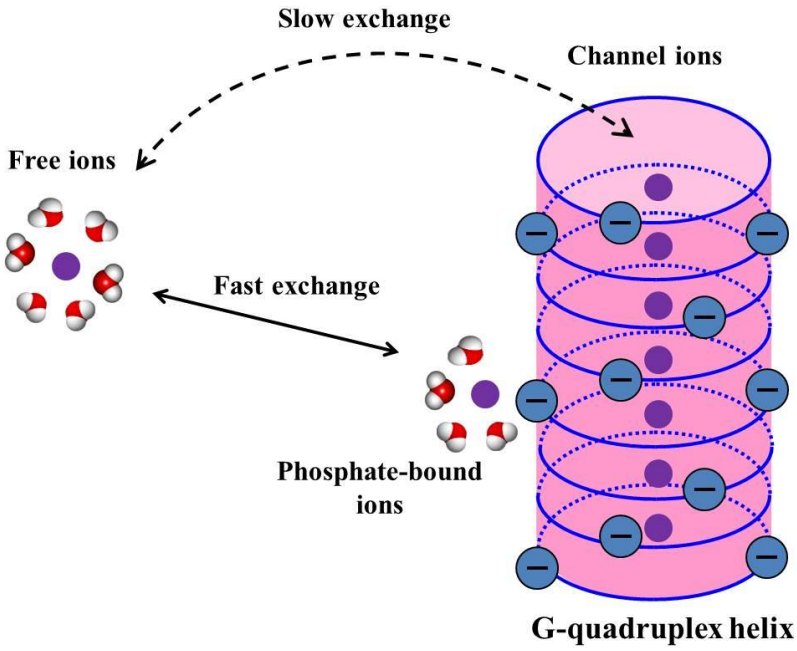

Figure 1-7. Ion exchange between channel and free ions in the bulk (slow exchange) and between phosphate-bound and free ions in the bulk (fast exchange). The cations are shown in purple. The cylinder represents the G-quadruplex.

Since the discovery of G-quadruplex structures, several experimental methods including X-ray crystallography and NMR have been used to obtain further structural information.[7,8,25-38] For the X-ray crystallography studies, X-ray fiber diffraction has been used as it is not possible to obtain single crystals of 5'-GMP. On the other hand, direct detection of light alkali metal ions cannot be achieved using X-ray fiber diffraction. Since the electron density of light atoms such as

Na^+ is similar to water molecules surrounding the G-quadruplex, recognition of these species is not attainable.

NMR spectroscopy is a powerful technique to investigate the structural information of biological systems including G-quadruplex DNA structures. However, solution NMR studies of quadrupolar nuclei (nuclear spin quantum number larger than 1/2) is less common than NMR studies of spin-1/2 nuclei such as ^{1}H and ^{13}C.[28,39,40] This is due to poor intrinsic resolution and sensitivity of quadrupolar nuclei. Despite these difficulties, solution NMR method has been used by Laszlo and co-workers[25-28] to acquire alkali metal NMR spectra in G-quadruplex structure formed by 5'-GMP. Braunlin et al.[31,32] used ^{23}Na NMR spectroscopy to examine the binding environments of Na^+ ions in G-quadruplex DNA structures. They reported that the signal corresponding to the tightly bound ions is not observable. It was mainly due to the line broadening and the small fraction of strongly bound ions. Other research groups developed NMR methods based on surrogate spin-1/2 probes such as $^{15}NH_4^+$ and $^{205}Tl^+$ in solution NMR experiments.[41-44] This method is an alternative to direct probing of metal ions in G-quadruplex DNA.

In 2002, direct detection of alkali metal ions in 5'-GMP was reported by Wu et al.[33-35] using solid-state NMR techniques. Using ^{23}Na Magic Angle Spinning (MAS) NMR, they identified three distinct peaks and assigned them to the channel, phosphate-bound, and free Na^+ ions in the bulk. The chemical shift for the channel Na^+ ions residing inside the G-quadruplex appeared at -17 ppm. The observed ^{23}Na chemical shift was in accordance with the ^{23}Na NMR signal of a nonactin-Na complex in which the Na^+ ion is coordinated by eight oxygen atoms.[45] These authors proposed possible reasons as to why previous studies had failed to observe channel ions. Use of low magnetic fields along with the line broadening problems misled the earlier

researchers to conclude that the channel ion is not observable. Later on, these authors reported the first direct solution NMR detection of the channel alkali metal ions inside G-quadruplex structures.[36] Wu and Ida[37] also developed solution and solid-state NMR methods to study alkali metal ion-binding to G-quadruplex structures formed by three DNA oligomers: $d(TG_4T)$, $d(G_4T_3G_4)$, and $d(G_4T_4G_4)$. In this study, they directly observed alkali metal ions tightly bound to G-quadruplex DNA structures in solution. Furthermore, they showed the first direct NMR evidence of Na^+ ions in the diagonal T4 loop region of $d(G_4T_4G_4)$ G-quadruplex structure. They also estimated a 220 μs residence lifetime for the Na^+ ions in the loop region, while the residence lifetime for Na^+ ions inside the channel was almost two orders of magnitude longer. Recently, Wu and Kwan[38] determined the helical structure of Na_2(5'-GMP), which was unknown for several decades. They proposed a right-handed four-stranded helical structure. Weak molecular interactions rather than the phosphodiester bonds link the G-quartets to one another.

Paris and co-workers[46] reported the value of 30 ns as the lifetime of phosphate-bound Na^+ ions of 5'-GMP quadruplex structure. The residence time of ammonium ions in a quadruplex structure of $[d(G_4T_4G_4)]_2$ has been determined as 250 ms by Feigon et al.[42] A Na^+ NMR study of the same type of quadruplex, however, yielded a value of 250 μs as the lifetime of Na^+ ions.[32] These studies, unfortunately, did not report lifetime of any ion inside the channel. As mentioned earlier, channel ions play important structural roles in G-quadruplex DNA structures. Therefore, a good understanding of the dynamic nature of alkali metal ion-binding in these structures will be of great importance.

1.4 Computational studies of G-quadruplex structures

Besides many experimental techniques, computational methods in particular molecular dynamics (MD) simulations have been extensively used to study structural and dynamical properties of biological systems including G-quadruplex DNA. MD simulation can provide detailed insight into each individual atom motion over time as well as its correlation with other particles in the system. It is a powerful technique to address specific structural questions during the simulation. The first biological MD simulation studies were performed on the bovine pancreatic trypsin inhibitor almost 35 years ago.[47] The dynamics of a folded protein was studied by solving the equations of motion for atoms with an empirical potential energy function. MD method has been extensively used in the thermodynamic fluctuation studies of protein molecules,[48] dynamics of ligand-protein binding,[49] protein dynamics in aqueous solution,[50] enzyme design,[51] and the mechanism of selectivity in channels[52,53] and enzymes.[53] Also, broad free energy calculations have been performed on protein-ligand interactions,[54] drug design,[55,56] signal transduction,[57] peptide folding,[58,59] membrane protein association,[60] and ion transport along various types of membranes and ion channels.[61-63] Although various free energy calculation methods have been employed to determine the energetics and dynamics of different types of ion channels including KcsA,[52,64-67] Gramicidin,[68-75] voltage-gated,[76] ligand-gated,[77] carbon nanotube,[78-82] and peptide nanotubes,[83-85] to the best of our knowledge, free energy method has not been used to evaluate energetic properties of ion transport along a G-quadruplex DNA channel.

Several researchers have employed MD simulations to study G-quadruplex structures.[86-90] The dynamical properties of $[d(G_3T_4G_3)]_2$ quadruplex in the presence of K^+ ions were investigated using NMR and MD simulation methods.[91] This study confirmed the importance of

cations in stabilizing the quadruplex structures. Also, it revealed that in the presence of K^+ ions the thymine loops are more flexible than in the presence of Na^+ ions. Therefore, different structures can be formed for K^+ and Na^+ ions. Furthermore, grooves and interiors of the loops were determined as the most probable regions for hydration. Sponer et al.[92] performed MD simulations on G-quadruplex DNA structures of $d(TG_4T)$ and $d(G_4T_4G_4)$. They obtained very stable structures in the nanosecond time scales when Na^+ ions are present in the channel. The structures could be preserved after removing a few cations from the channel. The channel ions could exchange with the bulk ions without significant structural deformation. Complete removal of the channel ions destabilized the structure within a few picoseconds. In general, the cavities became occupied by water molecules after removing the channel ions, but this was not enough to stabilize the G-quadruplex structures. The MD simulation studies of $d(G_4T_4G_4)$ demonstrated that thymine residues in the loop are capable of coordinating to a Na^+ ion that has left the channel. In a separate study, Sponer et al.[89] investigated the formation of cation-stabilized G-quadruplex DNA structure of $d(GGGG)_4$ using MD simulations. Because of the limited accessible time scales of the simulations, they probed the stability of possible intermediate assemblies that could be involved in the formation of the G-quadruplex structure. They proposed "cross-like" two-stranded assemblies as the nucleation center to initiate G-quadruplex formation. They also estimated the relative free energies for these structures using molecular mechanics Poisson-Bolzmann surface area (MM-PBSA) methodology.[93,94] In another study, Sponer et al.[95] performed MD simulations to investigate the structural behavior and the ion interactions in G-quadruplex DNA structures containing mixed guanine/cytosine and also all-guanine residues. They demonstrated that the interaction of cations with all-guanine model results in a stable structure, whereas in the mixed guanine/cytosine model, the quartets are quite tolerated and do not interact considerably with the

cations. They also identified several temporary structures in the thymine loops that coordinate with the cations. Several potential pathways were recognized within the thymine loop region for the channel ion exchange with the bulk. In a recent study, Golovin et al.[96] investigated the cation-binding to 15-mer thrombin-binding quadruplex DNA aptamer. They concluded that cation-binding is a multiple pathway process which is highly dependent on the type of cation. The individual ion-binding events are controlled by the connecting loops of the aptamer in which they stabilize the molecules in the absence of the bound ions, control the pathway of the ion into the stem, and stabilize the ions by closing their entrance gates to the quadruplex. For the first time, they observed a full spontaneous exchange between the internal cations and the bulk. Cang et al.[97] studied the effect of loop length and sequence as well as the G-tract length on the folding and stability of G-quadruplex DNA. Whereas a single nucleotide dT loop mostly forms parallel structures, a dinucleotide dT2 loop prefers the formation of parallel and antiparallel structures. Trinucleotide or longer loops assist the formation of anti-parallel structures. In general, the long loops are more stable compared to the short loops. Single nucleotide loop structures have a tendency to form edge-type loops over narrow grooves. Chowdhury and Bansal[98] studied the ion dynamics in a parallel d(G7) quadruplex structure in the presence of Na^+ and K^+ ions in the channel. Na^+ ions exhibit greater mobility within the quadruplex cavity than K^+ ions. In the absence of the channel ions in the initial structure, the structure did not change significantly until a Na^+ ion entered the channel in addition to a few water molecules. In the absence of channel ions, the O6 atoms of guanine carbonyl groups became farther apart to minimize the repulsion forces between them. A few hydrogen bonded water molecules were observed within the grooves. In an MD study by Cavallari et al.[99], the migration and stability of the metal ions were studied in G4-wires of different lengths and with different cation concentrations. They concluded that the

cations stabilize the structures. Also, a short G-wire structure unfolds quickly if there is no ion in the channel or if the ion size is too small (e.g., Li^+). Water molecules occupied several backbone and channel sites as well as the channel entrance sites in these structures. In a recent study, Sponer et al.[100] reviewed the application of molecular dynamics simulation method to understand several structural and dynamical features of nucleic acids.

Several researchers have performed MD simulation studies to examine ion exchange between the free and groove states in G-quadruplex structures.[101-110] An MD simulation study for a triplex DNA indicated a strong interaction between counter ions and water molecules around them. Water molecules and counter ions were found to be more ordered in the presence of triplex DNA. Also a slow mobility was found for the counter ions and water molecules close to the triplex.[101] Different models showed that the structure of the minor groove and the positions of the ions around it are related.[102,111] In an AATT dodecamer, a narrow minor groove is formed in the presence of Na^+ ions. MD simulation studies of a d(TATAGGCCTATA) duplex indicated that the amount of fluctuations is the same in minor grooves of a G-tract and an A-tract, but fewer ion interactions exist in the former one. In general, the interactions between ions and minor grooves affect the width of the minor groove. When the ions neutralize the phosphate charges, a narrow minor groove is formed, whereas in the absence of the neutralizing ions, a wide minor groove is preferred.

Although there are several MD simulation studies in the literature about the G-quadruplex structures, one can rarely find a specific MD simulation study of ion dynamics between the channel and the bulk. Because of the critical roles of the channel ions in the stabilization of the G-quadruplex structures, a full investigation of the dynamics, energetics, and mechanism of their transport along the G-quadruplex DNA channel is required.

1.5 Organization of the thesis

The objective of this thesis is to use MD simulation methods to investigate ion dynamics along G-quadruplex DNA channels. A brief theory of MD simulations is given in Chapter 2 with an emphasis on free energy calculation techniques such as adaptive biasing force (ABF) and Umbrella sampling (US) methods. In Chapter 3, we use the ABF method to study the energetics and dynamics of Na^+, K^+, and NH_4^+ ions along the $[d(TG_4T)]_4$ G-quadruplex DNA channel. In Chapter 4, we employ the ABF method to obtain free energy profiles for Na^+, K^+, and NH_4^+ ion movement along the $[d(G_3T_4G_4)]_2$ channel. In Chapter 5, the US technique is employed to examine the energetics of Na^+ and K^+ ion movement along the $[d(TG_4T)]_4$ G-quadruplex DNA channel. In Chapter 6, NH_4^+ ion transport along the $[d(G_3T_4G_4)]_2$ channel is investigated. Finally, Chapter 7 provides concluding remarks and suggestions for future work.

Chapter 2

Molecular dynamics simulation

2.1 Why molecular dynamics simulations?

Computer simulations are useful techniques to understand molecular structures and gain new and deeper insight into the microscopic interactions. They can link the microscopic properties of the systems to macroscopic experimental observations. Nowadays, computer simulations have become a substantial part of the mathematical modeling of many natural systems. Among numerous computational techniques available to examine biological systems, molecular dynamics (MD) simulations can provide detailed information about time dependent behaviors of molecular systems.[112,113] These include many structural and conformational features as well as dynamical properties and energetics of molecular processes.

The MD method has been widely used to determine structures along with X-ray, NMR and other experimental techniques. Sometimes, the method could be used to investigate the systems under extreme conditions that could not be reached experimentally. The MD method was first developed in 1957 to study the interactions of hard spheres.[114] Later on, it was used by Rahman[115] in 1964 to simulate a liquid argon system. In 1974, Rahman and Stillinger[116] published an MD simulation study on a system of 216 water molecules. In 1977, MD was used to obtain structural information about a folded globular protein (bovine pancreatic trypsin inhibitor).[47] Today, MD simulation methods are widely used for particular purposes such as structural and dynamical studies of DNA, protein-protein, protein-ligand, and drug design problems.

2.2 The MD simulation method

The MD simulation technique is used to calculate the equilibrium and dynamic properties of a system of atoms. It can predict the behavior of simple systems such as ideal gases as well as complex systems such as living cells.[112,113] In MD simulations, an overall force is calculated for each atom and classical equations of motion are employed to move the system trajectory over time.[112,113] By using statistical mechanics and thermodynamics, macroscopic properties of the system could be calculated. In this regard, MD connects theoretical and experimental methods. In fact, MD simulation is "running experiments on computers".[112,113] MD simulations are particularly important for analysis of complex systems in which experimental methods may be difficult to apply.

In a typical system containing *N* atoms, the total Hamiltonian $H(\vec{q},\vec{p})$ consists of both kinetic and potential energies. Such a Hamiltonian could be adequately described by the position $\vec{q} = (\vec{q}_1, \vec{q}_2, \dots, \vec{q}_N)$ and the momentum $\vec{p} = (\vec{p}_1, \vec{p}_2, \dots, \vec{p}_N)$ of all atoms in the system:[112,113]

$$H(\vec{q},\vec{p}) = K(\vec{p}) + U(\vec{q}) \qquad [2.1]$$

where $K(\vec{p})$ and $U(\vec{q})$ represent the kinetic and potential energies of the system, respectively. The kinetic energy could be further written as:

$$K(\vec{p}) = \sum_{\alpha=1}^{N} \sum_{i=1}^{M} \frac{\vec{p}_{i\alpha}^{2}}{2m_{i\alpha}} \qquad [2.2]$$

where N is the number of molecules, M is the number of atoms in each molecule, $m_{i\alpha}$ is the atomic mass of atom i in molecule α, and $p_{i\alpha}$ is the momentum of this atom. The classical equations of motion are expressed as follows:

$$\frac{d\vec{q}_{i\alpha}}{dt} = \frac{\partial H}{\partial \vec{p}_{i\alpha}} = \frac{\vec{p}_{i\alpha}}{m_{i\alpha}} \quad [2.3]$$

$$\frac{d\vec{p}_{i\alpha}}{dt} = -\frac{\partial H}{\partial \vec{q}_{i\alpha}} = -\frac{\partial U}{\partial \vec{q}_{i\alpha}} \quad [2.4]$$

Equation 2.4 can be re-written in terms of the force:

$$\vec{F}_{i\alpha} = -\frac{\partial U}{\partial \vec{q}_{i\alpha}} = m_{i\alpha}\frac{d^2\vec{q}_{i\alpha}}{dt^2} = m_{i\alpha}\vec{a}_{i\alpha} \quad [2.5]$$

where $a_{i\alpha}$ corresponds to the acceleration of atom i in molecule α.

In each step of a typical MD simulation, the overall force on each atom is calculated and then the positions of all atoms are advanced based on the force acting on them and the simulation time step. Apparently, the force calculation is the heart of each MD simulation. Therefore, one needs to develop a model that can accurately reflect the real forces on the atoms.

2.3 Equations of motion

2.3.1 The Verlet algorithm

In MD simulations, forces on all atoms are calculated and then used to advance the positions of all atoms using the defined time step for the simulation. This process continues for a long enough time so that the sampling of the system is adequate to produce statistically

meaningful results (see Figure 2-1). The most important step after the force calculation is to employ a fast accurate integration algorithm to advance atomic positions. Such an algorithm must be time reversible. That is, if one could reverse the time, the system should go back to its initial state.

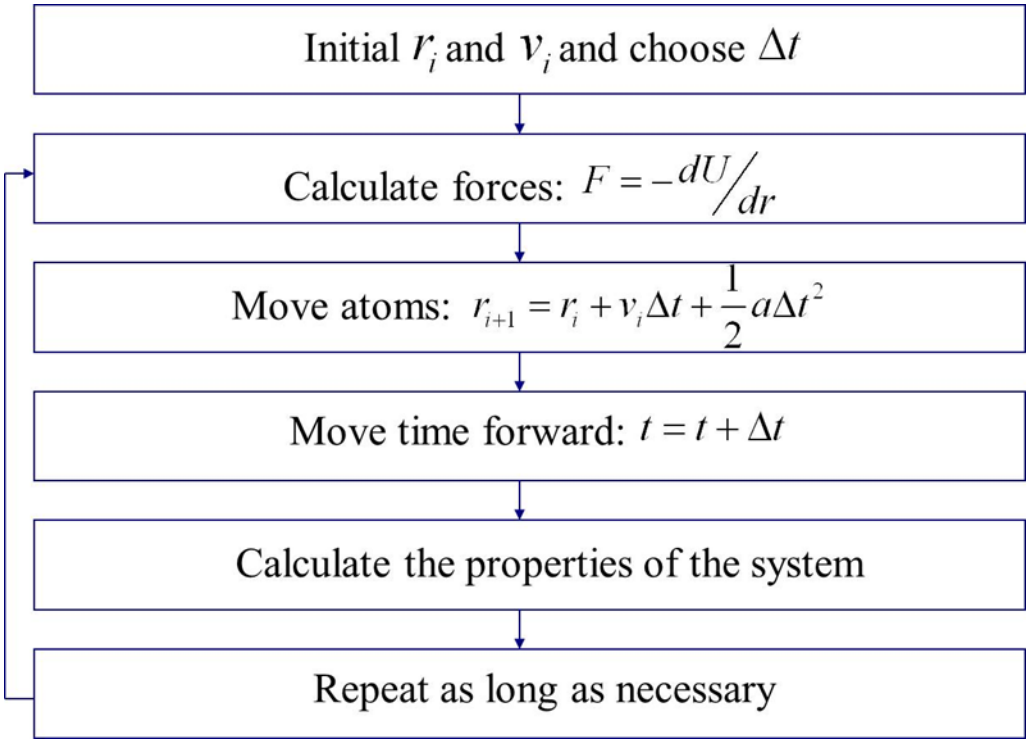

Figure 2-1. MD simulation algorithm.

There are several MD simulation algorithms available in the literature.[112,113] One of the most widely used algorithms is the Verlet algorithm which is based on a Taylor expansion of the positions of the atoms at time *t*:[112,113]

$$\vec{r}(t+\Delta t) = \vec{r}(t) + \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)\Delta t^2 + \cdots$$

$$\vec{r}(t-\Delta t) = \vec{r}(t) - \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)\Delta t^2 - \cdots \quad [2.6]$$

where $\vec{r}$ is the position of the atom, Δt is the time step of the simulation, $\vec{v}$ is the velocity, and $\vec{a}$ is the acceleration. The acceleration is calculated from:

$$\vec{a}(t) = \frac{\vec{F}(t)}{m} \quad [2.7]$$

where $\vec{F}(t)$ is the overall force on each atom at time t and m is the atom mass. In practice, the velocities are not used to advance the positions. By adding the two parts of Equation 2.6, the effect of velocities is cancelled. Also, by choosing a reasonably short time step, higher order derivatives in the two parts of Equation 2.6 could be ignored to produce the below equation:

$$\vec{r}(t+\Delta t) = 2\vec{r}(t) - \vec{r}(t-\Delta t) + \vec{a}(t)\Delta t^2 \quad [2.8]$$

In the above equation the error is in the order of Δt^4. So, in general, choosing a small time step could considerably decrease the algorithm error. It is critical to select a proper time step. The time step should be small enough so that only negligible deviations happen for ignoring the higher order terms. On the other hand, it should be large enough so that an adequate sampling could be achieved in a reasonable amount of CPU time. In Equation 2.8 the new position at time $t+\Delta t$ could be predicted if one knows the current position at time t, the old position at time $t-\Delta t$, and of course, the force at the current time. The Verlet algorithm is time reversible in a sense that Δt and $-\Delta t$ play symmetric roles. The other advantage of this algorithm is that it offers stable

energy conservation. Although velocities are not explicitly used to advance atom positions, they can be readily calculated to yield properties such as kinetic energy:

$$\vec{v}(t) = \frac{\vec{r}(t+\Delta t)-\vec{r}(t-\Delta t)}{2\Delta t} \quad [2.9]$$

For this equation the error is in the order of Δt^3. The Verlet algorithm has several alternative versions that use explicit velocity for solving the equations of motion. The most commonly used ones are Velocity-Verlet and Leap-frog. The Velocity-Verlet algorithm has two steps:

$$\vec{r}(t+\Delta t) = \vec{r}(t) + \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)\Delta t^2$$

$$\vec{v}(t+\Delta t) = \vec{v}(t) + \tfrac{1}{2}\Delta t[\vec{a}(t) + \vec{a}(t+\Delta t)] \quad [2.10]$$

It can be shown that these two equations are equivalent forms of the normal Verlet algorithm. In the first step of this algorithm, new atom positions are calculated and then the velocities at the half time step are calculated using the following equation:

$$\vec{v}\left(t+\tfrac{1}{2}\Delta t\right) = \vec{v}(t) + \tfrac{1}{2}\vec{a}(t)\Delta t \quad [2.11]$$

The next step is to calculate force and acceleration at $t+\Delta t$ and then the velocities are advanced using the following equation:

$$\vec{v}(t+\Delta t) = \vec{v}\left(t+\tfrac{1}{2}\Delta t\right) + \tfrac{1}{2}\vec{a}(t+\Delta t)\Delta t \quad [2.12]$$

This algorithm has some technical advantages. For example, it makes the transition from NVE (constant-energy, constant-volume) ensemble to NVT (constant-temperature, constant-volume) ensemble easier by only scaling the velocities in the simulation cell.

2.3.2 The Leap-Frog algorithm

Other alternatives to the Verlet algorithm are also available. While simple truncation of the Taylor expansion might be erroneous, not reversible, and not area preserving, other simple methods based on Taylor expansion are practical. The Leap-Frog algorithm[117] calculates the velocities at half time steps and uses them to predict the new system trajectory. This algorithm can be derived from the Verlet algorithm:

$$v\left(t-\frac{\Delta t}{2}\right)=\frac{r(t)-r(t-\Delta t)}{\Delta t}$$
$$v\left(t+\frac{\Delta t}{2}\right)=\frac{r(t+\Delta t)-r(t)}{\Delta t} \qquad [2.13]$$

By combining these equations, the new atom positions can be found:

$$r(t+\Delta t)=r(t)+\Delta t v(t+\frac{\Delta t}{2}) \qquad [2.14]$$

Also, the following equation can be used to update the velocities:

$$v\left(t+\frac{\Delta t}{2}\right)=v\left(t-\frac{\Delta t}{2}\right)+\Delta t\frac{F(t)}{m} \qquad [2.15]$$

The Leap-Frog algorithm produces exactly the same trajectory as the Verlet algorithm. However, in the Leap-Frog algorithm, velocities and positions are not known at the same time. Consequently, kinetic and potential energies could be defined at different times and one cannot calculate the total energy at any time step.

2.4 Constraint dynamics

Ideally, long time scales are desirable in any MD simulation since slow processes can be probed. However, because of the error associated with the integration of equations of motion, one needs to minimize the time step to reduce the error. In practice, the time step should be chosen in such a way that it is reasonably shorter than the time scale of the fastest motion in the system. Usually, the fastest motion is associated with the bond stretching in a molecule. Several methods exist to solve the time scale problem. One approach is based on multiple time steps[118] that uses short and long time steps for fast and slow motions of the system, respectively. Another alternative is to freeze the fast motions in the system by using constraint dynamics. The most commonly used approaches to constrain atomic motion in the system are SHAKE[120] and its velocity version, RATTLE.[121] In the SHAKE algorithm the equations of motion are solved as usual, but one extra constraint force is added to each atom:

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i + \vec{g}_i \qquad [2.16]$$

where $\vec{F}_i$ and $\vec{g}_i$ are the real and constraint forces, respectively. As a result, the Verlet algorithm needs to be modified to include this extra force:

$$\vec{r}_i(t+\Delta t) = \vec{r}_i'(t+\Delta t) + \frac{\Delta t^2}{m_i}\vec{g}_i(t) \qquad [2.17]$$

where $\vec{r}_i'(t+\Delta t)$ is the position without the constraint. Such a scheme could be used to fix all bond lengths since the bond stretching is the fastest motion in the system and in classical MD simulations, processes including bond dissociation or association are not considered. This constraint force should be pointed along the bond direction and needs to follow the Newton's third law:

$$\vec{g}_i = \sum_j \lambda_{ij}\vec{r}_{ij} \quad \text{and } \lambda_{ij} = \lambda_{ji} \qquad [2.18]$$

where *i* represents the atom of interest and *j* stands for the atom that is bonded to atom *i*. The undetermined time dependent Lagrange multiplier is shown by λ_{ij}. By combining Equations 2.17 and 2.18 we will get:

$$\vec{r}_{ij}(t+\Delta t) = \vec{r}_{ij}'(t+\Delta t) - \sum_{h\neq j}\frac{\Delta t^2}{m_h}\lambda_{hi}\vec{r}_{hi}(t) + \frac{\Delta t^2\left(m_i+m_j\right)}{m_i m_j}\lambda_{ij}\vec{r}_{ij}(t) - \sum_{k\neq i}\frac{\Delta t^2}{m_k}\lambda_{jk}\vec{r}_{jk}(t) \qquad [2.19]$$

Here, the index *h* is over all atoms except atom *j* and atom *j* does not bond to atom *i*. Index *k* is over all the atoms except for atom *i* and atom *i* bonds to atom *j*. The term $\vec{r}_{ij}'(t+\Delta t)$ is the bond when there is no constraint. If the square modulus of both sides of the equation is taken, we will have:

$$\left|\vec{r}_{ij}(t+\Delta t)\right|^2 = \left|\vec{r}_{ij}(t)\right|^2 = d_{12}^2 \qquad [2.20]$$

where d_{12} is the bond distance. As a result, we will have a series of quadratic equations to solve for λ_{ij}. These equations are solved numerically and the solutions are used in Equation 2.18. Such an algorithm is easily expandable to include angular motions as well.[120]

The SHAKE algorithm applies well to the Verlet since the latter algorithm only includes positions and accelerations. By modifying the SHAKE algorithm, it is possible to produce its velocity based version known as RATTLE. Practically, two steps need to be taken in RATTLE. First, a constraint should be added to Equation 2.10:

$$\vec{r}_i(t+\Delta t) = \vec{r}_i'(t+\Delta t) + \frac{\Delta t^2}{2m_i}\vec{g}_i^r(t)$$

$$\vec{v}_i\left(t+\frac{\Delta t}{2}\right) = \vec{v}_i'\left(t+\frac{\Delta t}{2}\right) + \frac{\Delta \mathrm{t}}{2m_i}\vec{g}_i^r(t) \qquad [2.21]$$

Similarly, these equations are iteratively solved. The velocities at half time steps are then calculated using the second part of Equation 2.21. In the second step, velocities are evolved further:

$$\vec{v}_i(t+\Delta \mathrm{t}) = \vec{v}_i'(t+\Delta \mathrm{t}) + \frac{\Delta \mathrm{t}}{2m_i}\vec{g}_i^v(t+\Delta t) \qquad [2.22]$$

Generally, SHAKE and RATTLE are used as simple approaches to fix the motions of fast degrees of freedom.

2.5 Definition of the potential terms

Force is equal to the gradient of the potential with respect to the atom position. So, by knowing the potential of the system and its dependence on the atom position, the force can be readily calculated. The potential of the system is often divided into two parts:

$$U = U^{intra} + U^{inter} \quad [2.23]$$

where the first part is the intramolecular potential and the second part is the intermolecular potential. The intramolecular potential is usually defined as the potential between the atoms that are "close" to each other within a molecule. It consists of several individual terms such as stretching (U^{st}), bending (U^{bend}), improper torsion (U^{imp}), and torsion (U^{tor}):

$$U^{intra} = U^{st} + U^{bend} + U^{imp} + U^{tor} \quad [2.24]$$

Stretching potentials are usually defined by a simple harmonic potential:

$$U^{st}(r) = k_s(r - r_e)^2 \quad [2.25]$$

where k_s, r, and r_e are the stretching force constant, the instantaneous distance between the two interacting atoms, and the equilibrium distance between them, respectively (see Figure 2-2). It is assumed that r oscillates around its equilibrium value r_e. As a result, when the two atoms become closer or farther compared to the equilibrium distance, a high energy will be experienced. Because the harmonic potential is a "sharp" potential model, it is expected to reasonably produce

the correct potential at distances around the equilibrium distance. In fact, the more accurate potential form is the Morse potential[119] which is defined as:

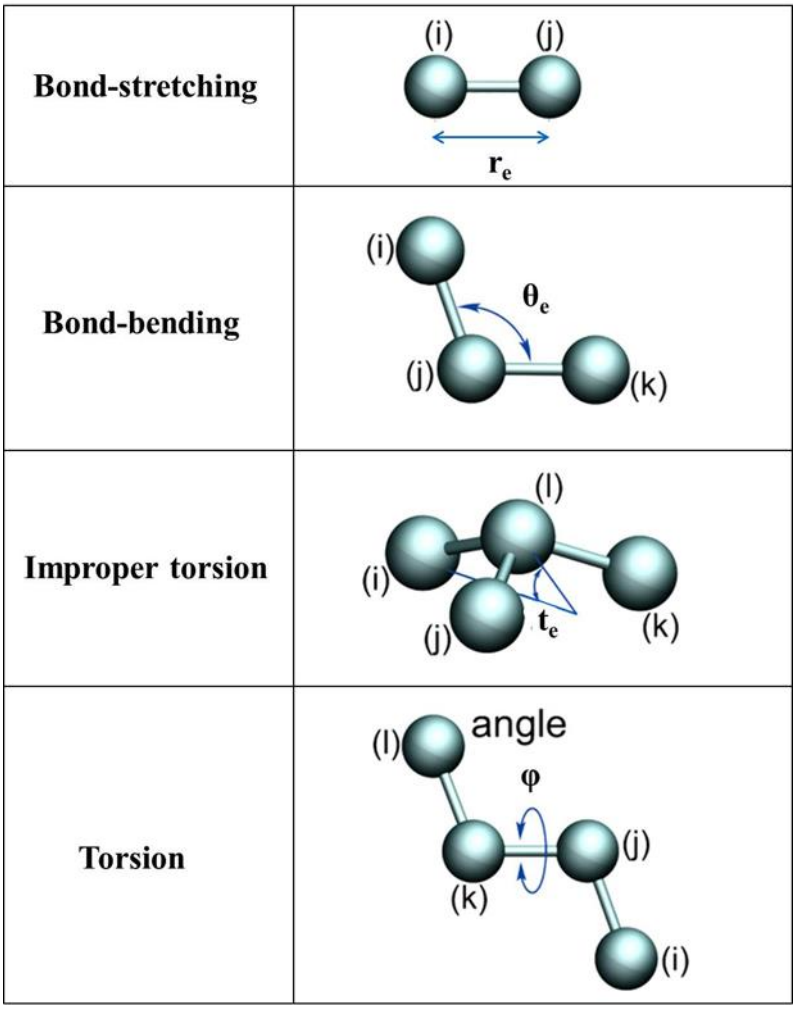

Figure 2-2. A schematic display of intramolecular potentials. (Figures reproduced from http://www.mbnexplorer.com/users-guide/4-energy-and-force-calculation/43-molecular-echanics-potential), December 2012.

$$U^{st}(r) = D_e[1 - e^{-\alpha(r-r_e)}]^2 \qquad [2.26]$$

where D_e is the bond dissociation energy and α is the inverse of the potential width. Such a potential resembles closely the experimental one as shown in Figure 2-3. Therefore, in principle, Morse potential is one of the most accurate potential forms to model molecular stretching. However, it is computationally expensive because it includes an exponential term. The advantage of the harmonic model is that it is simple and computationally efficient. In general, as atoms do not deviate significantly from their equilibrium positions, a harmonic potential is sufficiently accurate, as shown in Figure 2-3. In MD simulations most of the bond distances are held fixed by using algorithms such as SHAKE[120] or RATTLE[121] and stretching potentials are only used for specific cases. In some cases, cubic and quartic terms are added to the harmonic potential to improve accuracy.

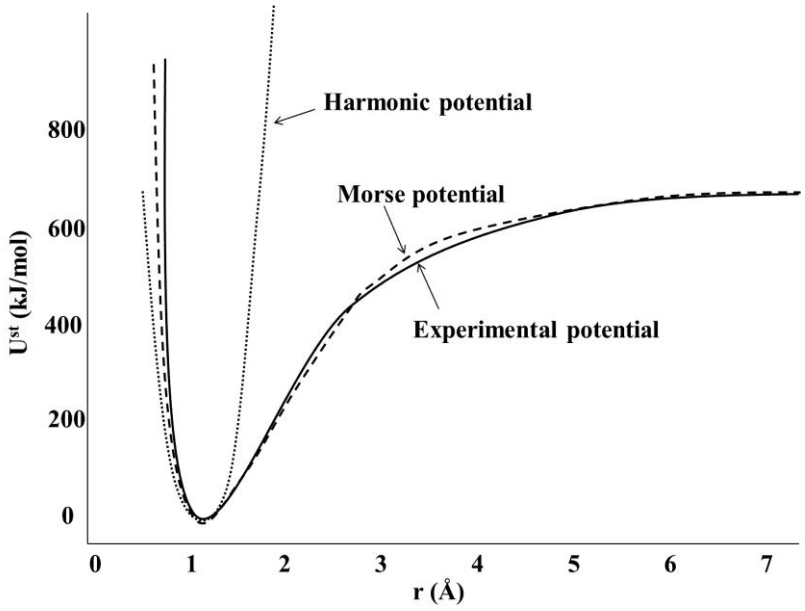

Figure 2-3. Typical comparison between Morse, harmonic, and experimental potentials for bond stretching.

As illustrated in Figure 2-2, bending potential is defined as the potential that keeps the bond angle between three consecutive atoms in the molecule at around the equilibrium value. It could be described by a simple harmonic potential:

$$U^{bend}(\theta) = k_\theta(\theta - \theta_e)^2 \qquad [2.27]$$

where k_θ, θ, and θ_e represent the bending force constant, the angle, and the equilibrium angle. Similar to bond stretching, in general, each bond angle exhibits small fluctuation around the equilibrium value and a simple harmonic potential is often considered accurate enough.

The third term in the intramolecular potential in Equation 2.7 represents the energy cost for out of plane motion. Such a term is mostly necessary for planar molecules and chiral centers to conserve the planarity and chirality during the simulation. Usually a simple harmonic behavior could also be employed to model improper torsion in the molecules:

$$U^{imp}(t) = k_t(t - t_e)^2 \qquad [2.28]$$

where k_t, t, and t_e are the improper torsion force constant, out-of-plane angle, and the corresponding equilibrium value as defined in Figure 2-2.

While stretching, bending, and improper torsion could be modeled using a simple harmonic potential function, the behavior of a torsional potential is very complex. In fact, in a typical molecule there exist dihedral angles than can simply rotate the whole range of motion (360º) at room temperature. The energy costs for rotation are typically lower than those for other modes of motion mentioned earlier. On the other hand, these low barrier motions are associated with significant changes in atomic positions. So, in contrast to other potentials in which we have

high energy barriers and small amplitude motions, here we need to model low energy barriers with large amplitude motions. Rotational energy barriers typically show repetitive behaviors that suggest they could be modeled using *sin* or *cos* terms. There are several potential models that can be used in the literature to describe torsional energy barriers. One of them is the Ryckaert-Bellemans model:[122]

$$U^{tor}(\varphi) = \left(\sum_{i=0}^{6} C_i \cos^i(\varphi + \varphi_i)\right) \quad [2.29]$$

where C_i is the linear parameter to be fitted, φ is the instant dihedral angle and φ_i is the nonlinear phase shift to be fitted. Such a potential is flexible enough to model complex torsional behavior. Another commonly used potential function is:

$$U^{tor}(\varphi) = \frac{1}{2}\sum_{\{i\}} V_i\left[1 + (-1)^{i+1}\cos(i\varphi + \varphi_i)\right] \text{ with i = 0, 1, 2, 3…} \quad [2.30]$$

where φ, φ_i, and V are the dihedral angle, the dihedral angle corresponding phase shift, and the coefficients in the Fourier series, respectively. In this Fourier series, a set of periodicities are specific to each dihedral torsion. Most of the force fields such as OPLS[123] and CHARMM[124] use this equation with less than four terms.

Intermolecular potentials could be defined as a sum of two individual terms: electrostatic and Lennard-Jones potentials:

$$U^{inter} = U^{el} + U^{LJ} \quad [2.31]$$

Electrostatic potential is defined by the Coulombs law:

$$U^{el} = \sum_{i\alpha<j\beta} \frac{Q_{i\alpha}Q_{j\beta}}{4\pi\varepsilon_0 r_{i\alpha j\beta}} \quad [2.32]$$

where $Q_{i\alpha}$ is the atomic point charge of the atom i in molecule α and ε_0 is the dielectric constant. Typically, these charges are fixed during the MD simulation although in some cases charge may change during the simulation. The electrostatic potential is calculated between all of the atoms of separate molecules or the atoms within a molecule that do not share any intramolecular potential. The Lennard-Jones potential is a steep potential curve consisting of a repulsion and an attraction term:

$$U^{LJ} = \sum_{i\alpha<j\beta} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{i\alpha j\beta}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{i\alpha j\beta}}\right)^{6}\right] \quad [2.33]$$

where i and j are atoms belonging to two different molecules α and β. ε_{ij} is the well depth, σ_{ij} is the interatomic distance in which potential energy is zero, and r_{ij} is the distance between the two atoms (see Figure 2-4). During the simulation, for every atom i, the summation goes over all of the other atoms in the system that do not share any intramolecular potential with atom i. In the above equation, ε_{ij} and σ_{ij} are the combined Lennard-Jones energy and length parameters, respectively. These parameters are specific to any atom in any molecule to accurately reproduce the real potential energies. Usually, these parameters are parameterized to fit a set of experimental data such as diffusion coefficients and experimental infra-red (IR) frequencies.[125] Note that these parameters are not only defined for one atom, but they are defined when one atom type interacts with another atom type. So, there is a need to combine ε and σ of interacting atoms to produce ε_{ij} and σ_{ij}. Usually the Lorentz-Berthelot mixing rule[112] is used to combine atomic Lennard-Jones parameters:

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}, \varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \quad [2.34]$$

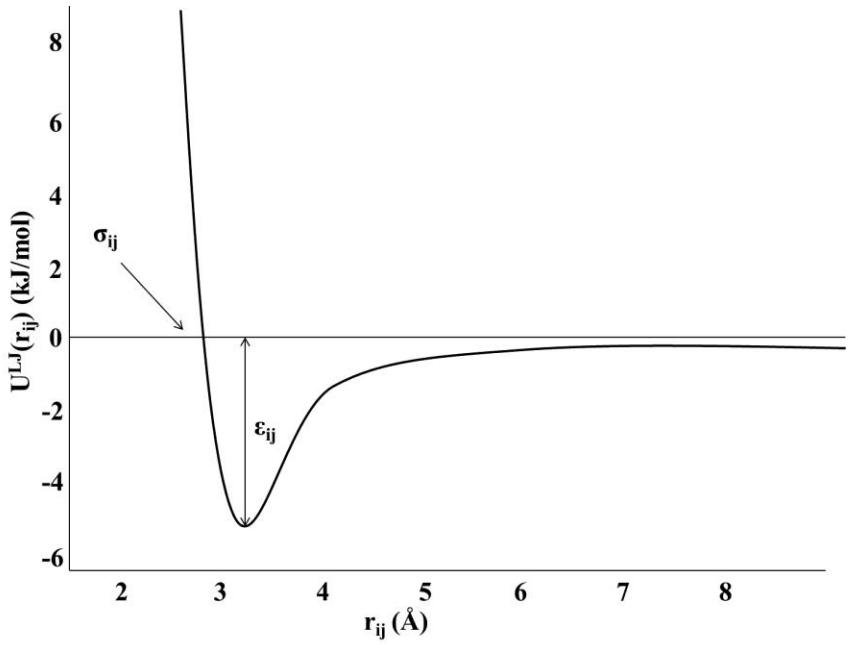

Figure 2-4. A schematic display of Lennard-Jones potential.

2.6 Periodic boundary conditions

Real chemical systems are composed of "moles" of molecules. However, because of the technical issues such as storage limitation or CPU speed, computer simulations are limited to a fewer number of molecules. Usually the number of atoms in MD simulations range from hundreds to hundred thousands. The atoms are usually confined in a space called the simulation cell. The simulation cell could adopt different shapes such as cubic, sphere or hexagonal. For such a small number of atoms in a simulation cell, a large portion of the atoms will be positioned

at the surface of the cell.[126] This will cause artificial effects at the boundaries of the simulation cell because the surface atoms do not experience similar forces as the atoms in the bulk. For real systems, the number of atoms is much larger and the ratio of the surface to the bulk atoms is close to zero. Therefore, for a real system the surface effects are supposed to be minimal, whereas for a simulation cell the effects could considerably alter the collected properties of the system. One way to overcome this problem is to increase the size of the simulation cell. However, this will slow down the calculations and often is not feasible. The second method is to apply a periodic boundary condition on the simulation cell. In this method, each simulation cell is replicated in all three dimensions of the space infinitely. As a result, an infinite 3D lattice will be formed. During the simulation if any atom moves outside of its simulation cell, its periodic image will move into the simulation cell from the opposite side. A 2D representation of such a process is shown in Figure 2-5.[112,113] Assuming the central box as the original simulation cell, if one atom leaves the cell from the right, its image from the left simulation cell will be added. In this regard, the simulation cell is practically infinite and each atom "feels" the bulk environment.

To apply the periodic boundary conditions, special considerations should be taken regarding the force calculations. For instance, an atom in a simulation cell should not interact with its own image in another simulation cell. Moreover, in an infinite lattice there exists infinite number of atoms and practically it is impossible to consider all of the atomic interactions. In order to resolve these issues, we need to modify the atom-atom interaction representation in the simulation cell. Here, we need to employ our physical knowledge about the way the atoms interact. For instance, Lennard-Jones interactions are known to be sharp short-range interactions.

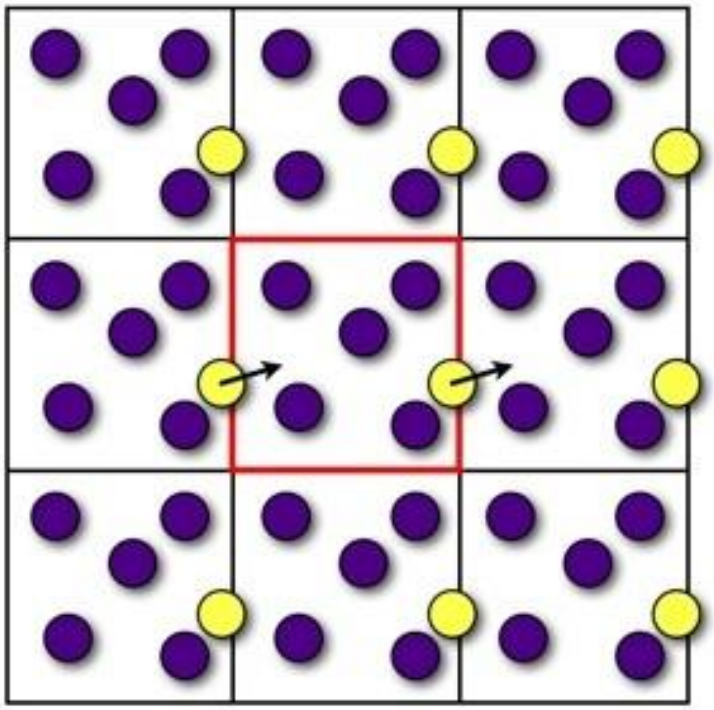

Figure 2-5. 2D representation of periodic boundary condition. The blue and the yellow circles represent the atoms remaining within and leaving the simulation cell. (Figure reproduced from http://matdl.org/repository/view/matdl:857), December 2012.

Figure 2-6 shows the interatomic potential for two typical atoms. At small interatomic distances this potential is highly repulsive resulting in a high positive potential energy. By increasing the distance the potential energy falls into the negative attractive region. Such region usually expands around 3-4 to 8-10 Å. As can be seen, the potential energy curve almost reaches a plateau at zero energy. At such a distance the interaction between two atoms is practically zero. That is, two atoms that are separated by approximately 10 Å, do not "feel" each other. As a result, one can just consider the interaction between atom pairs closer than 10 Å for Lennard-Jones force

calculations. A distance beyond which short-range interactions disappear is known as the cut-off radius (r_c). In an MD simulation, inter atomic distances between atom pairs are first calculated. If they fall within the cut-off radius, force will be calculated. A cut-off radius should satisfy two important conditions. First, it should be large enough so that only negligible interactions are ignored. Second, it should be smaller than the half of the simulation cell length to avoid the interaction between atoms with their own images. This way, the interaction will only be considered between atoms residing in the sphere with the radius of $r_c < \frac{L}{2}$,where L is the length of a cubic simulation cell.

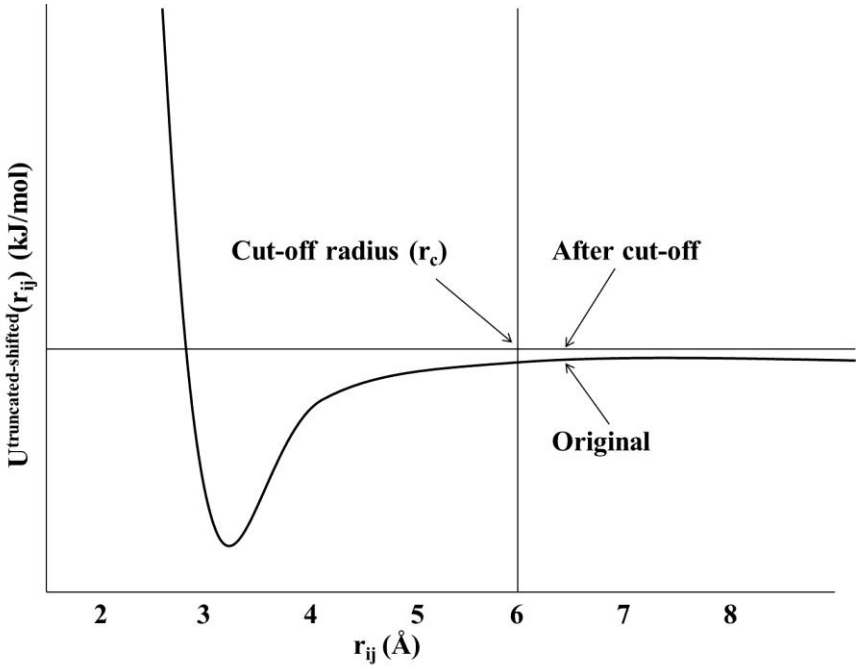

Figure 2-6. Definition of the cut-off distance for the Lennard-Jones potential.

Although this treatment is really helpful, it may cause other problems. As can be seen from Figure 2-6, having the radius cut-off causes the interatomic potentials to be non-zero and zero at distances smaller and larger than r_c, respectively. Mathematically, this causes a discontinuity at r_c. Since force is the derivative of the potential, it will jump to infinity at r_c. To overcome this problem, one needs to redefine the potential term:

$$U^{truncated-shifted}(r_{ij}) = U^{LJ}(r_{ij}) - U^{LJ}(r_c) \qquad \text{when } r_{ij} \leq r_c$$
$$U^{truncated-shifted}(r_{ij}) = 0 \qquad \text{when } r_{ij} \geq r_c \qquad [2.35]$$

With this potential function, both the potential and force of the system will change smoothly. This effect is similar to shifting the whole potential of the system by a certain value.

2.7 Ewald summation

In contrast to Lennard-Jones short-range interactions, electrostatic interactions are long-range. They decay with the inverse of the distance, while short-range interactions decay with r^{-6}. As a result, they should be treated differently from short-range interactions. In fact, electrostatic interactions constitute the most time consuming part of an MD simulation. In a cubic simulation cell, the electrostatic potential is defined as:

$$U^{Coulomb} = \frac{1}{2}\sum_{i=1}^{N} Q_i \Phi(\vec{r}_i) \qquad [2.36]$$

where N is the total number of charged atoms, Q_i is the partial atomic charge, and $\vec{r}_i$ is the atomic position, respectively. $\Phi(\vec{r}_i)$ is defined as the electrostatic potential at the position of atom i:

$$\Phi(\vec{r}_i) = \sum_{j=1}^{N} \sum_{\vec{n}} \frac{Q_j}{|\vec{r}_{ij}+\vec{n}L|} \quad [2.37]$$

where L is the simulation cell length and $\vec{n} = (n_x, n_y, n_z)$ with integer values for n_i that contains the positions of atoms in the replicated simulation cells. The condition in which $i = j$ is not counted in the summation since it represents the interaction of an atomic charge with itself. In contrast to the short-range potentials, the interaction with other replicated cells is included in the calculation of $\Phi(\vec{r_i})$ for long-range potentials. In this regard, the summation will not converge because it depends on the order that each term in Equation 2.20 is added.

Ewald summation is a solution to avoid this problem.[127] This summation adds all of the long-range interactions between infinite number of particles and all of the periodic images. In this method one assumes that each atomic charge is surrounded by a screening charge distribution of opposite sign and equal magnitude. Such a charge distribution is normally chosen to be a Gaussian distribution. Figure 2-7 shows the evaluation of Ewald summation. Ewald summation decomposes the long-range interactions into three separate terms:

$$U^{el} = U^r + U^k + U^0 \quad [2.38]$$

The first term consists of all interactions between atomic charges and also the screening potentials. This real space summation is defined as:

$$U^r = \frac{1}{2} \sum_{i \neq j}^{N} \sum_{\vec{n}} Q_i Q_j \frac{erfc(\sqrt{\alpha}|\vec{r}_{ij}+\vec{n}L|)}{|\vec{r}_{ij}+\vec{n}L|} \quad [2.39]$$

where $erfc(x)$ is the complementary error function and α is a parameter to adjust the width of the Gaussian distribution.

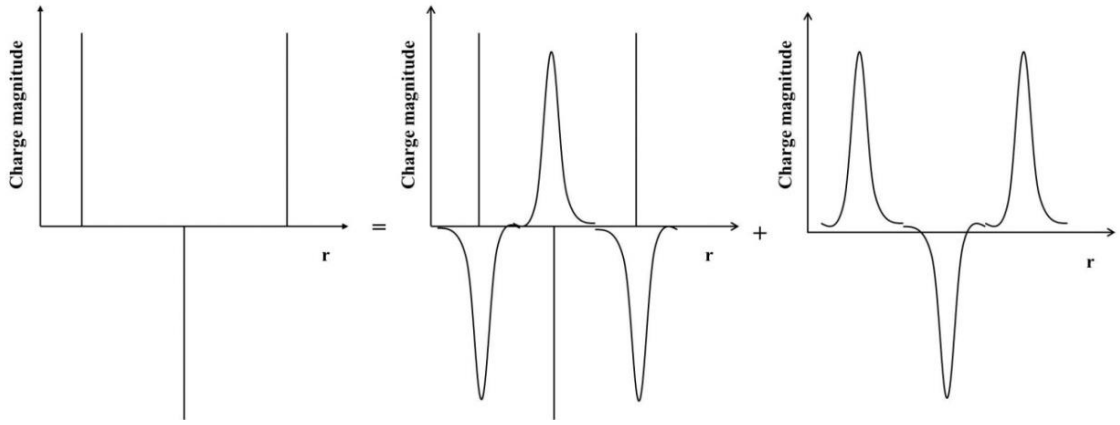

Figure 2-7. Representation of the charge distribution in Ewald summation. Vertical lines and Gaussian curves represent point charges and screening charge distributions, respectively.

The second term in Equation 2.21 neutralizes the Gaussian distribution term. This reciprocal space summation is defined as:

$$U^k = \frac{1}{2\pi L^3} \sum_{i,j=1}^{N} \sum_{\vec{k}\neq 0} Q_i Q_j \frac{4\pi}{k^2} \exp(-\frac{k^2}{4\alpha^2}) \cos(\vec{k}.\vec{r}_{ij}) \qquad [2.40]$$

where $\vec{k} = (\frac{2\pi n_x}{L}, \frac{2\pi n_y}{L}, \frac{2\pi n_z}{L})$. The last term in Equation 2.21 cancels out the effect of all interactions between artificial counter-charges with themselves:

$$U^0 = -\sqrt{\frac{\alpha}{\pi}} \sum_{i=1}^{N} Q_i^2 \qquad [2.41]$$

Finally, the Ewald summation can be written as:

$$U^{el} = \frac{1}{2}\sum_{i\neq j}^{N}\sum_{\vec{n}} Q_i Q_j \frac{erfc(\sqrt{\alpha}|\vec{r}_{ij}+\vec{n}L|)}{|\vec{r}_{ij}+\vec{n}L|} + \frac{1}{2\pi L^3}\sum_{i,j=1}^{N}\sum_{\vec{k}\neq 0} Q_i Q_j \frac{4\pi}{k^2}\exp(-\frac{k^2}{4\alpha^2})\cos(\vec{k}.\vec{r}_{ij}) - \sqrt{\frac{\alpha}{\pi}}\sum_{i=1}^{N} Q_i^2 \quad [2.42]$$

In this equation $\vec{k}$ and α are parameters to be adjusted and should be selected carefully. Usually, they are chosen so that the Ewald summation converges rapidly. Also, α should be adjusted in a manner that the first term in the Ewald summation only needs the central simulation cell and not any images. In the derived Ewald summation it is assumed that the simulation cell is cubic. For non-cubic simulation cells additional corrections may be needed to account for the shape of the cell.[112,113]

2.8 Free energy calculations

MD simulations provide useful information about the dynamical behavior of the system at the atomic level. An analysis of the trajectory data of the simulation can explain many structural features including bonds, angles, and dihedrals as well as hydrogen bonding between atoms. Although MD simulations can always provide useful information regarding the mechanism and dynamics of the system, a majority of the experimental characteristics can be best understood from the free energy information of the system.

Free energy calculations are broadly used to investigate the energetics of many phenomena such as protein folding, protein-protein interactions, and protein-ligand binding as well as ion transport in ion channels. The free energy of the system can describe how probable it

is to find the system in a given state. To be more convenient, the free energy difference between two states is usually calculated. The free energy difference between various states as a function of a particular reaction coordinate is referred to as the potential of mean force (PMF) profile. Typically, a PMF profile may have a global minimum in addition to multiple local minima. Although different approaches are now available to obtain free energy information for various systems, accurate estimation of free energy in some large biological systems is still challenging. The main issue is how to overcome large free energy barriers and obtain a reasonable sampling that represents a uniform distribution along the reaction pathway. In general, the free energy difference between two states can be expressed as:[128]

$$\Delta A = A_1 - A_0 = -k_B T ln \frac{P_1}{P_0} \qquad [2.43]$$

where *0* and *1* are the initial and final states, k_B is the Boltzmann's constant, T is the absolute temperature, and P_0 and P_1 are the probabilities of finding the system in states *0* and *1*, respectively. Apparently, in order to find the free energy difference between states *0* and *1*, the system has to overcome the barrier between them. On the time scales typically accessible by brute-force classical MD, it may be impossible to overcome all energy barriers, particularly, when large barriers are present. Poor sampling along the reaction pathway will produce unreliable free energy estimation. In order to improve the sampling, several techniques have been developed. The methodological background of some free energy calculation techniques will be discussed in the following sections.

2.8.1 Thermodynamic integration

Thermodynamic integration (TI)[129-134] is a general method of calculating a PMF profile between two states of the system. This method calculates the free energy difference between states *0* and *1* by integrating the average force along the reaction pathway. TI includes constrained MD methods such as adaptive biasing force (ABF)[135,136] and steered molecular dynamics (SMD).[137-141] Here, the Hamiltonian of the system, is defined as a function of a continuous parameter, λ. This parameter can adopt different values between *0* to *1* corresponding to states *0* and *1*, respectively. In practice, states *0* and *1* are connected by a series of states having discrete but close values of λ_i. For each state, the system is allowed to relax and then the average force is calculated by integrating the force between states *0* and *1*:

$$\Delta A^{TI}(0 \rightarrow 1) = \int_{\lambda=0}^{\lambda=1} \langle \frac{\partial H}{\partial \lambda} \rangle_{\lambda} \, d\lambda \approx \sum_{i=1}^{n-1} \langle \frac{\partial H}{\partial \lambda} \rangle_{\lambda_i} \qquad [2.44]$$

where H is the Hamiltonian of the system and $< >_{\lambda}$ is the ensemble average at each value of λ. This method is always associated with large errors if the adjacent steps connecting states *0* and *1* are structurally different. In fact, the transition between states *0* and *1* should be divided into enough steps so that a transition could be considered smooth. Luckily, free energy of the system is a state function. This facilitates the process of choosing a trajectory that connects states *0* and *1*. Any physical or non-physical pathway could be taken to bring the system from state *0* to state *1*. However, in practice, because of the limitations imposed by the second law of thermodynamics this transition process should be carried out along a reversible pathway. The accuracy of TI highly depends on sampling. Since these methods deterministically explore the phase space,

inadequate sampling will result in erroneous results. Accuracy of this method could be monitored by measuring the convergence of repeated runs of the calculations.[142,143]

Among several methods to calculate the free energy such as free energy perturbation (FEP),[144] umbrella sampling (US),[113,145] ABF, and metadynamics,[146] we will describe the ABF and US methods in the following sections as these two methods will be extensively used in this thesis.

2.8.2 The ABF method

TI in association with unconstrained MD and average force can be used to obtain the PMF profiles. PMF can be defined as the work required to bring two solvated particles from infinity to a contact distance.[147] This could be generalized to any type of reaction coordinate, λ. The free energy as a function of the reaction coordinate can be presented as:

$$A(\lambda) = -(\frac{1}{\beta}) \ln P(\lambda) + A_0 \qquad [2.45]$$

where $A(\lambda)$ is the free energy at state λ, $P(\lambda)$ is the probability distribution to find the system at state λ along the reaction coordinate, and A_0 is a constant. $P(\lambda)$ can be defined as:

$$P(\lambda) = \int \delta[\lambda - \lambda(X)]\, exp[-\beta H(X, P_x)]\, dX dP_x \qquad [2.46]$$

The first derivative of the free energy with respect to the reaction coordinate can be written as:[148]

$$\frac{dA(\lambda)}{d\lambda} = \langle \frac{\partial \upsilon(X)}{\partial \lambda} \rangle_{\lambda} \quad [2.47]$$

where $\upsilon(X)$ is the potential energy. Equation 2.46 can also be written as two separate parts including configurational and kinetic terms:

$$P(\lambda) = \int det(J)\, exp[-\beta v(q,\lambda)]\, dq \int exp[-\beta \mathrm{T}(P_x)]\, dP_x \quad [2.48]$$

where $\upsilon(q,\lambda)$ is the potential energy and is a function of the reaction coordinate, λ and the generalized coordinates, q. $T(P_x)$ and $det(J)$ stand for the kinetic energy and the determinant of the Jacobian for the inverse transformation from generalized to Cartesian coordinates, respectively. Using Equation 2.45, the derivative of the free energy with respect to the reaction coordinate can be expressed as below:

$$\frac{dA(\lambda)}{d\lambda} = -\frac{1}{\beta} \frac{d \ln P(\lambda)}{d\lambda} = -\frac{1}{\beta} \frac{1}{P(\lambda)} \frac{dP(\lambda)}{d\lambda} \quad [2.49]$$

Using Equation 2.48, one can find $\frac{dP(\lambda)}{d\lambda}$ as below:

$$\frac{dP(\lambda)}{d\lambda} = \int exp[-\beta v(q,\lambda)] \left\{ -\beta\, det(J) \frac{\partial v(q,\lambda)}{\partial \lambda} + \frac{\partial\, det(J)}{\partial \lambda} \right\} dq \times \int exp[-\beta \mathrm{T}(P_x)]\, dP_x \quad [2.50]$$

Now, by combining Equations 2.49 and 2.50, the kinetic term vanishes and Equation 2.49 can be written by defining an integral over λ^* and a Dirac function as below:

$$\frac{dA(\lambda)}{d\lambda} = -\frac{1}{\beta} \frac{1}{P(\lambda)} \int \exp[-\beta \upsilon(q,\lambda^*)] \delta(\lambda^* - \lambda) \times \left\{ -\beta\, det(J) \frac{\partial v(q,\lambda^*)}{\partial \lambda} + \frac{\partial det(J)}{\partial \lambda} \right\} dq d\lambda^* \quad [2.51]$$

The back transformation into Cartesian coordinates, results in the following equation:

$$\frac{dA(\lambda)}{d\lambda} = -\frac{1}{\beta}\frac{1}{P(\lambda)}\int \exp[-\beta v(X)]\delta[\lambda(x)-\lambda)] \times \left\{-\beta\frac{\partial v(X)}{\partial \lambda} + \frac{1}{det(J)}\frac{\partial det(J)}{\partial \lambda}\right\} dX \qquad [2.52]$$

And finally, the derivative of the free energy with respect to the reaction coordinate can be shown in terms of configurational averages at constant λ:

$$\frac{dA(\lambda)}{d\lambda} = \langle\frac{\partial v(X)}{\partial \lambda} - \frac{1}{\beta}\frac{\partial \ln det(J)}{\partial \lambda}\rangle_{\lambda} = -\langle F_{\lambda}\rangle_{\lambda} \qquad [2.53]$$

where $<F_{\lambda}>_{\lambda}$ is the average force acting along the reaction coordinate. In the ABF simulations, the first step is to define the reaction coordinate. The reaction coordinate itself could be divided into a number of non-overlapping windows. This technique is very useful to improve the sampling in the systems in which different regions are separated by high energy barriers. Often, in theses systems some areas along the reaction pathway are inaccessible or hard to be sampled smoothly during the simulations. Therefore, by dividing the reaction coordinate into small pieces, one can perform several simulations simultaneously. This guarantees faster convergence of force as well as reliable sampling histograms. Another parameter to be determined is the number of bins within each window. This determines the resolution of the free energy profiles. It is important to select a number that provides a satisfactory resolution while the computation time is still affordable. The more number of bins requires the more number of average force calculations which can highly slow down the calculations.

The ABF method is based on the calculation of the average force along the reaction coordinate. In practice, the instantaneous force is accumulated in small bins of finite size $\delta\lambda$ and the average force is calculated as shown in Equation 2.53. The average force is then canceled out

by applying a biasing force with the equal magnitude and opposite sign of the calculated average force. This provides the opportunity for the system to overcome the barrier even if it is located in a minimum along the reaction coordinate:

$$F^{ABF} = \nabla_x \tilde{A} = -\langle F_\lambda \rangle_\lambda \nabla_x \lambda \qquad [2.54]$$

Here, F^{ABF} and $\tilde{A}$ are the biasing force applied by the ABF method and the current estimate of the free energy of the system, respectively. During the simulation, the estimate of $\nabla_x \tilde{A}$ is continuously updated.

To avoid the large fluctuations of the force once the system just locates at the new position, the ABF force is not applied until a user defined threshold is reached. Ideally, the ABF force is equal to the average force exerted on the ion so that the evolution of the system along the reaction pathway is only determined by the diffusion properties and only the instantaneous force on the system remains. The simulation should be continued until the force converges and the sampling becomes smooth along the reaction coordinate. Figure 2- shows a schematic display of the simulation setup in the ABF method.

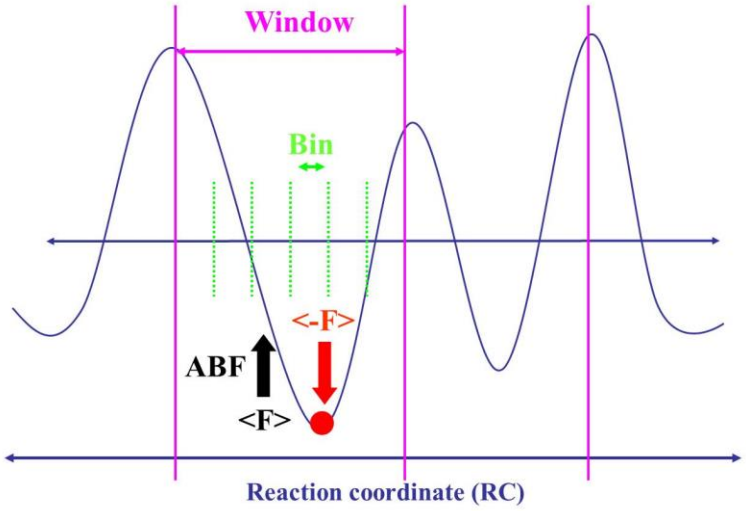

Figure 2-8. A schematic illustration of the ABF method. The windows and the bins are distinguished by solid and dotted vertical lines, respectively.

2.8.3 The US method

Free energy is the driving force for physical or chemical processes. The free energy calculation provides information about the stability of two states and the energy barrier between them. The key, however, is to explore the phase space sufficiently. Poor sampling will lead to inaccurate results. The free energy of system A is related to the canonical partition function Q in the following fashion:

$$A = -\frac{1}{\beta}\ln Q \qquad [2.55]$$

where $\beta = (K_B T)^{-1}$. In NPT (constant-temperature, constant-pressure) ensembles, G is used for free energy instead of A. In the condensed phase where the pressure of the system hardly changes,

the difference between *A* and *G* is very small. Since exploring the entire phase space requires infinite time, there is a need to define a continuous parameter, λ, that connects the two states. In principle, λ could be a variable of one-, two- or higher dimensions. λ is often defined by geometrical properties such as distance, dihedral angle, or the difference between the root-mean-square-deviations (RMSD) from two reference states. By defining the trajectory connecting the two states, one may also define the probability distribution along λ that includes all other degrees of freedom:

$$Q(\lambda) = \frac{\int \delta[\lambda(r)-\lambda]\, exp[(-\beta E) d^N r]}{\int exp[(-\beta E) d^N r]} \qquad [2.56]$$

where $Q(\lambda)\, d\lambda$ is the probability of finding the system at the interval of $d\lambda$ around λ. In MD simulations, we cannot integrate phase-space equations. Alternatively, assuming the ergodic behavior of the system, only a path connecting the two states should be sufficiently sampled. As a result, $A(\lambda)$ can be obtained from MD simulations by monitoring the probability distribution of the system along λ.

The US method is based on the comparison of the probability density function of the system at each point along the pathway connecting states *0* and *1*. One practical consideration is that because of the rough and complex potential energy surface of the system, it is impossible to sample the entire pathway from states *0* to *1* sufficiently. In fact, areas with low energy will be sampled more, while regions with high repulsive forces will be explored with much less frequencies. Such high energy regions in the phase space (significantly larger than k_BT) are supposed to be experienced less and they constitute "rare" events. To obtain a complete profile, rare events must also be sampled.

Among several methods to improve sampling efficiency, introducing an additional force as a new degree of freedom has been shown to be effective.[146,149-152] Alternative methods include modifying the Hamiltonian of the system to reduce the energy barrier. In the US method, the reaction coordinate is divided into several individual windows in which a bias potential is added to keep the system within each desired window. The probability distributions of finding the system in each window are ultimately combined using methods such as weighted histogram analysis method (WHAM)[153-155] or umbrella integration.[156,157] The bias potential should be chosen carefully. It should be strong enough to let the system experience high energy states and week enough to provide a reasonable sampling.[158,159] Also, the bias potential should be chosen so that two adjacent windows exhibit sufficient overlap and could energetically be connected to each other. The bias potential, w_i, could adopt different values for different windows:

$$U^b(r) = U^u(r) + w_i(\lambda) \qquad [2.57]$$

where *b* and *u* stand for biased and unbiased potentials. The biased probability distribution for each window is:

$$P_i^b(\lambda) = \frac{\int exp\{-\beta[U(r)+w_i(\lambda'(r))]\}\delta[\lambda'(r)-\lambda]d^N r}{\int exp\{-\beta[U(r)+w_i(\lambda'(r))]\}d^N r} \qquad [2.58]$$

To calculate the PMF we should use unbiased distributions:

$$P_i^u(\lambda) = \frac{\int exp[-\beta U(r)]\delta[\lambda'(r)-\lambda]d^N r}{\int exp[-\beta U(r)]d^N r} \qquad [2.59]$$

Eventually, the free energy of the system can be obtained from the following equation:

$$A_i(\lambda) = -(\frac{1}{\beta}) \ln P_i^b(\lambda) - w_i(\lambda) + F_i \quad [2.60]$$

where

$$F_i = -(\frac{1}{\beta}) \ln\langle exp[-\beta w_i(\lambda)]\rangle \quad [2.61]$$

The only requirement for successful application of the US method is sufficient sampling, which highly depends on the choice of appropriate reaction coordinate and bias potential. An ideal bias potential should provide smooth sampling along the reaction coordinate. If one has prior knowledge about the shape of the free energy profile, the negative of the free energy could be used as the bias potential. Unfortunately, the PMF profile is not known ahead in most cases and one should use other forms of bias potentials. Typically, in each window along the reaction coordinate, a harmonic bias potential is applied to keep the system within the window. This bias potential has the following form:

$$w_i(\lambda) = \frac{k_i}{2}(\lambda - \lambda_i^{ref})^2 \quad [2.62]$$

where *k* is the force constant and λ_i^{ref} is the center point of the window *i*. This harmonic potential is simple and efficient. The choice of *k* is critical and a decision has to be made on the magnitude of *k* before running the simulation. Normally, when there is no prior information about the shape of the PMF profile, a reasonable guess for *k* can be based on available data for similar systems. Then the overlap between two adjacent windows should be examined after simulation data are

produced. If the windows do not overlap sufficiently, the *k* value should be reduced. On the other hand, if the system deviates significantly from its reference point in each window, the value of *k* should be increased. In the US method, one can also use different *k* values for different windows. A sufficient overlap between two adjacent windows is a necessity if one wants to apply WHAM for final probability distribution analysis, whereas it is not necessary in the umbrella integration method.[156,157]

2.8.4 WHAM

Assuming no added bias potential, the high energy regions of the system may never be sampled during the MD simulations. On the contrary, the low energy areas that represent the stable states of the system will be sampled frequently. The insufficient sampling of some regions between states *0* and *1* will result in inaccurate PMF profiles. As discussed, adding bias potentials to each window can improve the sampling efficiency and is expected to produce a uniform sampling. In a typical window, w_i, the system movement is limited by the magnitude of the bias potential. Since a harmonic potential is a steep function, it is unlikely that the system moves to regions very far from the reference point of window *i*. As a result, the sampling will happen more or less around the reference point of window *i* and the probability of finding the system will be reduced as the deviation from reference point increases (see Figure 2-9). For the adjacent window, w_{i+1}, a similar shape of the probability distribution is expected although it highly depends on the system energy at that specific window and the magnitude of the bias potential. Again, at areas around the reference point of w_{i+1} a higher probability distribution is expected.

In a WHAM analysis, two distributions from the two adjacent windows should efficiently overlap. This will help to improve the sampling of the areas far from the reference points of the

adjacent windows. Eventually, an approximately uniform sampling along the chosen reaction coordinate will yield relative free energy results.

From statistical mechanics one can simply link the associated free energy of each state to its probability distribution. However, such distributions are supposed to be unbiased, while the distributions obtained from added bias potentials are biased. WHAM is a tool to obtain free energy profiles from biased probability distributions. In fact, WHAM calculates the unbiased probability distributions from biased probability distributions produced in trajectory data during the simulation. The unbiased probability distribution of each window can be written in terms of the biased probability:

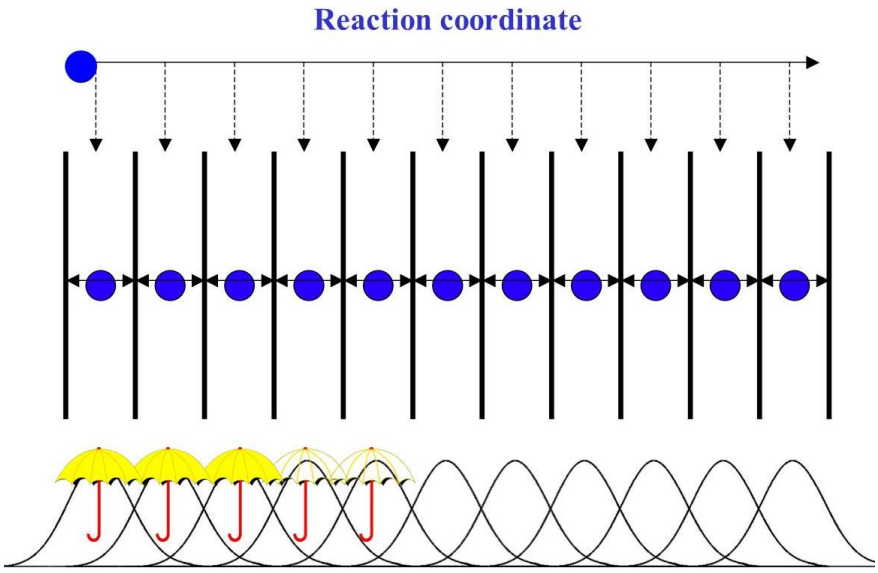

Figure 2-9. A schematic diagram of the US method. The windows are distinguished by vertical lines. A dotted arrow shows the center of each window. A blue circle represents the atom within each window. (Figure reproduced from http://www.bevanlab.biochem.vt.edu/Pages/ Personal/justin/ gmx-tutorials/umbrella/05_pull.html), December 2012.

$$P_i^u(\lambda) = e^{\beta(w_i(\lambda)-f_i)}P_i^b(\lambda) \quad [2.63]$$

where $\beta = \frac{1}{k_B T}$, $P_i^u(\lambda)$ is the unbiased probability of window *i*, $w_i(\lambda)$ is the bias potential of window *i*, P_i^b is the biased probability of window *i*, and f_i is the free energy obtained by adding the bias potential and is equal to:

$$e^{-\beta f_i} = \int e^{-\beta w_i(\lambda)}P(\lambda)d\lambda = \langle e^{-\beta w_i(\lambda)}\rangle \quad [2.64]$$

Apparently, these two equations cannot be directly used to calculate the unbiased probability. WHAM uses a linear combination of unbiased probabilities:

$$P(\lambda) = \sum_{i=1}^{N} c_i(\lambda)P_i(\lambda) = \sum_{i=1}^{N} c_i(\lambda)(e^{\beta(w_i(\lambda)-f_i)}\, P_i^b(\lambda)) \quad [2.65]$$

where $c_i(\lambda)$ is the weight of the probability distribution in each window. Now, one has to find $c_i(\lambda)$ values. These parameters are first subjected to normalization:

$$\sum_{i=1}^{N} c_i(\lambda) = 1 \quad [2.66]$$

Moreover, they are defined by minimizing the variance in the $P(\lambda)$ with respect to the weights:

$$\frac{\partial \sigma^2[P(\lambda)]}{\partial c_i} = 0 \qquad \text{when i=1…N} \quad [2.67]$$

Therefore, the resulting weights are as below:

$$c_i(\lambda) = \frac{n_i e^{-\beta(w_i(\lambda)-f_i)}}{\sum_{j=1}^{N} n_j e^{-\beta(w_j(\lambda)-f_i)}} \quad [2.68]$$

where n_i is the number of samples in the window i. However, f_i itself depends on the $P(\lambda)$. This inter-dependence should be solved iteratively until the convergence is reached:

$$P(\lambda) = \frac{\sum_{i=1}^{N_{sims}} n_i(\lambda)}{\sum_{i=1}^{N_{sims}} N_i \exp\left(\frac{\left[F_i - U_{bias,i}(\lambda)\right]}{K_B T}\right)} \quad [2.69]$$

$$F_i = -K_B T \ln\left(\sum_{\lambda_{bias}} P(\lambda)\, exp\left(\frac{-U_{bias,i}(\lambda)}{K_B T}\right)\right) \quad [2.70]$$

where F_i is an undetermined parameter that should be calculated by adjusting the various free energies of the two adjacent overlapping windows. To solve such equations iteratively and achieve consistency, an adequate number of points on the reaction coordinates are needed. If the number of points, especially in the overlapped area, is not enough, the produced PMF profile would be inaccurate. Figure 2-10 shows a comparison between sufficient and insufficient overlaps and the corresponding PMF profiles.[128] Ideally, the bias potential should keep the system within each window, while providing enough overlap between two adjacent windows. Also, enough number of windows should be chosen so that the overall reaction coordinate is covered with a reasonable overlap between two adjacent windows.

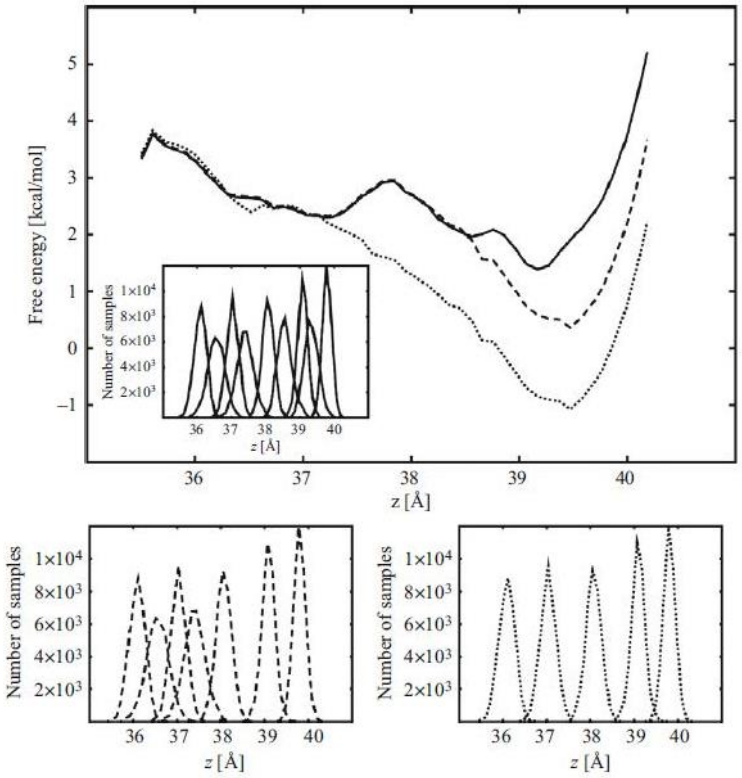

Figure 2-10. Comparison between a good (solid line), a moderate (dashed line), and a poor (dotted line) sampling and the corresponding PMF profiles. (Figure reproduced from Ref. 128).

Chapter 3

Adaptive biasing force MD studies of ion movement in $[d(TG_4T)]_4$

3.1 Introduction

As mentioned in Chapter 1, the common feature of all G-quadruplex structures is the stacking G-quartets that form a channel-like central pore. Formation and stability of these channel-like structures depend critically on the presence of certain cations such as Na^+, K^+, and NH_4^+ inside the central channel.[10,160-163] Inside the channel, each ion is fully dehydrated and coordinated to the pore-lining guanine carbonyl O6 atoms. While the modes of ion-binding inside G-quadruplex channels have been established by many X-ray crystallographic and NMR studies, much less is known about the energetic properties of ion translocation through a G-quadruplex DNA channel. In this regard, detailed information about the movement of NH_4^+ ions through G-quadruplex channels can be obtained from ^{15}N NMR studies as originally demonstrated by Feigon, Hud, and co-workers[41-43] and more recently by Plavec and co-workers.[164-170] Other ions such as Na^+, K^+, Rb^+, Tl^+, and Ca^{2+} are much more difficult to study and, as a result, only a few scattered NMR studies have been reported in the literature.[31,32,36,37,44,171-174] All these previous studies have shown that ion movement in G-quadruplex DNA generally occurs on a time scale between milliseconds and seconds. As ion movements on such a long time scale are always associated with steep features in the free energy landscape, it is rather difficult to achieve satisfactory sampling in the configuration space by classic brute-force MD simulations, although we should note that conventional MD simulations have been used previously to generate useful insight into various aspects of G-quadruplex DNA.[87,92,95,99,175-179] While extensive computational studies can be found in the literature dealing with ion transport through ion channels such as K^+

ion channels,[52,128,180-183] cyclic peptide nanotubes,[83-85] and carbon nanotubes,[78-82] modern free energy methods have not been used to explicitly evaluate the energetic properties of ion movement through any G-quadruplex DNA channel. In this chapter, we employ the adaptive biasing force (ABF) method in unconstrained MD simulations to obtain, for the first time, the 3D free energy landscapes for ion transport (Na^+, K^+, and NH_4^+) through the $[d(TG_4T)]_4$ G-quadruplex DNA channel. As illustrated in Figure 3-1, $[d(TG_4T)]_4$ is a parallel-stranded tetramolecular G-quadruplex consisting of four stacked G-quartets.[169]

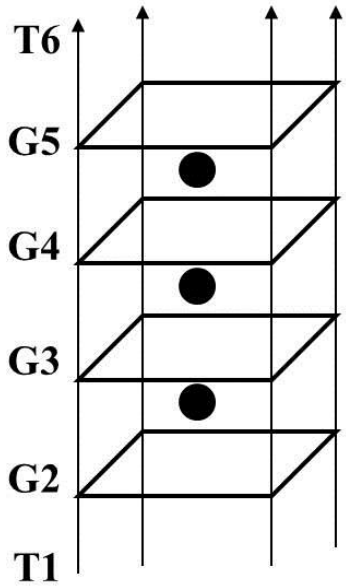

Figure 3-1. G-quadruplex DNA structure of $[d(TG_4T)]_4$. T and G represent thymine and guanine residues, respectively. A square indicates a G-quartet and a circle represents a channel cation. The arrows indicate the DNA strand direction from 5' to 3'.

3.2 Computational details

3.2.1 Model structure generation

An initial model of the $[d(TG_4T)]_4$ G-quadruplex was generated from the NMR structure of a ligand-quadruplex complex (PDB code 2JT7)[184] by removing the ligand molecules and then adding three ions into the channel pore. The G-quadruplex structure was also neutralized by adding 17 counter ions. Each system was immersed in a box of TIP3P water molecules[112,113] extending up to 10 Å from the solute in each direction. Each of the systems was optimized and equilibrated using multiple minimization steps. A 10000-step minimization of water and counter ions was carried out on each system with solute and channel ions restrained by a force constant of 500 kcal/mol/Å^2. Subsequently, a series of successive 20 ps restrained minimizations were performed with the channel ions restrained by 500 kcal/mol/Å^2 and solute restrained by 300, 100, 10, and 0 kcal/mol/Å^2 force constants. The positions of the channel ions were relaxed during four energy minimizations (20 ps each) in which the force constant acting on channel ions was gradually reduced from 500 to 300, 100, 10, and 0 kcal/mol/Å^2. The systems were heated to 310 K over 50 ps at constant pressure while maintaining a force constant of 50 kcal/mol/Å^2 on solute and channel ions. These restraints were removed in three stages (50 ps each) from 50, 30, and 0 kcal/mol/Å^2 at 310 K. Eventually, the restraints were fully removed during a 50 ps equilibration.

All MD simulations were performed in the isothermic-isobaric ensemble.[112,113] Langevin dynamics and a Langevin piston algorithm[112,113] were used to keep the temperature and pressure at 310 K and 1 atm, respectively. The particle mesh Ewald (PME) method[112,113] of calculating long-range electrostatic interactions was used with a cut-off of 9 Å. Rattle was applied to constrain the bonds containing hydrogen. A time step of 2 fs was used to integrate the equations of motion. All

MD simulations were performed using the program NAMD2.7b3[177,185] and the CHARMM27 force field[186,187] at the SHARCNET computing facilities on HP servers running CentOS 5.x with Xeon 2.83 GHz processors (each having 16 GB of RAM). Typically, eight processors were used for each calculation. The program VMD 1.8.7[188] was used for visualization and analysis. Figure 3-2 shows the minimized $[d(TG_4T)]_4$ G-quadruplex DNA structures used in this study.

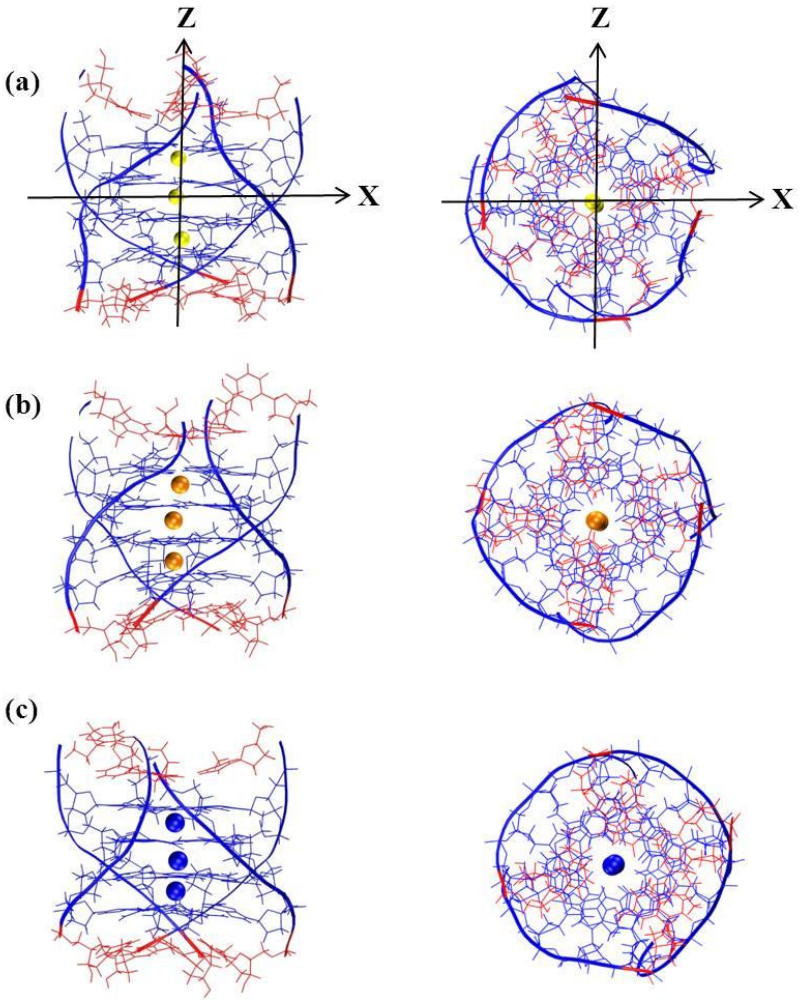

Figure 3-2. (Left) Side and (right) top views of the minimized structures of (a) Na^+, (b) K^+, and (c) NH_4^+ forms of $[d(TG_4T)]_4$. The arrows represent the directions of Z and X axes.

3.2.2 ABF computations

The ABF method was used to determine the free energy profiles. In this method, an average force experienced by an ion at any point along the reaction coordinate (λ) is calculated. The average force is accumulated in bins along λ and is continuously updated as the simulation progresses. A biasing force with the equal magnitude and opposite sign of the average force is then applied to the ion. The application of this biasing force allows the ion to overcome existing free energy barriers along λ. In a typical ABF simulation, application of the biasing force is initiated after accumulation of 800 samples in each bin.

To obtain a complete potential of mean force (PMF) profile for ion movement, it is necessary to separate the ion movement into two regions. The first region is inside the channel. The second region is from the channel exit/entrance point to bulk. In the first region, we chose the reaction coordinate along the channel axis (Z) of $[d(TG_4T)]_4$ and performed simulations in three separate non-overlapping windows, each being 4 Å in width. In each window, the channel axis is defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms from each of the two adjacent G-quartets. The ion movement in the plane norm to the dynamic axis was confined by a harmonic potential (k = 100 kcal/mol/Å^2). In some cases, two additional windows were calculated to extend the range of Z to $+12 \geq Z \geq -12$ Å. Within each window the force acting on the ion was averaged in 0.1 Å bins. Typical production runs of 3 ns in each window were performed. In some cases, MD simulations were extended to 20 ns. The final PMF profile was obtained through integration of the average forces over λ. In the second region, we performed ABF-MD simulations along eight reaction pathways within the XY plane for a series of Z positions. For each Z position, we chose (X,Y) = (1,0), (−1,0), (0,1), (0,−1), (1,1), (−1,1), (1,−1), and (−1,−1). A 2D PMF profile was reconstructed from these 1D data. Weak harmonic restraints

(with a force constant of 3 kcal/mol/Å^2) were applied to the C1' atoms of guanine residues for the two central G-quartets.

3.3 Results and discussion

3.3.1 Force convergence assessment

To test force convergence, we performed ABF-MD simulations over 20 ns. As an example, the distance RMSD of the backbone atoms with respect to the initial structure for the central window (defined between G3 and G4) is shown in Figure 3-3. The overall backbone structure rarely deviates more than 3.5 Å from the initial structure which is mostly due to the thymine backbone movements at both ends of the $[d(TG_4T)]_4$ structure. This suggests that the core G-quadruplex structure is highly stable after only a few hundreds of ps with deviation less than 2.5 Å from the initial structure. No considerable core backbone perturbation is observed after 1 ns. Figure 3-4 shows the force distribution over the entire $[d(TG_4T)]_4$ channel as a function of simulation time. Although the force profiles in some regions do not change significantly over time, they differ by 2-3 kcal/mol/Å between 3 and 20 ns when the ion is located close to the central G-quartets (G3 and G4). This deviation could be attributed to the fact that only a single ion is present in this model and the overall structure may slightly move away from the original configuration. It appears that the central region of the channel responds more to the presence of only one ion because the deviations are about 3 kcal/mol/Å at Z = ±0.75 Å, they reduce to 2 kcal/mol/Å at Z = ±2.5 Å, and finally they become negligible toward the outer regions of the channel. Within the range of accuracy of our data, a force deviation of 2-3 kcal/mol/Å is acceptable at the moment considering the fact that there is only one ion in the channel. Since we

are studying ion transport along G-quadruplex channels for the first time, we start with the simple single-ion models and will extend our investigations to the multiple-ion models which more accurately represent real situations. Also, most of our studies will be performed for no longer than 3 ns as continuing up to 20 ns will only change the results by 2-3 kcal/mol/Å for the central regions.

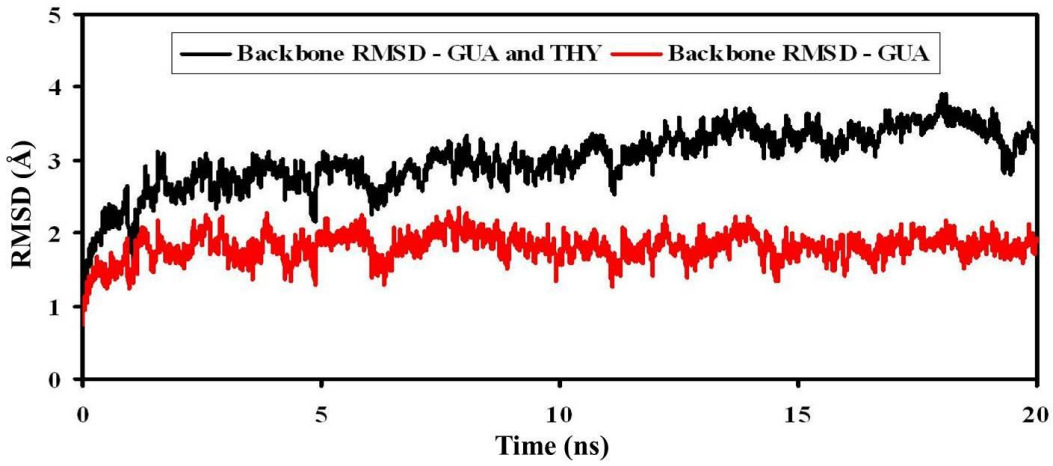

Figure 3-3. RMSD of the backbone for guanine (GUA) and thymine (THY) residues (black) and guanine residue (red) with respect to the initial structure in a typical ABF-MD simulation for the $[d(TG_4T)]_4$ G-quadruplex.

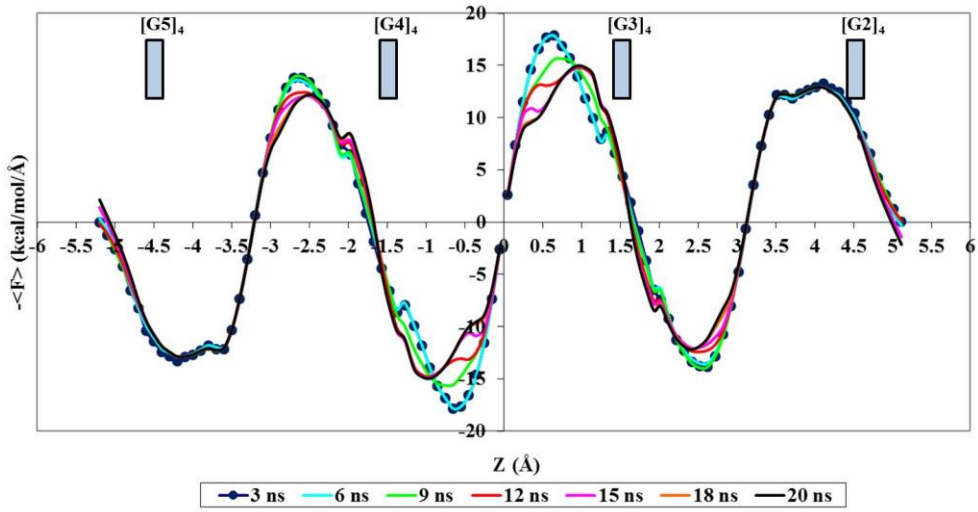

Figure 3-4. Convergence of forces in the K^+ form of $[d(TG_4T)]_4$. The results of 3 ns simulation are marked by dark blue circles.

3.3.2 Force anti-symmetrization

As the two ends of $[d(TG_4T)]_4$ are not very different, it would be a valid approximation to assume that both ends should produce similar force profiles while ions are passing through the channel. To test this idea, a comparison between the original (black) and anti-symmetrized (purple) force distributions for Na^+, K^+, and NH_4^+ ion movement along the channel axis (Z) of $[d(TG_4T)]_4$ is shown in Figure 3-5. The good agreement between the two profiles validates the anti-symmetrization assumption in the present system. However, as mentioned in the previous section, the presence of only a single ion in the channel could deviate the force by 2-3 kcal/mol/Å from the original value. Therefore, the force anti-symmetrization can produce more accurate results because the average force between similar positions in the anti-symmetrized profile will

represent a closer profile to the real situation in case the force at one end is significantly distorted in the single-ion model.

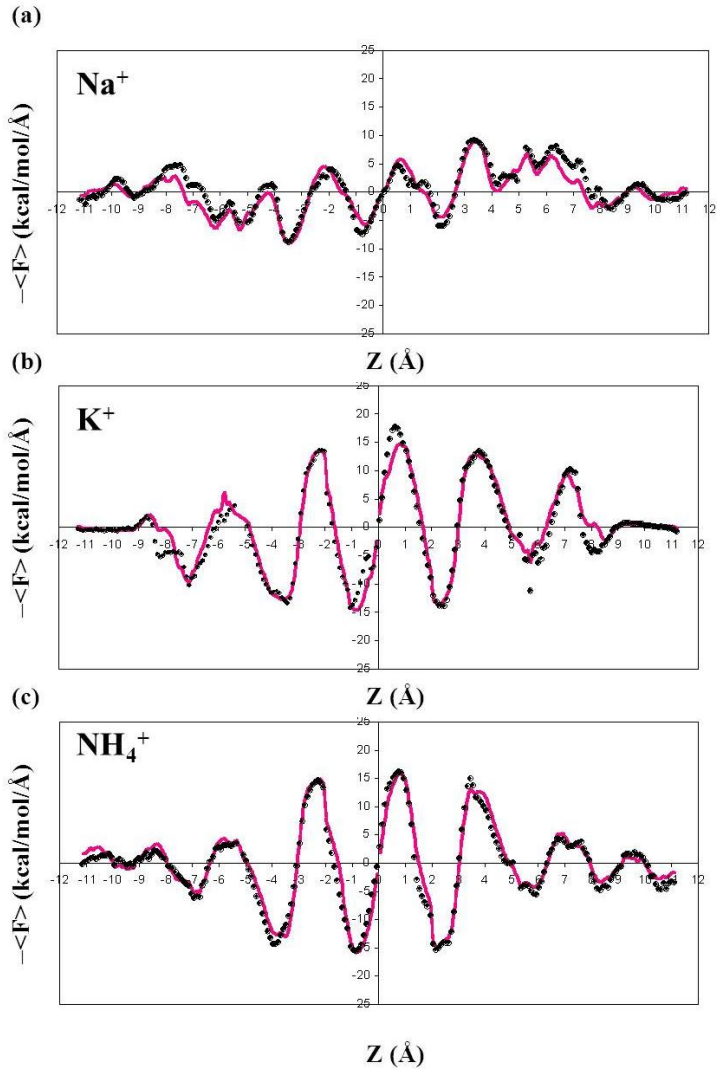

Figure 3-5. Comparison between the original (black) and anti-symmetrized (purple) force distributions for (a) Na^+, (b) K^+, and (c) NH_4^+ ion movement along the channel axis (Z) of $[d(TG_4T)]_4$.

3.3.3 PMF profiles for single ion movement

To investigate the free energy profile for ion movement through the $[d(TG_4T)]_4$ G-quadruplex channel, we divided the ion movement into two regions. One region is within the channel pore between the two exit/entrance points. The other is from the channel exit/entrance points to the bulk. In the discussion that follows, we examine these two regions separately.

Inside the narrow channel pore, as the ion movement is confined, it is natural to define the reaction coordinate for ion movement to be along the channel axis (the Z axis; see Figure 3-2(a)). Average forces on the targeted ion were calculated using the ABF method. Figure 3-6 shows the PMF profiles for the movement of single Na^+, K^+, and NH_4^+ ions inside the channel. The ABF-MD computational results suggest that, while both K^+ and NH_4^+ ions experience a free energy barrier of roughly 13–15 kcal/mol for moving between adjacent channel binding sites, the corresponding barrier for Na^+ movement is significantly smaller, about 4–5 kcal/mol. This computed free energy barrier for NH_4^+ movement within the $[d(TG_4T)]_4$ channel is in reasonably good agreement with those reported by Plavec and co-workers[167] for NH_4^+ movement within a unimolecular G-quadruplex formed by $d[G_4(T_4G_4)_3]$, approximately 12–15 kcal/mol. Although there has been no direct experimental data in the literature with respect to the free energy barriers for Na^+ and K^+ movement within a G-quadruplex channel, the computed results for these ions are understandable on the basis of their ionic radii (Na^+: 0.95 Å; K^+: 1.33 Å; NH_4^+: 1.43 Å). Therefore, while large K^+ and NH_4^+ ions must "squeeze" through a G-quartet to reach the adjacent channel binding site, Na^+ can be considered to "diffuse" nearly continuously through the channel. This result is also consistent with the observation made by Plavec and co-workers[164] that the movement of NH_4^+ ions inside the $[d(G_4T_4G_4)]_2$ G-quadruplex channel is considerably faster for the NH_4^+/ Na^+ mixed form than for the pure NH_4^+ form.

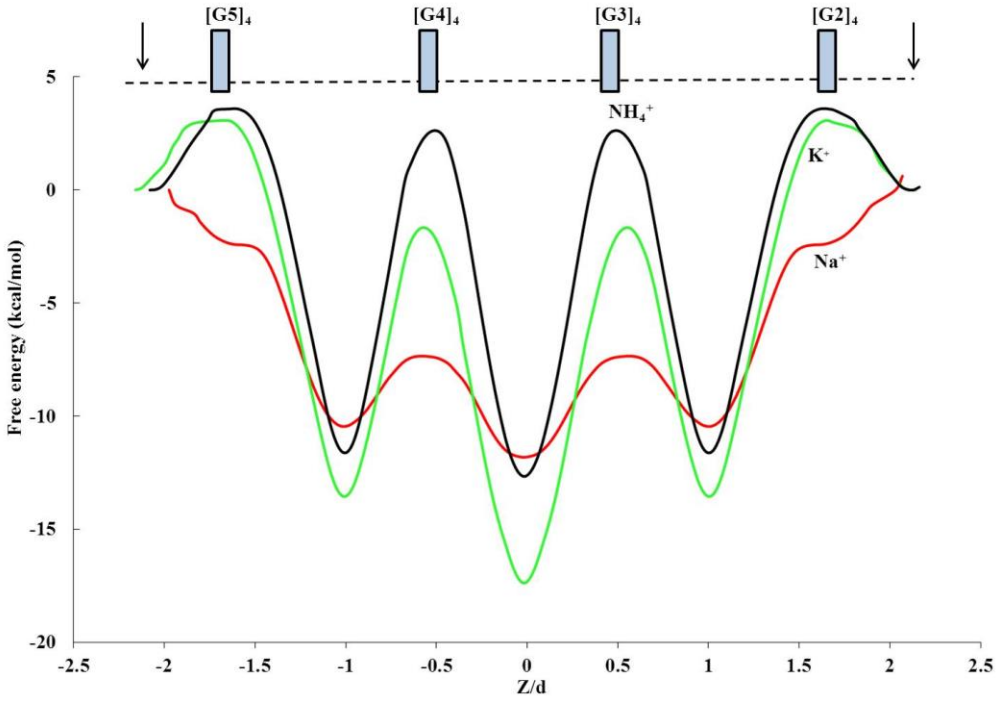

Figure 3-6. PMF profiles for ion movement inside the $[d(TG_4T)]_4$ G-quadruplex channel. The arrows indicate the exit/entrance points of the channel. For easy comparison, a normalized reaction coordinate (Z/d) is used in the diagram where d is the average separation between adjacent channel binding sites (d = 2.71, 2.95, and 3.09 Å for the Na^+ (red), K^+ (green), and NH_4^+ (black) forms of the G-quadruplex, respectively).

It is important to point out that, while it is always assumed in the literature that ions move in and out of the narrow G-quadruplex channel through the entrance points at the two ends, there has never been experimental proof that ions would not "leak" out from the side wall of the G-quadruplex channel. To test this hypothesis, we obtained the PMF profiles for the "sideways" ion movement within the $[d(TG_4T)]_4$ channel. As seen in Figures 3-7 and 3-8, for the sideways movement within the channel, Na^+, K^+, and NH_4^+ ions would encounter a free energy barrier of 50–60 kcal/mol, which is significantly larger than those along the channel axis. It is also interesting to note that all three ions experience essentially the same barrier for sideways

movement within the channel. Thus if the channel ions were to "leak" through the channel wall, all three ions should exhibit the same dynamics, which is obviously in contradiction with experimental observations. Thus, our computational results provide a definite proof that channel ions (Na^+, K^+, and NH_4^+) do not exit the channel through the channel wall.

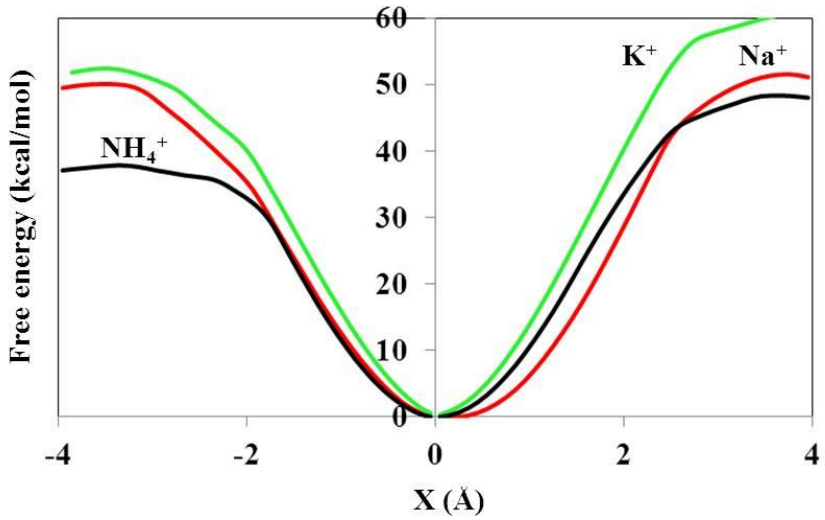

Figure 3-7. Typical 1D PMF profiles for the "sideways" ion movement inside the $[d(TG_4T)]_4$ channel (Z = 0 and along the X axis).

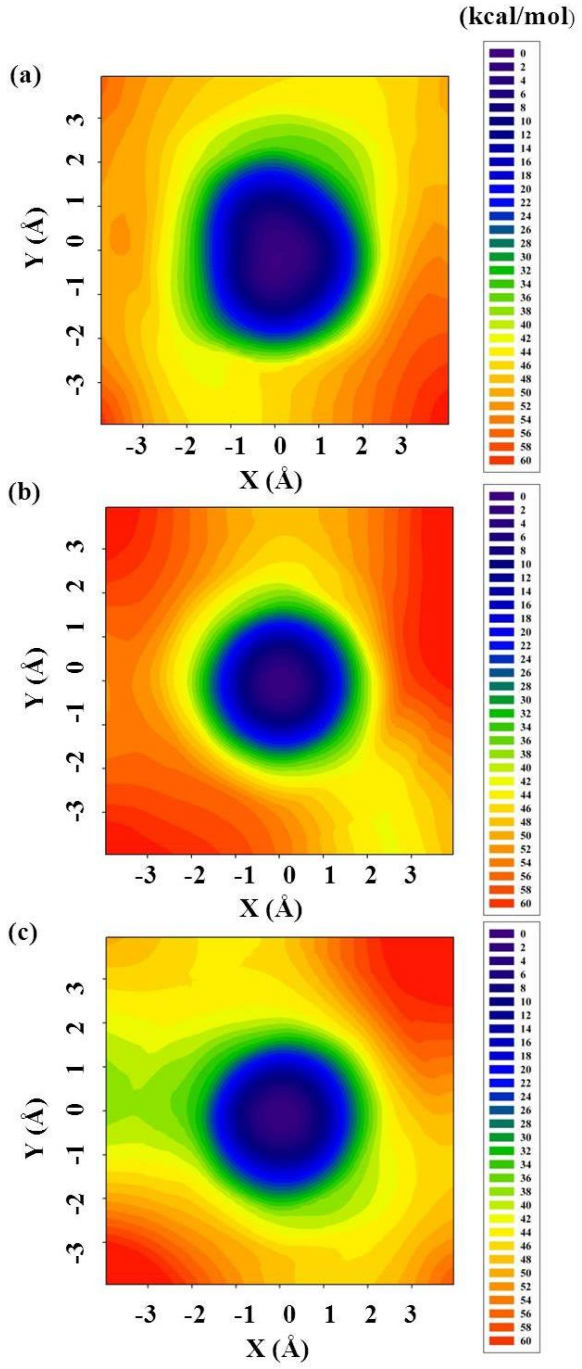

Figure 3-8. 2D PMF profiles for "sideways" (a) Na^+, (b) K^+, and (c) NH_4^+ movement in the XY plane within the channel (Z = 0 Å).

Now we turn our attention to the second region of ion movement to investigate how an ion, after reaching the channel exit/entrance point, would move into the bulk. In this region, as there is no natural confinement for ion movement, we decided to perform a 2D PMF mapping for a series of Z positions in XY plane. In particular, for each Z position, we performed ABF calculations for eight directions within the XY plane and then reconstructed a 2D PMF topographic profile. Figures 3-9 and 3-10 show 2D PMF profiles for ion movement in the XY plane away from the channel axis (X = 0 and Y = 0) toward bulk. Z is about 2.3 and 1.5 Å from the 5'- and 3'-end G-quartets, respectively. In general, the free energy landscapes in this region are quite flat and have a barrier of only about 4 kcal/mol for Na^+, K^+, and NH_4^+ movement. Now combining the computational results obtained for both regions (see Figures 3-6 and 3-9), we conclude that the free energy barrier for ion movement from the channel site to the bulk is approximately 20 and 14 kcal/mol for K^+/ NH_4^+ and Na^+, respectively. Recently, Sket and Plavec[169] reported that the exchange rate constant for NH_4^+ movement between channel sites and bulk in $[d(TG_4T)]_4$ is approximately 40 s^{-1} at 25˚C, corresponding to a free energy barrier of 15 kcal/mol if the Eyring equation is used.[189]

$$k = \frac{k_B T}{h} e^{\frac{-\Delta G^{\ddagger}}{RT}} \qquad [3.1]$$

where k, k_B, T, h, $\Delta G^{\ddagger}$ and R are rate constant, Boltzmann constant, temperature, Planck's constant, Gibbs energy of activation and gas constant, respectively. For the $d(G_3T_4G_4)_2$ G-quadruplex, Sket and Plavec[166] performed variable temperature NMR experiments and obtained a free energy barrier of approximately 16 kcal/mol at 27˚C for the movement of a NH_4^+ ion from one of the channel sites to the bulk. It should be noted that the experimental determination of

exchange rate constants (thus a free energy barrier) for NH_4^+ movement from a channel site to the bulk is sometimes complicated by the fact that proton exchange between NH_4^+ and solvent (water) can be facile. Nonetheless, our computational results appear to be consistent with the limited experimental data. For Na^+, Wu and co-workers[36] reported that the residence time of channel Na^+ ions in $[d(TG_4T)]_4$ G-quadruplex is more than 0.5 ms at 25˚C. Using magnetic relaxation dispersion (MRD) experiments, Snoussi and Halle[171] determined the residence time of Na^+ ions inside the $[d(G_3T_4G_3)]_2$ G-quadruplex channel to be 0.6–1.0 ms at 27˚C. On the basis of the Eyring equation (Equation 3.1) our computed free energy barrier of 14 kcal/mol for Na^+ is in qualitative agreement with the aforementioned experimental residence times determined for channel Na^+ ions in G-quadruplexes.

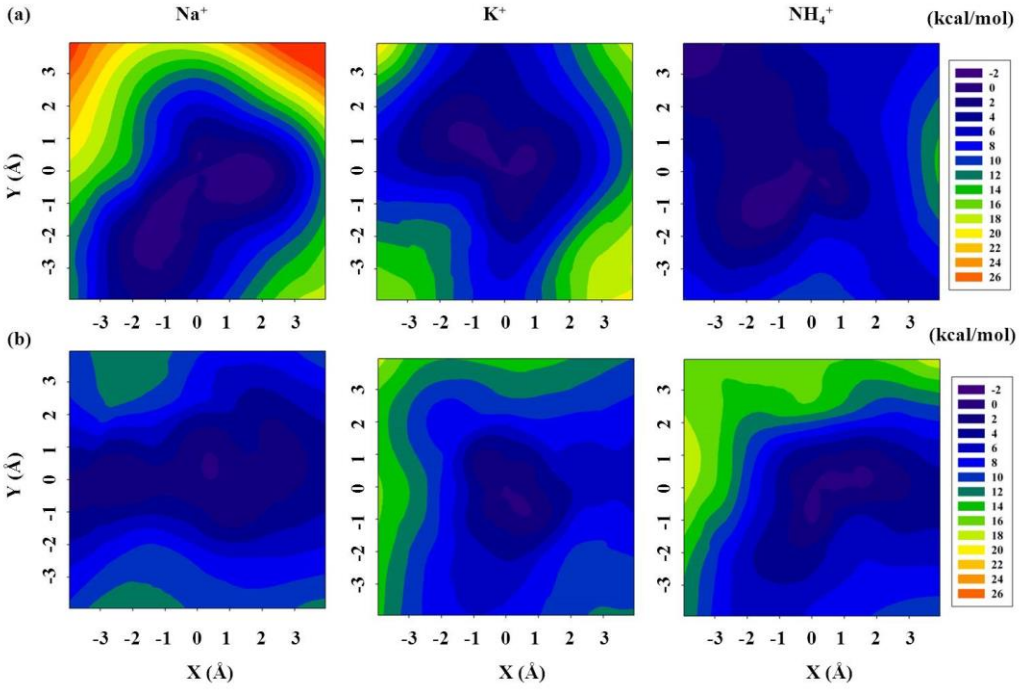

Figure 3-9. 2D PMF profiles for ion movement in the XY plane away from the channel axis (X = 0 and Y = 0) toward bulk. Z is 2.3 Å from the (a) 5′- and (b) 3′-end G-quartets.

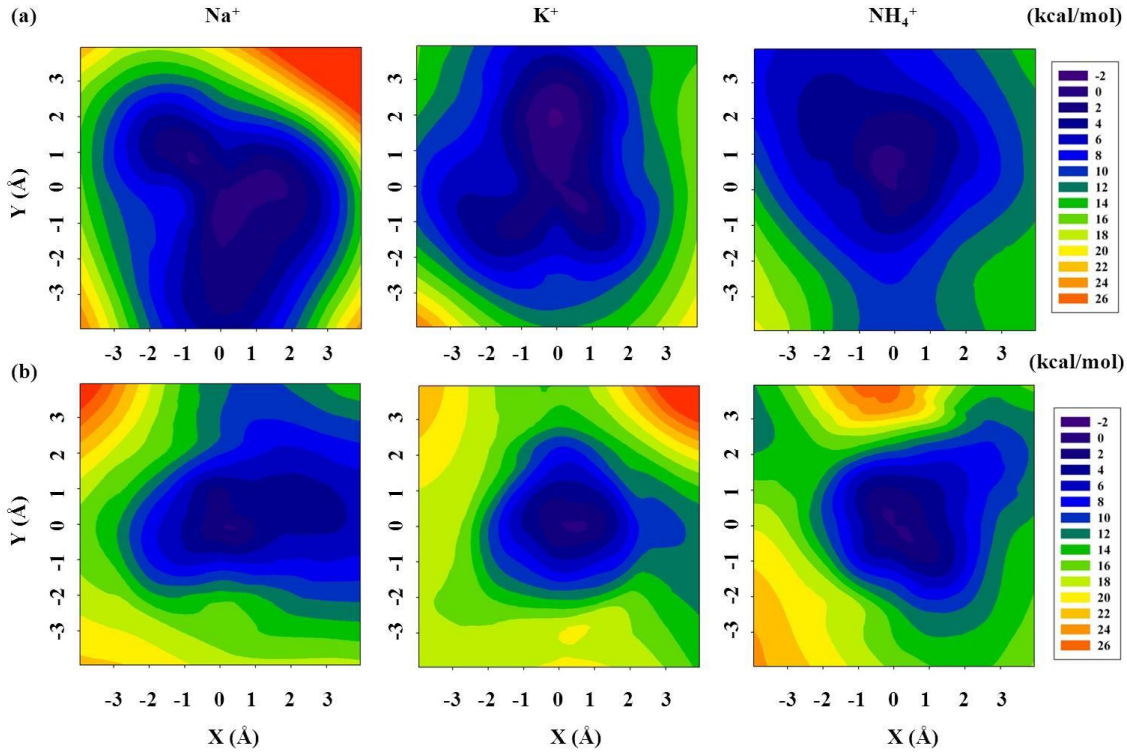

Figure 3-10. 2D PMF profiles for ion movement in the XY plane away from the channel axis (X = 0 and Y = 0) toward bulk. Z is 1.5 Å from the (a) 5′- and (b) 3′-end G-quartets.

Now that we have obtained the energetic properties for ion movement through a G-quadruplex DNA channel, it is worth comparing them with those found in K^+ ion channels (e.g., KcsA). The ion coordination geometries are strikingly similar at the binding sites in these two systems. Yet, our computational results suggest much higher free energy barriers, 14–20 kcal/mol, for an ion to move through a G-quadruplex channel than that seen in the KcsA channels, 2–3 kcal/mol.[52] This implies that different mechanisms for ion translocation are operative in the two systems. While we are still investigating the exact mechanism for ion transport through a G-quadruplex, the high free energy barriers found in this system are most

likely due to the fact that the G-quadruplex channel pore is much more rigid than that in the KcsA selectivity filter. As the flow of K^+ ions through a G-quadruplex channel has never been directly measured, the computational results reported in this chapter provide some key insights into the process. Meanwhile, we should point out that, as the computed free energy barrier for Na^+ movement within the G-quadruplex channel is only 4–5 kcal/mol, it may still be possible to utilize a G-quadruplex DNA as an artificial Na^+ ion channel. Of course, the challenge is how to reduce the free energy barriers at the channel exit/entrance points.

3.3.4 Hydration state analysis

Ions residing inside a G-quadruplex channel are expected to be fully dehydrated and coordinated to eight carbonyl oxygen atoms from the guanine bases, whereas they are fully hydrated in the bulk. A transition between these two hydration states occurs when an ion leaves or enters the channel. In our ABF simulations, we can readily monitor the hydration state of the targeted ion as it moves along the reaction coordinate, the Z axis. First, we need to obtain some benchmark values for free ions in bulk aqueous solution. Figure 3-11 shows the calculated radial distribution functions of Na^+, K^+, and NH_4^+ ions in bulk, which suggest that the first coordination shell of Na^+, K^+, and NH_4^+ ions consists of 5.94, 6.55, and 6.54 water molecules, respectively. For Na^+ and K^+ ions, 4.4-5.4 and 5.9-6.4 are found experimentally as the first coordination numbers, respectively.[190,191] The calculated CHARMM27 values are 5.9 and 6.9, respectively.[192] For NH_4^+ ion, 7.9-8.1 and 7.9-8.4 are reported as the first coordination numbers found experimentally and from MD calculations, respectively.[193-195]

In these calculations, the cut-off cation-water radii used for Na^+ and K^+ ions are 3.05 and 3.35 Å, respectively. These numbers correspond to the first minimum in each of the radial

distribution function diagrams shown in Figure 3-11. We then used these values to calculate the number of water molecules around each cation when it resides inside the channel moving along the reaction coordinate. For NH_4^+ ion, we obtained a cut-off cation-water radius of 3.68 Å. This value is slightly lower than the literature value, 4 Å.[195] In a separate analysis, we used 4 Å as the cut-off cation-water radius to calculate the first coordination shell for NH_4^+ ion in bulk. We found that the first coordination shell of NH_4^+ ion in bulk consists of 8.26 water molecules which is in agreement with the literature value.[193-195]

Figure 3-12 depicts the hydration profile for the Na^+, K^+, and NH_4^+ ions moving in a range of $-12 \leq Z \leq +12$ Å along the Z axis of the $[d(TG_4T)]_4$ channel. The results show that ions are completely dehydrated inside the channel (between the end G-quartets, $[G5]_4$ and $[G2]_4$). However, we also observed that occasionally water molecules enter into the channel to be within the first coordination shell of ions, resulting in a fractional number of coordination water molecules around the ion. Interestingly, Plavec and co-workers[170] did recently find NMR evidence for water molecules to reside inside a G-quadruplex channel. As seen in Figure 3-12, as the ion exits the channel, it quickly becomes hydrated. It is particularly interesting to note the situations occurring near the channel exit/entry points ($6.0 \geq Z \geq 5.0$ Å and $-5 \geq Z \geq -6.0$ Å). In each of these interfacial regions, while Na^+, K^+, and NH_4^+ ions are still coordinated to four carbonyl O6 atoms from the end G-quartet on one side, they start to receive on average approximately one, one, and three water molecules on the opposite side. Similar situations were previously seen in the crystal structure of the K^+ form of $[d(G_4T_4G_4)]_2$ G-quadruplex[13] as well as in a ^{23}Na NMR study of the Na^+ form of the same sequence.[37] Beyond these interfacial regions, each ion becomes nearly fully hydrated, approaching its hydration state found in the bulk. Using the literature cut-off cation-water radius (4 Å) we obtained an overall similar hydration profile for

NH_4^+ ion moving along the channel as that seen in Figure 3-12 when a cut-off of 3.68 Å was used. The only difference is that the number of hydration water molecules approaches 8 at Z = ±12 Å. This suggests that once the NH_4^+ ion has completely left the channel, it gradually becomes hydrated and eventually behaves like a bulk NH_4^+ ion. [193-195]

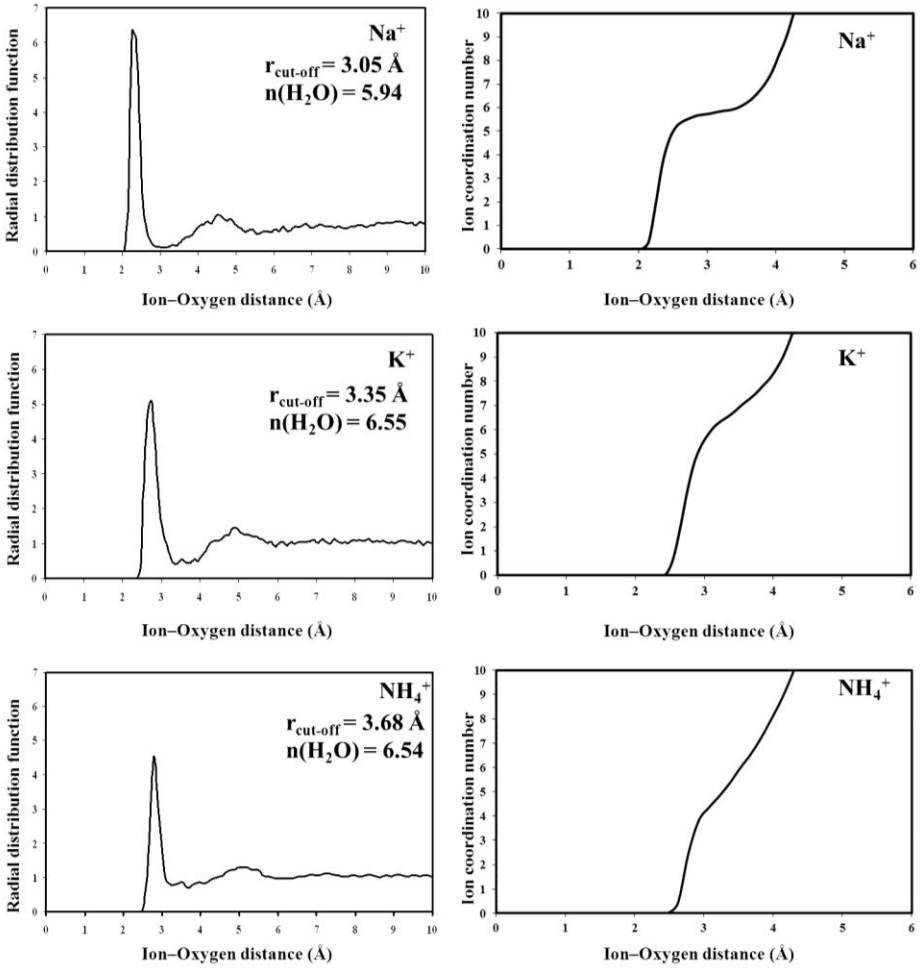

Figure 3-11. Hydration state analysis for Na^+, K^+, and NH_4^+ in bulk. (Left) Radial distribution function between Na^+, K^+, and NH_4^+ ions and oxygen of water molecules and (right) number of coordination water molecules.

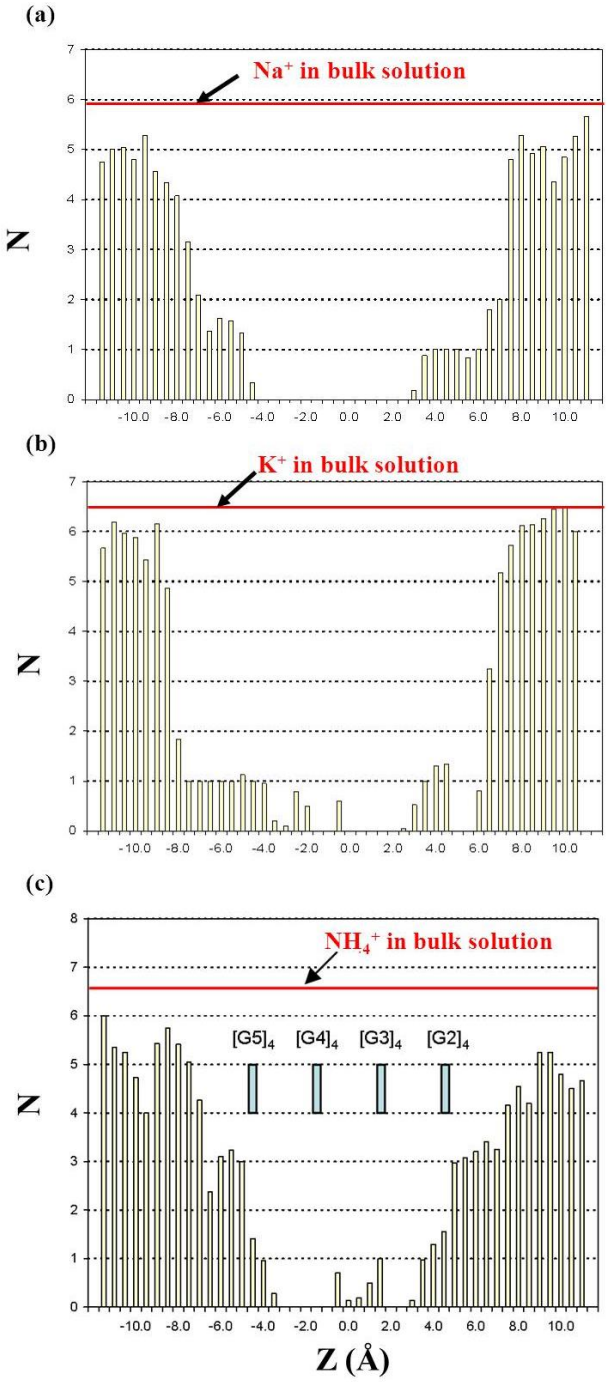

Figure 3-12. Average number of hydration water molecules (N) in the first coordination shell of the targeted (a) Na^+, (b) K^+, and (c) NH_4^+ ions along the $[d(TG_4T)]_4$ G-quadruplex channel axis.

3.3.5 Ion movement in $[d(TG_4T)]_4$ channel containing multiple Na^+ ions

So far, we have considered movement of single Na^+, K^+, or NH_4^+ ions through the G-quadruplex $[d(TG_4T)]_4$ channel. However, a G-quadruplex channel may contain multiple ions. The results of single-ion studies cannot be completely generalized to multiple-ion systems as the single-ion model does not include ion-ion interactions. In this section, we will attempt to extend our investigations to multiple-ion systems. In the following models, the initial minimized and equilibrated structures were taken from our previous regular MD simulations as shown in Figure 3-2. Also, the ion movement in the plane norm to the dynamic axis was confined by a harmonic potential (k = 100 kcal/mol/Å^2). The average force acting on the ion was accumulated in 0.1 Å bins. A production run of 3 ns was performed for each model. The positions of C1' atoms of central guanine residues were fixed by a harmonic potential constant of k = 3 kcal/mol/Å^2. Table 3-1shows the summary of the following models.

Table 3-1. Summary of the models to study ion movement in $[d(TG_4T)]_4$ channel containing multiple Na^+ ions.[a]

Model	Dynamics axis	Ion studied	Movement of other ions
1	G3 and G4	Ion 2 (-1.6 to 1.6 Å)	Ions 1 and 3 between their adjacent G-quartets
2	G2 and G3	Ion 1 (2.7 to 6.9 Å)	Ions 2 and 3 are fixed
3	G2 and G3	Ion 1 (1.6 to 5.8 Å)	Ions 2 and 3 between their adjacent G-quartets
4	G3 and G4	Ion 2 (-1.7 to 1.7 Å)	Ions 1 and 3 (-6.5 to 6.5 Å)
5	G3 and G4	Ion 2 (-1.7 to 1.7 Å)	Ions 1 and 3 fully unconstrained
6	G2 and G3	Ion 1 (1.6 to 5.8 Å)	Ion 3 (-9.0 to 15.0 Å)
7	G2 and G3	Ion 1 (1.6 to 5.8 Å)	Ion 3 fully unconstrained
8	G2 and G3	Ion 1 (1.6 to 5.8 Å)	Ion 2 (-9.0 to 15.0 Å)
9	G2 and G3	Ion 1 (1.6 to 5.8 Å)	Ion 2 fully unconstrained

[a] The positions of C1' atoms of central guanine residues were fixed by a harmonic potential constant of k = 3 kcal/mol/Å^2.

Model 1:

In this model we studied ion transport of one Na^+ ion along the $[d(TG_4T)]_4$ channel containing three Na^+ ions. The distance between adjacent Na^+ ions was about 3 Å initially and the Na^+ ions were allowed to move in ranges shown in Figure 3-13(a). The ABF method was used to investigate the average force experienced by ion 2 from RC = -1.6 to 1.6 Å. The reaction coordinate was chosen along the channel axis of $[d(TG_4T)]_4$, which was defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms from the two central G-quartets (G3 and G4). Figure 3-13(b) shows the average force experienced by ion 2 in this model. For comparison, the result for the single-ion model is also shown. In this diagram the zero average force at the origin in the single-ion model corresponds to the position in which the ion is located in the middle of $[G3]_4$ and $[G4]_4$. The maximum and minimum peaks occur when the ion moves close to the G-quartets and the force returns to almost zero when the ion moves within the G-quartet. In Model 1, although the average force shows a similar behavior at the origin, it displays a quite different picture once the ion moves toward either $[G3]_4$ or $[G4]_4$. The absolute value of the force increases at this point and it does not approach zero. In order to understand what may cause this, we examined the positions of all Na^+ ions using the trajectory data of the simulations (see Figure 3-13(c)). During the simulation, the movement of the three ions is correlated. That is, although ions are free to move within their defined regions, they prefer to follow a correlated pattern. At the moment, we are not certain what may cause this correlated pattern of movement. The trajectory data also show that, when ion 2 is located at RC = -1.5 Å, it is 3 Å away from ion 3, and 4.5 Å from ion 1. Thus, ion 2 experiences a stronger repulsion from ion 3 than from ion 1 resulting in a net positive force on ion 2. The same discussion is valid to describe the net negative

force experienced by ion 2 when it is at RC = 1.5 Å. At the origin, ions 1 and 3 are 4 Å apart from ion 2 and therefore, the net force on ion 2 is almost zero, similar to the single-ion model.

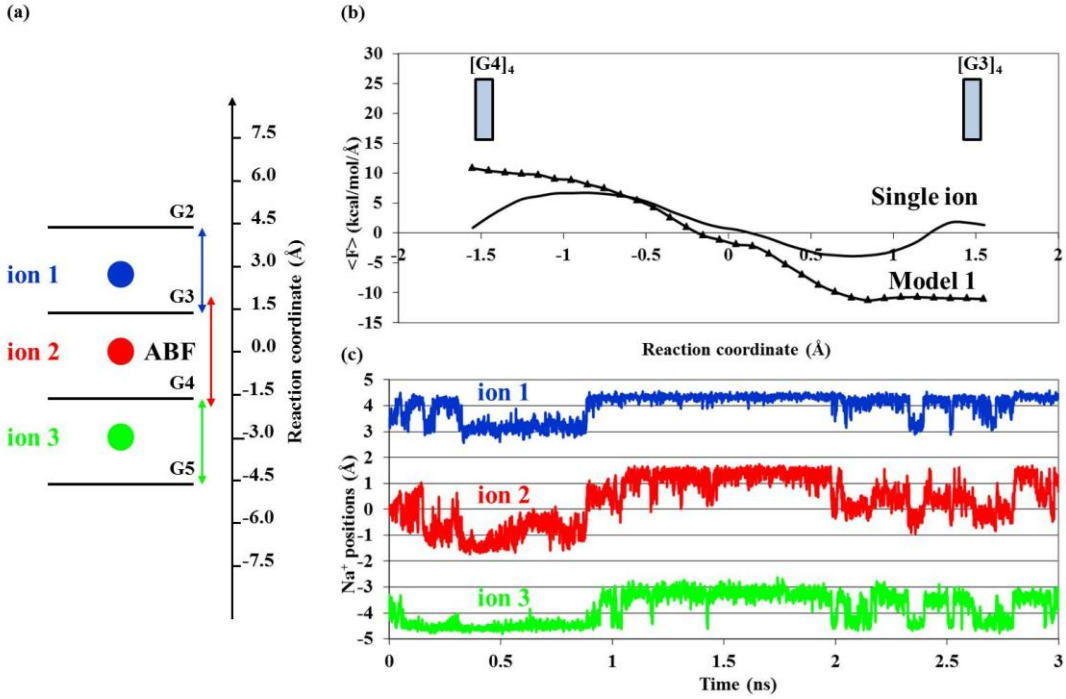

Figure 3-13. (a) A schematic display of multiple Na^+ ions in $[d(TG_4T)]_4$ channel (Model 1). The blue, red, and green colors represent ions 1, 2, and 3, respectively. A double arrow indicates the range of allowed ion movement. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4). (b) Average force experienced by Na^+ ion in a single-ion model (solid line) and multiple-ion Model 1 (triangle). (c) Na^+ positions for ions 1, 2, and 3 during the simulation in Model 1.

Model 2:

In this model we used the ABF method to investigate the average force experienced by ion 1. The positions of ions 2 and 3 were fixed during the simulation by a harmonic potential (k = 100 kcal/mol/Å^2), while ion 1 was allowed to move between RC = 2.7 and 6.9 Å (see Figure

3-14(a)). The reaction coordinate was chosen along the channel axis of $[d(TG_4T)]_4$, which was defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms of G2 and G3. Figure 3-14(b) shows the average force experienced by ion 1 in this model. For comparison, the result for the single-ion model is also shown. Since the positions of ions 2 and 3 are fixed, there is not much movement for any of them during simulation. This can be checked from trajectory data of ion positions. The distance between Na^+ ions 1 and 2 during the simulation is presented in Figure 3-14(c). Interestingly, ion 1 exits the channel briefly then returns to the original position for the most of the simulation time. As seen from Figure 3-14(b), in the single-ion model, the zero average force close to RC = 2.6 Å corresponds to the position in which the ion is located between $[G2]_4$ and $[G3]_4$. The maximum and minimum peaks occur when the ion moves close to $[G2]_4$ and $[G3]_4$ and the force returns to almost zero again when the ion moves within the G-quartet. In Model 2, when ion 1 is at RC = 3 Å, the distance between ions 1 and 2 is 3 Å and thus ion 1 experiences a strong positive repulsion force. When ion 1 moves toward $[G2]_4$, away from ion 2, the shape and the magnitude of the average force in Model 2 is not very different from that seen in the single-ion model, suggesting that the ion-ion repulsion becomes negligible in the G-quadruplex systems when the ion-ion distance exceeds 4.5 Å.

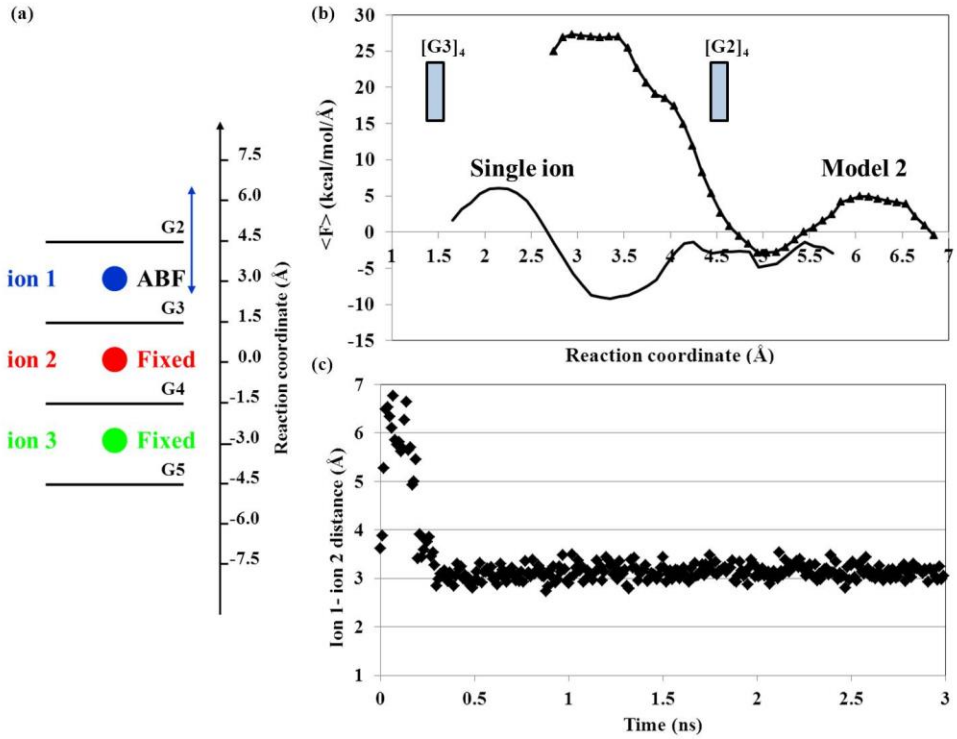

Figure 3-14. (a) The schematic display of multiple Na^+ ions in $[d(TG_4T)]_4$ channel (Model 2). The blue, red, and green colors represent ions 1, 2, and 3, respectively. The double arrow indicates the range of allowed movement for ion 1. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4). (b) Average force experienced by Na^+ ion in a single-ion model (solid line) and multiple-ion Model 2 (triangle). (c) Ion 1- ion 2 distance during the simulation in Model 2.

Model 3:

In this model we examined the average force experienced by ion 1 between RC = 1.6 and 5.8 Å. The range of movement for Na^+ ions is shown in Figure 3-15(a). The reaction coordinate was chosen along the channel axis of $[d(TG_4T)]_4$, which was defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms of G2 and G3. Figure 3-15(b) shows the average force experienced by ion 1 in this model. For comparison, the result for the single-ion model is

also shown. The positions of the three Na^+ ions during the simulation are presented in Figure 3-15(c). The same correlation for ion movement as in Model 1 can be observed. The average force diagram shows a larger force experienced by ion 1 at about RC = 3 Å compared to the single-ion model. The amount of shift is smaller than in Model 2 though because in Model 2, ion 2 is fixed and imposes a large repulsion force on ion 1. Ion 2 in Model 3 has more space to move, therefore, it applies a smaller repulsion force. At RC = 1.5 Å the overall repulsion impact avoids the average force to reach zero in this model compared to the single-ion model. At RC = 5.8 Å the distance between ions 1 and 2 is approximately 5 Å and this produces the same profile as that observed in the single-ion model. Generally, the range of average force distribution is smaller than in Model 2 because ions have more freedom to move in Model 3 and this reduces the repulsion effects.

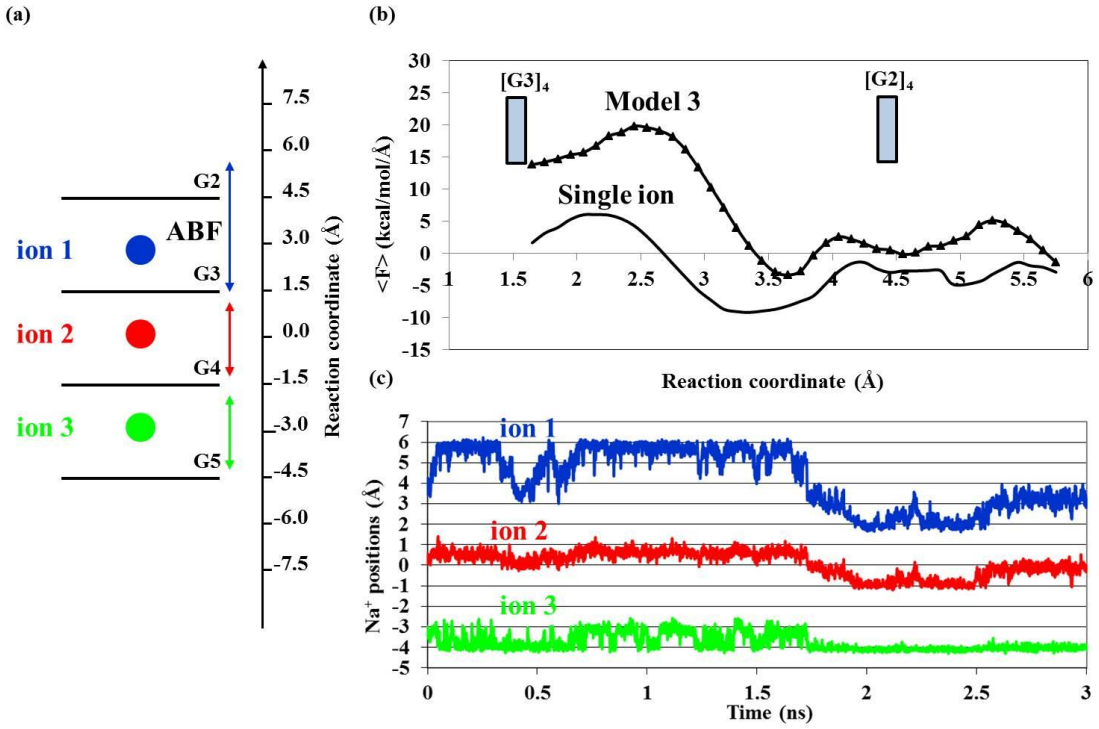

Figure 3-15. (a) The schematic display of multiple Na^+ ions in $[d(TG_4T)]_4$ channel (Model 3). The blue, red, and green colors represent ions 1, 2, and 3, respectively. A double arrow indicates the range of allowed ion movement. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4). (b) Average force experienced by Na^+ ion in a single-ion model (solid line) and multiple-ion Model 3 (triangle). (c) Na^+ positions for ions 1, 2, and 3 during the simulation in Model 3.

Model 4:

In this model we examined the average force experienced by ion 2 in a range of RC = -1.7 to 1.7 Å. Ions 1 and 3 were allowed to move from RC = -6.5 to 6.5 Å (see Figure 3-16(a)). The reaction coordinate was chosen along the channel axis of $[d(TG_4T)]_4$, which was defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms from the two central G-quartets (G3 and G4). Figure 3-16(b) shows a similar average force experienced by ion 2 in this

model. For comparison, the result for the single-ion model is also shown. The positions of the three Na^+ ions during the simulation are presented in Figure 3-16(c). The same correlation for ion movement as in Model 1 can be observed. Because of the large range of allowed movement for ions 1 and 3, there will be more space around ion 2. As can be seen, ions 1 and 3 keep a minimum distance of 4 Å to ion 2. So, the repulsion force on ion 2 is almost canceled during the simulation and it will experience about the same magnitude of force as a single-ion model even at RC = ±1.5 Å. The small bump at the right hand side of the average force graph in Model 4 is most likely because of the poor sampling in this area.

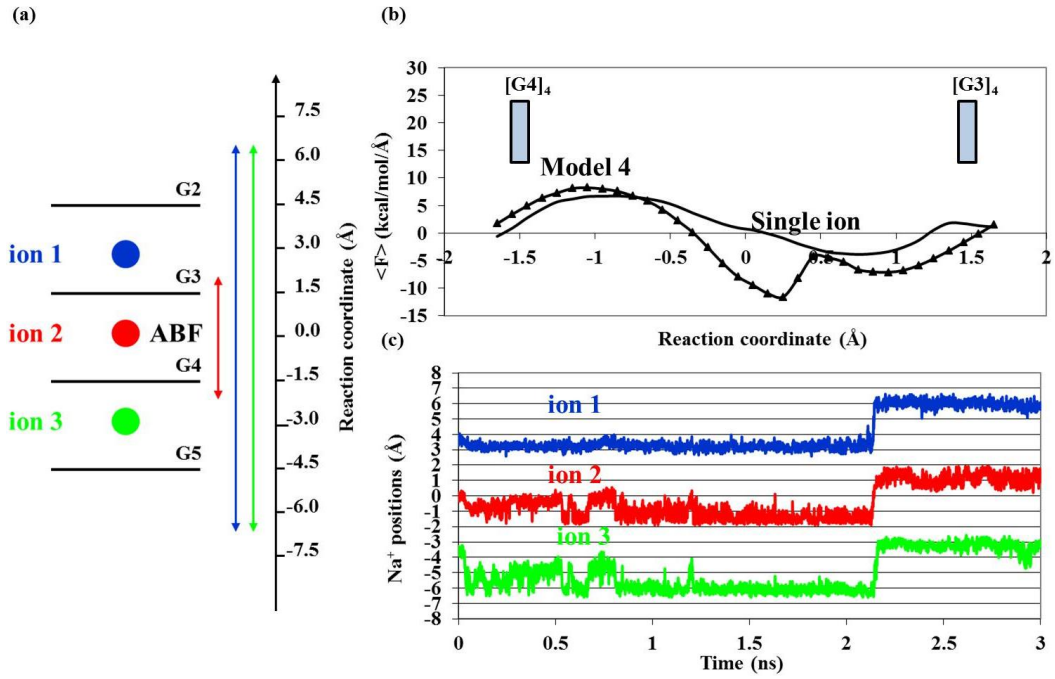

Figure 3-16. (a) The schematic display of multiple Na^+ ions in $[d(TG_4T)]_4$ channel (Model 4). The blue, red, and green colors represent ions 1, 2, and 3, respectively. A double arrow indicates the range of allowed ion movement. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4). (b) Average force experienced by Na^+ ion in a single-ion model (solid line) and multiple-ion Model 4 (triangle). (c) Na^+ positions for ions 1, 2, and 3 during the simulation in Model 4.

Model 5:

In this model we examined the average force experienced by ion 2 in a range of RC = -1.7 to 1.7 Å, while ions 1 and 3 were fully unconfined (see Figure 3-17(a)). The reaction coordinate was chosen along the channel axis of $[d(TG_4T)]_4$, which was defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms from the two central G-quartets (G3 and G4). Figure 3-17(b) shows the average force experienced by ion 2 in this model. For comparison, the result for the single-ion model is also shown. The shape and the magnitude of the average force are similar in both models showing that there is no significant repulsion between ions in Model 5. The positions of all Na^+ ions during the simulation are presented in Figure 3-17(c). The ions are moving in correlation although ions 1 and 3 are completely unconstrained. This may be related to the fact that the channel prefers to stay almost occupied as long as there is enough space for all ions. It seems that just providing enough space between the ions to avoid a huge repulsion can most likely produce a profile close to the single-ion model. At RC = ±1.5 Å, ion 2 is at least 4 Å apart from the other two ions and experiences an almost zero average force within the G-quartet, similar to that seen in the single-ion model.

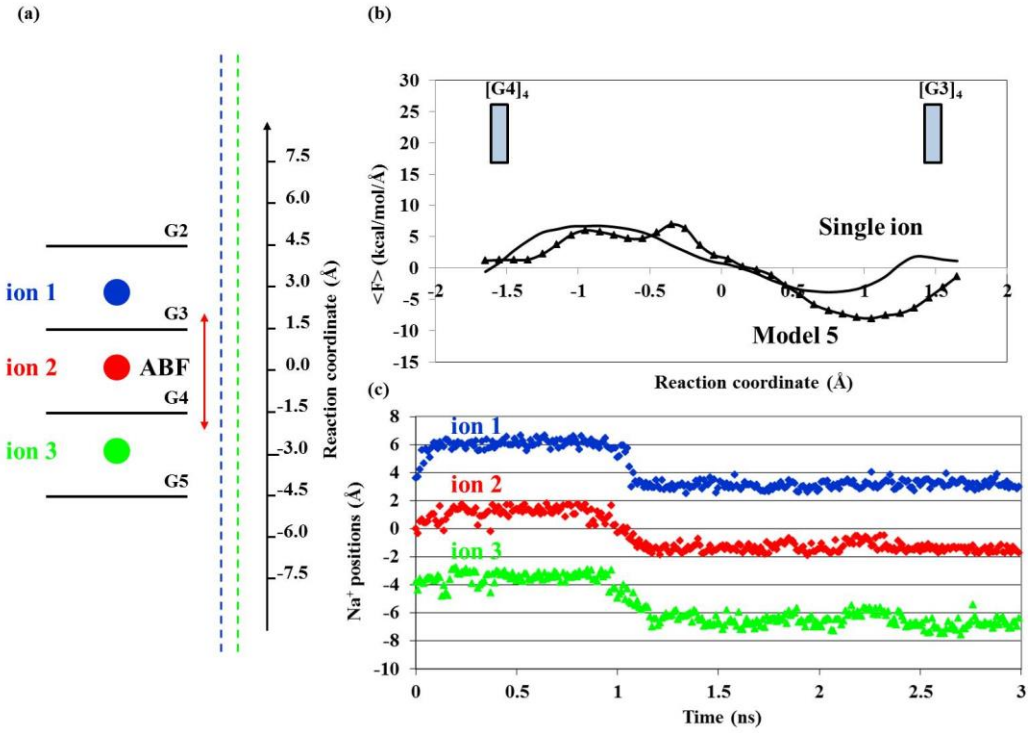

Figure 3-17. (a) The schematic display of multiple Na^+ ions in $[d(TG_4T)]_4$ channel (Model 5). The blue, red, and green colors represent ions 1, 2, and 3, respectively. The double arrow and the dotted lines indicate the range of allowed ion movement. A dotted line represents the unconstrained range of movement. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4). (b) Average force experienced by Na^+ ion in a single-ion model (solid line) and multiple-ion Model 5 (triangle). (c) Na^+ positions for ions 1, 2, and 3 during the simulation in Model 5.

Model 6:

In this model we studied ion transport of one Na^+ ion along the $[d(TG_4T)]_4$ channel containing two Na^+ ions. The initial minimized and equilibrated structures were prepared after removing the central Na^+ ion from the structure shown in Figure 3-2. The distance between the center of mass of O6 atoms of two subsequent G-quartets was almost 3 Å initially. The ABF method was used to investigate the average force experienced by ion 1 in a range of RC = 1.6 to

5.8 Å while ion 3 was allowed to move from RC = -9.0 to 15.0 Å (see Figure 3-18(a)). The reaction coordinate was chosen along the channel axis of $[d(TG_4T)]_4$, which was defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms from G2 and G3. Figure 3-18(b) shows the average force experienced by ion 1 in this model. For comparison, the result for the single-ion model is also shown. The positions of both Na^+ ions during the simulation are presented in Figure 3-18(c). Although ion 3 is free to move in a large range, the ions are moving in correlation. There is no significant repulsion between ions in this model, so the shape and the magnitude of the average force are similar in both models as shown in Figure 3-18(b). At RC = 1.5 Å, ion 1 is at least 5 Å apart from ion 3 and experiences an almost zero average force within the G-quartet similar to that seen in the single-ion model.

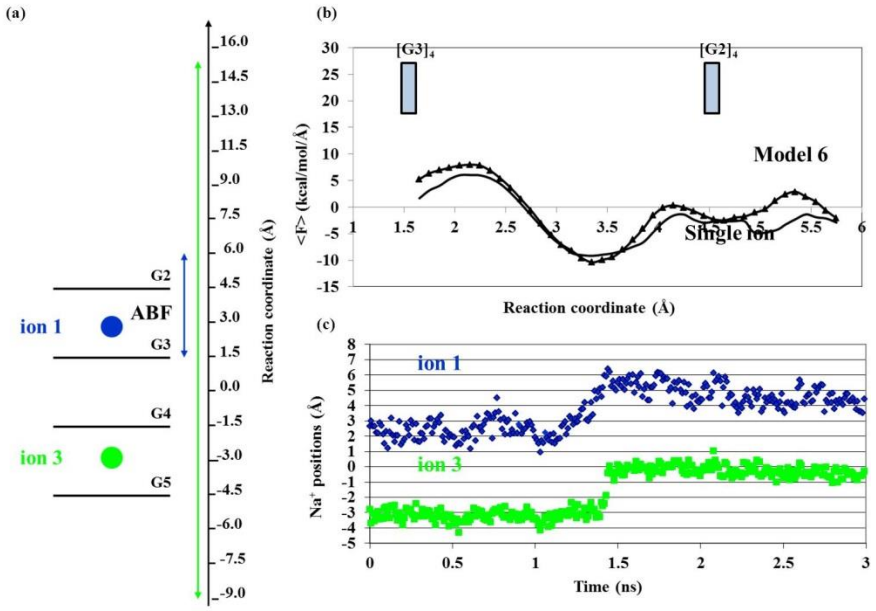

Figure 3-18. (a) The schematic display of multiple Na^+ ions in $[d(TG_4T)]_4$ channel (Model 6). The blue and green colors represent ions 1 and 3, respectively. A double arrow indicates the range of allowed ion movement. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4). (b) Average force experienced by Na^+ ion in a single-ion model (solid line) and multiple-ion Model 6 (triangle). (c) Na^+ positions for ions 1 and 3 during the simulation in Model 6.

Model 7:

In this model we examined the average force experienced by ion 1 similar to Model 6, expect for the ion 3 that was entirely unrestricted (see Figure 3-19(a)). Figure 3-19(b) shows the average force experienced by ion 1 in Model 7. For comparison, the result for the single-ion model is also shown. The shape and the magnitude of the average force on ion 1 are close to the single-ion model. Also, there is no considerable difference between the average force profiles in Models 6 and 7. This basically suggests that there is even no need to let the other ion be entirely unrestricted, just providing enough space to avoid a huge repulsion can most likely result in a profile similar to that seen in the single-ion model. At RC = 1.5 Å, ion 1 is at least 5 Å apart from ion 3 and experiences an almost zero average force within the G-quartet, similar to the single-ion model. In Model 7, there is not much correlation between ion movements as shown in Figure 3-19(c).

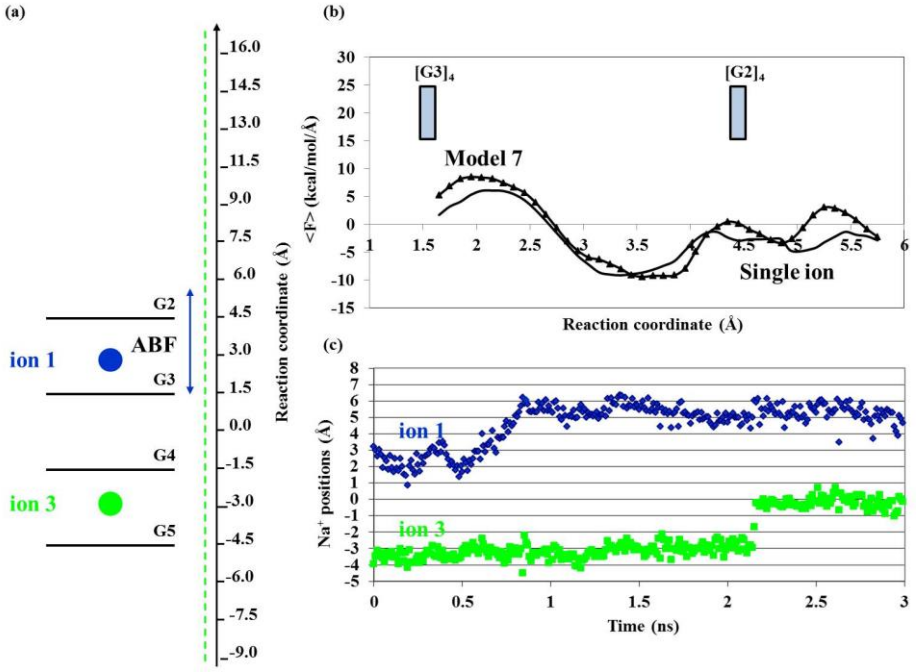

Figure 3-19. (a) The schematic display of multiple Na^+ ions in $[d(TG_4T)]_4$ channel (Model 7). The blue and green colors represent ions 1 and 3, respectively. The double arrow and the dotted line indicate the range of allowed ion movement. The dotted line represents the unconstrained range of movement. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4). (b) Average force experienced by Na^+ ion in a single-ion model (solid line) and multiple-ion Model 7 (triangle). (c) Na^+ positions for ions 1 and 3 during the simulation in Model 7.

Model 8:

In this model we studied ion transport of one Na^+ ion along the $[d(TG_4T)]_4$ channel after removing the lower Na^+ ion from the structure shown in Figure 3-2. The ABF method was used to investigate the average force experienced by ion 1 in a range of RC = 1.6 to 5.8 Å while ion 2 was allowed to move from RC = -9.0 to 15.0 Å (see Figure 3-20(a)). The reaction coordinate was chosen along the channel axis of $[d(TG_4T)]_4$, which was defined by a "dynamic axis" linking the

two centers of mass of the four O6 atoms from G2 and G3. Figure 3-20(b) shows the average force experienced by ion 1 in this model. For comparison, the result for the single-ion model is also shown. In Model 8, the two ions are closer in the channel compared to Models 6 and 7. However, as ion 2 is allowed to move in a large range there is no significant repulsion between ions in this model. Therefore, the shape and the magnitude of the average force in this model are close to the single-ion model and also to Models 6 and 7. This further confirms that just providing enough space to avoid a huge repulsion can most likely result in a profile close to that seen in the single-ion model. At RC = 1.5 Å, ion 1 is at least 5 Å apart from ion 2 and the average force on it reaches almost zero within the G-quartet, similar to the profile of the single-ion model. As shown in Figure 3-20(c), there is not much correlation between ion movements in this model.

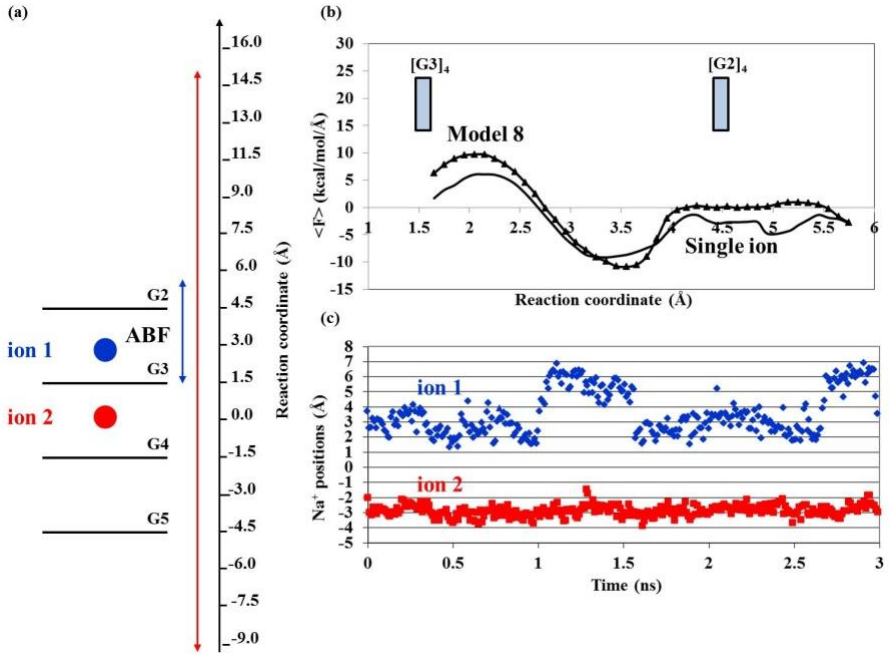

Figure 3-20. (a) The schematic display of multiple Na^+ ions in $[d(TG_4T)]_4$ channel (Model 8). The blue and red colors represent ions 1 and 2, respectively. A double arrow indicates the range of allowed ion movement. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4). (b) Average force experienced by Na^+ ion in a single-ion model (solid line) and multiple-ion Model 8 (triangle). (c) Na^+ positions for ions 1 and 2 during the simulation in Model 8.

Model 9:

In this model we examined the average force experienced by ion 1 similar to Model 8, however, ion 2 was not restricted (see Figure 3-21(a)). Figure 3-21(b) shows the average force experienced by ion 1 in this model. For comparison, the result for the single-ion model is also shown. Since ion 2 is allowed to move all over the reaction coordinate, there is no significant repulsion between ions in this model. The shape and the magnitude of the average force in this model are close to the single-ion model and also to Models 6, 7, and 8. At RC = 1.5 Å, ion 1 is at

least 5 Å apart from ion 2 and the average force on it reaches almost zero within the G-quartet. So, the overall picture is not very different from the single-ion model. Similar to Model 8, there is not much correlation between ion movements in this model (see Figure 3-21(c)).

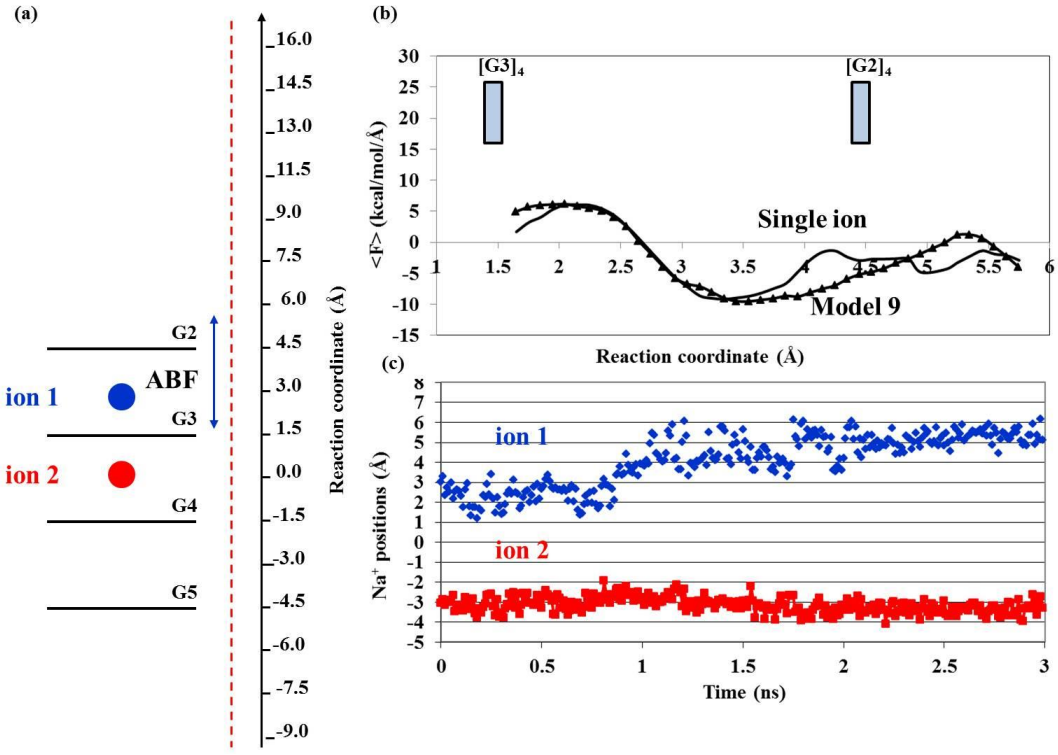

Figure 3-21. (a) The schematic display of multiple Na^+ ions in $[d(TG_4T)]_4$ channel (Model 9). The blue and red colors represent ions 1 and 2, respectively. The double arrow and the dotted line indicate the range of allowed ion movement. The dotted line represents the unconstrained range of movement. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4). (b) Average force experienced by Na^+ ion in a single-ion model (solid line) and multiple-ion Model 9 (triangle). (c) Na^+ positions for ions 1 and 2 during the simulation in Model 9.

3.4 Conclusion

In this chapter, we have obtained a complete picture about the free energy landscapes for movement of single Na^+, K^+, and NH_4^+ ions through the $[d(TG_4T)]_4$ G-quadruplex DNA channel. This is the first time that the energetic and dynamic aspects of ion movement in G-quadruplex DNA have been investigated quantitatively by ABF-MD simulations. The computed results are in qualitative agreement with the very limited experimental data available in the literature. We believe that this general ABF-MD approach can be applied to other G-quadruplex structures. We also examined several models for Na^+ ion movement in the $[d(TG_4T)]_4$ G-quadruplex channel containing multiple ions. The computational results suggest that ion-ion repulsion is important up to when two ions are within 4.5 Å inside a G-quadruplex channel. However, further investigation is required to obtain a complete understanding of the seemingly concerted movement of ions within the channel.

Chapter 4

Adaptive biasing force MD studies of ion movement in $[d(G_3T_4G_4)]_2$

4.1 Introduction

In the previous chapter, we established the validity of the adaptive biasing force (ABF) MD simulations in evaluating free energy landscapes for ion movement (Na^+, K^+, and NH_4^+) in G-quadruplex DNA structures. Here, we apply this computational method to investigate the NH_4^+ ion movement inside a bimolecular G-quadruplex DNA structure formed by $d(G_3T_4G_4)$, denoted as $[d(G_3T_4G_4)]_2$. As shown in Figure 4-1, this G-quadruplex structure consists of a three G-quartet stem, one edge-type and one diagonal loop.[165] Inside the stem, two NH_4^+ ions are present. A few years ago, Sket and Plavec[166] observed that the two bound NH_4^+ ions exit the stem in a distinctly asymmetrical fashion. In particular, they found that the NH_4^+ ion transport rate is ten times faster for the "upper" NH_4^+ ion than for the "lower" NH_4^+ ion. This was the first case that ion movement through a G-quadruplex is shown to be asymmetrical. In addition, they observed that the NH_4^+ ion never crosses the central G-quartet. Thus they concluded that this particular G-quadruplex does not exhibit ion-channel-like properties for NH_4^+. In this context, a fundamental question should be asked about the structural basis for such an asymmetrical ion movement. Furthermore, is there any general principle that allows one to predict whether a particular G-quadruplex may or may not exhibit ion-channel-like properties? As there are very few cases in which reliable experimental data are available with respect to ion movement in G-quadruplex DNA, we decided to use the ABF-MD computational approach to gain some insights into the asymmetrical NH_4^+ ion movement in this important G-quadruplex structure. The main objectives of this chapter are

(1) to further evaluate the accuracy of the ABF approach in obtaining free energy barriers for ion movement in G-quadruplexes and (2) to apply this computational approach to examine the very unusual ion movement in $[d(G_3T_4G_4)]_2$ G-quadruplex DNA.

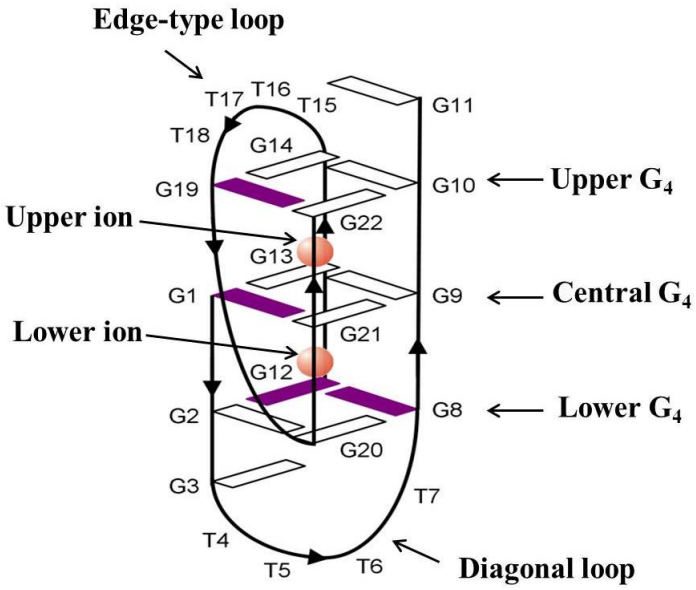

Figure 4-1. Schematic diagram showing the folding of $[d(G_3T_4G_4)]_2$ G-quadruplex and the locations of the two bound NH_4^+ ions (balls) inside the G-quadruplex stem. Filled and open rectangles are the guanine residues in syn and anti conformations, respectively.

4.2 Computational details

4.2.1 Model structure generation

Solution NMR structure of $[d(G_3T_4G_4)]_2$ (PDB code 1U64)[165] was used as the initial structure. Two ions were manually added to the two channel binding sites. All DNA systems were charge neutralized by adding 17 counter ions. The systems were immersed in a box of

TIP3P water molecules extending up to 10 Å from the solute in each direction. The systems were optimized and equilibrated using multiple initial minimization and dynamics runs. A 10000-step minimization of water and counter ions was carried out with solute and channel ions restrained by a force constant of 500 kcal/mol/Å^2 for each system. The systems were then subjected to a series of 20 ps restrained minimizations with the channel ions restrained by 500 and solute restrained by 300, 100, 10, and 0 kcal/mol/Å^2. The channel ions were relaxed during four step energy minimizations (20 ps each) in which the force constant on channel ions was decreased from 500 to 300, 100, 10, and 0 kcal/mol/Å^2. The systems were heated to 298 K over 50 ps at constant pressure with a force constant of 50 kcal/mol/Å^2 maintained for the solute and channel ions. These restraints were then scaled down in three stages (50 ps each) with the solute and channel ions restrained by 50 and 30 kcal/mol/Å^2 at the first step, 50 and 0 kcal/mol/Å^2 at the second step, and 30 and 0 kcal/mol/Å^2 at the final step. Eventually, the restraints were fully removed in the course of a 50 ps equilibration. Three different forms (NH_4^+, K^+, and Na^+) of $[d(G_3T_4G_4)]_2$ were prepared in the same fashion. Separate 20 ns regular MD simulations were performed to examine the dynamics of different parts of $[d(G_3T_4G_4)]_2$. In some MD simulations, two bound ions were fixed at their binding sites inside the channel by a harmonic potential (k = 100 kcal/mol/Å^2). RMSD calculations were performed with respect to the average structure. All MD simulations were performed in the isothermic-isobaric ensemble. Langevin dynamics and a Langevin piston algorithm were used to maintain the temperature and pressure at 298 K and 1 atm, respectively. The particle mesh Ewald (PME) method of calculating long-range electrostatic interactions was employed with a cut-off of 9 Å. Rattle was applied to constrain the bonds containing hydrogen atoms. A time step of 2 fs was used to integrate the equations of motion. All MD simulations

were performed by using the program NAMD2.7b3 and the CHARMM27 force field. The program VMD 1.8.7 was used in the visualization and RSMD analysis.

4.2.2 ABF computations

We divided the reaction pathway along the $[d(G_3T_4G_4)]_2$ stem axis into four windows in a total width of approximately 28 Å. Two central windows each being 4 Å wide were defined such that they include the upper and lower G-quartets, respectively, while the central G-quartet is covered by both windows. Two outer windows each being 10 Å wide were defined to extend the reaction coordinate into both edge-type and diagonal loops. The reaction coordinate for each window was defined along the central axis connecting the centers of mass of O6 atoms of two consecutive G-quartets in the same window. This axis was dynamically updated during the ABF simulation. The vector between the ion and the midpoint of the axis was projected onto the axis and considered as the reaction coordinate. For all windows, a cylindrical confining potential (k = 100 kcal/mol/Å^2) was applied to restrict ion movement in the plane norm to the dynamic axis. Within each window the force acting on the ion was averaged in 0.1 Å sized bins. Application of ABF was initiated after accumulation of 800 samples in individual bins. In some ABF simulations, harmonic restraints with a force constant of 3 kcal/mol/Å^2 were applied to fix C_1' atoms of the guanine residues. Production runs in each window were typically continued for 3-6 ns. In order to obtain 2D potential of mean force (PMF) profiles, we performed ABF simulations along eight reaction pathways within the XY plane for a series of Z positions. The eight directions were (X,Y) = (1,0), (−1,0), (0,1), (0,−1), (1,1), (−1,1), (1,−1), and (−1,−1). A 2D PMF map was reconstructed from these 1D data.

4.3 Results and discussion

4.3.1 PMF profiles for single ion movement

As we illustrated in the previous chapter, it is often useful to consider ion movement in a G-quadruplex DNA structure in two separate regions. One is within the G-quadruplex stem (or channel) and the other is from the stem entrance/exit point to bulk. In the discussion that follows, we treat these two regions separately. Figure 4-2 shows the PMF profile for a single NH_4^+ ion movement within the $[d(G_3T_4G_4)]_2$ G-quadruplex stem obtained from "unconstrained" ABF simulations. Here, it is important to point out that in the unconstrained ABF simulation runs, while biasing forces are constantly being applied to the NH_4^+ ion, the G-quadruplex DNA structure itself is completely unconstrained (*vide infra*). As seen in Figure 4-2, the shape of the PMF profile is clearly asymmetrical. In particular, while the free energy barrier for a NH_4^+ ion to cross the upper G-quartet is approximately 14 kcal/mol in order to exit the stem, the corresponding barrier to cross the lower G-quartet is higher, 17 kcal/mol. It is also important to note that the NH_4^+ ion movement between the two binding sites inside the stem is also high, ~17 kcal/mol.

The PMF profile for a single K^+ ion movement within the $[d(G_3T_4G_4)]_2$ stem is similar to that seen for NH_4^+ as can be seen in Figure 4-3. In contrast, Na^+ ions can move through the G-quadruplex stem much more easily, experiencing free energy barriers of only 5-7 kcal/mol (also see Figure 4-3). These observed similarity between NH_4^+ and K^+ and discrepancy between NH_4^+/K^+ and Na^+ are in good agreement with our previous ABF results on a parallel-stranded $[d(TG_4T)]_4$ G-quadruplex described in Chapter 3.

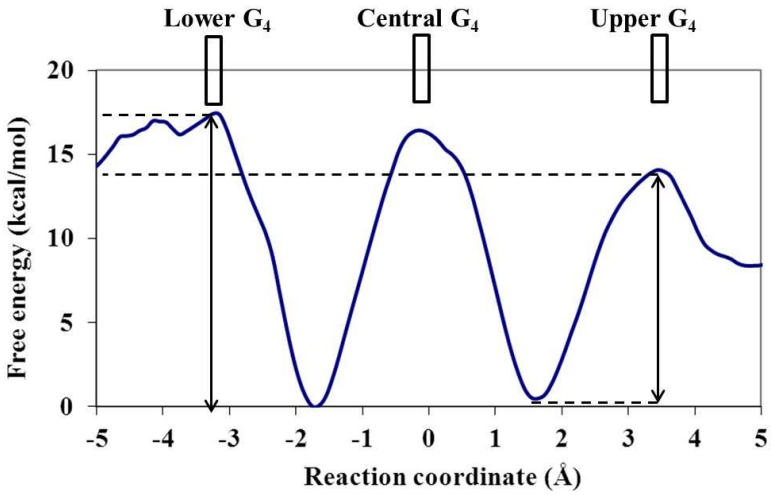

Figure 4-2. PMF profile for single NH_4^+ ion movement inside the $[d(G_3T_4G_4)]_2$ G-quadruplex stem.

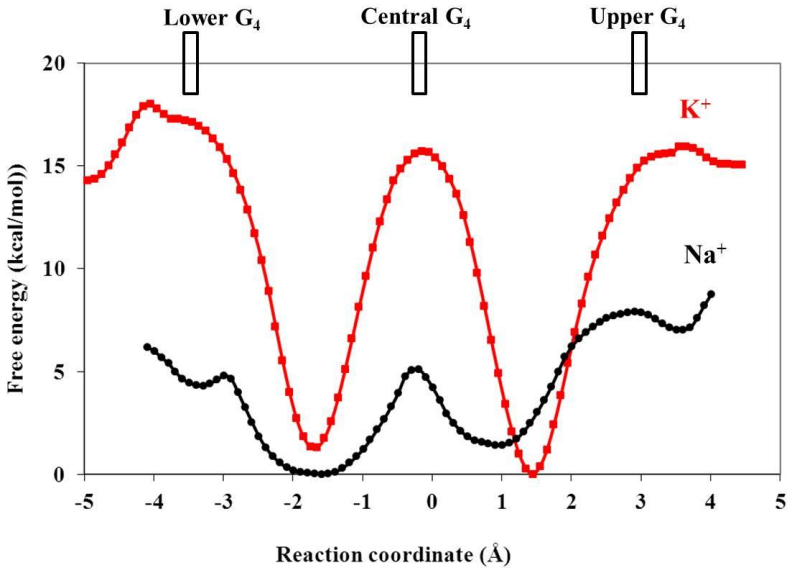

Figure 4-3. PMF profiles for single K^+ and Na^+ ion movement inside the $[d(G_3T_4G_4)]_2$ G-quadruplex stem.

In the aforementioned ABF simulations, we found that the force converges rather rapidly. Typically, 1 ns ABF simulations can produce accurate PMF profiles for ion movement. We also noticed that in some cases, long ABF simulations (e.g., 10-20 ns) in fact produced erroneous results. The origin of the problem comes from the fact that, in each ABF simulation run, we assumed single NH_4^+ ion occupancy inside the G-quadruplex stem. If the NH_4^+ ion is kept in a particular region for a long period of time, the whole G-quadruplex structure may become unstable. This may be particularly true when we sample high energy regions. To overcome this problem, in our previous chapter we added weak constraints on C1′ atoms (a force constant of only 3 kcal/mol/Å^2) to prevent the G-quadruplex from falling apart in the absence of bound NH_4^+ ions. Here, we refer these simulations as "constrained" ABF simulations. Our previous results showed that such weak constraints on C1′ atoms indeed do not significantly alter the free energy barrier for NH_4^+ ion to cross a G-quartet, as C1′ atoms are somewhat remote from the center of a G-quartet and their movements are not strongly coupled to that of a guanine base.

To further investigate the reproducibility of the unconstrained ABF results shown in Figure 4-2, we performed a complete set of constrained ABF simulations in which the positions of four C1′ atoms in each G-quartet were confined by harmonic potentials with a weak force constant of 3 kcal/mol/Å^2, similar to what we did in the previous study. The simulation results are summarized in Table 4-1. It is clear that the asymmetrical nature of the stem is also reflected in these constrained ABF simulations. For example, when the constraint is on the lower G-quartet, the barrier to cross the upper G-quartet is 11 kcal/mol. In comparison, when the constraint is on the upper G-quartet, the barrier to cross the lower G-quartet is 16 kcal/mol. It is also clear that, for each series of data shown in Table 4-1, the G-quartet on which the constraint is being applied exhibits the highest energy barrier.

Table 4-1. Free energy barriers for single NH_4^+ ion movement inside $[d(G_3T_4G_4)]_2$ G-quadruplex stem obtained from ABF simulations under various conditions.

	Free energy barrier (kcal/mol)[a]		
Constraint	lower G_4	central G_4	upper G_4
C1′ atoms on lower G_4	**20**	17	11
C1′ atoms on central G_4	17	**19**	15
C1′ atoms on upper G_4	16	16	**16**
None	17	17	14

[a]Boldface numbers highlight the energy barriers to cross the constrained G-quartets.

We also did the same constrained simulation test for the parallel-stranded $[d(TG_4T)]_4$ G-quadruplex. The results suggest that the NH_4^+ ion movement inside the $[d(TG_4T)]_4$ stem is essentially symmetrical (see Table 4-2). This further proves that the asymmetry observed for NH_4^+ movement in $[d(G_3T_4G_4)]_2$ is real. A comparison between unconstrained and constrained ABF simulations for both G-quadruplex sequences leads to the following conclusions. First, the unconstrained ABF simulations yield reliable PMF profiles for ion movement in $[d(G_3T_4G_4)]_2$. Second, the weak constraint on C1′ atoms tends to increase the energy barrier of crossing the constrained G-quartet by 2-3 kcal/mol for NH_4^+ and K^+ ion movement, but has negligible effects for Na^+ ion movement.

Table 4-2. Free energy barriers for single NH_4^+ ion movement inside a parallel-stranded $[d(TG_4T)]_4$ G-quadruplex stem obtained from ABF simulations under the listed conditions.[a]

	Free energy barrier (kcal/mol)[b]			
Constraint	$[G5]_4$	$[G4]_4$	$[G3]_4$	$[G2]_4$
C1′ atoms on $[G5]_4$	**18**	14	14	11
C1′ atoms on $[G4]_4$	14	**17**	14	14
C1′ atoms on $[G3]_4$	17	13	**16**	12
C1′ atoms on $[G2]_4$	12	15	16	**16**

[a]The residue numbering in $d(TG_4T)$ DNA sequence is 5′-T1-G2-G3-G4-G5-T6-3′.

[b]Boldface numbers highlight the energy barriers to cross the constrained G-quartets.

Once the NH_4^+ ion exits the quadruplex stem, its movement is no longer confined to the stem (channel) axis. As a result, it is necessary to consider its movement in all directions. In practice, we used ABF simulations to perform a 2D free energy mapping in a plane perpendicular to the axis, i.e., the XY plane. As seen from Figure 4-4, the 2D free energy landscapes for NH_4^+ ion movement are rather flat as opposed to that seen inside the stem shown in Figure 4-2. It is also important to note that, although the 2D maps look very different for the two ends, the actual free energy costs for the NH_4^+ ion to reach bulk from the two stem exit points are essentially the same, ca. 2 kcal/mol. This is a somewhat surprising observation. A common argument found in the literature is that, as the diagonal loop is "in the way" of ion passage to bulk, the energy barrier must be much higher for the NH_4^+ ion to exit the quadruplex stem from the end containing such a diagonal loop than from the end containing other loops such as edge-type loops. The results shown in Figure 4-4, suggest a quite different picture. In fact, it appears that the NH_4^+ ion, once exiting the stem, can always find a "sideway" direction which imposes a minimum free energy cost, regardless of the type of loop at the stem end. In other words, the NH_4^+ ion does not need to

get close to any loop. The loop topology determines only the direction at which the NH_4^+ ion travels toward bulk.

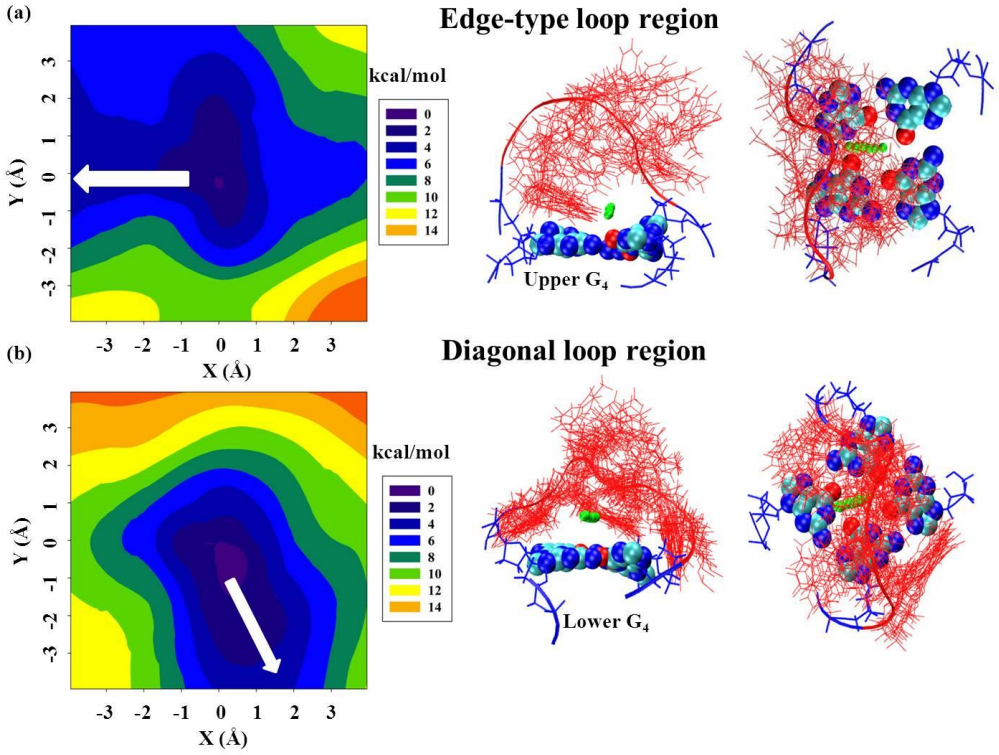

Figure 4-4. (Left) 2D free energy landscapes, (middle) side, and (right) top views of the single NH_4^+ ion movement in the edge-type and diagonal loop regions of $[d(G_3T_4G_4)]_2$ G-quadruplex. The least resistant pathway for sideway NH_4^+ ion movement is marked by an arrow in each 2D map and a string of green balls in side and top views. In both side and top views, twenty simulation frames of the loop residues are shown to indicate the scope of loop movement.

Now we have generated a complete picture about single NH_4^+ ion movement from its binding site inside the quadruplex stem to bulk. By combining ABF simulation results shown in Figures 4-2 and 4-4, we can estimate the free energy barriers for the lower and upper NH_4^+ ions to reach bulk to be 19 ± 2 and 16 ± 2 kcal/mol, respectively. These computed values are in

qualitative agreement with the experimental results reported by Sket and Plavec for the same G-quadruplex sequence.[166]

4.3.2 Regular MD simulations

To supplement the aforementioned ABF simulations, we performed a regular 20 ns unconstrained MD simulation to further monitor the NH_4^+ ion movement in $[d(G_3T_4G_4)]_2$ G-quadruplex structure. Figure 4-5 shows several snap shots of the simulation trajectories. Initially, the two bound NH_4^+ ions are in their equilibrium positions within the quadruplex stem, each being sandwiched between two G-quartets. The simulation shows that while the upper NH_4^+ ion very quickly (in approximately 100 ps) exits the stem by passing through the upper G-quartet, the lower NH_4^+ ion remains at the binding site for as long as 8 ns. In the end, the lower NH_4^+ ion exits the stem through the lower G-quartet and the whole G-quadruplex structure becomes unstable (data not shown). Remarkably, during the entire period of 8 ns before the lower NH_4^+ exits the stem, it never once crosses the central G-quartet. This indicates that the energy barrier to cross the central G-quartet is prohibitively high, which is consistent with the ABF simulation results discussed earlier. In this context, ion movement between two adjacent sites within a G-quadruplex stem has been seen in previous MD simulations.[31,171,196] More recently, Reshetnikov and co-workers[96] even observed a concerted ion movement in which a free ion from bulk enters the G-quadruplex stem from one side, while simultaneously expelling the bound ion into bulk from the other side. Although caution must be exercised to avoid overinterpretation of the regular MD simulation results, we believe that the observed contrast in ion movement between the two bound NH_4^+ ions within the aforementioned 20 ns MD simulation does provide additional support for the free energy results obtained from ABF simulations.

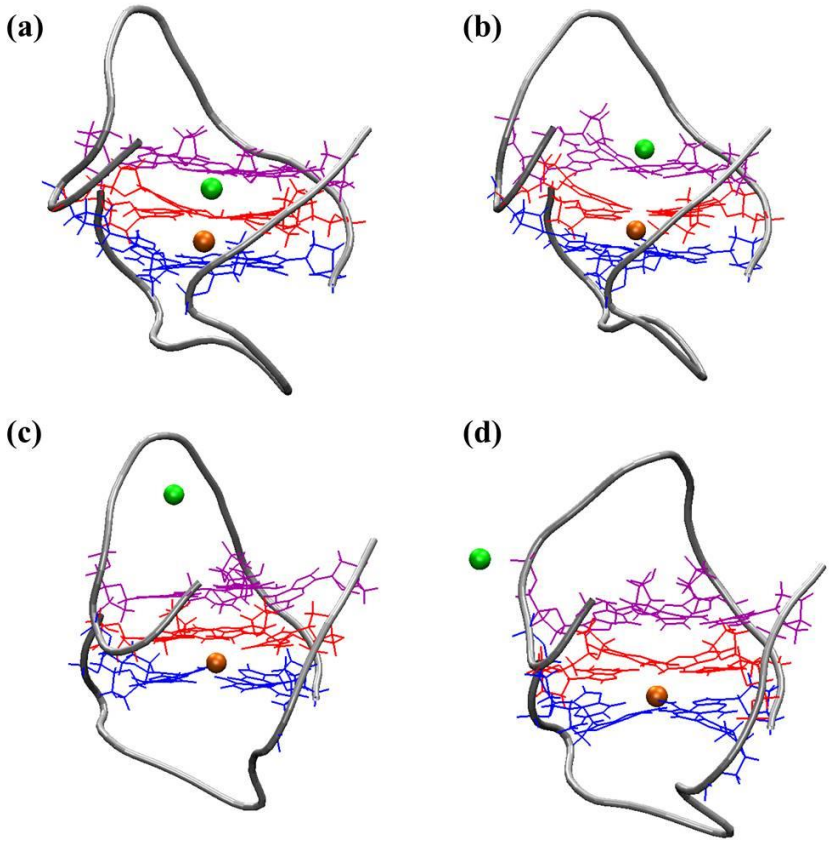

Figure 4-5. Representative frames from a 20 ns unconstrained MD simulation run for [d($G_3T_4G_4$)$_2$] G-quadruplex. Initially two bound NH_4^+ ions are inside the quadruplex stem. (a) less than 100 ps, (b) 100-700 ps, (c) 700 ps-4.9 ns, and (d) 4.9-8.2 ns. For clarity two different colors are used to distinguish the two NH_4^+ ions.

To further investigate whether the rigidity of a G-quartet is, in any way, related to its overall structural dynamics, we performed another 20 ns regular MD simulation in which the two bound NH_4^+ ions were fixed at their equilibrium binding sites inside the [d($G_3T_4G_4$]$_2$ G-quadruplex stem. This simulation allows us to analyze dynamics for different parts of the G-quadruplex structure including stem backbone, diagonal and edge-type loops. Figure 4-6 displays

the RMSD analysis of this simulation run. It is seen that with two bound NH_4^+ ions in place the whole quadruplex structure is very stable. Several observations are worth mentioning. First, the two end G-quartets are less stable (showing larger fluctuations) than the central one. This is probably due to the fact that while each of the end G-quartets interacts with only one bound NH_4^+ ion, the central G-quartet is stabilized simultaneously by both NH_4^+ ions. Second, the stem backbone is also very stable. Third, the diagonal loop is the most flexible.

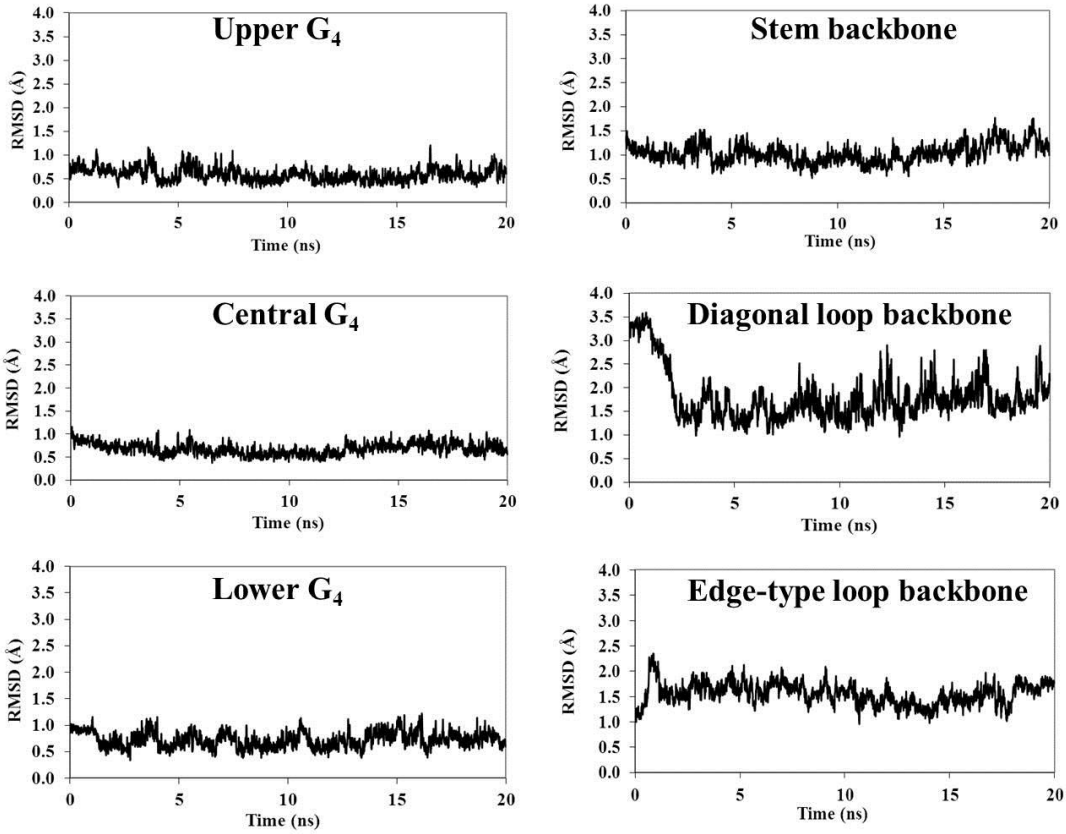

Figure 4-6. RMSD results for different parts of the $[d(G_3T_4G_4)_2]$ G-quadruplex during a 20 ns regular MD simulation run in which the two bound NH_4^+ ions are constrained at their binding sites.

4.3.3 Stiffness of a G-quartet is determined by base stacking: A hypothesis

Although our ABF simulations have reproduced reasonably well the experimental results, they do not automatically point to the structural factor that determines the various free energy barriers for NH_4^+ ion movement. One can imagine that the energy barrier for NH_4^+ ion to cross a G-quartet inside a G-quadruplex stem must be related to the stiffness or flexibility of that G-quartet. This is because the G-quartet in question must open, at least partially, in order to allow NH_4^+ ion to pass through, and the opening of a G-quartet would destroy partially the base stacking. Therefore, a rigid G-quartet must impose a high energy barrier for a passing NH_4^+ ion, whereas a dynamical or flexible G-quartet should be less difficult to move through. For the G-quartet inside a G-quadruplex stem, its stiffness or flexibility appears to be dominated by the base stacking interactions between the G-quartet in question and adjacent G-quartets. Here, we hypothesize that inside a G-quadruplex stem the base stacking arrangement around a particular G-quartet determines its stiffness or flexibility. Of course, for a G-quartet appearing at the end of a G-quadruplex stem, base stacking from loop residues, if any, may also contribute to its stiffness.

Now let us examine base stacking interactions inside a G-quadruplex stem. As Webba da Silva[197] pointed out, there are four possible types of intrastrand base stacking according to the glycosidic bond angle (GBA) propagation along the 5′→3′ direction: 5′-syn-anti, 5′-syn-syn, 5′-anti-syn, and 5′-anti-anti. Recently, Cang et al.[198] performed free energy calculations that show the following relative stability trend for these four types of base stacking: syn-anti > anti-anti > anti-syn > syn-syn. In particular, G-quartets in the syn-anti stacking mode may be more stable than those in the anti-anti stacking by as large as 16 kcal/mol. Close inspection of the $[d(G_3T_4G_4)]_2$ structure reveals that while the base stacking between the upper and central G-quartets is of the anti-anti type, the lower and central G-quartets are in the syn-anti base stacking

mode. These two types of base stacking are illustrated in Figure 4-7. In the syn-anti base stacking mode, the adjacent five-membered rings of the guanine bases are nearly perfectly on top of one another, allowing very sufficient base stacking. In contrast, the two bases are offset in the anti-anti base stacking mode. This immediately suggests that the lower G-quartet must be much more rigid than the upper G-quartet. This prediction is in full agreement with both the ABF simulation and experimental results. Our hypothesis offers an explanation as to why the NH_4^+ ion movement in the $[d(G_3T_4G_4)]_2$ quadruplex stem exhibits a directional asymmetry. Furthermore, several other experimental observations seem to support our hypothesis. First, the NH_4^+ ion movement in the parallel-stranded $[d(TG_4T)]_4$ quadruplex stem appears to exhibit the fastest transport rates,[169] and the NH_4^+ ion movement is symmetrical, as we discussed earlier. This is because all G-quartets in $[d(TG_4T)]_4$ are involved in the anti-anti base stacking.[199,200] Second, the NH_4^+ ion movement in the unimolecular $[d(G_4(T_4G_4)_3)]$ G-quadruplex stem is rather slow.[167] This is because the GBA propagation for the stem is 5′-syn-anti-syn-anti, which makes both end G-quartets very rigid. Third, Plavec and co-workers[168] reported that the NH_4^+ ion movement throughout the antiparallel $[d(G_4T_3G_4)]_2$ G-quadruplex stem is too slow to be reliably detected. We note that not only the two end G-quartets in $[d(G_4T_3G_4)]_2$ are involved in the most stable syn-anti stacking they also participate in significant base stacking from the edge-type T_3 loops. As a result, the two very stiff end G-quartets slow down the NH_4^+ ion flow throughout the quadruplex stem.

Our hypothesis not only provides new insights into the structural factor that controls ion movement in G-quadruplexes, but also offers a guideline for the design of G-quadruplex ion channels. For example, the hypothesis predicts that those G-quadruplexes with all G-quartets being in the anti-anti stacking mode likely exhibit ion channel behaviors. As mentioned earlier, this GBA arrangement occurs in the parallel-stranded $[d(TG_4T)]_4$. The same anti-anti base

stacking is also seen in the crystal structure of the K^+ form of four repeat human telomeric DNA $d(AG_3(T_2AG_3)_3)$.[201] We should emphasize, however, that our hypothesis about G-quartet stiffness is perhaps only valid in the context of movement of large ions such as NH_4^+ and K^+. For Na^+, it is not clear whether base stacking interactions would make a significant contribution to the energy barrier for ion passage. The currently available experimental and computational data appear to suggest that Na^+ ions pass through all G-quadruplex channels with essentially the same free energy barrier of ca. 5 kcal/mol. This also suggests that while some G-quadruplex structures may not be channel-like for NH_4^+ and K^+ ions, they always allow Na^+ ions to pass through with ease.

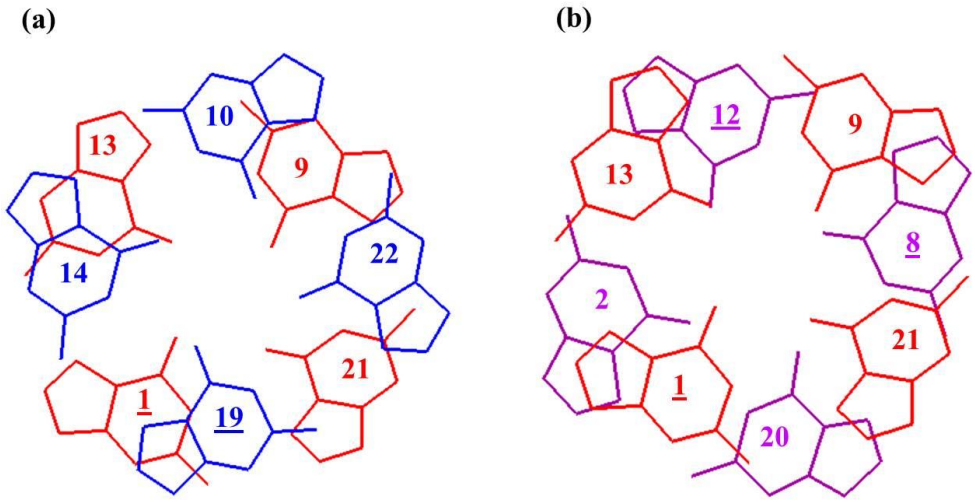

Figure 4-7. (a) 5′-Anti-anti base stacking between the upper and central G-quartets and (b) 5′-syn-anti base stacking between the lower and central G-quartets in the $[d(G_3T_4G_4)]_2$ G-quadruplex stem. The underlying residues are in syn conformation.

4.3.4 Ion movement in $[d(G_3T_4G_4)]_2$ channel containing multiple NH_4^+ ions

In this section, we will investigate ion movement along the $[d(G_3T_4G_4)]_2$ channel containing multiple NH_4^+ ions. Several models will be used. The initial minimized and equilibrated structure was taken from our previous regular MD simulations as discussed in Section 4.2.1. In all simulations, the ion movement in the plane norm to the dynamic axis was confined by a harmonic potential (k = 100 kcal/mol/Å^2). The average force acting on the ion was accumulated in 0.1 Å bins. A production run of 2 ns was performed for each model. To ensure a satisfactory sampling distribution along the reaction coordinate the ion pathway was divided into two windows with the same width. When we studied multiple-ion models in Chapter 3, we used the average force profiles to illustrate the consequence of repulsion forces. In this chapter, we will use the free energy profiles to describe the consequence of the repulsion force on the energetics of ion movement along the channel. Table 4-3 shows the summary of the following models.

Table 4-3. Summary of the models to study ion movement in $[d(G_3T_4G_4)]_2$ channel containing multiple NH_4^+ ions.[a]

Model	Dynamics axis	Fixed C1'	Ion studied	Movement of other ion
1	U and C	Unconfined	U (1.7 to 5.7 Å)	L fixed (k = 100 kcal/mol/Å^2)
2	L and C	Unconfined	L (-1.7 to -5.7 Å)	U fixed (k = 100 kcal/mol/Å^2)
3	U and C	Constrained	U (1.7 to 5.7 Å)	L fixed (k = 50 kcal/mol/Å^2)
4	L and C	Constrained	L (-1.7 to -5.7 Å)	U fixed (k = 50 kcal/mol/Å^2)

[a] U, C, and L represent upper, central, and lower states, respectively.

Model 1:

In this model we studied ion transport of the upper NH_4^+ ion along the $[d(G_3T_4G_4)]_2$ channel containing two NH_4^+ ions. The position of the lower NH_4^+ ion was fixed during the simulation by a harmonic potential (k = 100 kcal/mol/Å^2). The ABF method was used to obtain the PMF profile for the upper NH_4^+ ion movement from RC = 1.7 to 5.7 Å (see Figure 4-8). The reaction coordinate was chosen along the channel axis of $[d(G_3T_4G_4)]_2$, which was defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms from the upper and central G-quartets. C1' atoms of guanine residues were unrestrained. The PMF profile for the upper NH_4^+ ion movement is shown in Figure 4-8. According to this profile, the upper NH_4^+ ion experiences a free energy barrier of 8 kcal/mol to cross the upper G-quartet. As mentioned earlier, in the single-ion model with the same configuration, the corresponding barrier is 14 kcal/mol. Since in the current model, there is a NH_4^+ ion located at the lower binding site, ions repel each other. As a result, the upper ion will be kicked out and this provides a source of energy for the ion passage through the upper G-quartet. At the moment, it is not obvious how precise the magnitude of the barrier is, but the overall impact of introducing another ion in the channel makes sense as a lower energy barrier is observed in the multiple-ion model.

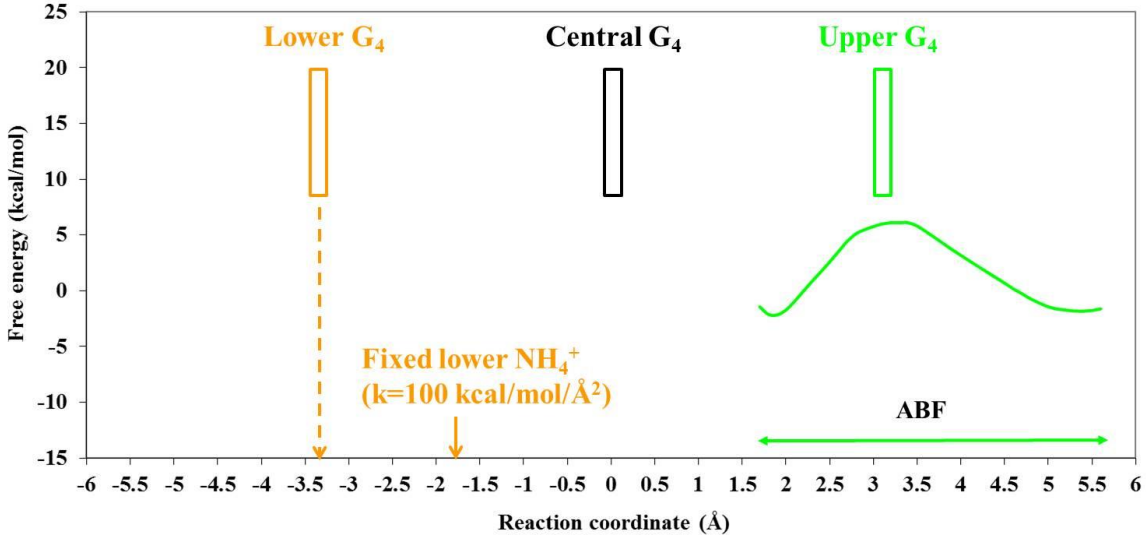

Figure 4-8. The schematic display and the PMF profile for the upper NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 1). The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet. The green and the orange colors represent the upper and lower ions, respectively. The green double arrow indicates the range of movement for the upper NH_4^+ ion.

Model 2:

In this model we used the ABF method to obtain the PMF profile for the lower NH_4^+ ion movement from RC = -1.7 to -5.7 Å (see Figure 4-9). The position of the upper NH_4^+ ion was fixed during the simulation by a harmonic potential (k = 100 kcal/mol/Å^2). The reaction coordinate was chosen along the channel axis of $[d(G_3T_4G_4)]_2$, which was defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms from the lower and central G-quartets. C1' atoms of guanine residues were unrestrained. The PMF profile for the lower NH_4^+ ion movement is shown in Figure 4-9. According to this profile, the lower NH_4^+ ion experiences a free energy barrier of 16 kcal/mol to cross the lower G-quartet. As mentioned earlier, in the single-ion model with the same configuration, the corresponding barrier is 17 kcal/mol. Since in

the current model, a NH_4^+ ion is located at the upper binding site, there is a repulsion force between ions in the channel as discussed in Model 1. Therefore, we expect to see a lower energy barrier on the lower G-quartet compared to the single-ion model. However, the barriers in these two models differ by only 1 kcal/mol, much smaller than expected from ion-ion repulsion.

To check if the PMF profile for this model is reliable, we analyzed the sampling distribution for the lower ion movement along the reaction pathway during the simulation. As discussed in the single-ion model, there are some specific structural features at the lower end of $[d(G_3T_4G_4)]_2$. These could make the sampling in this region challenging. Figure 4-10 shows the sampling distribution after 0.5, 1, 1.5, and 2 ns simulations. Although for RC = -1.75 to -2.75 Å, the sampling is not as smooth as the other regions, adequate number of samples has been taken along the reaction coordinate after 0.5 ns. As a result, the unusual high energy barrier observed at the lower end could not be due to the sampling in this region. At the moment, we do not fully understand what may cause this, so we need to perform more tests to further investigate this problem. It should be noted that the sudden jump between the samples at about RC = -3.5 Å is due to dividing the windows into two at this point and does not affect our interpretations.

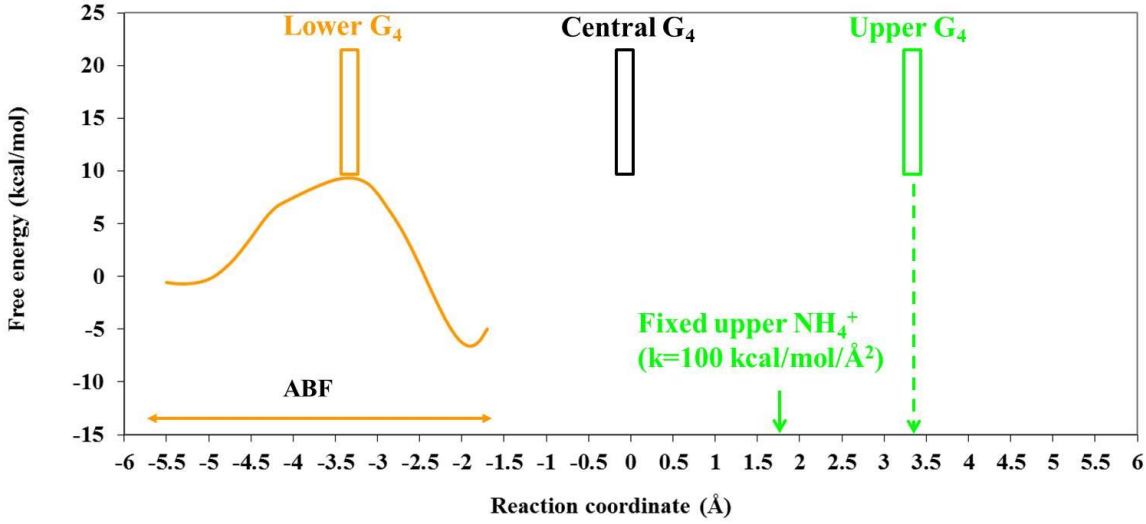

Figure 4-9. The schematic display and the PMF profile for the lower NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 2). The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet. The green and the orange colors represent the upper and lower ions, respectively. The orange double arrow indicates the range of movement for the lower NH_4^+ ion.

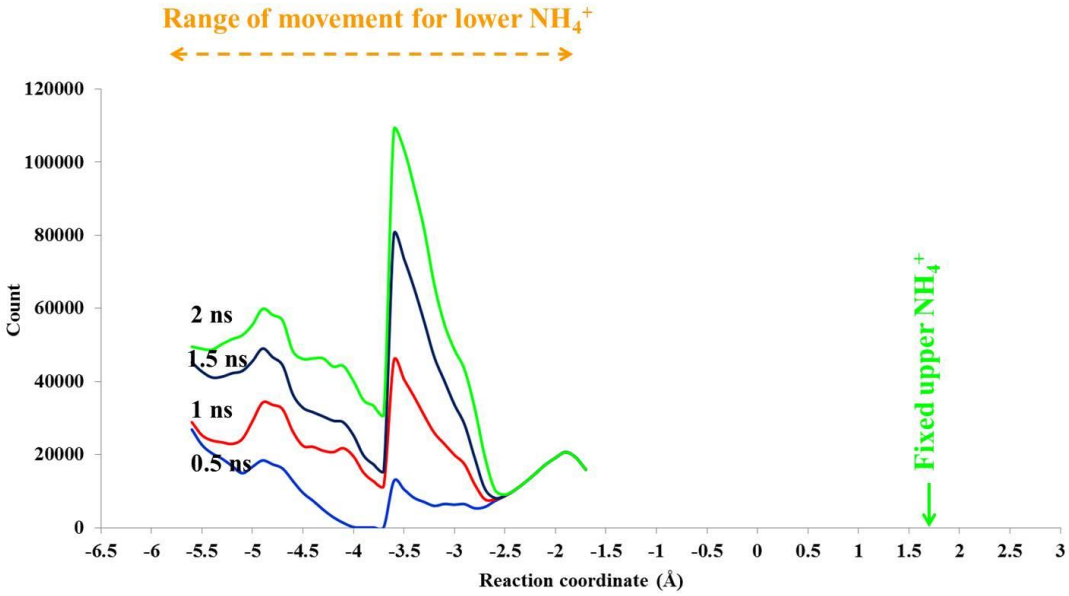

Figure 4-10. Sampling distribution of the lower NH_4^+ ion along the $[d(G_3T_4G_4)]_2$ channel in the multiple-ion Model 2 after 0.5, 1, 1.5, and 2 ns. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

Model 3:

In this model we studied ion transport of the upper NH_4^+ ion movement along the $[d(G_3T_4G_4)]_2$ channel in a system similar to Model 1. However, in this model we particularly focused on the effect of reducing the harmonic potential that is responsible to fix the lower ion. The idea was to allow the lower ion to adjust itself partially if there is too much repulsion between two ions while it is still constrained inside the channel. The position of the lower NH_4^+ ion was fixed during the simulation by a harmonic potential (k = 50 kcal/mol/Å^2) (see Figure 4-11). The positions of C1' atoms of all guanine residues were fixed by a harmonic potential (k = 3 kcal/mol/Å^2). The PMF profile for the upper NH_4^+ ion movement is shown in Figure 4-11.

According to this profile, the upper NH_4^+ ion experiences a free energy barrier of 5.6 kcal/mol to cross the upper G-quartet. As mentioned earlier, in the single-ion model, the corresponding barrier is 14 kcal/mol when none of the C1' atoms are constrained. We also discussed how introducing constraint on C1' atoms can increase the barriers on G-quartets. Since in the current model, a NH_4^+ ion is located at the lower binding site, ions will repel each other. Therefore, the upper ion will be kicked out and is expected to encounter a lower free energy barrier to cross the upper G-quartet. However, since the positions of C1' atoms are fixed in the current model, the barrier should slightly increase because the G-quartet should partially open during NH_4^+ ion passage and the constraints on C1' atoms make this more difficult. On the other hand, we reduced the force on the lower NH_4^+ ion by 50 kcal/mol/Å^2. This provides more freedom for the lower NH_4^+ ion movement in case ions repel each other. So, the repulsion effects on the upper NH_4^+ ion will be slightly compensated.

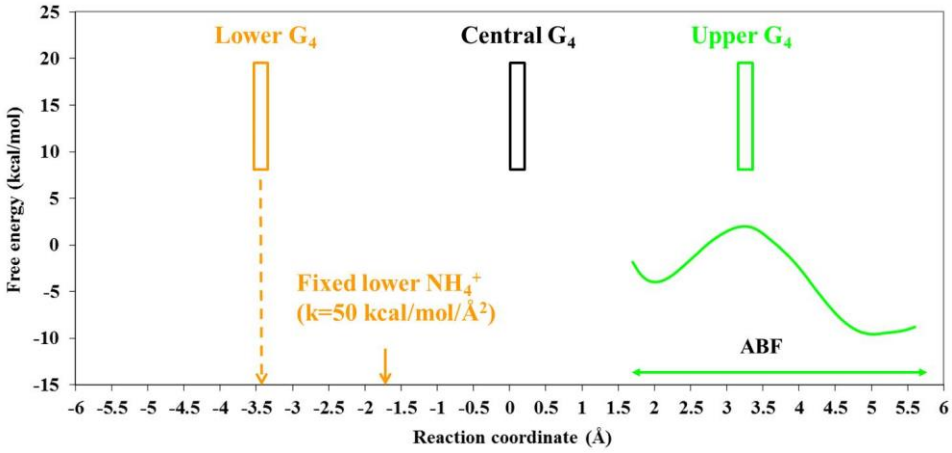

Figure 4-11. The schematic display and the PMF profile for the upper NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 3). The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet. The green and the orange colors represent the upper and lower ions, respectively. The green double arrow indicates the range of movement for the upper NH_4^+ ion.

Model 4:

In this model we studied the lower NH_4^+ ion movement from RC = -1.7 to -5.7 Å while the position of the upper NH_4^+ ion was fixed by a lower harmonic potential compared to Model 2 (k = 50 kcal/mol/Å^2) (see Figure 4-12). The PMF profile for the lower NH_4^+ ion movement is shown in Figure 4-12. According to this profile, the lower NH_4^+ ion experiences a free energy barrier of 14 kcal/mol to cross the lower G-quartet. As mentioned earlier, in the single-ion model, the corresponding barrier is 17 kcal/mol when none of the C1' atoms are constrained. As discussed in Model 3, the presence of a neighbor ion in the channel which is fixed by a lower harmonic potential will reduce the free energy barrier for the passage of the other ion. On the other hand, fixing the positions of C1' atoms will make the G-quartets less flexible during ion passage. Considering all these, 14 kcal/mol as the energy barrier for the lower NH_4^+ ion passage through the lower G-quartet seems reasonable qualitatively. According to the specific structural features and the rigidity of the lower end, we expect a higher sensitivity at this end to the configuration changes. However, a similar response to the new conditions was observed between both channel ends.

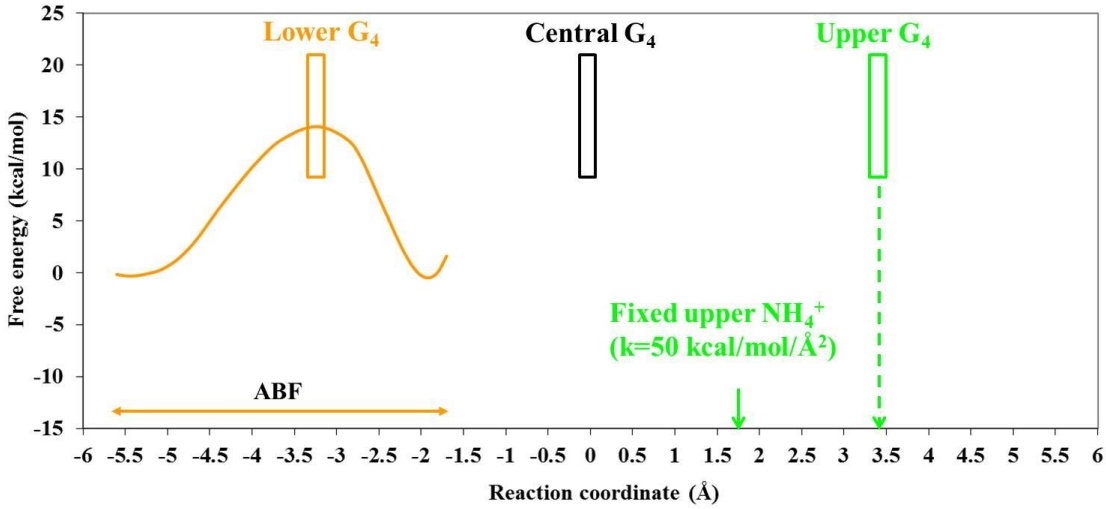

Figure 4-12. The schematic display and the PMF profile for the lower NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 4). The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet. The green and the orange colors represent the upper and lower ions, respectively. The orange double arrow indicates the range of movement for the lower NH_4^+ ion.

4.4 Conclusion

In this chapter we performed extensive regular MD and ABF-MD simulations to obtain information about the energetics of NH_4^+ ion movement along the $[d(G_3T_4G_4)]_2$ G-quadruplex channel. The computations reproduced qualitatively the experimental results. That is, the free energy barriers for a bound NH_4^+ ion to exit from the two ends of the $[d(G_3T_4G_4)]_2$ G-quadruplex stem differ by 3-4 kcal/mol, which corresponds to approximately a factor 10-100 in transport rates at 298 K. The simulations also confirmed that the energy barrier to cross the central G-quartet is high. The 2D free energy maps suggest that once the NH_4^+ ion exits the G-quadruplex stem, it faces a very shallow free energy landscape that allows it to easily reach bulk, regardless the actual loop topology at the end. We further hypothesized that the stiffness of a G-quartet is

primarily determined by the base stacking around it. Therefore, the stiffness of a G-quartet should follow the known relative stability trend for base stacking: syn-anti > anti-anti > anti-syn > syn-syn. For a G-quartet inside the quadruplex stem, base stacking interactions from adjacent G-quartets are important. For an end G-quartet, one should consider base stacking from both the inner G-quartet and possibly loop residues. This hypothesis appears to be consistent with currently available dynamical information about ion movement within G-quadruplex DNA structures. We also used several models to study NH_4^+ ion movement along the $[d(G_3T_4G_4)]_2$ channel containing multiple ions. In general, a lower energy barrier was observed for ion transport through both end G-quartets when the neighboring site is occupied. Although the asymmetry in ion transport rate between the two sites was also observed in the multiple-ion models, it is desirable to have an independent method to verify the free energy barriers obtained by the ABF-MD method.

Chapter 5

Umbrella sampling MD studies of ion movement in $[d(TG_4T)]_4$

5.1 Introduction

In the previous two chapters, we used the adaptive biasing force (ABF) method to obtain free energy profiles for Na^+, K^+, and NH_4^+ ion movement along G-quadruplex channels. The free energy barriers for ion movement especially for K^+ and NH_4^+ along G-quadruplex channels are considerably higher than those seen in some other ion channels such as K^+ ion channel. Therefore, it is more challenging to achieve a smooth sampling profile in G-quadruplexes. In the ABF method, average forces are accumulated in bins along the reaction coordinate and in each bin a biasing force, which is equal to the estimated average force in the same bin but of the opposite direction, is applied during simulation. This operation allows the ion to overcome steep free energy barriers along the reaction coordinate, so that the ion can move along the entire channel. As our ABF-MD work is the first and the only study so far to examine potential of mean force (PMF) profiles for ion transport in G-quadruplexes, it is highly desirable to evaluate the accuracy of the computational results by using an independent MD method. To this end, there are several other advanced MD techniques such as umbrella sampling (US),[113,145] steered molecular dynamics (SMD)[137-141], and metadynamics[146] that can be used to obtain free energy information about the system. These methods use different procedures to overcome free energy barriers and improve sampling along the reaction coordinate. In this chapter, we decided to use the US method to study ion movement along G-quadruplex DNA channels. This is the first time that the US method is used to tackle the ion movement problem in DNA systems as explained in Chapter 2. US is a well-established technique that uses a bias potential to keep the ion at the center of a

defined window ensuring sufficient sampling in this window during the simulation. In an US simulation, the overall reaction coordinate is divided to several windows in which the simulations can be simultaneously run. Ideally the bias potential should be chosen such that the ion in each simulation remains within the defined window and also there is a reasonable overlap between two adjacent windows. Here, we will first investigate how US simulations can be applied to G-quadruplex DNA and then obtain PMF profiles for ion transport along the $[d(TG_4T)]_4$ channel containing either a single ion or multiple ions.

5.2 Computational details

5.2.1 Model structure generation

Similar to the ABF study of $[d(TG_4T)]_4$, an NMR structure of $[d(TG_4T)]_4$ (PDB code 2JT7)[184] was used as the initial model. The same MD minimization and equilibration procedures as described in Chapter 3/Section 3.2.1 were used to prepare the minimized structures. All MD simulations were performed in the isothermic-isobaric ensemble. Langevin dynamics and a Langevin piston algorithm were used to keep the temperature and pressure at 310 K and 1 atm, respectively. The particle mesh Ewald (PME) method of calculating long-range electrostatic interactions was used with a cut-off of 9 Å. Rattle was applied to constrain the bonds containing hydrogen. A time step of 2 fs was used to integrate the equations of motion. All MD simulations were performed using the program NAMD2.7b3 and the CHARMM27 force field at the SHARCNET computing facilities on HP servers running CentOS 5.x with Xeon 2.83 GHz processors (each having 16 GB of RAM). Typically eight processors were used for each

calculation. The program VMD 1.8.7 was used for visualization and analysis. The minimized $[d(TG_4T)]_4$ G-quadruplex DNA structures used in this study are shown in Figure 3-2.

5.2.2 US computations

In the US method, the overall reaction coordinate should be divided into several windows and the ion should be kept within each window by a bias potential. In order to find the optimal bias potential that could guarantee the ion movement inside each defined window and could also produce enough overlap between two neighboring windows, we performed simulations for the $[d(TG_4T)]_4$ structure in the presence of one K^+ ion in the channel and found k = 100 kcal/mol/Å^2 to be a reasonable bias potential. The details of this testing procedure are presented later in the following section.

All US simulations were performed using the program NAMD2.7b3 and the CHARMM27 force field at the SHARCNET computing facilities. The final PMF profiles were obtained after unbiasing the ion distribution using the weighted histogram analysis method (WHAM) program.

5.3 Results and discussion

5.3.1 Determination of optimal bias potential constants

To find out the optimal bias potential constant for US simulation in G-quadruplex systems, we performed four sets of simulations using different bias potential constants to evaluate the K^+ ion movement along the central section of the $[d(TG_4T)]_4$ channel. In all simulations, we chose the reaction coordinate along the channel axis (Z) of $[d(TG_4T)]_4$ and performed simulations

in 13 separate windows, each being 0.25 Å in width. Bias potential constants of k = 25, 50, 100, and 125 kcal/mol/Å^2 were used to hold the ion within each window. The channel axis was defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms from each of the two central G-quartets (G3 and G4 as defined previously in Figure 3-1). The ion movement in the plane norm to the dynamic axis was confined by a harmonic potential (k = 100 kcal/mol/Å^2). Within each window the force acting on the ion was averaged in 0.1 Å bins. Typical production runs of 2 ns for each window were performed. Weak harmonic restraints (with a force constant of 3 kcal/mol/Å^2) were applied to the C1' atoms of guanine residues for the two central G-quartets. Figures 5-1 and 5-2 show the window overlaps and the PMF profiles from these US simulations.

First, let us examine the window overlaps shown in Figure 5-1. There are three distinct regions: (a) RC = 0.00 or 3.00 Å, (b) RC = 1.50-1.75 Å, and (c) other regions. In the region of RC = 0.00 or 3.00 Å, ion positions essentially did not change during simulations and there are good overlaps between adjacent windows for all four bias potentials. In the regions of RC = 1.50-1.75 Å, a bias potential of k = 25 kcal/mol/Å^2 is not strong enough to keep the ion within each window. This is not acceptable because if the ion is not located at the proper position, the number of counts will significantly decrease as can be seen in Figure 5-1. In this case, the probability of finding the ion will also decrease and a huge and perhaps an unreal energy barrier at this point will be calculated. For RC = 0.00 Å if we compare four k values, we can see that the full width at the half maximum (FWHM) is 0.2, 0.2, 0.2, and 0.2 Å for all four k values. The width is 0.5, 0.5, 0.5, and 0.6 Å for k = 125, 100, 50, and 25 kcal/mol/Å^2, respectively. This shows that for low energy barrier the ion will most likely stay at the defined window center even by using a small k.

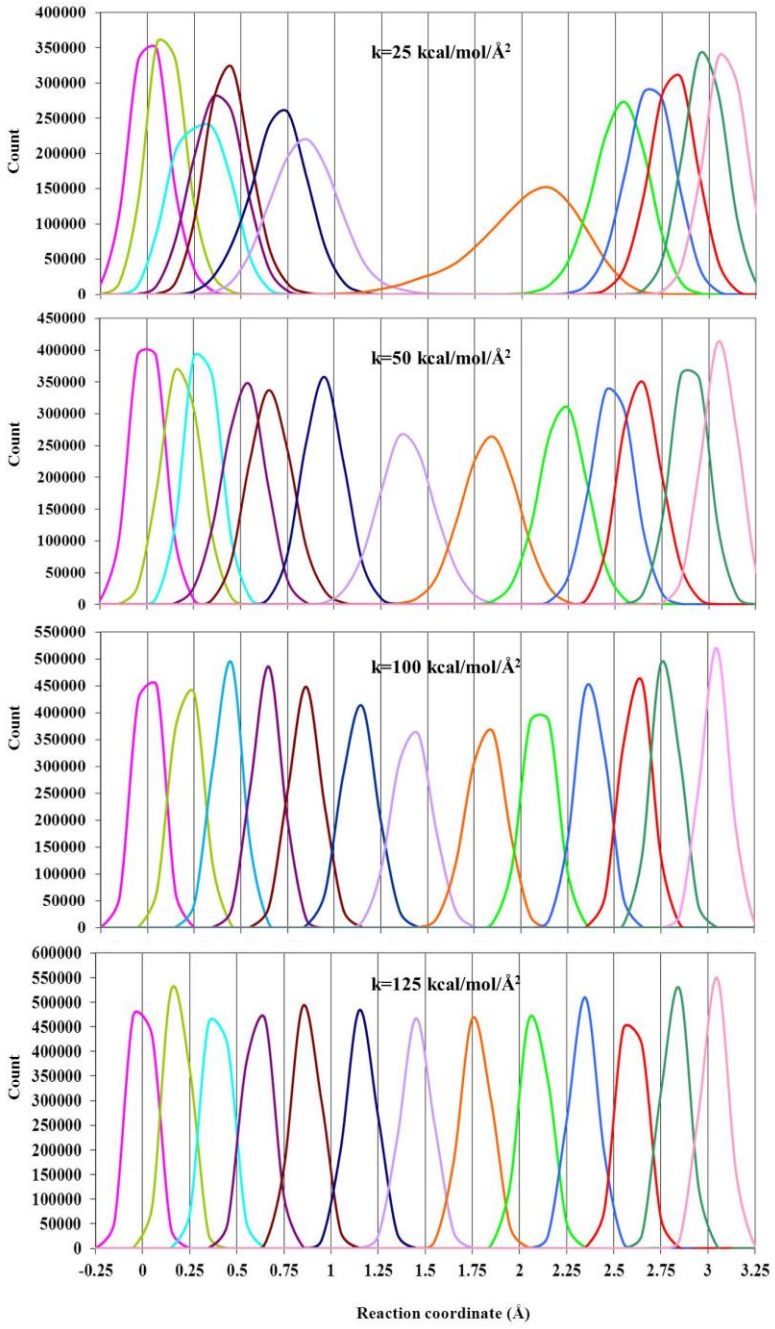

Figure 5-1. Window overlaps for single K^+ ion movement along the $[d(TG_4T)]_4$ channel using k = 25, 50, 100, and 125 kcal/mol/Å^2. The vertical lines correspond to the center of windows in the US simulations.

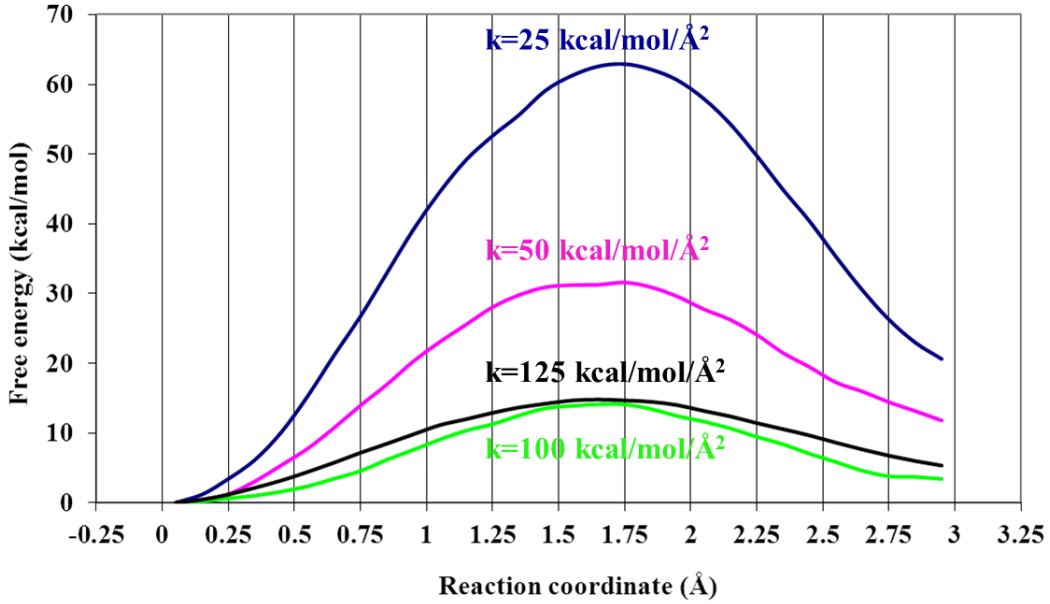

Figure 5-2. PMF profiles for single K^+ ion movement along the $[d(TG_4T)]_4$ channel using k = 25, 50, 100, and 125 kcal/mol/Å^2. The vertical lines correspond to the center of windows in the US simulations.

It looks like by using a k even as small as 25 kcal/mol/Å^2, the ion will still remain at RC = 0.00 Å with a larger width that is easier for increasing the overlap with the adjacent window. This suggests that one can safely use smaller k at lower energy barrier regions to increase the overlap while the ion is still located at the expected window center. To further test this idea, we examined RC = 3.00 Å that is still a good example of low energy barrier. FWHM is 0.1, 0.1, 0.3, and 0.3 Å and width is 0.5, 0.5, 0.6, and 0.7 Å for k = 125, 100, 50, and 25 kcal/mol/Å^2, respectively. The ion position shifts slightly away from the defined window center by decreasing k, but is in general covering the RC = 3.00 Å. For high energy barriers (e.g., RC = 1.50 Å), even by applying k = 125 and 100 kcal/mol/Å^2, the ion is not perfectly centered at RC = 1.50 Å. There is a tendency for the

ion to shift to the left side because the barrier is lower at that point compared to RC = 1.50 Å. This effect is more noticeable for k = 25 kcal/mol/Å^2 in which the ion does not experience RC = 1.50 Å anymore. The same problem happens for RC = 1.75 Å in which the ion shifts to the right side as we decrease k from 125 to 25 kcal/mol/Å^2. Although keeping the ion at defined window center is more difficult for higher energy barriers, the curves in general are wide in these regions, whereas for low energy barriers they are narrow. For RC = 1.50 Å, FWHM is 0.2, 0.2, 0.3, and 0.4 Å and the width is 0.6, 0.6, 0.8, and 1.1 Å for k = 125, 100, 50, and 25 kcal/mol/Å^2, respectively. This suggests that for higher energy barriers, the overlap between two adjacent windows occurs more easily than for the lower energy barriers. So, at high energy barriers, there are two issues to be considered: the ion position and the overlap. We have to make sure that k is strong enough to keep the ion at the defined window center. For k = 25 kcal/mol/Å^2, no overlap can be found because k is not strong. So, at RC = 1.50 and 1.75 Å, the ion shifts to the left and right, respectively and these two windows never get the chance to overlap. FWHM for RC = 1.75 Å is 0.15, 0.2, 0.4, and 0.6 Å and width is 0.55, 0.65, 0.8, and 1.5 Å for k = 125, 100, 50, and 25 kcal/mol/Å^2, respectively.

In general, in regions of low energy barriers, using small k can increase overlap between two adjacent windows while the ion will most likely remain within the defined window. Thus, one can decrease the number of windows because the overlap is enough to cover the entire range of reaction coordinate. For high energy barriers, k should be strong enough to keep the ion at the defined window center, but still guarantee a good overlap between two neighboring windows. This is broadly discussed as "guided umbrella sampling" or GUS method as one of the practical solutions to save computer time when running several simulations.[202] The authors studied a PMF profile and they compared the results of GUS with experimental data. The main problem with this

approach is that one should already have a PMF profile to test the results of GUS or at least a good idea about the barriers of the structure is required. This results in selecting a proper bias potential for each window and finally obtaining a reliable PMF profile. Also, to get a trustworthy PMF profile from GUS, one may have to try several sets of bias potentials for different windows to find the best combination that produces a reasonable PMF profile. This also is structure dependent. So, in general, if one performs this method to find the optimal bias potential for each window in a structure and then uses these data for several simulations on the same structure, he will save a considerable amount of time. Otherwise, he may spend a lot of time to find the optimal bias potential for running only a few simulations. Based on our studies, k = 100 and 125 kcal/mol/Å^2 both produce similar PMF profiles while there is sufficient overlap between adjacent windows and the ion remains within the defined windows. In our simulations, we use k = 100 kcal/mol/Å^2 unless mentioned.

5.3.2 US simulations of $[d(TG_4T)]_4$ containing a single ion

In this section, we applied the US method to obtain a complete PMF profile for single Na^+ and K^+ ion movement along the entire $[d(TG_4T)]_4$ channel. In each structure, we defined 81 windows each of 0.25 Å width and placed the ion in the center of each window using k = 100 kcal/mol/Å^2. Figure 5-3 shows the window overlap analysis for these simulations. It is clear that since good overlaps were observed between adjacent windows throughout the entire region, the resultant PMF profiles should be reliable. Figure 5-4 shows the anti-symmetrized PMF profiles obtained from the US and ABF methods for single Na^+ and K^+ ion movement along the $[d(TG_4T)]_4$ channel. The two methods produced the same shape and similar energy barriers for ion transport through the channel. The minor difference between the two PMF profiles for Na^+ is

within the accuracy of the results. Similar to the ABF results, the US results indicate that while a K^+ ion experiences a free energy barrier of approximately 13–15 kcal/mol for moving across a G-quartet, the corresponding barrier for Na^+ ion movement is significantly smaller, about 4–5 kcal/mol. As mentioned before, these results make perfect sense considering the difference between ionic radii of Na^+ and K^+ (Na^+: 0.95 Å; K^+: 1.33 Å). A G-quartet needs to open slightly when K^+ is passing through. In contrast, passage of a Na^+ ion causes much less structural changes.

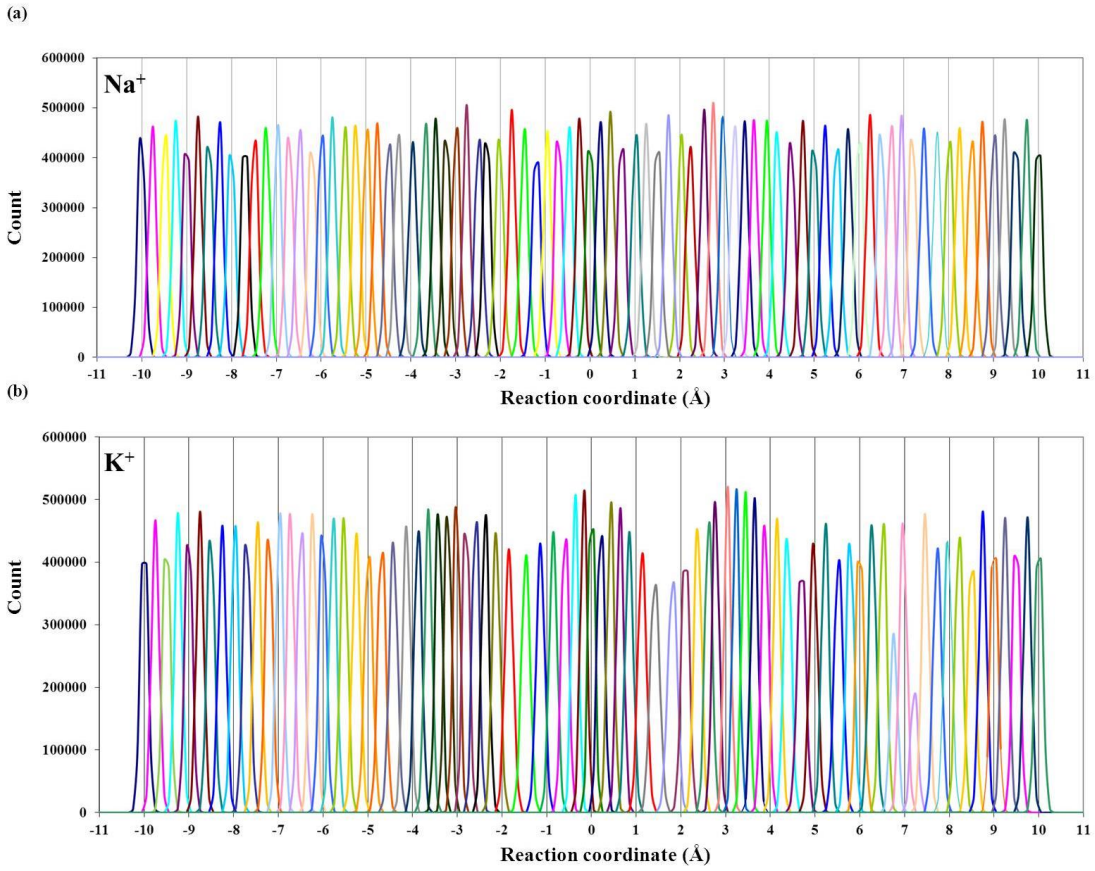

Figure 5-3. Window overlaps for (a) single Na^+ and (b) single K^+ ion movement along the $[d(TG_4T)]_4$ channel using k = 100 kcal/mol/Å^2.

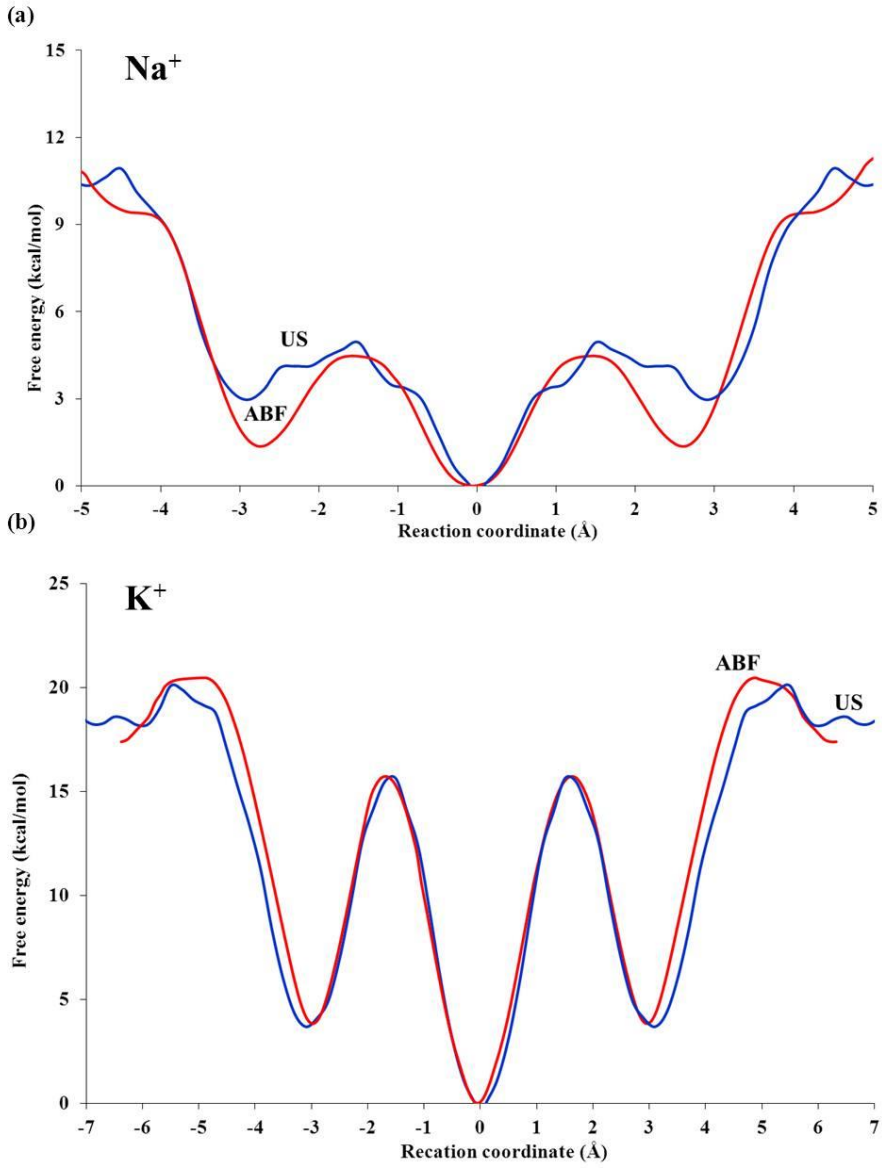

Figure 5-4. Comparison between US and ABF anti-symmetrized PMF profiles for (a) single Na^+ and (b) single K^+ ion movement along the $[d(TG_4T)]_4$ channel. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4).

5.3.3 US simulations of $[d(TG_4T)]_4$ containing multiple ions

The US method has been extensively used by Roux and co-workers[52] to study the multiple ion energetics of ion conduction in the K^+ ion channel. These authors reported a model in which the ion transmission is mainly controlled by diffusion process. That is, the strong ion-channel attraction is compensated with the electrostatic ion-ion repulsion. They showed that the K^+ ion conduction process in the K^+ ion channel involves a switch between two and three ions in the structure. After studying free energy maps, they proposed two main mechanisms that direct the ion passage through the channel. In the first mechanism, there are two K^+ ions in the channel once the third one gets close from one side and pushes the other two forward to make space for itself to enter the channel. The repulsion force between ions in the channel kicks the ion at the other end out and finally produces a channel containing two ions. As a result, K^+ ion only experiences 2-3 kcal/mol free energy barrier while passing through the channel. In the other mechanism, however, the two K^+ ions in the channel first move up and create a vacancy for the third one. Then, the third ion approaches the channel to fill this position and one ion leaves the channel from the other end. The free energy for ion transport along channel in this case is 3-4 kcal/mol which again suggests a low energy barrier. The key in both mechanisms proposed for K^+ ion conduction in this structure, is that the ion-channel attraction is significantly compensated by the strong ion-ion repulsion between K^+ ions. Obviously, this force is much stronger for shorter distances between ions.

Our single-ion studies in the previous section did not consider the interactions among different ions in a fully occupied channel. Although single-ion simulations provide useful qualitative information about ion transport energetics, they do not fully represent the complete picture of ion movement along a G-quadruplex channel. The main issue is how to treat the

multiple ion occupancy problem. As a first step, we use several simple models to examine multiple Na^+ and K^+ ion movement along the $[d(TG_4T)]_4$ channel using the US method. In all models discussed in the following section, a bias potential of k = 100 kcal/mol/Å^2 was applied to maintain ions at their positions within defined windows at 0.25 Å increments. Also, the distance between ions was kept constant where needed by fixing ions at their positions using the bias potentials. Table 5-1 shows the summary of the following models.

Table 5-1. Summary of the models to study ion movement in $[d(TG_4T)]_4$ channel containing multiple ions.

Model	Dynamics axis	C1' atoms fixed	Ion	Distance of adjacent ions (Å)
1	G3 and G4	All guanines	K^+	4
2	Varies	Central guanines	K^+	3.25
3	Varies	Central guanines	Na^+	3
4	G3 and G4	Central guanines	K^+	3.25
5	G3 and G4	Central guanines	K^+	4
6	G3 and G4	All guanines	K^+	3.25
7	G3 and G4	All guanines	K^+	3.25 (ion 2 removed)
8	G3 and G4	All guanines	K^+	Varies (ion 1 removed)

Model 1: multiple K^+ ion movement

In this model we studied the concerted ion transport of four K^+ ions along the $[d(TG_4T)]_4$ channel. The distance between K^+ ions was kept at 4 Å during the simulations. Overall, we performed 17 individual simulations in which the channel axis in each window was defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms of G3 and G4 as defined previously in Figure 3-1. The positions of C1' atoms of all guanine residues were fixed by a harmonic potential constant of k = 3 kcal/mol/Å^2. Figure 5-5 shows the initial and final states of the simulations in Model 1. Also, the individual free energy profiles for all four ions are shown in Figure 5-6. Apparently, four energy barriers can be distinguished due to the K^+ ion movement along the channel. Although the overall shape of the PMF profiles for the inner ions look similar to the single-ion model, the outer ions represent a quite different pattern. Particularly for ion 4, the ion has to overcome a continuously increasing barrier to be able to enter the channel. The results of this model will be fully analyzed and compared with the following models.

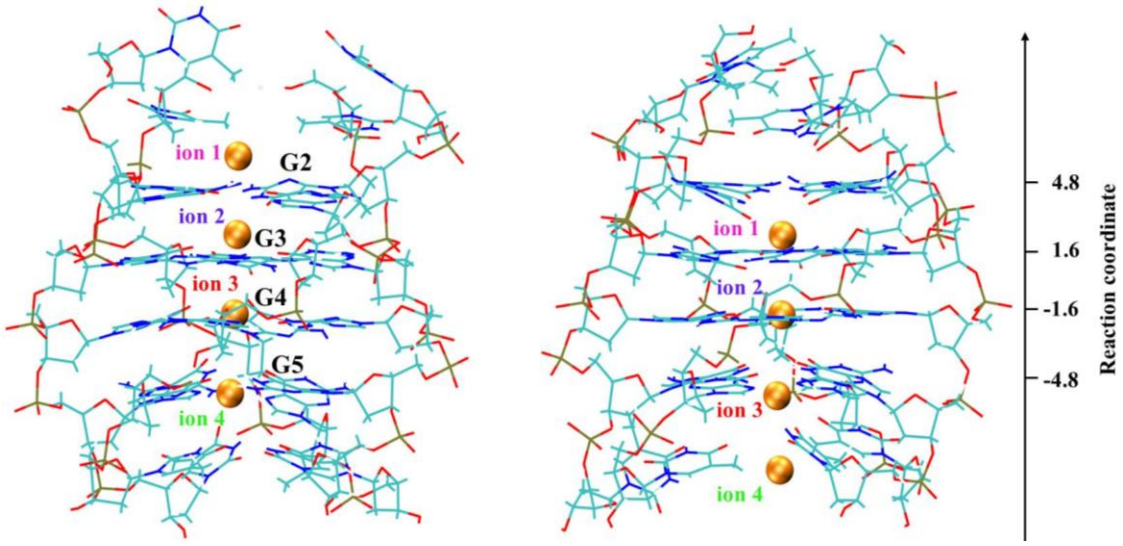

Figure 5-5. (Left) Initial and (right) final states of Model 1. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4).

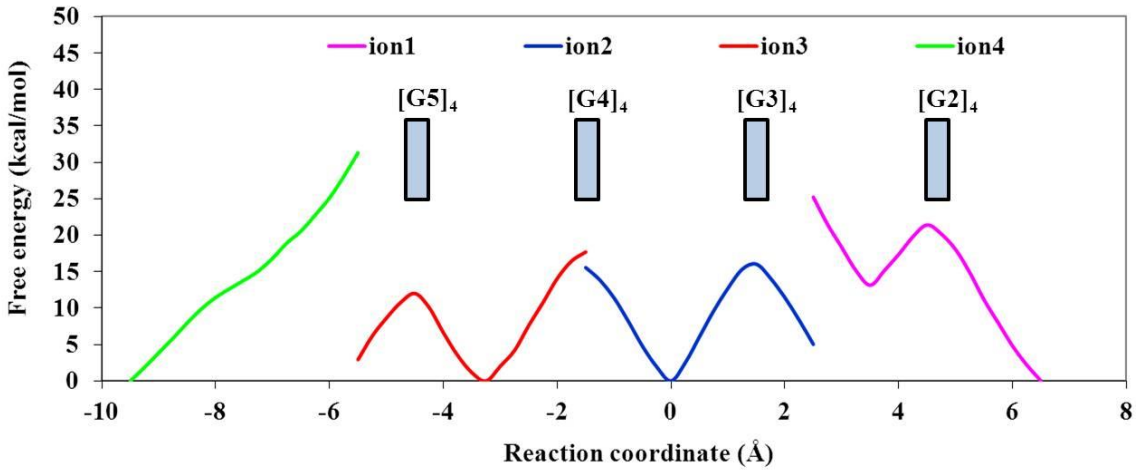

Figure 5-6. PMF profiles for multiple K^+ ion movement in $[d(TG_4T)]_4$ channel (Model 1). Individual PMF profiles correspond to K^+ ions with the same colors. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets ($[G3]_4$ and $[G4]_4$).

Model 2: multiple K^+ ion movement

In this model, we examined if the choice of dynamic axis used in Model 1 is reliable. We studied the concerted ion transport of four K^+ ions along the $[d(TG_4T)]_4$ channel. The distance between K^+ ions was kept at 3.25 Å during simulations. This distance is calculated from the minimized and equilibrated structure of $[d(TG_4T)]_4$ in the presence of K^+ ions which was previously presented in Figure 3-2. Overall, we performed 13 individual simulations in which the channel axis in each window was defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms from two adjacent G-quartets. For each simulation the closest two G-quartets to the ion were chosen to define the dynamic axis. That is, we introduced three different dynamic axes in this model as a combination of G2 and G3, G3 and G4, and finally G4 and G5 as defined previously in Figure 3-1. The positions of C1' atoms of central guanine residues were fixed by a harmonic potential constant of k = 3 kcal/mol/Å^2. Figure 5-7 shows the initial and final states of the simulations in Model 2. Also, the individual free energy profiles for all four ions are shown in Figure 5-8. It seems that the energy barriers for the movement of ions 1 and 4 have disappeared, but their overall continuously increasing barriers are produced again. This indicates that observing such profile for these ions is not related to the choice of dynamic axis. The PMF profiles for the inner ions look similar to the single-ion model and those of Model 1. The results of this model will be discussed and compared with the other models later.

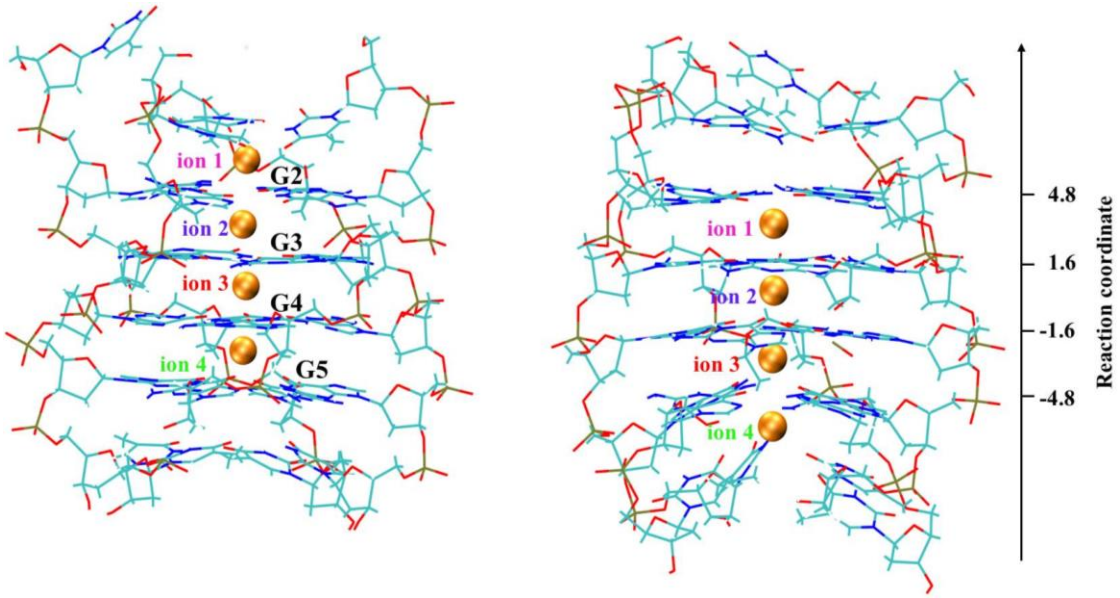

Figure 5-7. (Left) Initial and (right) final states of Model 2. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4).

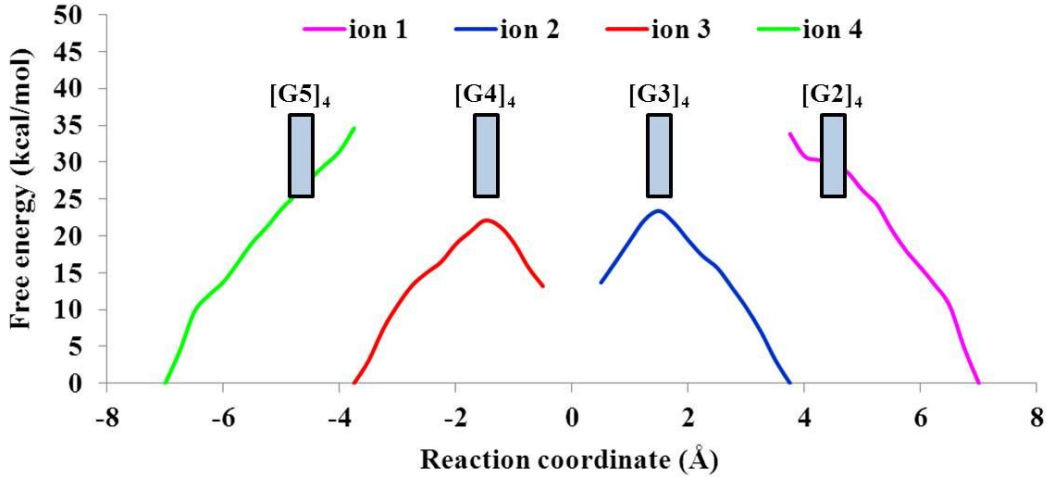

Figure 5-8. PMF profiles for multiple K^+ ion movement in $[d(TG_4T)]_4$ channel (Model 2). Individual PMF profiles correspond to K^+ ions with the same colors. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets ($[G3]_4$ and $[G4]_4$).

Model 3: multiple Na^+ ion movement

To further test our idea about the choice of dynamics axis, we examined ion transport of four Na^+ ions along the $[d(TG_4T)]_4$ channel similar to Model 2, however, the distance between adjacent Na^+ ions was kept at 3 Å during simulations. Figure 5-9 shows the initial and final states of the simulations in Model 3. Also, the individual free energy profiles for all four ions are shown in Figure 5-10. The overall shape of the PMF profiles does not look very different from Model 2. Even though the size of Na^+ is smaller than K^+, still a considerable energy barrier should be overcome by ions to move from outside into the channel. This is further evidence that the problem of observing different shapes in PMF profiles for ions 1 and 4 is not related to the choice of dynamic axis. The results of this model will be analyzed in the next models.

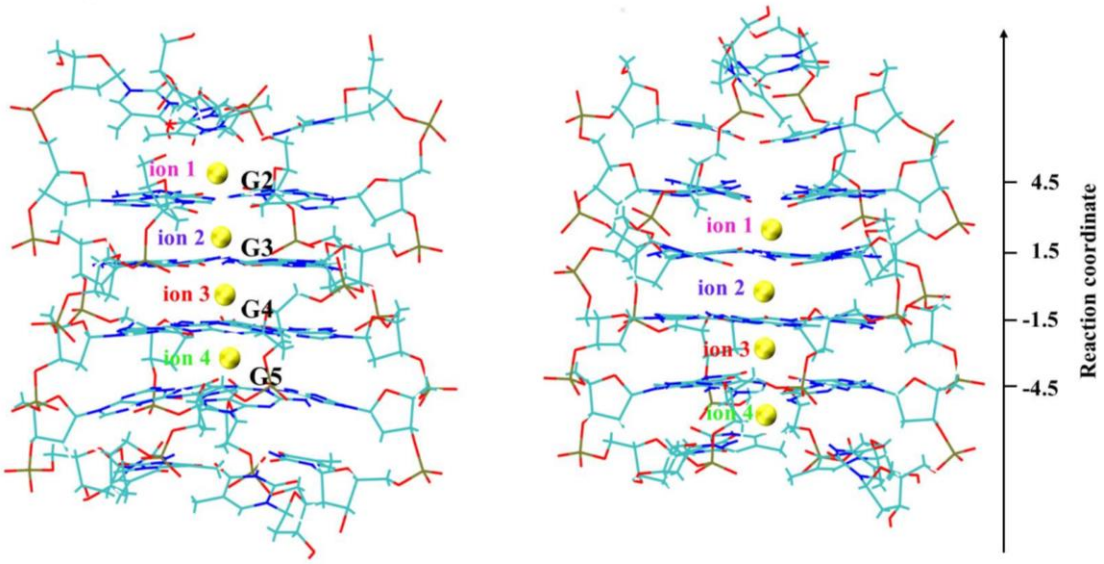

Figure 5-9. (Left) Initial and (right) final states of Model 3. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4).

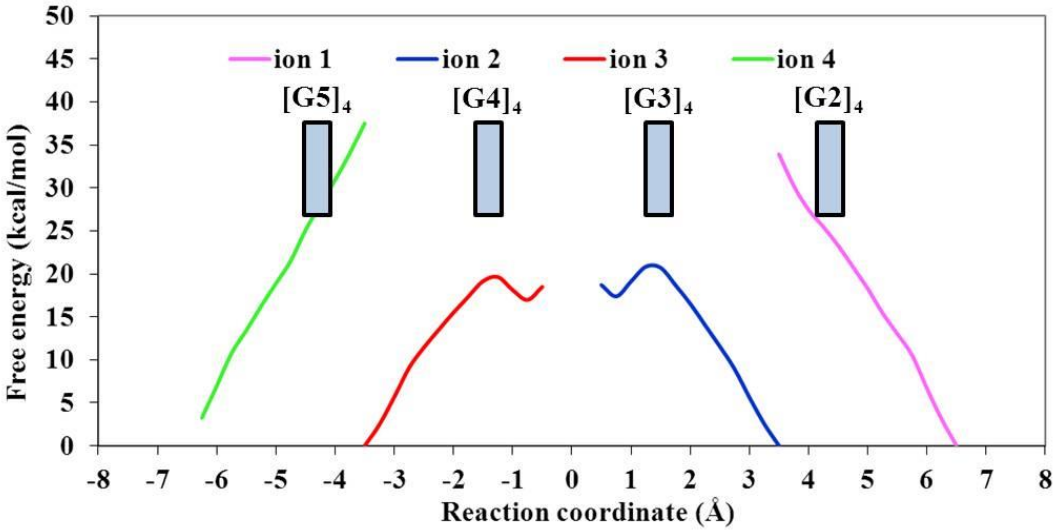

Figure 5-10. PMF profiles for multiple Na^+ ion movement in $[d(TG_4T)]_4$ channel (Model 3). Individual PMF profiles correspond to Na^+ ions with the same colors. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets ($[G3]_4$ and $[G4]_4$).

Model 4: multiple K^+ ion movement

Since using three different dynamic axes to define the ion pathway in Models 2 and 3 did not provide additional information, we decided to follow the same dynamics axis definition as in Model 1. In Model 4, we studied the concerted ion transport of four K^+ ions along the $[d(TG_4T)]_4$ channel. The distance between K^+ ions was kept at 3.25 Å during simulations. Overall, we performed 14 individual simulations in which the positions of C1' atoms of central guanine residues were fixed by a harmonic potential constant of k = 3 kcal/mol/Å^2. Figure 5-11 shows the initial and final states of the simulations in Model 4. Also, the individual free energy profiles for

all four ions are shown in Figure 5-12. This model did not also change the form of the PMF profiles for the movement of outer ions noticeably. Following, all models we introduced so far will be analyzed and the results will be compared.

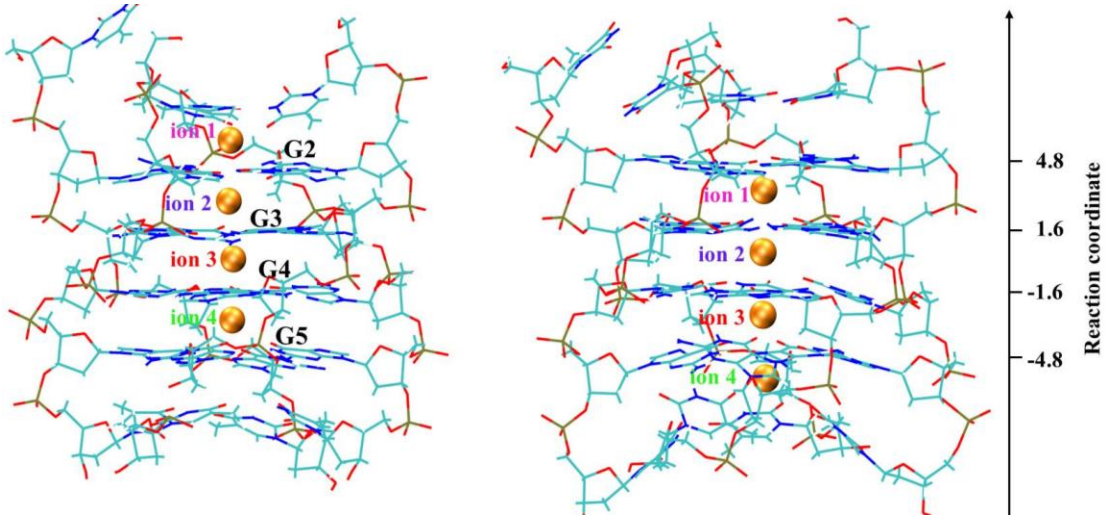

Figure 5-11. (Left) Initial and (right) final states of Model 4. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4).

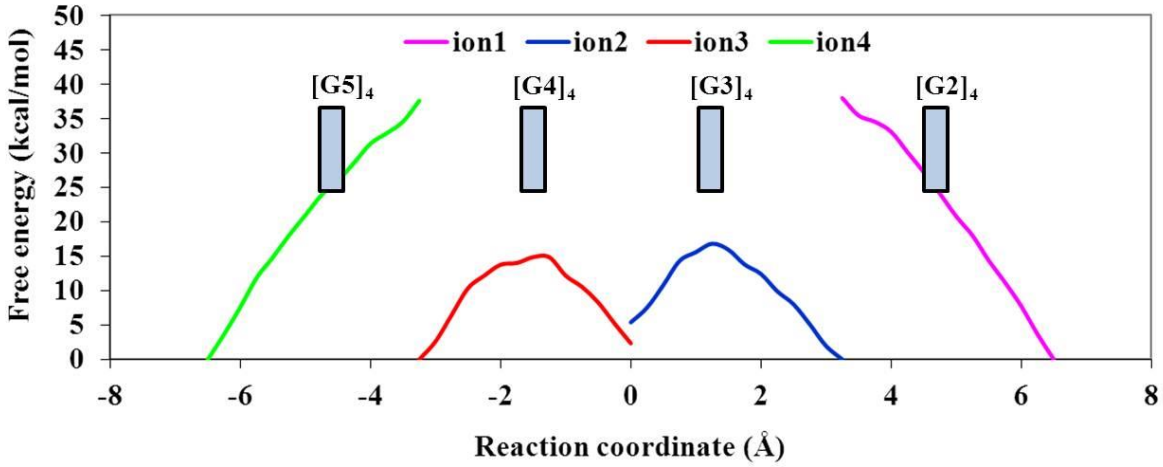

Figure 5-12. PMF profiles for multiple K^+ ion movement in $[d(TG_4T)]_4$ channel (Model 4). Individual PMF profiles correspond to K^+ ions with the same colors. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets ($[G3]_4$ and $[G4]_4$).

Models 5 and 6: multiple K^+ ion movement

In the previous models, we performed simulations under different initial conditions including the size of ions, the choice of dynamic axis, and the distance between ions. Although the PMF profiles are slightly different, they all show similar free energy barriers experienced by the outer ions. In order to get deeper insight into this process we visited Models 1 and 4 again. In both models, the same dynamic axis was defined. In Model 1, the distance between K^+ ions was 4 Å during the simulations, whereas in Model 4 ions were kept 3.25 Å apart. In addition, in Model 1, the positions of C1' atoms of all guanine residues were fixed, while in Model 4 only the positions of C1' atoms of the central guanine residues were fixed. We observed two bumps in

Model 1 for ions 1 and 4 (see Figure 5-6), but nothing significant in Model 4. Since in Model 4 the distance between ions is smaller than in Model 1, there is a possibility that the bumps have disappeared because the outer ions are kicked out by their neighbor ions. On the other hand, the bumps in Model 1 could be just due to fixing C1' atoms for outer G-quartets. To understand this, we ran two separate sets of simulations. In one case (Model 5) we designed the configuration as in Model 1, but fixed the positions of C1' atoms only for the central G-quartets by the same harmonic potential constant. In Model 6, we used the same definition as in Model 4, however, we fixed the positions of C1' atoms for all G-quartets by the same harmonic potential constant. Overall, we performed 17 and 14 individual simulations for Models 5 and 6, respectively. Figure 5-13 shows the initial and final states of the simulations in Model 5. The individual free energy profiles for all four ions in Model 5 are shown in Figure 5-14. Also, Figure 5-15 shows the initial and final states of the simulations in Model 6. The individual free energy profiles for all four ions in this model are shown in Figure 5-16. A comparison between Models 1 and 5 suggests that even removing constraint from C1' atoms does not change the overall shape and the magnitude of the free energy experienced by the outer ions. The bumps on outer G-quartets can still be observed clearly. A comparison between Models 4 and 6 results in the same conclusion. Adding constraint on outer G-quartets in Model 6 did not create a significant barrier. In conclusion, although fixing C1' atoms may introduce a small barrier, the magnitude of this barrier is almost negligible compared to the other effects. So, there should be other features responsible for creating such different PMF profiles. Now that we are certain the issue is not associated with the choice of dynamic axes and fixing or unfixing C1' atoms of guanine residues, we decided to remove one of the K^+ ions in the vicinity of one outer ion to study this phenomenon from other points of view.

Apparently, the first impact would be eliminating the repulsion force on this outer ion as there is no K^+ ion close to it. The simulation setup and details are presented in Model 7.

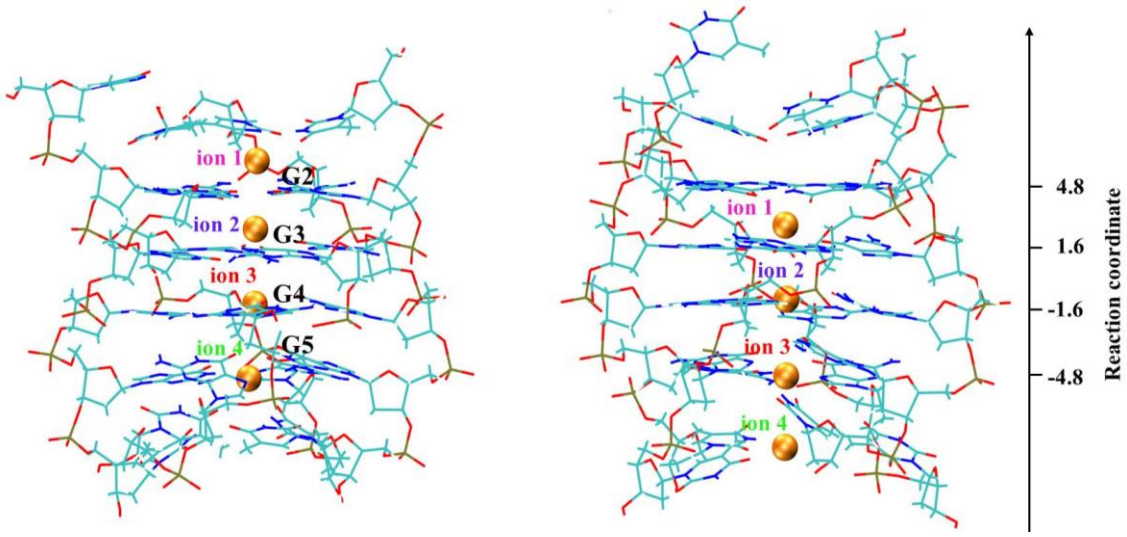

Figure 5-13. (Left) Initial and (right) final states of Model 5. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4).

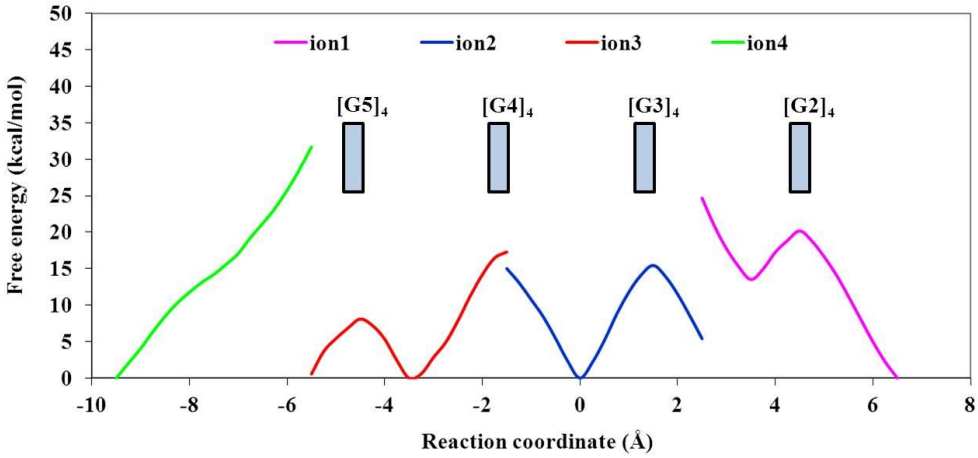

Figure 5-14. PMF profiles for multiple K^+ ion movement in $[d(TG_4T)]_4$ channel (Model 5). Individual PMF profiles correspond to K^+ ions with the same colors. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets ($[G3]_4$ and $[G4]_4$).

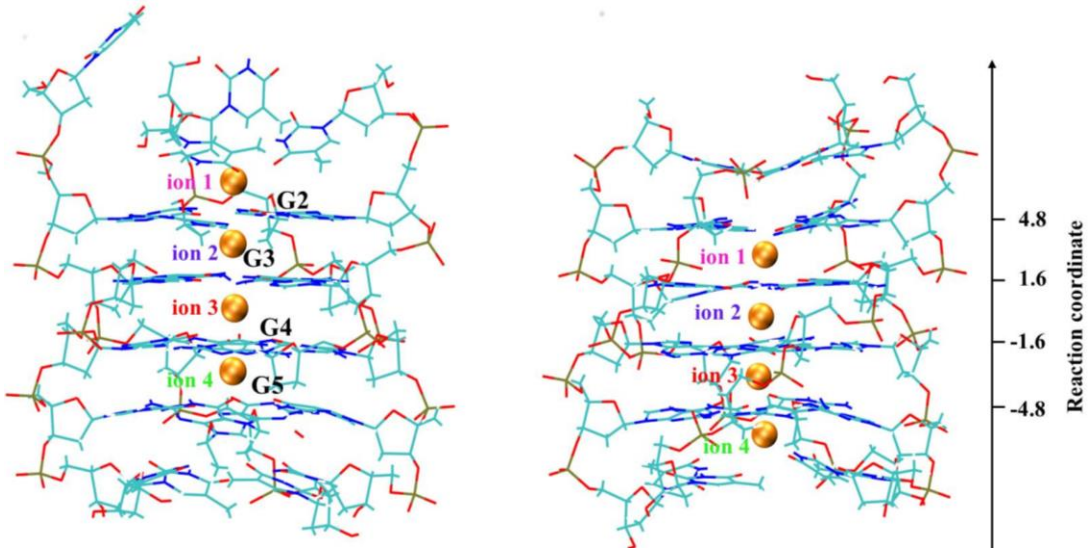

Figure 5-15. (Left) Initial and (right) final states of Model 6. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4).

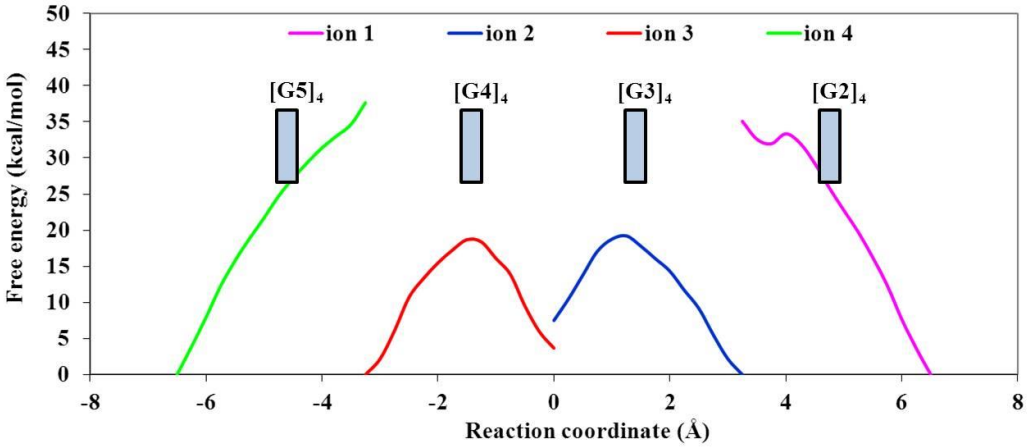

Figure 5-16. PMF profiles for multiple K^+ ion movement in $[d(TG_4T)]_4$ channel (Model 6). Individual PMF profiles correspond to K^+ ions with the same colors. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets ($[G3]_4$ and $[G4]_4$).

Model 7: multiple K^+ ion movement

In this model we studied the concerted ion transport of three K^+ ions along the $[d(TG_4T)]_4$ channel. We removed ion 2 compared to previous models to investigate how eliminating an ion will affect the repulsion force between ions and how it will change the free energy profiles. The distance between ions 1 and 3 and ions 3 and 4 was kept at 6.5 and 3.25 Å, respectively. Overall, we performed 14 individual simulations in which the channel axis links the two centers of mass of the four O6 atoms from the two central G-quartets (G3 and G4) as defined previously in Figure 3-1. The positions of C1' atoms of all guanine residues were fixed by a harmonic potential constant of k = 3 kcal/mol/Å^2. Figure 5-17 shows the initial and final states of the simulations in Model 7. Also, the individual free energy profiles for all four ions are shown in Figure 5-18.

Apparently, removing one K^+ ion from the channel has a significant influence on the pattern of free energy profiles particularly for ions near the empty space. Ions 1 and 3 show an entirely different picture in terms of the shape and the amount of free energies. Just a quick look suggests that the profile for ion 1 is similar to the inner ions in the previous models. Also, it seems that the same behavior as outer ion 1 is produced for ion 3 in Model 7. As only ions 1 and 3 are affected in this model, to understand the consequence of removing ion 2 and also to get more sense of our previous models, we compared the last two models (Models 6 and 7). These models have the same configuration except for ion 2 that is absent in Model 7.

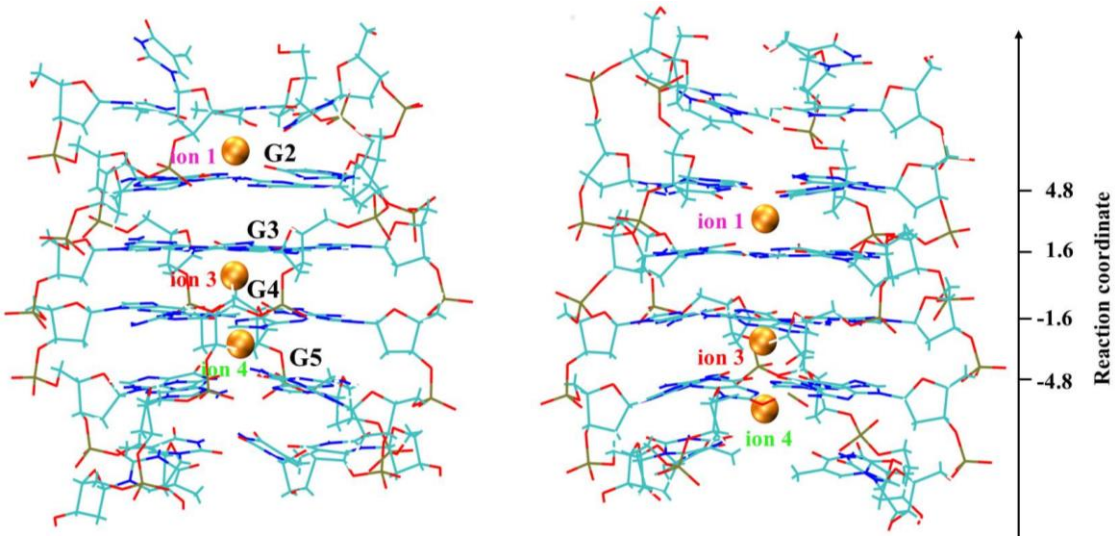

Figure 5-17. (Left) Initial and (right) final states of Model 7. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4).

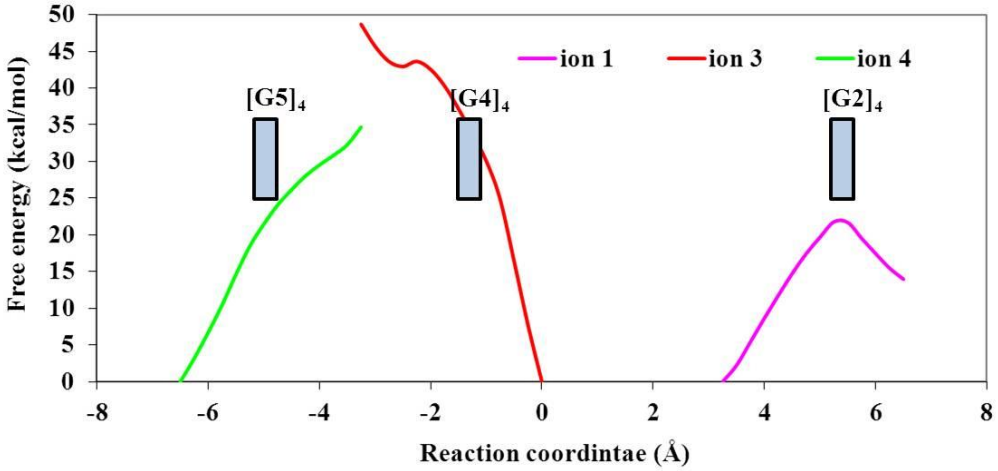

Figure 5-18. PMF profiles for multiple K^+ ion movement in $[d(TG_4T)]_4$ channel (Model 7). Individual PMF profiles correspond to K^+ ions with the same colors. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets ($[G3]_4$ and $[G4]_4$).

Figures 5-19 and 5-20 show the schematic display and the comparison between free energy profiles for the two models. The blue and the red colors represent Models 6 and 7, respectively. As Figure 5-20 indicates ion 4 shows a similar pattern in both models. This is not surprising as ion 4 is located in the vicinity of only one K^+ ion at its right hand side. So, the same effect should be experienced in both models. This could also be considered as an evidence for reproducibility of the results produced by these simulations. As mentioned, the most important indication from these simulations is the clear similarity between ion 1 in Model 6 and ion 3 in Model 7. The same correspondence exists between ion 3 in Model 6 and ion 1 in Model 7. The first step to appreciate these results is to understand the difference between the models as shown in Figure 5-20. Ion 1 in Model 6 has only one neighbor K^+ at its left hand side which is the same environment as

experienced by ion 3 in Model 7. Both of these ions encounter a huge energy barrier once they move from their initial toward their final state. Of course, they each have to pass through one G-quartet to reach the final state. The question, however, is why the energy barrier does not decrease once the ion has passed the G-quartet? Apparently, there is another factor involved in this process. Ion 3 in Model 6 and ion 1 in Model 7 have an entirely different environment around them. Ion 3 in Model 6 is surrounded by two K^+ ions from both sides, however, there is no ion nearby ion 1 in Model 7. What causes similar PMF profiles for these two ions then? Even the amount of energy barriers are almost the same. Moreover, their profiles are comparable to the single K^+ ion model. In fact, as there is no K^+ ion adjacent to ion 1 in Model 7, it could act the same as a single-ion model. On the other hand, as ion 3 in Model 6 is surrounded by the same type of ions with the same distances at both sides, the forces between these two ions and ion 3 will be canceled out. Therefore, although ion 3 in Model 6 seems to be situated among two other K^+ ions, it will not feel their influence and will show similar pattern as a single-ion model. So, basically these models suggest that K^+ ions do sense the repulsion from their neighbors as long as there is no other ion located at the same distance on the opposite side. This is the reason that ion 1 in Model 6 and ion 3 in Model 7 encounter a large resistance from ions 2 and 4, respectively, while crossing G-quartets. Because the distance between ions 1 and 2 in Model 6 and also between ions 3 and 4 in Model 7 remains constant during simulations, one expects a constant repulsion force between two ions. The green liens in Figure 5-20 show the difference between free energies in two models. The repulsion force is simply the derivative of the free energy difference between the two models with respect to the distance moved by each ion. This amount is negligible for ion 4 as the slope of the green line is only 0.9 kcal/mol/Å. However, for ions 1 and 3, the repulsion force calculated from the green lines is 14-16 kcal/mol/Å. Now that we are

certain about the effect of the repulsion force, we can understand the difference between PMF profiles for outer ions. To further test our idea, we ran another set of simulations in Model 8. In this model, the effect of repulsion force between K^+ ions will be examined using a different approach to move ions.

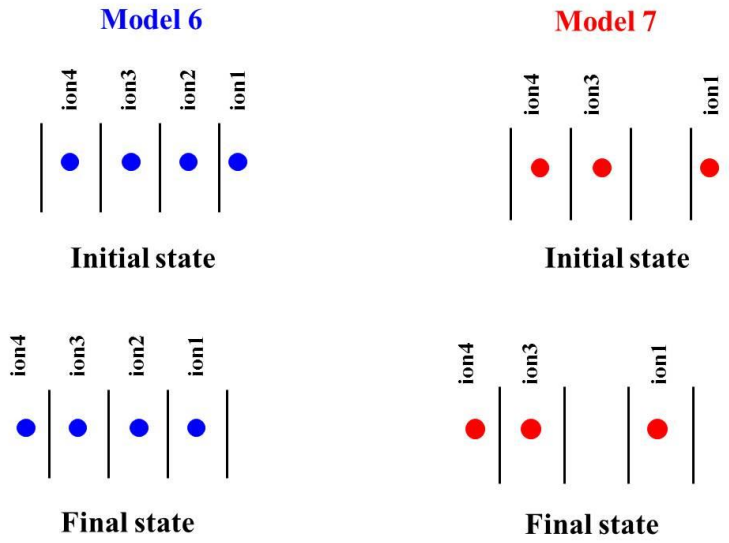

Figure 5-19. The schematic display for (left) multiple K^+ ions in $[d(TG_4T)]_4$ channel in Model 6 and (right) Model 7. The models represent the initial and final states of all ions during the simulations.

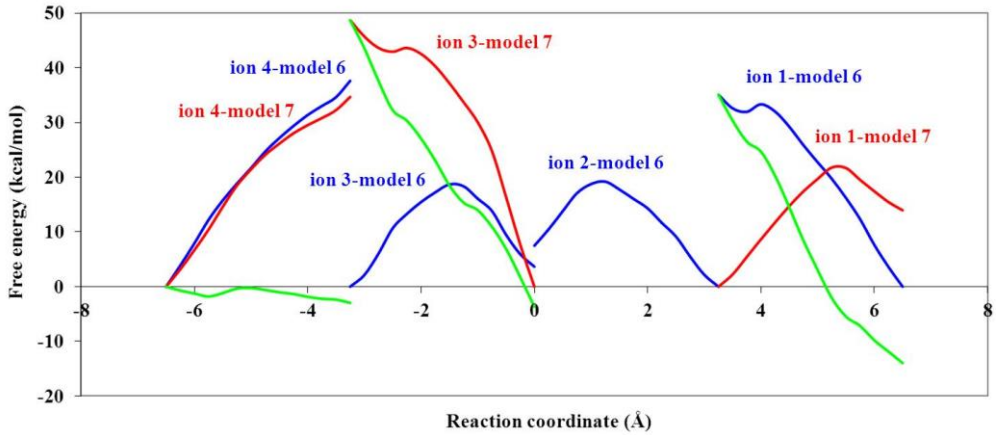

Figure 5-20. Comparison between free energy profiles for multiple K^+ ions in $[d(TG_4T)]_4$ channel in Model 6 (blue) and Model 7 (red) and the difference between the two models (green). The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets ($[G3]_4$ and $[G4]_4$).

Model 8: multiple K^+ ion movement

In Model 8, we further investigated the effect of repulsion forces on free energy profiles. In this model, we studied the concerted ion transport of two K^+ ions along the $[d(TG_4T)]_4$ channel in a structure containing three channel ions (ion 1 in previous models is removed here). We fixed the position of ion 3 in the middle of the central G-quartets (G3 and G4) by a bias potential of k = 100 kcal/mol/Å^2. We initially put ions 2 and 4 at a distance of 2.5 Å with respect to ion 3. We ran individual simulations each time moving ions 2 and 4, 0.25 Å away from their positions in the last simulation (toward outside) up to 6.5 Å distance eventually. Overall, we performed 17 individual simulations in which the channel axis links the two centers of mass of the four O6 atoms from the two central G-quartets (G3 and G4) as defined previously in Figure 3-1. The

positions of C1' atoms of all guanine residues were fixed by a harmonic potential constant of k = 3 kcal/mol/Å^2. Figure 5-21 shows the initial and final states of the simulations in Model 8. Also, the individual free energy profiles for all four ions are shown in Figure 5-22.

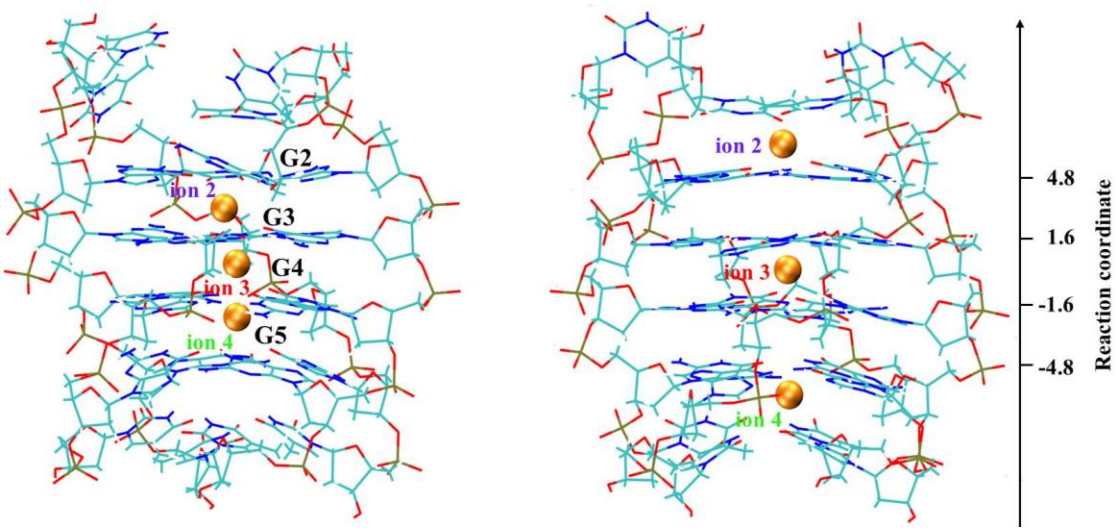

Figure 5-21. (Left) Initial and (right) final states of Model 8. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets (G3 and G4).

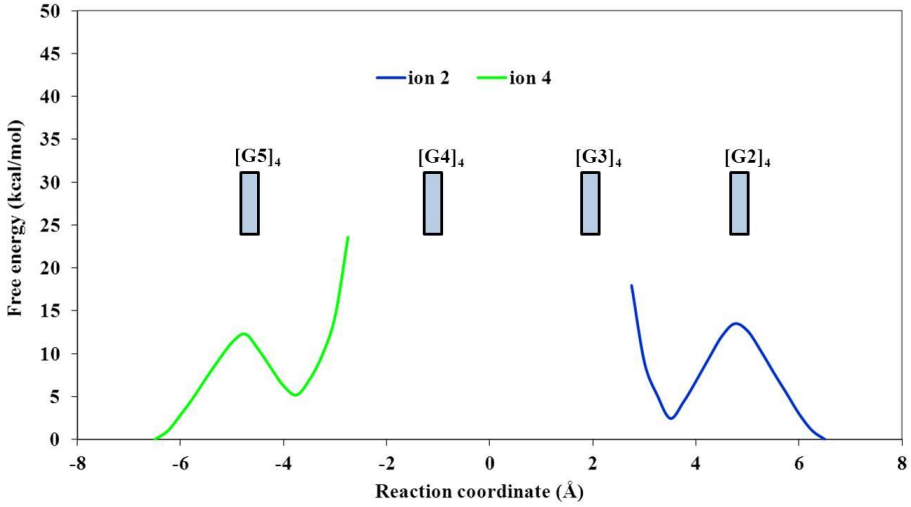

Figure 5-22. PMF profiles for multiple K^+ ion movement in $[d(TG_4T)]_4$ channel (Model 8). Individual PMF profiles correspond to K^+ ions with the same colors. The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets ($[G3]_4$ and $[G4]_4$).

In Model 8, ion 3 is located at the origin of the reaction coordinate. It is fixed during the simulations at its initial position and the other two ions are moving toward outside. The closest distance between either ion 2 or ion 4 with respect to ion 3 is 2.5 Å. In this situation, there is a significant repulsion between ions 2 and 4 with ion 3 as they are very close. This corresponds to a considerable energy barrier as can be observed in Figure 5-22. However, the amount of free energy continuously decreases at both sides as ions 2 and 4 get far from ion 3. This means that if any of ions 2 or 4 wants to move back toward ion 3 from the minima of either RC = 3.8 or -3.8 Å, it has to overcome the repulsion imposed by ion 3. As the effect of repulsion force highly depends on the distance between two ions, at larger distances it is not effective anymore. This happens at 3.8 Å distance between K^+ ions in our structure. Apparently, at distances larger than 3.8 Å the

interactions between K^+ ion and G-quartets are stronger and determine the shape of the rest of the energy profiles. That is, a similar profile to single-ion model can be distinguished when ions 2 and 4 are passing through the outer G-quartets. Even the amount of free energy barrier for ion transport is close to single-ion model. A perfect symmetry in free energy profiles for ions 2 and 4 is a good indication that the results are trustworthy because the two ends of $[d(TG_4T)]_4$ are not very different.

Since we already obtained knowledge about the repulsion force between ions, it is worth discussing this subject in this model too. Figures 5-23 and 5-24 show the schematic display of Model 8 and the comparison between free energy profiles for this model and a single K^+ ion. The blue and the red colors represent the single K^+ ion and the multiple-ions in Model 8, respectively. In Figure 5-24 the free energy barriers on outer G-quartets are equivalent to the single-ion model as ions 2 and 4 are far enough from ion 3 to be affected by it. So, the key parameter to determine the shape and the magnitude of the free energy on outer G-quartets is the interaction of K^+ ions with G-quartets. The green lines on Figure 5-24 represent the difference between single-ion model and Model 8. It is reasonable to divide these lines into two regions. One is the part close to ion 3 (from 2.5 to almost 3.8 Å distance with respect to ion 3 at both sides) in which the repulsion force experienced by ions 2 and 4 is strong. So, the slope of free energy difference (green line) with respect to the distance moved by ions 2 and 4 is larger and it becomes smaller when moving toward outside. In the second region, ions 2 and 4 are far enough from ion 3 and the slope is not significant. Therefore, the amount of repulsion force is much smaller once the ion moves close to both ends.

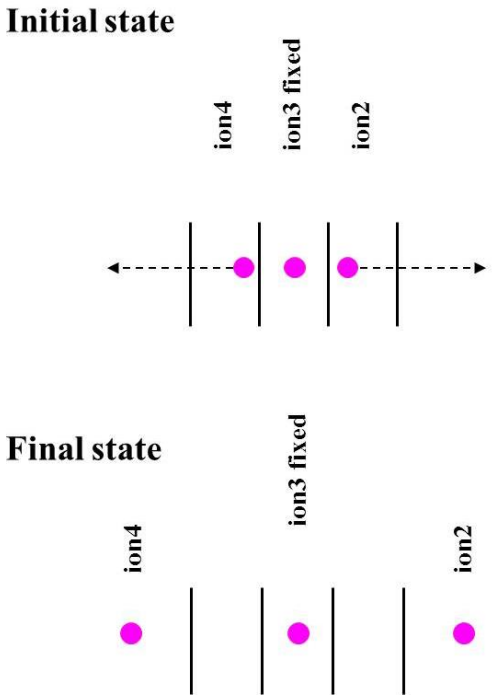

Figure 5-23. The schematic display for multiple K^+ ions in $[d(TG_4T)]_4$ channel for Model 8. The model represents the initial and final states of all ions during the simulations.

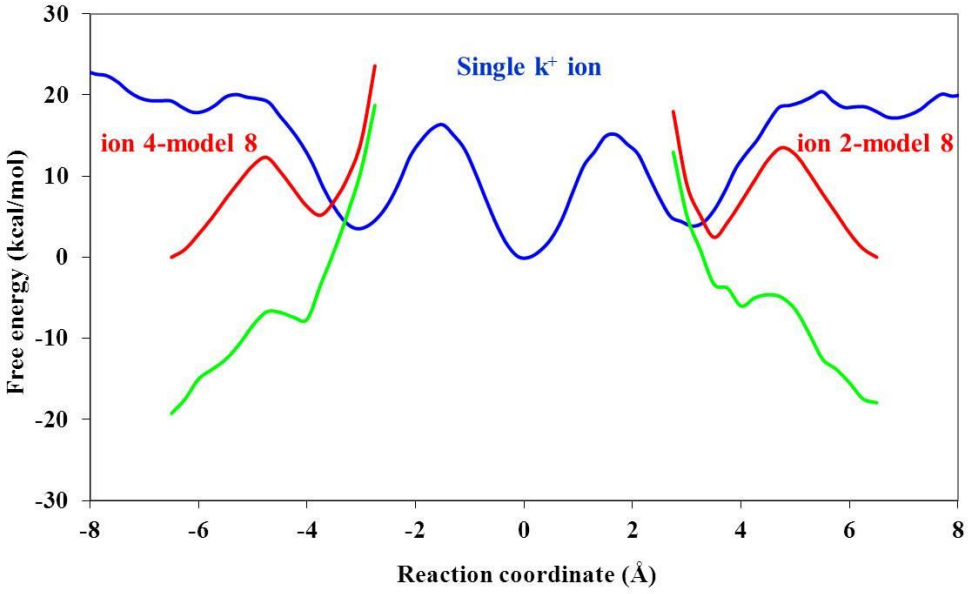

Figure 5-24. Comparison between free energy profiles for single (blue) and multiple K^+ ion Model 8 (red) and the difference between the two models (green). The origin of the reaction coordinate corresponds to the center of mass of eight O6 atoms from the two central G-quartets ($[G3]_4$ and $[G4]_4$).

5.4 Conclusion

In this chapter, we used the US method to investigate ion dynamics and energetics in several single- and multiple-ion models of G-quadruplex structure. This is the first time that the US method is applied to this important class of biological systems. The US results were in good agreement with those obtained by the ABF method. The US method was also extensively used to examine ion dynamics in multiple Na^+ and K^+ ion channels. A number of different parameters including various dynamic axes to define the reaction pathway, fixing and unfixing C1' atoms of

guanine residues, different types and sizes of ions were studied in addition to a series of different K^+-K^+ distances to move ions along the $[d(TG_4T)]_4$ channel. The appropriate bias potential to maintain ions at their expected positions during simulations was determined after studying several bias potentials. The optimal potential is the one that guarantees enough sampling for the ion at the center of the defined windows. Furthermore, there should be a reasonable overlap between adjacent windows to trust the PMF profiles. For low energy barriers, using smaller bias potential can increase the overlap between two adjacent windows. So, we can decrease the number of windows because the overlap will be enough to take care of the number of windows. For high energy barriers, bias potential should be strong enough to keep the ion at the defined window center (to make sure there is enough sampling from this region), but still guarantee a good overlap between windows. In all multiple-ion models, the outer ions mostly exhibited a different profile compared to the previous single-ion studies. The outer ion profiles were also different from the inner ions in most of the cases. We believe that the reason for observing a continuously increasing barrier experienced by the ion under investigation is the repulsion force between ions.

Chapter 6

Umbrella sampling MD studies of ion movement in $[d(G_3T_4G_4)]_2$

6.1 Introduction

In Chapter 4, we have used the adaptive biasing force (ABF) method to obtain free energy profiles for Na^+, K^+, and NH_4^+ ion movement along the $[d(G_3T_4G_4)]_2$ channel. The main goal was to investigate how asymmetrical features of the structure affect the potential of mean force (PMF) profiles for ion transport. Several models were introduced under different conditions to examine the ion dynamics in the $[d(G_3T_4G_4)]_2$ channel containing multiple NH_4^+ ions. In addition, in Chapter 5 we utilized the umbrella sampling (US) method to obtain free energy profiles for single and multiple ion movement along the $[d(TG_4T)]_4$ channel. As a bias potential is applied in this technique to keep the ion within defined windows, we discussed how to choose a proper bias potential to ensure the PMF profiles are reliable. k = 100 kcal/mol/Å^2 was chosen as a reasonable bias potential that guarantees a satisfactory sampling within each window while an acceptable overlap between two adjacent windows maintains. Using US as a powerful technique, we obtained useful information for ion movement (e.g., repulsion force between ions) in multiple-ion channels.

The main objective of this chapter is to employ US to examine ion energetics and dynamics along the $[d(G_3T_4G_4)]_2$ channel. As there are only two ion-binding sites inside this channel, it will be interesting to see how changing the simulation arrangement for one ion will affect the energy profile for the other one. We will also use our knowledge from single- and multiple-ion studies in Chapter 4 to further understand the role of ion-ion repulsion.

6.2 Computational details

6.2.1 Model structure generation

Solution NMR structure of $[d(G_3T_4G_4)]_2$ (PDB code 1U64)[165] was used as the initial structure. The same MD minimization and equilibration approach as described in Chapter 4/Section 4.2.1 was used to prepare the minimized structures. All MD simulations were performed in the isothermic-isobaric ensemble. Langevin dynamics and a Langevin piston algorithm were used to keep the temperature and pressure at 298 K and 1 atm, respectively. The particle mesh Ewald (PME) method of calculating long-range electrostatic interactions was used with a cut-off of 9 Å. Rattle was applied to constrain the bonds containing hydrogen. A time step of 2 fs was used to integrate the equations of motion. All MD simulations were performed using the program NAMD2.7b3 and the CHARMM27 force field. The program VMD 1.8.7 was used for visualization and data analysis.

6.2.2 US simulations

Following, we will use the US method to study the ion dynamics along the $[d(G_3T_4G_4)]_2$ channel. All US simulations were performed using the program NAMD2.7b3 and the CHARMM27 force field at the SHARCNET computing facilities. The final PMF profiles were obtained after unbiasing the ion distribution using the weighted histogram analysis method (WHAM) program.

6.3 Results and discussion

6.3.1 US simulations of $[d(G_3T_4G_4)]_2$ containing a single NH_4^+ ion

To perform US simulations for the single-ion model, we initially put the upper or lower NH_4^+ ion at the center of mass of O6 atoms of upper/central or lower/central G-quartets, respectively (as defined previously in Figure 4-1). The reaction coordinate was defined along the channel axis of $[d(G_3T_4G_4)]_2$. The channel axis for the upper and lower NH_4^+ ion movement was defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms from each of the two lower/central and upper/central G-quartets, respectively. To obtain PMF profiles the upper and lower NH_4^+ ions were moved from RC = 0.8 to 4.8 Å and RC = -0.8 to -4.8 Å, respectively, in separate sets of simulations (see Figure 6-1). To do so, we performed US simulations in 17 separate windows (each being 0.25 Å in width) for each ion in which a bias potential of k = 100 kcal/mol/Å^2 was used to hold the ion within each window. The ion movement in the plane norm to the dynamic axis was confined by a harmonic potential (k = 100 kcal/mol/Å^2). Within each window the force acting on the ion was averaged in 0.1 Å bins. Typical production runs of 2 ns for each window were performed. The positions of C1' atoms of all guanine residues were fixed by a weak harmonic potential constant of k = 3 kcal/mol/Å^2. Figure 6-2 shows the PMF profiles for the upper and lower NH_4^+ ion movement after 0.3 and 2 ns. As can be observed, after 0.3 ns, the upper and lower NH_4^+ ions require 16.6 and 20.5 kcal/mol to overcome the barriers from the G-quartets, respectively. However, the energy barriers change to 17.4 and 17.3 kcal/mol, respectively after 2 ns. It seems that 2 ns simulation removes the asymmetrical profile expected between two ends. Apparently, the upper NH_4^+ ion transport

along the $[d(G_3T_4G_4)]_2$ channel is less time dependent. This feature will be fully investigated in the following multiple-ion models.

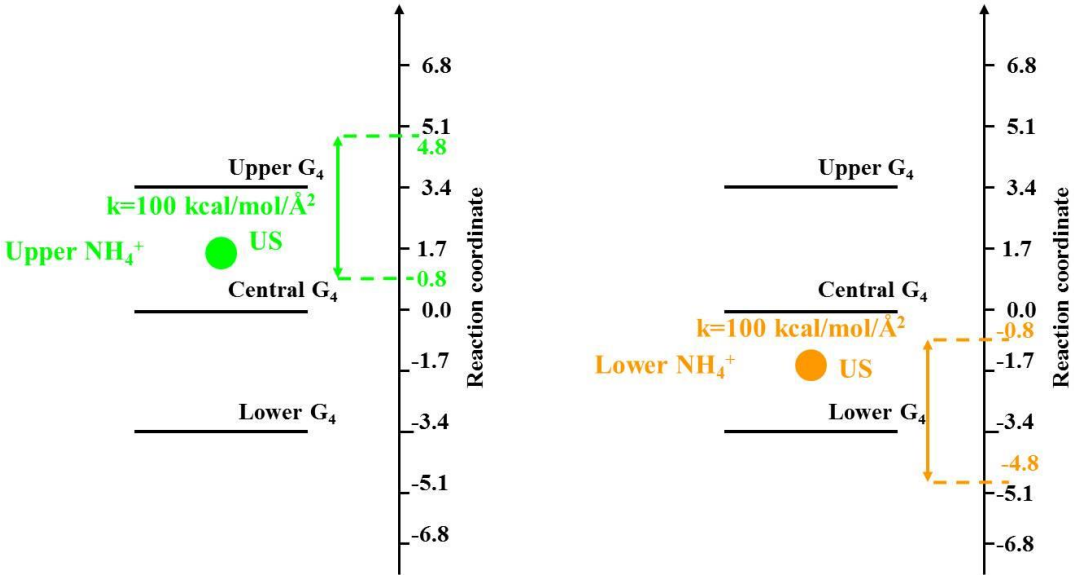

Figure 6-1. The schematic display of (left) the upper and (right) the lower NH_4^+ ion movement in a single-ion $[d(G_3T_4G_4)]_2$ channel. The green and the orange colors represent the upper and lower ions, respectively. A double arrow indicates the range of movement for an ion. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

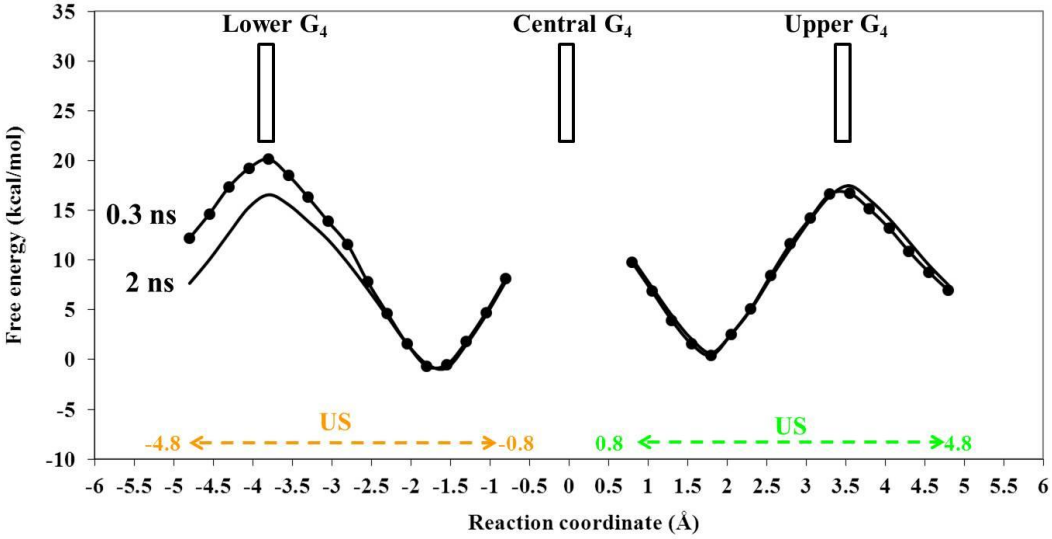

Figure 6-2. PMF profiles for the upper and lower NH_4^+ ion movement in a single-ion $[d(G_3T_4G_4)]_2$ channel after 0.3 (circle) and 2 ns (solid line) simulations. The green and the orange double arrows indicate the range of movement for the upper and lower NH_4^+ ions, respectively. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

Figure 6-3 shows the comparison between the PMF profiles for the single NH_4^+ ion movement along the $[d(G_3T_4G_4)]_2$ channel using the US and ABF methods. To obtain a complete PMF profile along the channel, we performed a new set of US simulations to cover the central region. The ABF simulation conditions are the same as US except that the positions of C1' atoms of guanine residues are fully unconstrained in ABF. In Chapter 4, energy barriers of 14, 17, and 17 kcal/mol were calculated for the NH_4^+ ion transport through the upper, central, and lower G-quartets, respectively. The calculated values from US simulations are 16.6, 19.5, and 20.5 kcal/mol for corresponding ion movements. The 2-3 kcal/mol higher energy barriers in this method are due to the fact that the positions of C1' atoms of guanine residues are fully confined. The new set of US simulations within the central region provides similar findings as in the ABF

method. The NH_4^+ ion experiences almost the same magnitude of free energy barrier as the lower end while crossing the central G-quartet.

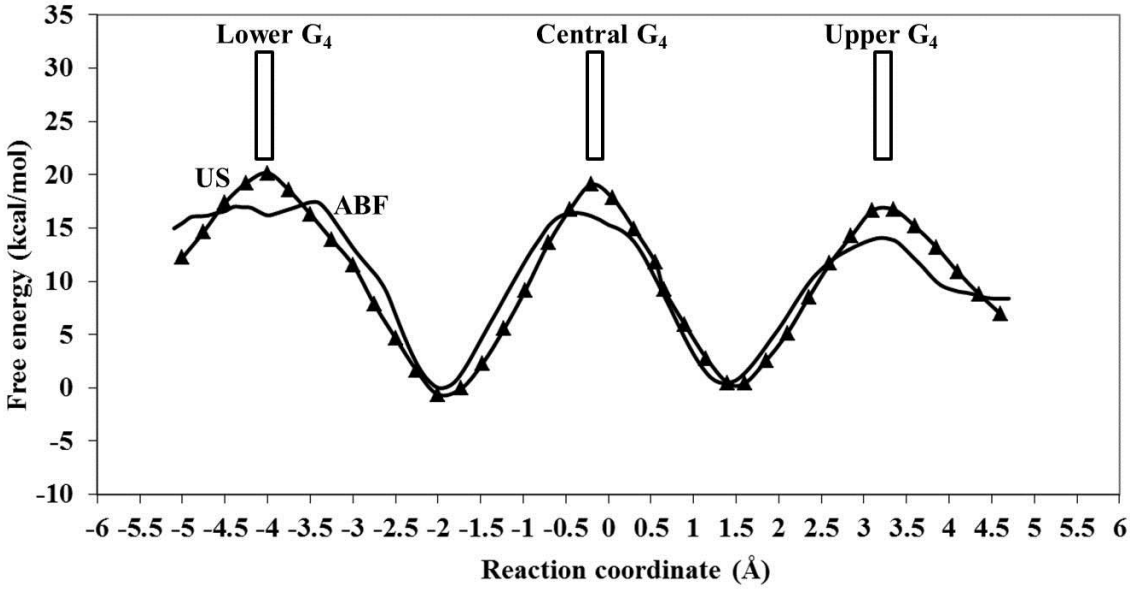

Figure 6-3. Comparison between PMF profiles using US (triangle) and ABF (solid line) methods for a single NH_4^+ ion movement along the $[d(G_3T_4G_4)]_2$ channel. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

6.3.2 US simulations of $[d(G_3T_4G_4)]_2$ containing multiple NH_4^+ ions

For the US studies of multiple-ion models, we initially put the upper and lower NH_4^+ ions at the centers of mass of O6 atoms of upper/central and lower/central G-quartets, respectively (as defined previously in Figure 4-1). The reaction coordinate was defined along the channel axis (Z) of $[d(G_3T_4G_4)]_2$. The channel axis for the upper and lower NH_4^+ ion movement was defined by a "dynamic axis" linking the two centers of mass of the four O6 atoms from each of the two

lower/central and upper/central G-quartets, respectively. For each multiple-ion model, the position of the upper or lower NH_4^+ ion (see Chapter 4 for definition) was fixed (denoted as "neighbor" ion) by different bias potentials. We performed US simulations in 17 separate windows (each being 0.25 Å in width) for each ion in which a bias potential of k = 100 kcal/mol/Å^2 was used to hold the ion within each window. The ion movement in the plane norm to the dynamic axis was confined by a harmonic potential (k = 100 kcal/mol/Å^2). Within each window the force acting on the ion was averaged in 0.1 Å bins. Typical production runs of 2 ns for each window were performed. Table 6-1 shows the summary of the following models.

Table 6-1. Summary of the models to study ion movement in of $[d(G_3T_4G_4)]_2$ containing multiple NH_4^+ ions.[a]

Model	Dynamics axis	Fixed C1'	Ion studied	Movement of other ions
1	L and C	Unconfined	U (0.8 to 4.8 Å)	L fixed (k = 100 kcal/mol/Å^2)
	U and C	Unconfined	L (-0.8 to -4.8 Å)	U fixed (k = 100 kcal/mol/Å^2)
2	L and C	Constrained	U (0.8 to 4.8 Å)	L fixed (k = 100 kcal/mol/Å^2)
	U and C	Constrained	L (-0.8 to -4.8 Å)	U fixed (k = 100 kcal/mol/Å^2)
3	L and C	Constrained	U (0.8 to 4.8 Å)	L fixed (k = 50 kcal/mol/Å^2)
	U and C	Constrained	L (-0.8 to -4.8 Å)	U fixed (k = 50 kcal/mol/Å^2)
4	L and C	Constrained	U (0.8 to 4.8 Å)	L fixed (k = 15 kcal/mol/Å^2)
	U and C	Constrained	L (-0.8 to -4.8 Å)	U fixed (k = 15 kcal/mol/Å^2)
5	L and C	Constrained	U (0.8 to 4.8 Å)	L fixed (k = 3 kcal/mol/Å^2)
	U and C	Constrained	L (-0.8 to -4.8 Å)	U fixed (k = 3 kcal/mol/Å^2)

[a] U, C, and L represent upper, central, and lower states, respectively.

Model 1:

In this model we studied ion transport of either the upper or lower NH_4^+ ion along the $[d(G_3T_4G_4)]_2$ channel while the neighbor ion was fixed at its initial position by a bias potential of k = 100 kcal/mol/Å^2. For example, as illustrated in Figure 6-4, when the upper NH_4^+ ion moved from RC = 0.8 to 4.8 Å, the lower NH_4^+ ion was fixed at RC = -1.7 Å. Figure 6-5 shows the PMF profiles obtained after 0.5 and 2 ns. The upper NH_4^+ ion experiences 4.7 kcal/mol free energy barrier to overcome the barrier while crossing the upper G-quartet. The amount of the required energy for the lower NH_4^+ ion passage through the lower G-quartet is 6.3 kcal/mol. As mentioned earlier, the two ends are different in this structure. As a result, we expect different energy barriers between two ends. In fact, the asymmetrical behavior is reproduced, but the magnitude of the energy barriers is significantly different from the ABF simulations. Under similar conditions, we obtained 8 and 16 kcal/mol for the energy barriers at the upper and lower ends, respectively using the ABF method. The discrepancy is more noticeable at the lower end. Here, we should emphasize that although we use ABF results for comparison, we are not still certain about the quantitative accuracy of these results. Particularly the lower end shows a similar energy barrier with the single-ion model, while we expect somewhat lower as discussed previously. The results of the US simulations in the current model show a lower energy barrier at both sides compared to the single-ion model. This is promising considering the effect of repulsion force on reducing the energy barriers.

After detailed trajectory analysis, we realized that the structure of the G-quartets, particularly the lower G-quartet, deviates from the initial structure during the simulation. The guanine residues move away from their initial planar shape specifically for the structures in which the ion is forced to locate within the G-quartet. There are some structures in which only three

guanine residues interact with the NH_4^+ ion while the other guanine residue is tilted and does not feel the NH_4^+ ion in its vicinity. These structures cannot fully represent the energetics of ion transport along the channel, because we are interested in obtaining the PMF profiles for a real situation in which the NH_4^+ ion interacts with all four guanine residues while crossing G-quartets. We calculated the PMF profiles for a shorter time scale (e.g., 0.5 ns) to reduce the chance of structural distortion. Interestingly, the energy barrier calculated for the upper ion movement did not change, whereas the lower ion required a higher energy to cross the lower G-quartet, ca. 8 kcal/mol. In general, our results show a lower barrier at both sides compared to the ABF simulations. So, we hypothesized that unfixing C1' atoms of guanine residues in this model may be the main reason for this significant discrepancy. While in ABF simulations, the allowed range of ion movement within each window inside the channel was typically 4 Å, it is only 0.25 Å for US simulations. This means that the ion in the ABF method has more opportunity to avoid unwanted structures, whereas in the US method the ion is forced to move within a small range and apparently responds more to other sources of freedom. To test this idea, we decided to introduce another model (Model 2) in which the positions of the C1' atoms of guanine residues will be constrained by a harmonic potential.

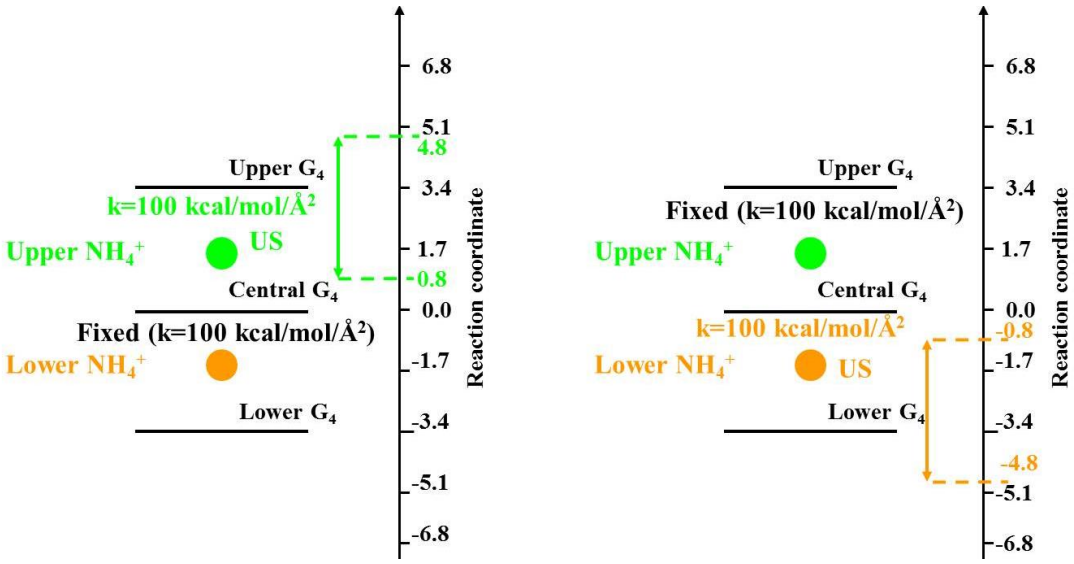

Figure 6-4. The schematic display of (left) the upper and (right) the lower NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 1). The green and the orange colors represent the upper and lower ions, respectively. A double arrow indicates the range of movement for an ion. The positions of neighbor ions are fixed by a bias potential of k = 100 kcal/mol/Å^2. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

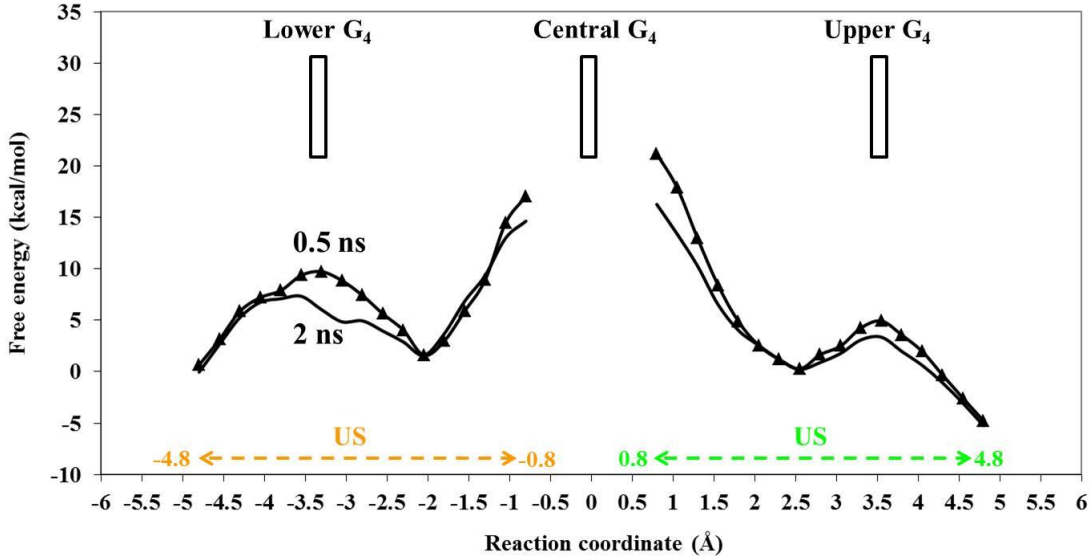

Figure 6-5. PMF profiles for the upper and lower NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 1) after 0.5 (triangle) and 2 ns (solid line) simulations. The green and the orange double arrows indicate the range of movement for the upper and lower NH_4^+ ions, respectively. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

Model 2:

In Model 2, the positions of C1' atoms of all guanine residues were fixed by a weak harmonic potential constant of k = 3 kcal/mol/Å^2. The simulation arrangement is shown in Figure 6-6. Figure 6-7 shows the PMF profiles for the upper and lower NH_4^+ ion movement after 0.3, 0.5, and 2 ns. The main objectives of these analyses are (1) to test if fixing C1' atoms can produce stable structures with reasonable PMF profiles and (2) if the energy profiles are time dependent. For the upper NH_4^+ ion transport through the upper G-quartet, 9.4, 9.4, and 8 kcal/mol energy is required after 0.3, 0.5, and 2 ns simulation, respectively. The PMF profiles show 15, 13.4, and 9.4 kcal/mol as the lower NH_4^+ ion energy barrier while passing through the lower G-quartet after 0.3, 0.5, and 2 ns, respectively. Several indications should be highlighted here. First, as expected the asymmetrical behavior is reproduced between two ends. Second, all time scales show a higher energy barrier for the NH_4^+ ion movement at the lower end. Third, the energy barriers are lower than the single-ion model. The magnitude of the energy barriers is also closer to the ABF results and the unusual low energy barriers are not observed in this model. Detailed trajectory analysis does not show a noticeable distortion from the initial structure. This confirms our hypothesis that fully unconstrained C1' atoms of guanine residues can mislead the simulations and produce unusual results. Similar to Model 1, the energy barrier at the upper end is not significantly time dependent, whereas the lower end shows a clear dependence on the time scale of the simulations. The energy barrier for the lower ion transport through the lower G-quartet decreases in both models as the simulation evolves. Although in model 2 we constrained the positions of all C1' atoms of guanine residues to ensure a stable structure, there are some simulation windows in which the ion is located within the G-quartet. Because of the size of NH_4^+ ion, these positions impose at least a partial opening for the G-quartets to let the ion either reside or pass through.

Forcing the NH_4^+ ion to undergo this situation for a long time may change the overall structure. This is more important for the lower end in which the overall structure is less flexible and therefore reacts more to the structural changes. So, in general, there are two issues to be considered when running US simulations. First, the stability of the initial structure and second the time scale of the simulation. Apparently, fixing the positions of C1' atoms of guanine residues guarantees a stable structure that could be maintained during short time scales of the simulation.

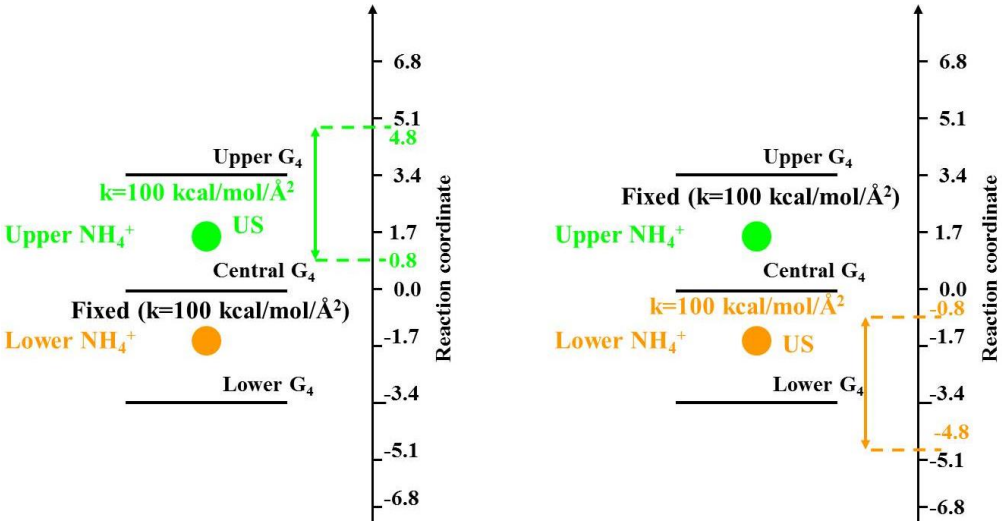

Figure 6-6. The schematic display of (left) the upper and (right) the lower NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 2). The green and the orange colors represent the upper and lower ions, respectively. A double arrow indicates the range of movement for an ion. The positions of neighbor ions are fixed by a bias potential of k = 100 kcal/mol/Å^2. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

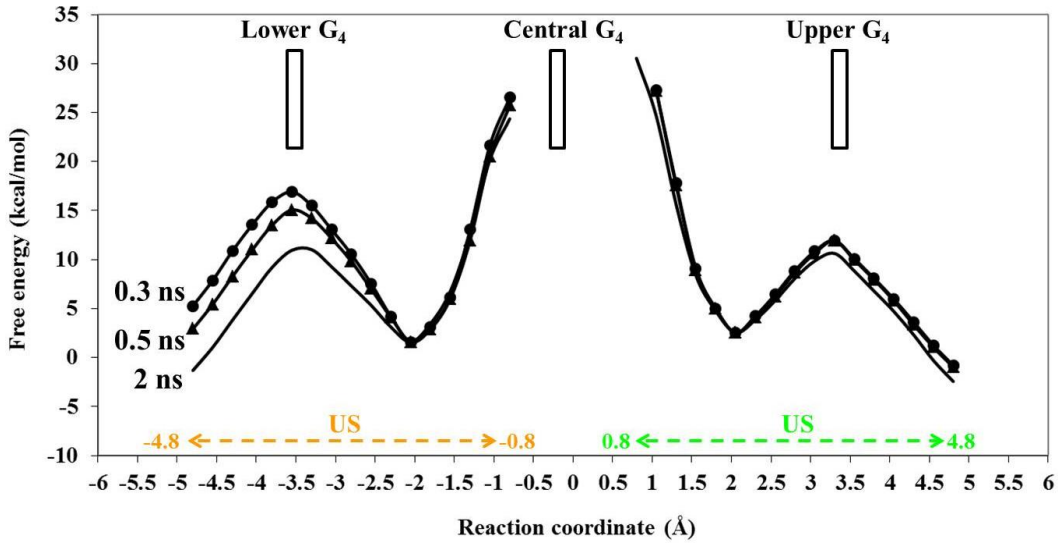

Figure 6-7. PMF profiles for the upper and lower NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 2) after 0.3 (circle), 0.5 (triangle), and 2 ns (solid line) simulations. The green and the orange double arrows indicate the range of movement for the upper and lower NH_4^+ ions, respectively. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

Model 3:

In Models 1 and 2, we fixed the neighbor ion at its initial position by a bias potential of k = 100 kcal/mol/Å^2. The reason to fix the neighbor ion is to make sure it maintains at its binding site and interacts with the ion under the US study. On the other hand, to represent a real situation, the neighbor ion should have the opportunity to adjust itself if too much repulsion exists between ions. Therefore, we should examine the consequence of applying a smaller harmonic potential on the neighbor ion. In model 3, we studied ion transport of the upper and lower NH_4^+ ions along the $[d(G_3T_4G_4)]_2$ channel while the neighbor ion was fixed at its initial position by a bias potential of k = 50 kcal/mol/Å^2 (see Figure 6-8). The positions of C1' atoms of all guanine residues were fixed by a weak harmonic potential constant of k = 3 kcal/mol/Å^2. Figure 6-9 shows the PMF profiles for the upper and lower NH_4^+ ion movement after 2 ns. The upper and lower NH_4^+ ion transport through their corresponding G-quartets requires 9.2 and 9.9 kcal/mol energy, respectively. These

energy barriers are in the same range as in Model 2. It appears that reducing bias potential on neighbor ion by 50 kcal/mol/Å^2 did not make a significant difference.

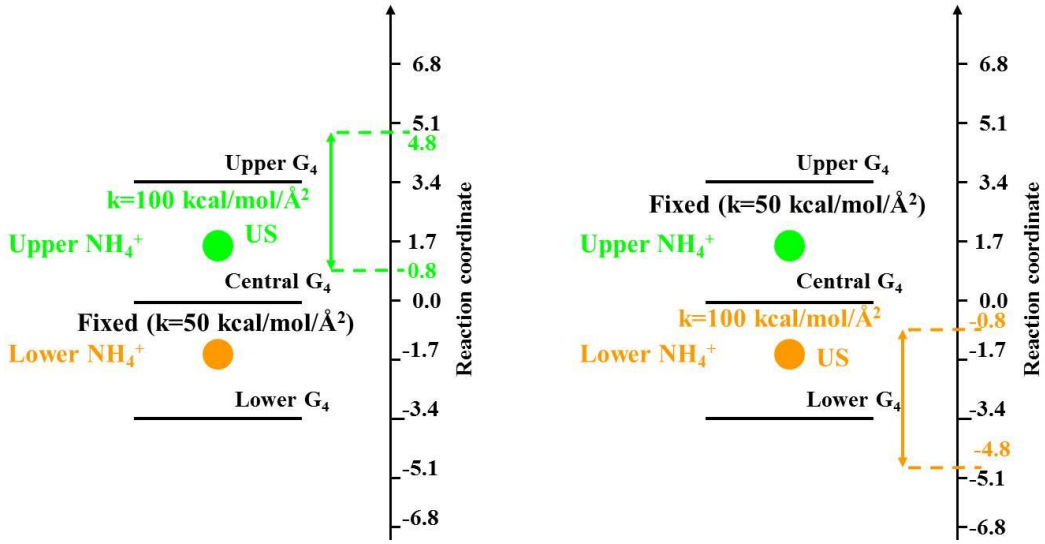

Figure 6-8. The schematic display of (left) the upper and (right) the lower NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 3). The green and the orange colors represent the upper and lower ions, respectively. A double arrow indicates the range of movement for an ion. The positions of neighbor ions are fixed by a bias potential of k = 50 kcal/mol/Å^2. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

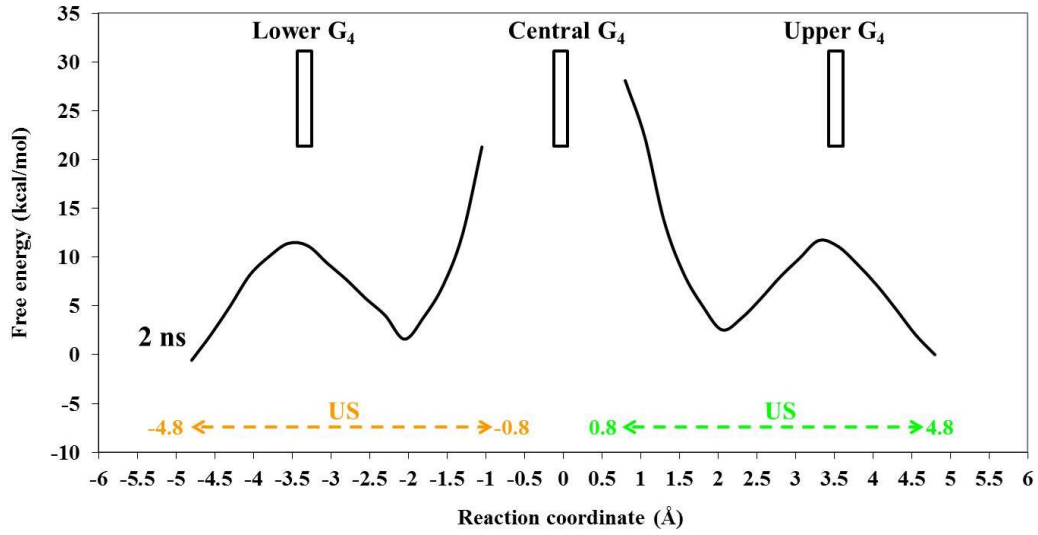

Figure 6-9. PMF profiles for the upper and lower NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 3) after 2 ns simulation. The green and the orange double arrows indicate the range of movement for the upper and lower NH_4^+ ions, respectively. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

Model 4:

Since reducing the bias potential on the neighbor ion in Model 3 did not provide a noticeable difference, we decided to further decrease the bias potential on the neighbor ion to k = 15 kcal/mol/Å^2 (see Figure 6-10). Figure 6-11 shows the PMF profiles for the upper and lower NH_4^+ ion movement after 2 ns. The energy barrier for the upper and lower NH_4^+ ion movement through their corresponding G-quartets is 8.7 and 10.3 kcal/mol, respectively. The difference between these energy barriers with those in Models 2 and 3, is less than 1 kcal/mol. So, even reducing bias potential on neighbor ion to 15 kcal/mol/Å^2 does not produce a different energy profile. Therefore, we decided to decrease the bias potential on the neighbor ion even further to see how the simulation responds.

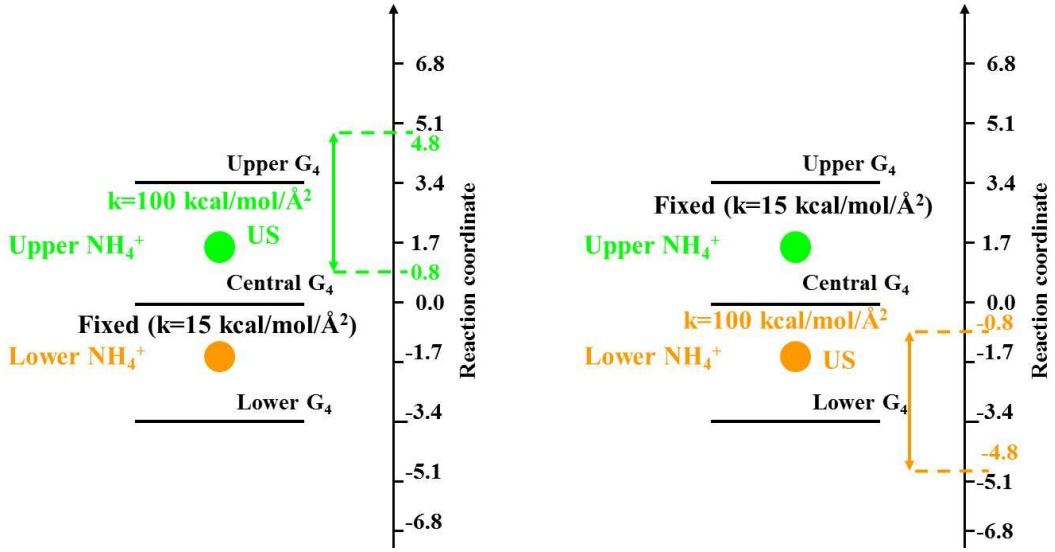

Figure 6-10. The schematic display of (left) the upper and (right) the lower NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 4). The green and the orange colors represent the upper and lower ions, respectively. A double arrow indicates the range of movement for an ion. The positions of neighbor ions are fixed by a bias potential of k = 15 kcal/mol/Å^2. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

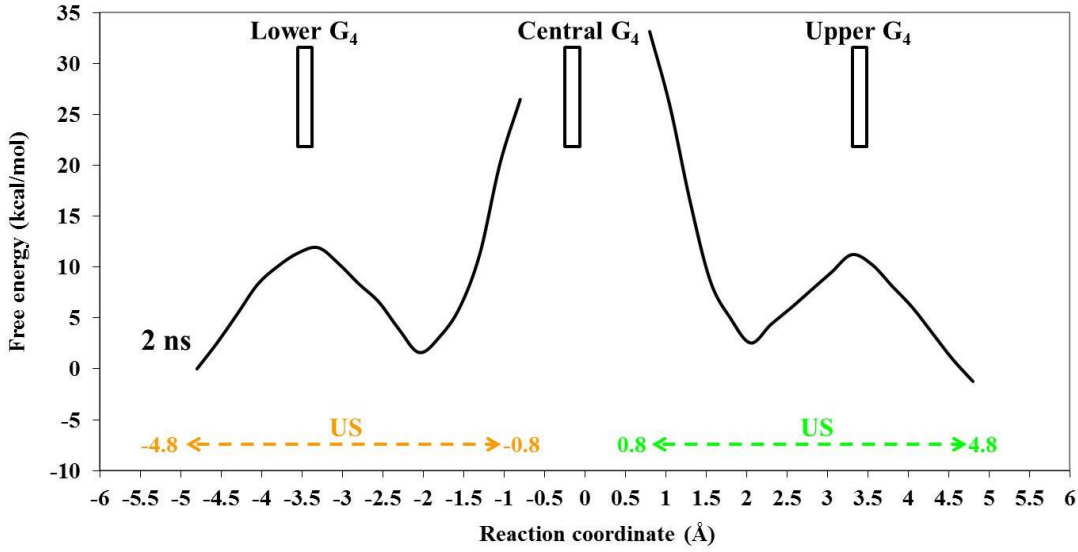

Figure 6-11. PMF profiles for the upper and lower NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 4) after 2 ns simulation. The green and the orange double arrows indicate the range of movement for the upper and lower NH_4^+ ions, respectively. The origin of the

reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

Model 5:

In this model, we further reduced the bias potential on the neighbor ion to k = 3 kcal/mol/Å^2 (see Figure 6-12). Figure 6-13 shows the PMF profiles for the upper and lower NH_4^+ ion movement after 0.3 and 2 ns. We initially calculated the PMF profiles after 2 ns simulation and obtained 8.9 and 7 kcal/mol as the energy barriers for the upper and lower NH_4^+ ion movement through their corresponding G-quartets. The barrier at the lower end differs from the previous models, while the upper end behaves the same. In addition, almost 2 kcal/mol difference is observed between the two ends although the trend is not as expected. To obtain additional information, we analyzed the results after 0.3 ns and obtained 8.2 versus 13 kcal/mol as the energy barriers for the upper and lower NH_4^+ ion movement through their corresponding G-quartets. Apparently, the expected asymmetrical behavior is observed between the two ends. Furthermore, in agreement with the previous results, the upper end shows similar energy barriers over time, whereas the energy for the lower NH_4^+ ion transport decreases as the simulation evolves. As mentioned, when running US simulations, one should be careful about the time scale of the simulations because for some windows the structure may deviate significantly over time.

It is important to mention that in all multiple-ion models the PMF profiles show the highest energy barrier when ions are at their closest distance. This is simply due to the repulsion force between two ions which is weakened by increasing the distance as can be seen from the free energy profiles.

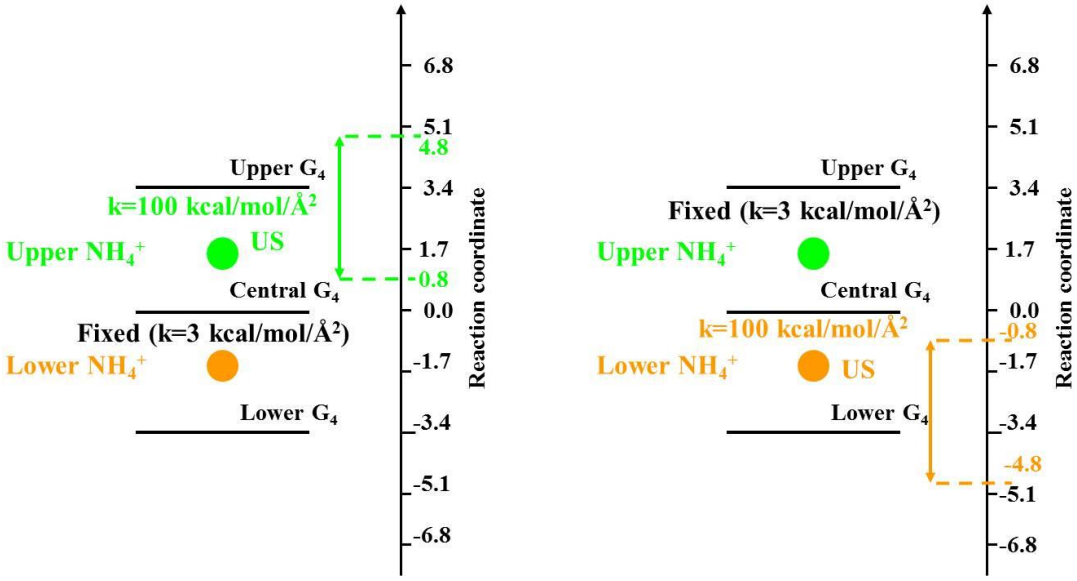

Figure 6-12. The schematic display of (left) the upper and (right) the lower NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 5). The green and the orange colors represent the upper and lower ions, respectively. A double arrow indicates the range of movement for an ion. The positions of neighbor ions are fixed by a bias potential of k = 3 kcal/mol/Å^2. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

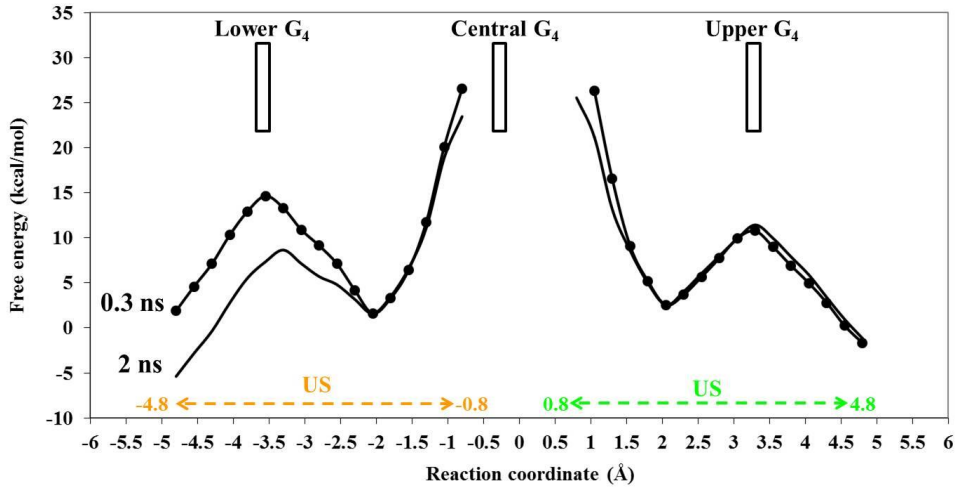

Figure 6-13. PMF profiles for the upper and lower NH_4^+ ion movement in a multiple-ion $[d(G_3T_4G_4)]_2$ channel (Model 5) after 0.3 (circle) and 2 ns (solid line) simulations. The green and the orange double arrows indicate the range of movement for the upper and lower NH_4^+ ions, respectively. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

Figure 6-14(a) shows the distribution histogram of the lower NH_4^+ as the neighbor ion along the reaction coordinate during 2 ns simulation while the upper NH_4^+ ion is under the US study. The histograms are obtained from the trajectory data for bias potentials of k = 3, 15, 50, and 100 kcal/mol/Å^2 on the neighbor ion while the upper ion is located at RC = 0.8, 2.05, and 3.3 Å. These positions correspond to two maxima and one minimum on the free energy profiles and could properly represent the total behavior of the structure. Similarly, Figure 6-14(b) exhibits the distribution histogram of the upper NH_4^+ as the neighbor ion along the reaction coordinate during 2 ns simulation while the lower NH_4^+ ion is under the US study. The histograms are obtained from the trajectory data for bias potentials of k = 3, 15, 50, and 100 kcal/mol/Å^2 on the neighbor ion while the lower ion is located at RC = -0.8, -2.05, and -3.55 Å. A common observation is that using a higher bias potential (e.g., k = 100 kcal/mol/Å^2) keeps the neighbor ion at its initial position with a smaller range of ion fluctuation. While for k = 100 kcal/mol/Å^2 the ion is always located around its initial position (RC = 1.7 Å), it deviates from the initial position for lower bias potentials particularly for k = 3 kcal/mol/Å^2. The shift is more significant for RC = 0.8 and -0.8 Å in which the distance between ions is the smallest. Interestingly, when the upper ion under the US investigation is located at the window center of 0.8 Å, the lower neighbor ion shifts to the left significantly to reduce the repulsion. Similar behavior can be observed for RC = -0.8 Å in which the upper neighbor ion shifts to the right. These effects are not considerable at other window centers since ions are far enough to repel each other significantly.

While the full width at the half maximum (FWHM) for k = 100 kcal/mol/Å^2 rarely exceeds from 0.1 Å, the ion experiences a larger range of movement using smaller bias potentials. Using k = 3 kcal/mol/Å^2 to fix the neighbor ion at the initial position appears to be a good choice

because the ion has the opportunity to fluctuate in a larger range up to 0.4 Å while it is still maintained within the binding site.

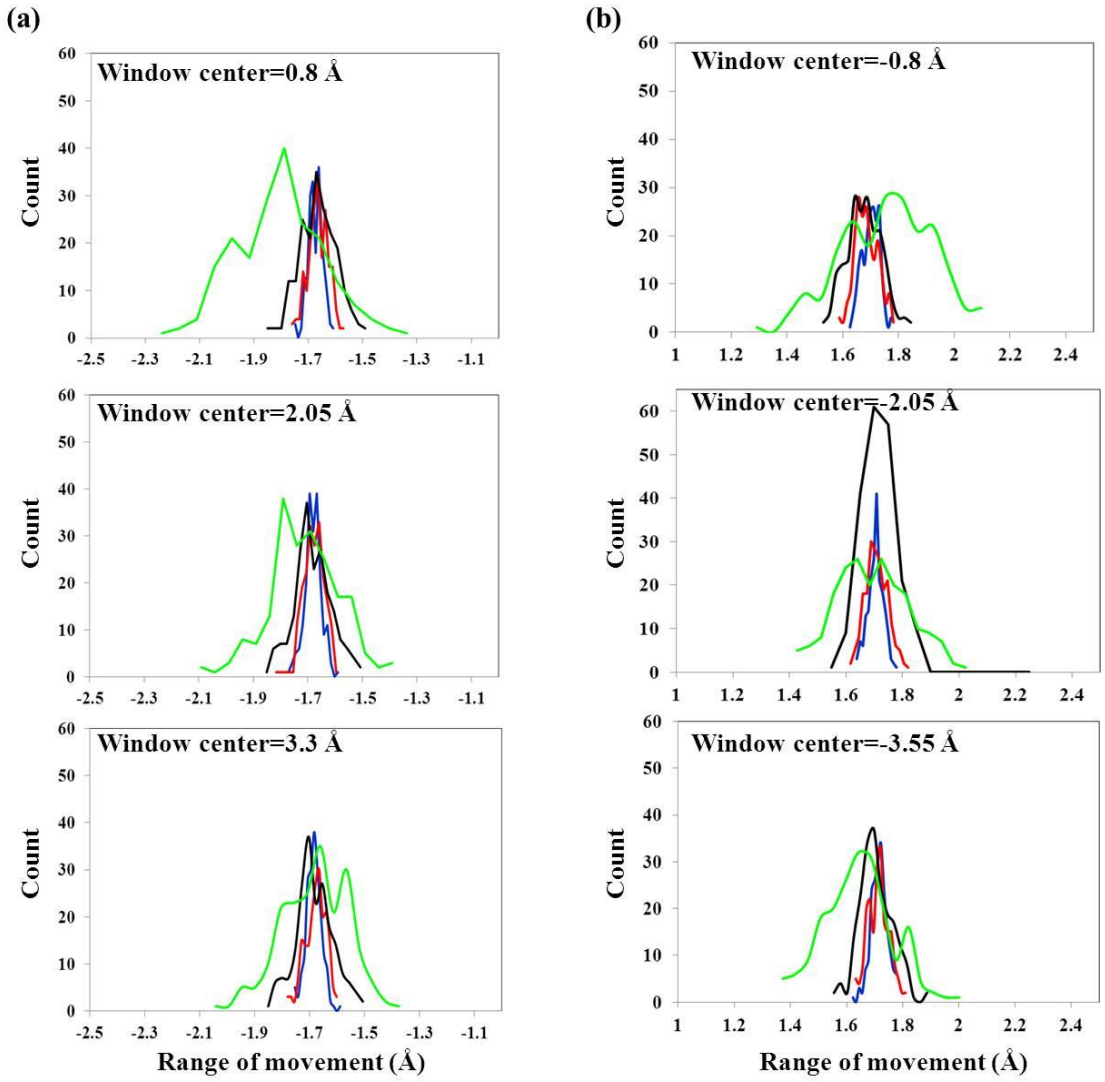

Figure 6-14. Distribution histograms of (a) the lower and (b) the upper NH_4^+ as the neighbor ion along the reaction coordinate during 2 ns simulation using bias potentials of k = 3 (green), 15 (black), 50 (red), and 100 (blue) kcal/mol/Å^2 while the upper NH_4^+ ion is located at the window centers of 0.8, 2.05, and 3.3 Å and the lower NH_4^+ ion is located at the window centers of -0.8, -2.05, and -3.55 Å.

In the previous chapters, we used the ABF and US methods to investigate the difference between several multiple- and single-ion models. We are particularly interested in understanding how the interaction between multiple ions affects the energetics of ion movement along the channel. As we concluded in Chapter 5, in general, when an ion is surrounded by two ions at the same distance, its PMF profile will be similar to the single-ion model. However, the PMF profile will differ from the single-ion model if the ion has only one neighbor. Two ions are located inside a $[d(G_3T_4G_4)]_2$ channel containing multiple ions. Each ion repels the other one so that they are both kicked out toward their corresponding G-quartets. As a result, both ions should require a lower amount of energy to overcome the barriers while passing through G-quartets. In a single-ion model, ions do not have any additional source of energy to pass the barrier, so a higher energy barrier is expected. Figure 6-15 shows the comparison between the PMF profiles for the NH_4^+ ion movement in a single-ion model and the multiple-ion Model 5 along the $[d(G_3T_4G_4)]_2$ channel after 0.3 ns. As expected, the energy barrier is reduced in the multiple-ion Model 5 by 8.4 and 7.5 kcal/mol for the upper and lower ends, respectively. Interestingly, the PMF profiles for the multiple-ion model shift slightly toward the upper and lower G-quartets. This is understandable because the repulsion force between the ions is lower once they are farther from each other.

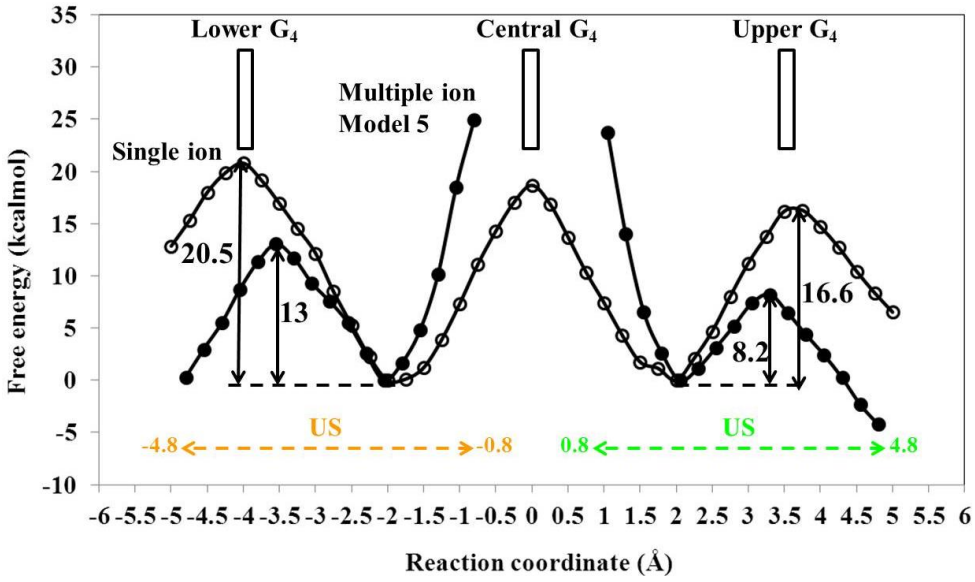

Figure 6-15. Comparison between PMF profiles for the NH_4^+ ion movement in a single-ion model (open circle) and the multiple-ion Model 5 (filled circle) along the $[d(G_3T_4G_4)]_2$ channel after 0.3 ns simulation. The green and the orange double arrows indicate the range of movement for the upper and lower NH_4^+ ions, respectively. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

6.3.3 2D US simulations of $[d(G_3T_4G_4)]_2$ containing multiple NH_4^+ ions

In the multiple ion models presented so far, we studied only one pathway for the ion movement along the channel. In fact, the energy profile for the movement of one ion was always calculated while the neighbor ion was kept fixed. However, there could be other pathways with lower energy barriers for the ion movement. In order to obtain a complete profile for the movement of ions independently, we performed 2D US simulations in which a total of 301 independent simulations were run with the entire simulation time of 90 ns. A bias potential of k = 100 kcal/mol/Å^2 was applied to move the lower NH_4^+ ion from RC = -5.0 to -1.75 Å while the upper NH_4^+ ion was moved starting from 3 Å above the lower NH_4^+ ion up to the final RC = 4.75

Å (with increments of 0.25 Å). Within each window the force acting on the ion was averaged in 0.1 Å bins. The positions of C1' atoms of all guanine residues were fixed by a weak harmonic potential constant of k = 3 kcal/mol/Å^2. The umbrella sampling simulations were unbiased using 2D-WHAM to generate full energy profiles of ion movements along the channel.

Figure 6-16 shows the 2D PMF profile for independent movement of upper and lower NH_4^+ ions along the channel. The blue and red colors on the map represent the lower and higher energy barriers, respectively. The X and Y axes show the position of the lower and upper NH_4^+ ions, respectively. The Y axis covers only a range of RC = 1.0 to 4.75 Å for the upper NH_4^+ ion movement. The corresponding structures to the three minimum energy regions on the map are shown. The dashed white line on the map shows the minimum energy pathway for NH_4^+ ion movement along the channel. Apparently, the energy barrier for the upper NH_4^+ ion to pass through the upper G-quartet is 6.4 kcal/mol, whereas the lower NH_4^+ ion should overcome 9.2 kcal/mol energy to pass through the lower G-quartet. In agreement with the previous models the asymmetric behavior between the two ends is reproduced in these calculations. Also, the barriers are significantly lower than those experienced by the single NH_4^+ ion. As mentioned before, the presence of a neighbor ion in the channel facilitates the movement of the other ion through the repulsion effect. Both ions experience a lower energy barrier since their neighbor ion kicks them out while they are passing through their corresponding G-quartets.

Although the 1D US studies provide useful information regarding the energetics of ion movement along the channel, these studies only cover one pathway from many other possible scenarios for the ion movement. The 2D US simulations provide the opportunity to study the movement of both ions independently and find the minimum energy pathway. The fact that the minimum energy pathway is tilted suggests that the movement of ions is correlated. In fact, once

the lower ammonium ion enters the channel from the lower end, it pushes the upper ion slightly toward up. Further movement of the lower ion along this direction kicks the upper ion out of the channel through the upper G-quartet. The driving force for this process is the repulsion force between the two ions which in fact reduces the energy barrier for the ion passage significantly.

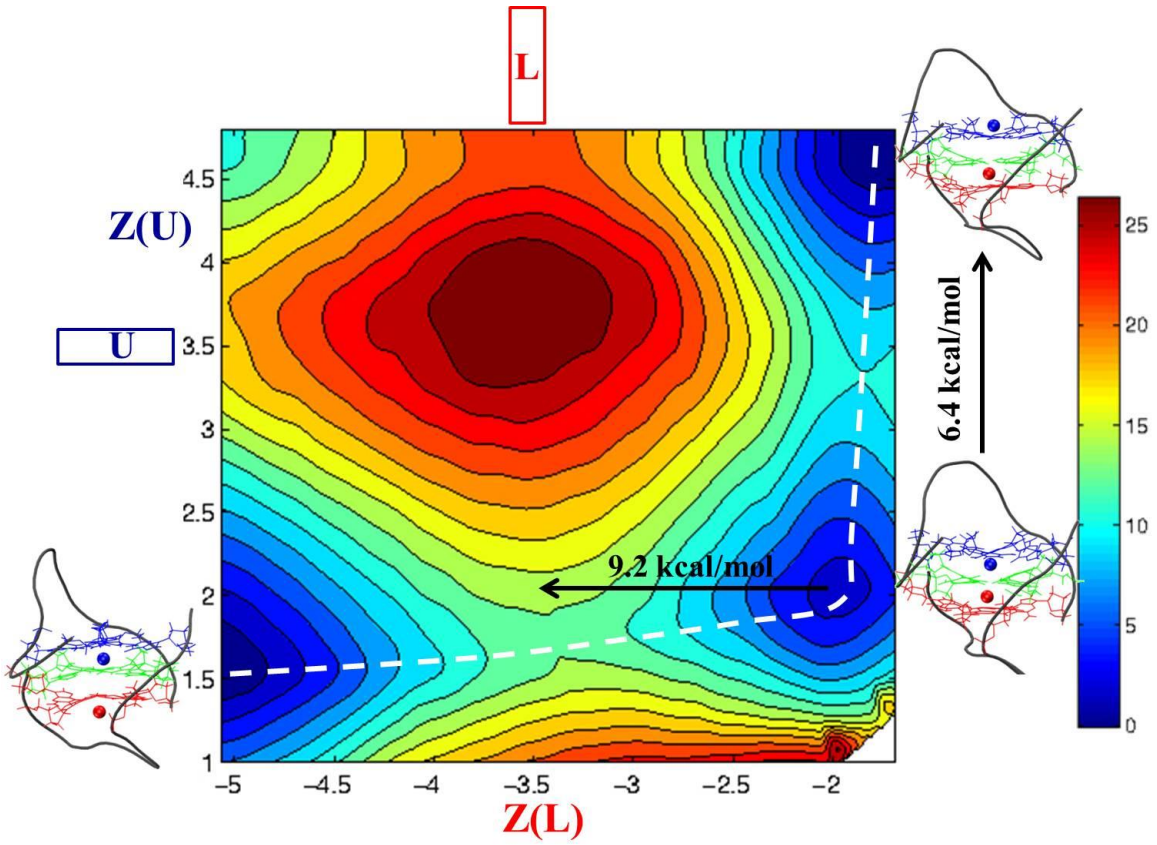

Figure 6-16. 2D PMF profile for the independent movement of upper and lower NH_4^+ ions along the $[d(G_3T_4G_4)]_2$ channel after 0.3 ns simulation. U and L represent the upper and lower G-quartets, respectively. Z(U) and Z(L) represent the positions of the upper and lower NH_4^+ ions, respectively. The origin of the reaction coordinate corresponds to the center of mass of the four O6 atoms of the central G-quartet.

6.4 Conclusion

In this chapter we employed the US method to study the NH_4^+ ion movement along the $[d(G_3T_4G_4)]_2$ channel using single- and multiple-ion models. For single-ion models, the US and ABF methods produced similar results. We particularly investigated the structural arrangement and the time dependence of the PMF profiles for the upper and lower NH_4^+ ion movement. Our results suggest that the structure is better maintained under the condition of fixed C1' atoms of guanine residues. In addition, performing the simulation for longer than 0.3 ns may change the structure specifically when the ion is located within the G-quartet. This may result in unusual energy profiles. The upper end seems more stable as it did not show significant time dependence. We also carried out several simulations in which the position of the neighbor ion was constrained by different bias potentials. While bias potentials of k = 100, 50, and 15 kcal/mol/Å^2 on the neighbor ion produce similar results, the structure shows a noticeable response to k = 3 kcal/mol and presents a clear asymmetrical behavior for the two ends of the channel. A comparison between a single- and a multiple-ion model clearly shows a lower barrier for both ends in the multiple-ion model. This further supports our previous discussions regarding the influence of the repulsion force to reduce the barriers while ions crossing G-quartets.

Chapter 7

Summary and conclusions

This thesis documents a comprehensive MD simulation study of ion dynamics inside G-quadruplex DNA channels. In particular, we have employed the ABF and US methods to investigate energetics of ion movement along two G-quadruplex DNA channels: $[d(TG_4T)]_4$ and $[d(G_3T_4G_4)]_2$. For the first time, we have obtained complete free energy profiles for the movement of Na^+, K^+, and NH_4^+ ions through the channel and for the channel exit/entrance points to the bulk. Because of the limitations of experimental techniques, there are not many studies published in the literature that examine ion dynamics along G-quadruplex DNA channels. Therefore, our research has produced a significant amount of new information about this process. This is the first time that these modern free-energy calculation methods are applied to G-quadruplex DNA. Our simulations showed that the energy barriers for ion movement through the G-quadruplex channels depend on the size of ions. While large K^+ and NH_4^+ ions must "squeeze" through a G-quartet to reach the adjacent channel binding site, Na^+ can be considered to "diffuse" nearly continuously through the channel. Interestingly, our results also indicated that the energetics of ion movement along G-quadruplex DNA channel is highly structural dependent. We found that $[d(TG_4T)]_4$ forms an almost symmetrical channel in which the energy barriers are comparable at both channel ends. In contrast, a considerable difference was observed for the $[d(G_3T_4G_4)]_2$ channel. This observation is consistent with the experimental data available in the literature. We proposed a new hypothesis that offers an explanation about the stiffness of a G-quartet. 2D PMF maps are obtained to investigate ion movement after reaching the channel exit/entrance points, while it moves into the bulk. Interestingly, almost flat free energy profiles are acquired for all ions

regardless of the ion size and also the topology of the loops in $[d(G_3T_4G_4)]_2$ structure. The hydration state analysis of the trajectory data revealed that the ions are almost completely dehydrated inside the channel and become fully hydrated after leaving the channel. Our ABF simulations have provided, for the first time, a certain proof that the movement of ions in G-quadruplex channels occurs along the channel and the ions do not exit the channel through the walls. The calculated free energy barrier for the "sideways" ion movement is within 50-60 kcal/mol, which is much larger than that along the channel axis.

Using the ABF and US methods, we have extensively investigated the ion-ion interactions and the repulsion forces between them in several multiple-ion models under different conditions. We discovered that for an ion surrounded by two other ions in the channel within the same distances, the repulsion forces will be cancelled and a similar free energy profile to the single-ion model will be obtained for the inner ion. However, the outer ions will be repelled by the inner ion toward their corresponding G-quartets and they will face a lower energy barrier to cross the G-quartets compared to the single-ion model.

Our US studies have provided significant information regarding the proper simulation arrangements and time scales that could produce reliable results. As in the US method, a bias potential should be applied to keep the ion within the defined window, there exist several unwanted structures in which the ion is forced to locate within the G-quartet. This is energetically unfavourable particularly for large ions such as NH_4^+. We have suggested to fix the positions of C1' atoms of guanine residues to avoid major structural distortion from the initial structure. Also, the simulations should be performed for no longer than 0.3 ns to ensure a good sampling while the structure is still maintained.

In the future, several simulations can be carried out to complement our MD studies for ion transport along G-quadruplex channels. First, our US simulations of NH_4^+ ion transport along $[d(G_3T_4G_4)]_2$ channel indicated a much lower energy barrier for the multiple-ion model compared to the single-ion model. In fact, they differ by 7-8 kcal/mol. Also, in general, the energy barrier for ion movement significantly depends on the size of the ion. For the single-ion model, the energy barrier for Na^+ ion to cross the G-quartet is about 5-7 kcal/mol. It would be interesting to see how the repulsion forces will reduce the barrier in multiple Na^+ ion models. Based on our previous conclusions, a flat free energy profile would not be surprising that in fact, resembles a Na^+ ion channel in G-quadruplex structures. Second, based on our US simulations, we have obtained appropriate conditions that guarantee the structural maintaining while producing reliable results. These useful outcomes in the US method can be used to investigate ion movement along $[d(TG_4T)]_4$ channel containing multiple ions. We expect to obtain (1) a similar energy profile to the single-ion model for the inner ions if they are surrounded by two other ions within the same distances, (2) a lower energy barrier for the outer ions compared to the single-ion model, (3) similar energy barriers for K^+ and NH_4^+ ions and smaller energy barriers for Na^+ ion, and (4) an almost flat PMF profile for the movement of Na^+ ion along the channel. Some of these results have been previously achieved from our ABF simulations and could be further verified by performing US simulations under our established proper conditions. Third, most of the G-quadruplex structures are known to be formed in water mostly because many G derivatives are insoluble in organic solvents.[10] However, there are some published researches in the literature that indicate lipophilic G derivatives can also form high-ordered aggregates that resemble ion channel behaviors.[203] The formation of these aggregates also requires the presence of metal ions. Our group has studied the G-quartet formation of a lipophilic guanosine derivative (2',3',5'-*O*-

triacetylguanosine, TAG) in the presence of trivalent lanthanide metal ions (La^{3+}, Eu^{3+}, Tb^{3+}, Dy^{3+}, Tm^{3+}).[204] Also, Ca^{2+} ion-binding to the G-quartet structure formed by the aggregation of TAG has been studied in our group using NMR method.[173] Our simulation studies have investigated ion dynamics for monovalent metal ions in water solution so far. The same methodologies as explained in previous chapters could be employed to examine the energetics of divalent and trivalent cations in both water and organic solutions. Fourth, all of our MD simulations have been performed under neutral condition. That is, since DNA has -1 charge per each phosphate, one cation has been added per one phosphate in the neutralization step. In order to investigate the role of counter ions in the ion movement along the channel, one could perform the simulations under higher salt concentrations. As extra ions will be provided, ion movement may be accelerated resulting in shorter residence time of ions inside the channel. Finally, we can use the same ABF and US methodologies to further investigate the energetics of ion dynamics along several other G-quadruplex channels considering the impact of nucleobase substitution, the effect of other loop topologies, the difference between parallel and antiparallel arrangements, as well as the difference between DNA and RNA nucleobases.

Bibliography

[1]Miescher, F. *Medicinischchemische Untersuchungen*. **1871,** *4*, 441-460.

[2]Rich, A. *Q. Rev. Biophys.* **2009**, *42*, 117-137.

[3]Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag New York Inc: New York, **1984**.

[4]Wang, A. H.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; Van Boom, J. H.; Van der Marel, G.; Rich, A. *Nature* **1979**, *282*, 680-686.

[5]Wu, Z.; Tian, T.; Yu, J.; Weng, X.; Liu, Y.; Zhou, X. *Angew. Chem. Int. Ed.* **2011,** *50*, 11962-11967.

[6]Bang, I. *Biochemistry Z* **1910,** *26*, 293-311.

[7]Ralph, R. K.; Connors, W. J.; Khorana, H. G. *J. Am. Chem. Soc.* **1962,** *84*, 2265-2272.

[8]Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. USA* **1962,** *48*, 2013-2018.

[9]Nikolova, E. N.; Euane, K.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. *Nature* **2011**, *470*, 498-502.

[10]Davis, J. T. *Angew. Chem. Int. Ed.* **2004,** *43*, 668-698.

[11]Miles, H.; Frazier, J. *Biochim. Biophys. Acta* **1964,** *79*, 216-220.

[12]Chantot, J. F.; Satocchi, M.; Guschlbauer, W. *Biochimie* **1971,** *53*, 347-354.

[13]Haider, S.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2002,** *320*, 189-200.

[14]Hazel, P.; Parkinson, G. N.; Neidle, S. *J. Am. Chem. Soc.* **2006,** *128*, 5480-5487.

[15]Schultze, P.; Macaya, R. F.; Feigon, J. *J. Mol. Biol.* **1994,** *235*, 1532-1547.

[16]Burge, S.; Parkinson, J. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res*. **2006,** *34*, 5402-5415.

[17]Kang, C.; Zhang, X.; Ratliff, R.; Moyzis, R.; Rich, A. *Nature* **1992**, *356*, 126-131.

[18]Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263-282.

[19]Zahler, A.; Williamson, J.; Cech, T.; Prescott, D. *Nature* **1991,** *350*, 718-720.

[20]McEachern, M.; Krauskopf, A.; Blackburn, E. *Annu. Rev. Genet.* **2000,** *34*, 331-358.

[21]Ma, L.; Iezzi, M.; Kaucher, M. S.; Lam, Y. F.; Davis, J. T. *J. Am. Chem. Soc.* **2006,** *128*, 15269-15277.

[22]Pinnavaia, T.; Miles, H.; Becker, E. *J. Am. Chem. Soc.* **1975,** *97*, 7198-7200.

[23]Pinnavaia, T.; Marshall, C.; Mettler, C.; Fisk, C.; Miles, H.; Becker, E. *J. Am. Chem. Soc.* **1978,** *100,* 3625-3627.

[24]Bouhoutsos-Brown, E.; Marshall, C. L.; Pinnavaia, T. J. *J. Am. Chem. Soc.* **1982,** *104,* 6576-6584.

[25]Paris, A.; Laszlo, P. *C. R. Acad. Sci. Paris Ser. D* **1978,** *286,* 717-719.

[26]Detellier, C.; Paris, A.; Laszlo, P. *C. R. Acad. Sci. Paris Ser. D* **1978,** *286,* 781-783.

[27]Borzo, M.; Laszlo, P. *C. R. Acad. Sci. Paris Ser. C* **1978,** *287,* 475-478.

[28]Delville, A.; Detellier, C.; Laszlo, P. *J. Magn. Reson.* **1979,** *34,* 301-315.

[29]Zimmerman, S. B. *J. Mol. Biol.* **1976**, *106,* 663-672.

[30]Lipanov, A. A.; Quintana, J.; Dickerson, R. E. *J. Biomol. Struct. Dyn.* **1990**, *8,* 483-489.

[31]Xu, Q. W.; Deng, H.; Braunlin, W. H. *Biochemistry* **1993,** *32,* 13130-13137.

[32]Deng, H.; Braunlin, W. H. *J. Mol. Biol.* **1996,** *255,* 476-483.

[33]Wong, A.; Fettinger, J. C.; Forman, S. L.; Davis, J. T.; Wu, G. *J. Am. Chem. Soc.* **2002,** *124,* 742-743.

[34]Wu, G.; Wong, A.; Gan, A.; Davis, J. *J. Am. Chem. Soc.* **2003,** *125,* 7182-7183.

[35]Wong, A.; Wu, G. *J. Am. Chem. Soc.* **2003,** *125,* 13895-13905.

[36]Wong, A.; Ida, R.; Wu, G. *Biochem. Biophys. Res. Commun*. **2005,** *337,* 363-366.

[37]Ida, R.; Wu, G. *J. Am. Chem. Soc.* **2008,** *130,* 3590-3602.

[38]Wu, G.; Kwan, I. C. M. *J. Am. Chem. Soc.* **2009,** *131,* 3180-3182.

[39]Hubbard, P. S. *J. Chem. Phys.* **1970**, *53,* 985-987.

[40]Bull, T. E. *J. Magn. Reson.* **1972,** *8,* 344-353.

[41]Hud, N. V.; Schultze, P.; Feigon, J. *J. Am. Chem. Soc.* **1998,** *120,* 6403-6404.

[42]Hud, N. V.; Schultze, P.; Sklenar, V.; Feigon, J. *J. Mol. Biol.* **1999,** *285,* 233-243.

[43]Feigon, J.; Butcher, S. E.; Finger, L. D.; Hud, N. V. *Methods Enzymol.* **2001,** *338,* 400-420.

[44]Basu, S.; Szewczak, A. A.; Cocco, M.; Strobel, S. A. *J. Am. Chem. Soc.* **2000,** *122,* 3240-3241.

[45]Wong, A.; Wu, G. *J. Phys. Chem. A* **2000,** *104,* 11844-11852.

[46]Borzo, M.; Detellier, C.; Laszlo, P.; Paris, A. *J. Am. Chem. Soc.* **1980,** *102,* 1124-1134.

[47]Levitt, M.; Warshel, A. *Nature* **1975,** *253,* 694-698.

[48]Cooper, A. *Proc. Natl. Acad. Sci. USA* **1976,** *73,* 2740-2741.

[49]Case, D. A.; Karplus, M. *J. Mol. Biol.* **1979,** *132,* 343-368.

[50]Levitt, M.; Sharon, R. *Proc. Natl. Acad. Sci. USA* **1988,** *85,* 7557-7561.

[51]Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Lin, J.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D.; Baker, D. *Nature* **2008,** *453,* 190-195.

[52]Berneche, S.; Roux, B. *Nature* **2001,** *414,* 73-77.

[53]Grayson, P.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2003,** *85,* 36-48.

[54]Ruiter, A. D.; Oostenbrink, C. *Curr. Opin. Chem. Biol.* **2011,** *15,* 547-552.

[55]McCarrick, M. A.; Kollman, P. *Methods Enzymol*. **1994,** *241,* 370-384.

[56]Durrant, J. D.; McCammon, J. A. *BMC Biology* **2011,** *9,* 1-9.

[57]Yang, Y.; Jin, Q.; Liu, H.; Yao, X. *J. Chem. Inf. Model.* **2011,** *51,* 680-692.

[58]Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. *Curr. Opin. Struct. Biol*. **2003,** *13,* 168-174.

[59]Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu. Rev. Phys. Chem.* **2007,** *58,* 57-83.

[60]Im, W.; Lee, J.; Kim, T.; Rui, H. *J. Comput. Biol.* **2009,** *30,* 1622-1633.

[61]Khalili-Araghi, F.; Gumbart, J.; Wen, P.; Sotomayor, M.; Tajkhorshid, E.; Schulten, K. *Curr. Opin. Struct. Biol*. **2009,** *19,* 128-137.

[62]Roux, B.; Schulten, K. *Structure* **2004,** *12,* 1343-1351.

[63]Gordon, D.; Krishnamurthy, V.; Chung, S. *J. Chem. Phys.* **2009,** *131,* 1-11.

[64]Allen, T. W.; Kuyucak, S.; Chung, S. *Biophys. J.* **1999,** *77,* 2502-2516.

[65]Chung, S.; Allen, T. W.; Kuyucak, S. *Biophys. J.* **2002,** *82,* 628-645.

[66]Crouzy, S.; Berneche, S.; Roux, B. *J. Gen. Physiol.* **2001,** *118,* 207-218.

[67]Burykin, A.; Schutz, C. N.; Villa, J.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2002,** *47, 265*-280.

[68]Roux, B.; Karplus, M. *J. Am. Chem. Soc.* **1993,** *115,* 3250-3262.

[69]Roux, B. *Biophys. J.* **1996,** *71,* 3177-3185.

[70]Allen, T. W.; Andersen, O. S.; Roux, B. *Proc. Natl. Acad. Sci.* **2004,** *101,* 117-122.

[71]Allen, T. W.; Andersen, O. S.; Roux, B. *Biophys. J.* **2006,** *90,* 3447-3468.

[72]Allen, T. W.; Andersen, O. S.; Roux, B. *Biophys. J.* **2006,** *124,* 251-267.

[73]Aqvist, J.; Warshel, A. *Biophys. J.* **1989,** *56,* 171-182.

[74]Chiu, S.; Novotny, J. A.; Jakobssontl, E. *Biophys. J.* **1993,** *64,* 98-108.

[75]Roux, B.; Karplus, M. *Annu. Rev. Biophys. Biomol. Struct.* **1994,** *23,* 731-761.

[76]Chowdhury, S.; Chanda, B. *J. Gen. Physiol.* **2012**, *139,* 3-17.

[77]Zhu, F.; Hummer, G. *Proc. Natl. Acad. Sci.* **2010,** *16,* 19814-19819.

[78]Joseph, S.; Mashl, R. J.; Jakobsson, E.; Aluru, N. R. *Nano Lett.* **2003,** *3,* 1399-1403.

[79]Dzubiella, J.; Allen, J. R.; Hansen, J. P. *J. Chem. Phys.* **2004,** *120,* 5001-5004.

[80]Peter, C.; Hummer, G. *Biophys. J.* **2005,** *89,* 2222-2234.

[81]Liu, H.; Murad, S.; Jameson, C. J. *J. Chem. Phys.* **2006,** *125,* 084713 (1-14).

[82]Beu, T. A. *J. Chem. Phys.* **2010,** *132,* 164513 (1-15).

[83]Asthagiri, D.; Bashford, D. *Biophys. J.* **2002,** *82,* 1176-1189.

[84]Hwang, H.; Schatz, G. C.; Ratner, M. A. *J. Phys. Chem. B* **2006,** *110,* 26448-26460.

[85]Dehez, F.; Tarek, M.; Chipot, C. *J. Phys. Chem. B* **2007,** *111,* 10633-10635.

[86]Reblova, K.; Spackova, N.; Koca, J.; Leontis, N. B.; Sponer, J. *Biophys. J.* **2004,** *87,* 3397-3412.

[87]Fadrna, E.; Spackova, N.; Stee, R.; Koca, J.; Cheatham III, T. E.; Sponer, J. *Biophys. J.* **2004,** *87,* 227-242.

[88]Spackova, N.; Cubero, E.; Sponer, J.; Orozco, M. *J. Am. Chem. Soc.* **2004,** *126,* 14642-14650.

[89]Stefl, R.; Cheatham III, T. E.; Spackova, N.; Fadrna, E.; Berger, I.; Koca, J.; Sponer, J. *Biophys. J.* **2003,** *85,* 1787-1804.

[90]Stefl, R.; Spackova, N.; Berger, I.; Koca, J.; Sponer, J. *Biophys. J.* **2001,** *80,* 455-468.

[91]Strahan, G. D.; Keniry, M. A.; Shafer, R. H. *Biophys. J.* **1998,** *75,* 968-981.

[92]Spackova, N.; Berger, I.; Sponer, J. *J. Am. Chem. Soc.* **1999,** *121,* 5519-5534.

[93]Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham III, T. E. *Acc. Chem. Res.* **2000,** *33,* 889-897.

[94]Srinivasan, J.; Cheatham III, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998,** *120,* 9401-9409.

[95]Spackova, N.; Berger, I.; Sponer, J. *J. Am. Chem. Soc.* **2001,** *123,* 3295-3307.

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

[97]Cang, X.; Sponer, J.; Cheatham III, T. E. *J. Am. Chem. Soc.* **2011,** *133,* 14270-14279.

[98]Chowdhury, S.; Bansal, M. *J. Biomol. Struct. Dyn.* **2000,** *18,* 11-28.

[99]Cavallari, M.; Calzolari, A.; Garbesi, A.; Di Felice, R.; *J. Phys. Chem. B* **2006,** *110,* 26337-26348.

[100]Sponer, J.; Cang, X.; Cheatham III, T. E. *Methods* **2012,** *57,* 25-39.

[101]Mohan, V.; Smith, P. E.; Pettitt, B. M. *J. Phys. Chem. B* **1993,** *97,* 12984-12990.

[102]Hamelberg, D.; Williams, L. D.; Wilson, W. D. *J. Am. Chem. Soc.* **2001,** *123,* 7745-7755.

[103]McConnell, K. J.; Beveridge, D. *J. Mol. Biol.* **2001,** *314,* 23-40.

[104]Mocci, F.; Saba, G. *Biopolymers* **2003,** *68,* 471-485.

[105]Rueda, M.; Cubero, E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2004,** *87,* 800-811.

[106]Cheng, Y.; Korolev, N.; Nordenskiold, L. *Nucleic Acids Res*. **2006,** *34,* 686-696.

[107]Feig, M.; Pettitt, B. M. *Biophys. J.* **1999,** *77,* 1769-1781.

[108]Long, H.; Kudlay, A.; Schatz, G. C. *J. Phys. Chem. B* **2006,** *110,* 2918-2926.

[109]Mocci, F.; Laaksonen, A.; Lyubartsev, A.; Saba, G. *J. Phys. Chem. B* **2004,** *108,* 16295-16302.

[110]Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L. *Proc. Natl. Acad. Sci. USA* **2004,** *101,* 14771-14775.

[111]Hamelberg, D.; McFail-Isom, L.; Williams, L. D.; Wilson, W. D. *J. Am. Chem. Soc.* **2000**, *122,* 10513-10520.

[112]Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids. 1 ed.;* Oxford University Press: New York, **1987**.

[113]Frenkel, D.; Smith, B. *Understanding Molecular Simulation: from Algorithms to Applications;* Academic Press: San Diego, **1996**.

[114]Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957,** *27,* 1208-1209.

[115]Rahman. A. *Phys. Rev.* **1964,** *136,* A405-A411.

[116]Rahman, A.; Stillinger, F. H. *Phys. Rev. A* **1974,** *10,* 368-378.

[117]Hockney, R. W.; Eastwood, J. W. *Computer Simulations Using Particles;* McGraw-Hill: New York, **1981**.

[118]Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992,** *97,* 1990-2001.

[119]Davankov, V. A. *Pure Appl. Chem.* **1997,** *69,* 1469-1474.

[120]Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Chem.* **1977,** *23,* 327-341.

[121]Andersen, H. C. *J. Comput. Chem.* **1983,** *52,* 24-34.

[122]Ryckaert, J. P.; Bellemand, A. *Chem. Phys. Lett.* **1975,** *30,* 123-125.

[123]Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996,** *118,* 11225-11236.

[124]MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Ngyyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.;

Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998,** *102,* 3586-3616.

[125]Karger, N.; Vargad, T.; Ludemann, H. D. *J. Chem. Phys.* **1990,** *93,* 3437-3444.

[126]Born, M.; Von Karman, T. *Phys. Z* **1912,** *13,* 297-309.

[127]Ewald, P. P. *Ann. Phys-Berlin.* **1921,** *64,* 253-287.

[128]Domene, C.; Furini, S. *Methods Enzymol.* **2009,** *466,* 155-177.

[129]Born, M. *Z Phys.* **1920,** *1,* 45-48.

[130]Kirkwood, J. G. *J. Chem. Phys.* **1935,** *3,* 300-313.

[131]Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989,** *156,* 472-477.

[132]van Gunsteren, W. F. *Methods for Calculation of Free Energies and Binding Constants: Successes and Problems*. In: van GunsterenWF, Weiner PK, eds. *Computer Simulation of Biomolecular Systems*. Vol. 1 Leiden: ESCOM; **1989,** 27.

[133]Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1991,** *95,* 1175-1188.

[134]Sprik, M.; Ciccotti, G. *J. Chem. Phys.* **1998,** *109,* 7737-7744.

[135]Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001,** *115,* 9169-9183.

[136]Henin, J.; Chipot, C. *J. Chem. Phys.* **2004,** *121,* 2904-2914.

[137]Gullingsrud, J. R.; Braun, R.; Schulten, K. *J. Comp. Phys.* **1999,** *151,* 190-211.

[138]Grubmuller, H.; Heymann, B.; Tavan, P. *Science* **1996,** *271,* 997-999.

[139]Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997,** *72,* 1568-1581.

[140]Evans, E.; Ritchie, K. *Biophys. J.* **1997,** *72,* 1541-1555.

[141]Balsera, M.; Stepaniants, S.; Izrailev, S.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *73,* 1281-1287.

[142]Beveridge, D. L.; Dicapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989,** *18,* 431-492.

[143]Mitchell, M. J.; McCammon, J. A. *J. Comput. Chem.* **1991,** *12,* 271-275.

[144]Zwanzig, R. W. *J. Chem. Phys.* **1954,** *22,* 1420-1426.

[145]Kastner, J. *WIREs Computational Molecular Science;* John Wiley & Sons, Ltd. Volume 1, **2011**.

[146]Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. USA* **2002,** *99,* 12562-12566.

[147]Chandler, D. *Introduction to Modern Statistical Mechanics;* Oxford University Press: Oxford, **1987**.

[148]Leimkuhler, B.; Chipot, C.; Elber, R.; Laaksonen, A.; Mark, A.; Schlick, T.; Schutte, C.; Skeel, R. *Comput. Sci. Eng.* **2006**, *49*, 185-211.

[149]Knight, J. L.; Brooks III, C. L. *J. Comput. Chem.* **2009,** *30,* 1692-1700.

[150]Kong, X.; Brooks III, C. L. *J. Chem. Phys.* **1996,** *105,* 2414-2423.

[151]Liu, Z.; Berne, B. J. *J. Chem. Phys.* **1993,** *99,* 6071-6077.

[152]Tidor, B. *J. Phys. Chem.* **1993,** *97,* 1069-1073.

[153]Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992,** *13,* 1011-1021.

[154]Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001,** *135,* 40-57.

[155]Grossfield, A. WHAM: the Weighted Histogram Analysis Method, http://membrane.urmc.rochester.edu/content/wham. December 2012.

[156]Kastner, J.; Thiel, W. *J. Chem. Phys.* **2006,** *124,* 234106 (1-7).

[157]Kastner, J.; Thiel, W. *J. Chem. Phys.* **2005,** *123,* 144104 (1-5).

[158]Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974,** *28,* 578-581.

[159]Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977,** *23,* 187-199.

[160]Neidle, S.; Parkinson, G. N.; *Curr. Opin. Struct. Biol.* **2003,** *13,* 275-283.

[161]Davis, J. T. *Angew. Chem.* **2004,** *116,* 684-716.

[162]Hud, N. V.; Plavec, J. *Quadruplex Nucleic Acids* (Eds.: Neidle, S.; Balasubramanian, S.), The Royal Society of Chemistry, Cambridge, UK, **2006,** 100-130.

[163]Engelhart, A. E.; Plavec, J.; Persil, O.; Hud, N. V. *Nucleic Acid—Metal Ion Interactions* (Ed.: Hud, N. V.), RSC Publishing, London, **2009,** 118-153.

[164]Sket, P.; Crnugelj, M.; Kozminski, W.; Plavec, J. *Org. Biomol. Chem.* **2004,** *2,* 1970-1973.

[165]Sket, P.; Crnugelj, M.; Plavec, J. *Nucleic Acids Res.* **2005,** *33,* 3691-3697.

[166]Sket, P.; Plavec, J. *J. Am. Chem. Soc.* **2007,** *129,* 8794-8800.

[167]Podbevsek, P.; Hud, N. V.; Plavec, J. *Nucleic Acids Res.* **2007,** *35,* 2554-2563.

[168]Podbevsek, P.; Sket, P.; Plavec, J. *J. Am. Chem. Soc.* **2008,** *130,* 14287-14293.

[169]Sket, P.; Plavec, J. *J. Am. Chem. Soc.* **2010,** *132,* 12724-12732.

[170]Zavasnik, J.; Podbevsek, P.; Plavec, J. *Biochemistry* **2011,** *50,* 4155-4161.

[171]Snoussi, K.; Halle, B. *Biochemistry* **2008,** *47,* 12219-12229.

[172]Gill, M. L.; Strobel, S. A.; Loria, J. P. *J. Am. Chem. Soc.* **2005,** *127,* 16723-16732.

[173]Kwan, I. C. M.; Wong, A.; She, Y. M.; Smith, M. E.; Wu, G. *Chem. Commun.* **2008,** *6,* 682-684.

[174]Wu, G.; Gan, Z.; Kwan, I. C. M.; Fettinger, J. C.; Davis, J. T. *J. Am. Chem. Soc.* **2011,** *133,* 19570-19573.

[175]Read, M. A.; Neidle, S. *Biochemistry* **2000,** *39,* 13422-13432.

[176]Chowdhury, S.; Bansal, M. *J. Phys. Chem. B* **2001,** *105,* 7572-7578.

[177]Sponer, J.; Spackova, N. *Methods* **2007,** *43,* 278-290.

[178]Haider, S.; Parkinson, G. N.; Neidle, S. *Biophys. J.* **2008,** *95,* 296-311.

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

[180]Aqvist, J.; Luzhkov, V. *Nature* **2000,** *404,* 881-884.

[181]Berneche, S.; Roux, B. *Proc. Natl. Acad. Sci. USA* **2003,** *100,* 8644-8648.

[182]Noskov, S. Y.; Berneche, S.; Roux, B. *Nature* **2004,** *431,* 830-834.

[183]Roux, B. *Curr. Opin. Struct. Biol.* **2002,** *12,* 881-884.

[184]Martino, L.; Virno, A.; Pagano, B.; Virgilio, A.; Micco, S. D.; Galeone, A.; Giancola, C.; Bifulco, G.; Mayol, L.; Randazzo, A. *J. Am. Chem. Soc.* **2007**, *129,* 16048-16056.

[185]Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005,** *26,* 1781-1802.

[186]Foloppe, N.; MacKerell Jr, A. D. *J. Comput. Chem.* **2000,** *21,* 86-104.

[187]MacKerell Jr, A. D.; Banavali, N. K. *J. Comput. Chem.* **2000,** *21,* 105-120.

[188]Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33-38.

[189]Eyring, H. *J. Chem. Phys.* **1935,** *3,* 107-115.

[190]Ansell, S.; Barnes, A. C.; Mason, P. E.; Neilson, G. W.; Ramos, S. *Biophys. Chem.* **2006,** *124,* 171-179.

[191]Soper, A. K.; Weckstrom, K. *Biophys. Chem.* **2006.** *124,* 180-191.

[192]Grossfield, A.; Ren, P.; Ponder, J. W. *J. Am. Chem. Soc.* **2003,** *125,* 15671-15682.

[193]Palinkas, G.; Radnai, T.; Hajdu, F. *Z. Naturforsch. Teil A* **1980,** *35,* 107-114.

[194]Palinkas, G.; Radnai, T.; Szasz, Gy. I.; Heinzinger, K. *J. Chem. Phys.* **1981,** *74,* 3522-3526.

[195]Dang, L. X. *Chem. Phys. Lett.* **1993,** *213,* 541-546.

[196]Gu, J.; Leszczynski, J. *J. Phys. Chem. A* **2000,** *104,* 6308-6313.

[197]Webba da Silva, M. *Chem. Eur. J.* **2007,** *13,* 9738-9745.

[198]Cang, X.; Sponer, J.; Cheatham III, T. E. *Nucleic Acids Res.* **2011,** *39,* 4499-4512.

[199]Laughlan, G.; Murchie, A. I. H.; Norman, D. G.; Moore, M. H.; Moody, P. C. E.; Lilley, D. M. J.; Luisi, B. *Science* **1994,** *265,* 520-524.

[200]Phillips, K.; Dauter, Z.; Murchie, A. I. H.; Lilley, D. M. J.; Luisi, B. *J. Mol. Biol.* **1997,** *273,* 171-182.

[201]Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002,** *417,* 876-880.

[202]Mills, M.; Andricioaei, I. *J. Chem. Phys.* **2008,** *129,* 114101 (1-13).

[203]Marlow, A. L.; Mezzina, E.; Spada, G. P.; Masiero, S.; Davis, J. T.; Gottarelli, G. *J. Org. Chem.* **1999**, *64,* 5116-5123.

[204]Kwan, I. C. M.; She, Y.M.; Wu, G. *Chem. Commun*. **2007**, *41,* 4286-4288.