COMPUTATIONAL STUDIES ON BIOLOGICALLY IMPORTANT MACROMOLECULES

Thesis submitted for the degree of
Doctor of Philosophy (Science)
in
Chemistry

by

ANGANA RAY

Department of Chemistry
University of Calcutta
2016

Dedicated
To My Late Father

Whose blessings have been
my support all through

Acknowledgements

So many people have encouraged and supported me throughout my research and during writing of this thesis. I would like to acknowledge their contribution by mentioning their names.

To begin with, I would like to express my earnest gratitude to my Ph.D. supervisor Prof. Dhananjay Bhattacharyya, for his guidance and constant support at each step of my research work. He not only provided me valuable suggestions from time to time but also worked alongside me with the same vigour. He has meticulously and with great care and attention, advised me in completing my Ph.D. work.

I am thankful to the ex-Directors, Prof. Milan K. Sanyal and Prof. Bikash Chakraborty; and the current Director, Prof. Ajit Kumar Mohanty, of Saha Institute of Nuclear Physics (SINP), for providing me with the amenities required during my research tenure. I am also thankful to the current Registrar, Mr. Anirban Banerjee, of SINP for his timely assistance.

I formally acknowledge Department of Atomic Energy, Govt. of India for my fellowship and funding for research work from projects viz.; Chemical and Biophysical Approaches for Understanding of Natural Processes (CBAUNP), Molecular Mechanism of Diseases and Drug Action (MMDDA), and Biomolecular Assembly, Recognition and Dynamics (BARD). I also acknowledge the computation facilities of Centre for Applied Mathematics and Computational Science (CAMCS). I also acknowledge the former Biophysics Division (currently Biophysics and Structural Genomics Division) of SINP where I worked during my initial days. Computational Science Division, SINP has been a fantastic workplace. Without the facilities of this division, performing high-end computation and storage of huge data would have been a tough task. I am grateful to all the members: Sumit Basu, Soumya Majumdar, Nandalal Sanpui, Gautam Datta, Deeptish Dey and Gautam Garai, of Computational Science Division for extending their cooperation. I am also thankful to the non-academic staff members of SINP, especially members of Accounts section, CARE-facility, Library and Canteen for their timely assistance.

I am thankful to my collaborators Prof. Sangam Banerjee of Surface Physics and Material Science Division and Prof. Kaushik Sengupta of Biophysics and Structural Genomics Division, of SINP; and Debarati De of Department of Chemistry, Vidyasagar College . I appreciate the support from Manindra Bera, Kaushik Bagani, Ankita Agarwal and Smruti Ranjan Sahoo. All of them have worked with me on different collaborative projects.

I am thankful to all my teachers of Post-M.Sc. and M.Sc. courses especially Prof. Dipak Dasgupta, Prof. Samita Basu, Prof. Rahul Banerjee and Prof. Munna Sarkar of SINP and Prof. Kamal Bhattacharyya, Prof. Debasis Mukhopadhyay, Prof. Swapan Chakraborti and Prof. Pinaki Chaudhury of Chemistry (Physical Chemistry) Department, University of Calcutta. I would specially mention Prof. Chaitali Mukhopadhyay and Prof. Tanushree Sanyal (Bala), Chemistry (Physical

Chemistry) Department, University of Calcutta for their constant motivation to pursue research and friendly assistance during my M.Sc. days.

The contributions of my colleagues and acquaintances at SINP have been invaluable. I am particularly obliged to my seniors, Swati di, Sukanya di, Sangeeta di and Manas da for mentoring me during my initial days of learning. They have always provided me valuable scientific inputs required for my work. I am grateful to Sanchita di for stepping up as an elder sister and providing valuable suggestions whenever I needed. A special acknowledgement note goes to my junior Debasish Mukherjee for his help and cooperation. I should mention that he has painstakingly done proof-reading of this thesis and helped me in preparing some of the images provided here. This acknowledgement would be incomplete without the mention of loving support, cooperation and encouragement received from my dear friends at SINP. I express my heartiest thanks to Amrita di, Shreyashi di, Avinanda, Pritha, Piyali, Banabithi, Subhas, Supratim and Satyabrata. I am thankful to Debashree for all the academic as well as much needed non-academic discussions we had. I am also thankful to Anupa for always being there and supporting me through my ups and downs at SINP. I am also thankful to my friends Sudipta Nandi, Soumavo Ghosh, Anindita Das, Antarip Halder and Srijeeta Talukder for their support.

My Ph.D. and this thesis would not have been possible without Dr. Nabakumar Bera, a friend, philosopher and guide. It was his constant motivation that I pursued higher academics. The appreciation and gratitude I owe to my family can never be expressed in words. I am thankful to my uncles Samir Saha, Rabindranath Saha and Ratan Saha for being supportive as a father figure, and my brother Arjun Ray for bearing all my tantrums. My mother Iti Ray has always showered her unconditional love and support that kept me going even in the most difficult times. Thank you for patiently bearing with my unreasonable expectations. I could never have achieved what I have without you.

I would like to conclude this acknowledgement by thanking all the authors, whose work I have cited in this thesis. I have been able to begin this pursuit only because of the ground that was already covered ahead of me.

Date:

_____

Angana Ray

List of Abbreviations:

A, Ade	Adenine
G, Gua	Guanine
C, Cyt	Cytosine
T, Thy	Thymine
U, Ura	Uracil
R, Pur	Purine
Y, Pyr	Pyrimidine
":"	Represents base pairing
"/" or "::"	Represents base pair stacking
"^"	Represents discontinuity in strand
H-bond	Hydrogen bond
bp	base pair
WC	Watson-Crick
W:W	Watson-Crick:Watson Crick base pair
W:W C	Watson-Crick:Watson Crick *Cis* base pair
W:W T	Watson-Crick:Watson Crick *Trans* base pair
W:H or H:W	Watson-Crick:Hoogsteen or Hoogsteen:Watson-Crick base pair
fs	femtosecond
ns	nanosecond
Å	Angstrom
2D	Two Dimension
3D	Three Dimension
MD	Molecular Dynamics
SMD	Steered Molecular Dynamics
RMSD	Root-mean-square Deviation
RMSF	Root-mean-square Fluctuation
QM	Quantum Mechanics
DFT	Density Functional Theory
GO	Graphene Oxide
BDBPZ	11-benzoyl-dibenzo[a,c]phenazine
NAD	Nicotinamide adenine dinucleotide

Table of Contents

Chapter I
Section I

Nucleic Acids
Structure, Properties and Biological Role

1.1 Nucleic Acids

Nucleic Acids are amongst the key biological macromolecules which are involved in a variety of processes like encoding, transmitting and expressing genetic information for all known forms of life. The nucleic acids were first discovered by Friedrich Miescher in1869 (Dahm, 2008); however, even at present major biological and medical research is being directed towards experimental and theoretic studies of these biopolymers. Nucleic acids were named by Richard Altman for their initial discovery inside the nucleus and for the presence of phosphate groups (Dahm, 2008). There are two types of nucleic acids: deoxyribonucleic acid, better known as DNA and ribonucleic acid, better known as RNA. The evidence that DNA, and no other molecule, transmits genetic information was initially provided by Avery, Macleod and McCarty in 1944 (Avery *et al.*, 1944) and later by Hershey and Chase in 1952 (Hershey & Chase, 1952). The double helical DNA structure was eventually revealed by Watson and Crick in 1953 (Watson & Crick, 1953), using X-ray fiber diffraction patterns generated by Franklin, Wilkins and associates (Franklin and Gosling, 1953; Wilkins *et al.*, 1953) along with the chemical evidence of complimentary bases provided by Chargaff (Chargaff *et al.*, 1950). The Meselson–Stahl experiment also called "The most beautiful experiment in biology" supported the hypothesis that DNA replication was semi-conservative (Meselson *et al.*, 1957). Unlike DNA, RNA is more often found in nature as a single-strand folded onto itself. In the central dogma of molecular biology, DNA→RNA→Protein (Crick, 1970), RNA was observed only as a passive transporter of genetic information, but discoveries in the past few years appreciating the numerous roles played by various functional and structured RNAs in many cellular processes ranging from protein synthesis to protein folding to gene expression and regulation have established their huge significance in the biological system (Nissen, 2000; Cech, 1990; Bartel, 2004; Tucker and Breaker, 2005; Samanta *et al.*, 2008). Although, until fairly recently less was known about RNA structure at atomic resolution in comparison to DNA, but this has changed with the advent of 'RNA era'.

1.1.1 Building Blocks of Nucleic Acids

Chemical degradation studies in the early years of twenty-first century on materials extracted from cell nuclei established that the high molecular-weight "nucleic acid" was actually composed of individual acid units, termed nucleotides **(Figure 1.1.1)**. These could be further cleaved to phosphate groups and distinct nucleosides. Nucleosides were subsequently identified as consisting of a sugar (D-(-) ribose or D-(-)-2'-deoxyribose) and a nitrogen-containing base (nucleobase). Thus, each repeating unit in a nucleic acid polymer comprises of a phosphate group, a sugar, and a nucleobase, all linked together by covalent bonds. The planar aromatic heterocyclic nitrogenous bases are of two types: pyrimidine (Pyr/Y: thymine (T) or cytosine (C) or uracil (U)) and purine (Pur/R: adenine (A) or guanine (G)) (Bloomfield *et al.*, 2000) **(Figure 1.1.2)**. In DNA the four bases: A, T, G and C are found while U instead of T is generally present in RNA, along with the three other bases. The sugar present in RNA is a ribose – an aldopentose in cyclic furanoside form **(Figure 1.1.1)** whereas, the sugar in DNA is 2-deoxy ribose, where the 2'-OH group of ribo-furanose ring is replaced by H-atom. A nucleobase is connected with one ring nitrogen to the anomeric center of the sugar (C1'), forming a nucleoside-adenosine, guanosine, cytidine, thymidine and uridine. In the case of the pyrimidines, the connecting nitrogen is N1 and all purines are connected via N9. The resulting linkage is called the β-glycosyl bond as in natural nucleic acids this glycosidic bond is always in β-orientation, *i.e.* the nucleobase is above the plane of the sugar when viewed onto the plane (*i.e.* on the same face of the plane as the 5' hydroxyl substituent). When nucleotides are polymerized into nucleic acid chain (polynucleotide chain) by chemical removal of water molecules, a sugar-phosphate backbone is formed. These sugar and phosphate groups are conserved throughout the nucleic acid backbone, while the bases incorporate the polymeric variations.

The C3'-hydroxyl group of the n^{th} nucleotide is joined to C5'-hydroxyl group of the $(n+1)^{th}$ nucleotide by a phosphodiester linkage forming the nucleic acid. A nucleic acid molecule with (n) number of nucleotides would have 4^n possible different arrangement of the four nucleotides. In addition to the four usual nucleotides, different functional RNA molecules like tRNAs, rRNAs, riboswitches, *etc*. often contain several other chemically modified bases, such as, Dihydrouridine, Pseudouridine, N,N-dimethylguanine, Inosine, *etc*. These are formed by different

post-transcriptional modification mechanisms, among which Pseudouridylation, 2′O-methylation and base methylation are the most abundant internal modifications (Auffinger and Westhof, 1998; Björk, 1995; Cavaillé and Bachellerie, 1998; Grosjean *et al.*, 1995; Hopper and Phizicky, 2003; Maden, 1990; Sprinzl *et al.*, 1998; Zhao and Yu, 2004). In DNA the most common modified base is 5-methycytosine (m^5C). Hypoxanthine (produced from adenine) and xanthine (produced from guanine) are two of the many modified bases created through mutagen presence (Nguyen *et al.*, 1992). Some very rare modified bases found in DNA are α-putrescinylthymine and 5-digydroxypentyluracil (Warren, 1980).

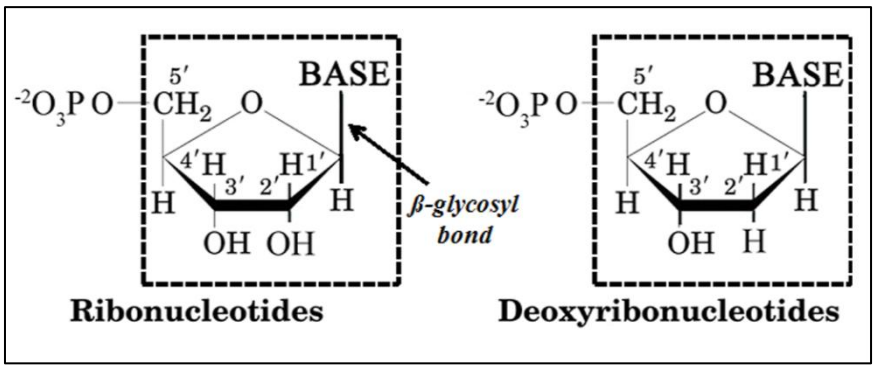


Figure 1.1.1: From Left to Right, the Ribonucleotides and Deoxyribonucleotides consisting of phosphate groups and the nucleosides (inside box). The nucleosides are formed by the base and sugar.

1.1.2 Sugar-Phosphate Backbone of Nucleic Acids

1.1.2.1 Phosphate Backbone

An oligonucleotide chain is formed by linking the 5'P of one nucleotide to the 3'OH of the neighbour nucleotide resulting in phosphodiester bonds. An important consequence of a phosphodiester linkage is that DNA and RNA molecules become directional- one end of the chain has free phosphate group, and the other contains a free –OH group **(Figure 1.1.3)**. Since the sugar-phosphate backbones are the same for every nucleotide, a nucleic acid molecule can simply be represented by sequence of the nucleobases from 5' to 3' direction. The bases A, T/U, G and C are synonymously used with their respective nucleotides, with the understanding that it is a convenient way to represent a complicated structure. For example, the sequence ApGpCpTpTpG has the 5' terminal adenosine nucleoside, with a free hydroxyl at its 5'-position and

the 3' end guanosine has a terminal 3'-hydroxyl group. Here "p" denotes the phosphate groups. Due to the connection of the nucleotides in DNA and RNA via phosphodiesters, each connecting unit bears a negative charge making the nucleic acids polyanionic. This imparts an inherent polarity to the DNA chain, which is important to keep the macromolecule soluble in water. It however also causes an enormous coulombic repulsion when two single strands come together to form a double strand. Thus, for the formation of soluble double strands metal ions closely associate with the polyanions to compensate the charge and establish electrical neutrality (Carell, 2004; Manning, 1978).

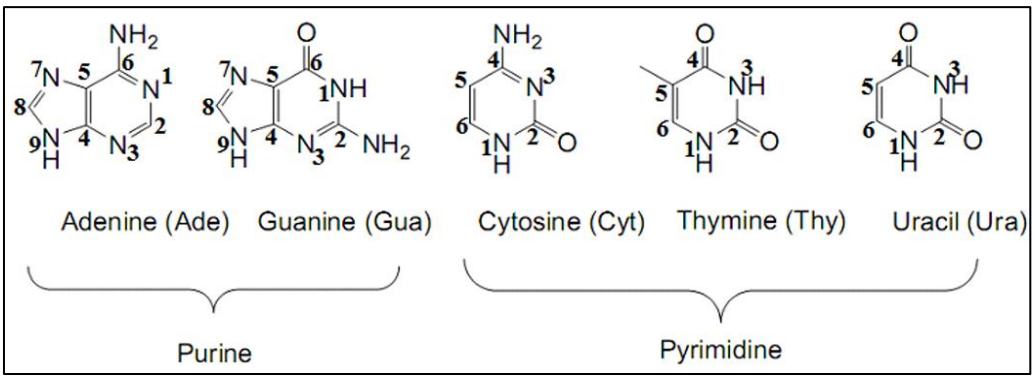


Figure 1.1.2: The nucleobases (Purine and Pyrimidine).

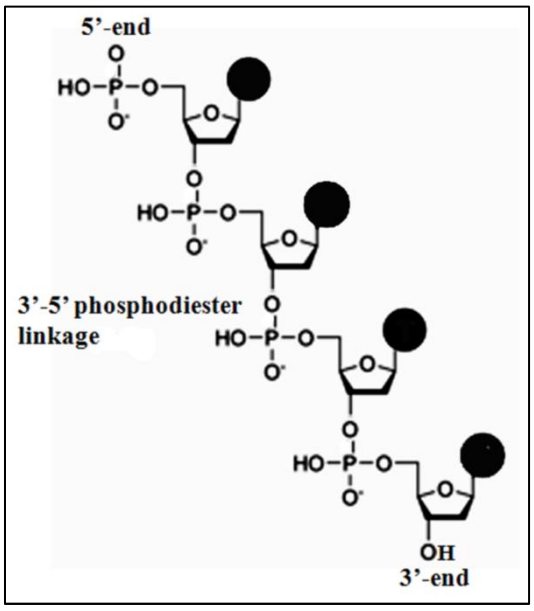


Figure 1.1.3: An oligonucleotide chain containing the phosphodiester linkage. The 5'-end contains free phosphate group and the 3'-end contains free hydroxyl group.

In molecules with rotational freedom about single bonds, all feasible torsion angles are not assumed but only certain sterically allowed conformations are

preferred. Hence, a structure can be described with torsion angle ranges. There exist six single covalent bonds in the nucleic acid backbone around which rotations are possible. These flexible bonds are P–O5', O5'–C5', C5'–C4', C4'–C3', C3'–O3' and O3'–P. The conformations of the sugar-phosphate backbone following the sequential numbering of atoms $P \rightarrow O_{5'} \rightarrow C_{5'} \rightarrow C_{4'}$ *etc.* is defined by torsion angles α, β, γ, δ, ε, ζ in alphabetical order following the IUPAC-IUB Nomenclature of 1983 (1983) (Dickerson, 1989) **(Figure 1.1.4)**. These phosphodiester backbone torsion angles are restricted to sterically allowed regions (Gorin *et al.,* 1995; Subirana and Faria, 1997) **(Figure 1.1.5a)** which are characteristic of different helical structures. The notations *gauche*$^+$ (g^+), *trans* (*t*) and *gauche*$^-$ (g^-) refer to dihedral angles corresponding to staggered conformations around 60º, 180º and -60º/300º, respectively (Saenger, 1984; Lavery, 2005) **(Figure 1.1.5a)**. An additional dihedral χ is defined around the glycosidic bonds C1'–N9 for purines and C1'–N1 for pyrimidines **(Figure 1.1.4)**. This dihedral represents the rotation of the nucleobase with respect to the sugar ring. If the O2 of pyrimidine or N3 of purine and O4' of sugar ring are positioned close to the bulky side of the bases, the form is called *syn* and if they are on opposite sides, it is called *anti* **(Figure 1.1.5b)**. In addition to these torsion angles, the endo-cyclic torsions about the single covalent bonds of the cyclic sugar also adopt various values **(Figure 1.1.4)**. Thus, the RNA (or DNA) backbone has eight degrees of freedom which makes RNA modeling, structure building and prediction a multidimensional problem of high complexity. In order to reduce the dimensionality of the RNA backbone in a physically reasonable way, a reduced representation of the nucleic acid backbone in terms of two pseudotorsion parametres was developed by Duarte and Pyle in 1998. The two pseudotorsion angles η and θ were defined as the torsions of $C4'_{i-1}$-P_i-$C4'_i$-P_{i+1} and P_i-$C4'_i$-P_{i+1}-$C4'_{i+1}$ respectively (Duarte and Pyle, 1998) **(Figure 1.1.4)**. The standard definitions of all the backbone dihedrals and pseudotorsion angles in a nucleic acid double helix are given in **Table 1.1.1.**

In a double helix, backbone dihedrals are often correlated and their torsional degrees of freedom are not independent of each other. Instances of a crankshaft motion leading to concerted alterations in α-γ dihedrals can often be observed (Srinivasan *et al.,* 1987). Correlated variations of ε and ζ resulting in polymorphism in B-DNA structure have also been observed (Becker and Wang, 1989; Clark *et al.*, 2000; Gupta *et al.,* 1980; Hartmann *et al.,* 1993; Schneider *et al.,* 1997). Moreover, the backbone shows highly correlated motions of χ, δ and ε where the torsional

degrees of freedoms are also related to the local motion of the base pairs (Beveridge *et al.*, 2004; Dixit *et al.*, 2005; El Hassan and Calladine, 1997; Packer and Hunter, 1998).

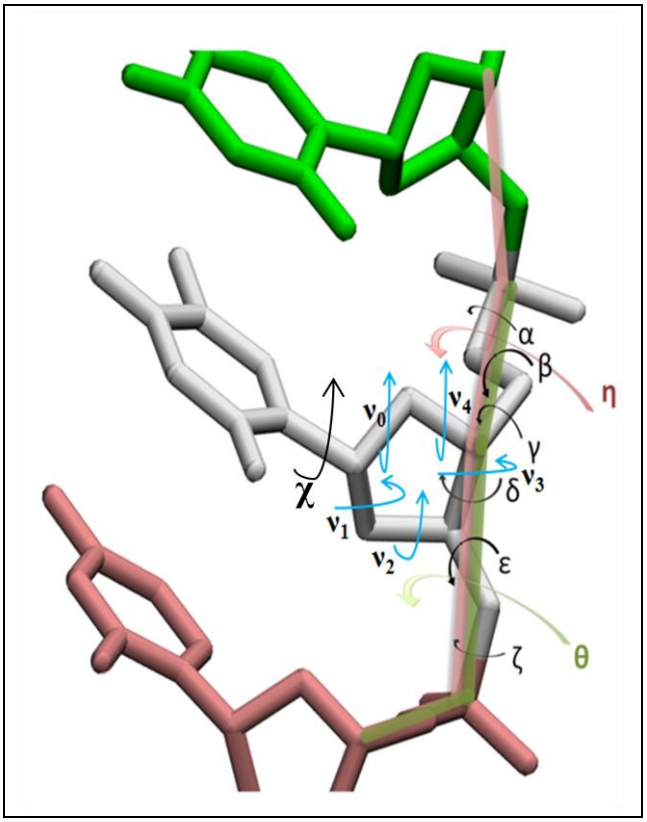
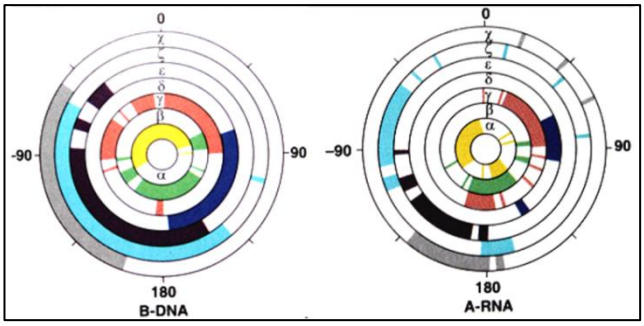


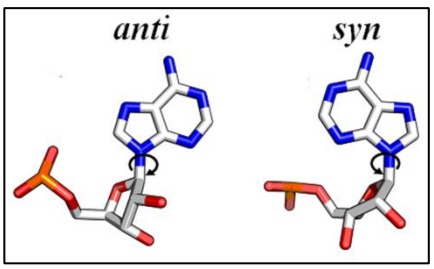
Figure 1.1.4: The conformations of the sugar-phosphate backbone as given by IUPAC.

Torsion angle	Atoms involved
α	$_{(n-1)}$O3'–P–O5'–C5'
β	P–O5'–C5'–C4'
γ	O5'–C5' –C4'–C3'
δ	C5'–C4'–C3'–O3'
ε	C4'–C3'–O3'–P
ζ	C3'–O3'–P–O5'$_{(n+1)}$
χ	O4'–C1'–N1–C2 (Y) O4'–C1'–N9–C4 (R)
ν_0	C4'–O4'–C1'–C2'
ν_1	O4'–C1' – C2'–C3'
ν_2	C1'–C2' – C3'–C4'
ν_3	C2'–C3' – C4'–O4'
ν_4	C3'–C4'–O4'–C1'
η	C4'$_{n-1}$-P$_n$-C4'$_n$-P$_{n+1}$
θ	P$_n$-C4'$_n$-P$_{n+1}$-C4'$_{n+1}$

Table 1.1.1: Details of the backbone torsion angles.



(a)



(b)

Figure 1.1.5: (a) The allowed regions of phosphodiester backbone torsion angles in B-DNA and A-RNA (Left to Right). (b) The *anti* and *syn* orientations of glycosidic bond.

1.1.2.2 Sugar Conformations

The five-membered furanose ring systems in DNA/RNA are never planar. Instead they are puckered in either envelope (E) or twist (T) forms **(Figure 1.1.6a)**. The ring puckering arises from the effect of non-bonded interactions between substituent groups at the four carbon atoms. When atoms are displaced from these three- or four-atom planes of the ring on the same side as C5', the structure attains *endo* conformation; and displacement on the opposite side forms *exo* conformation (Saenger, 1984). Given the chemical environment within nucleic acid structures, two puckers turn out to be most stable – *C2'-endo* in DNA and *C3'-endo* in RNA **(Figure 1.1.6b)**, *i.e.* C2' and C3' are the atoms out-of-palne in the respective cases. C2'-endo is not possible in RNA double helix because of the 2'-hydroxyl group in ribose moiety. The *O4'-endo* is also observed sometimes; but the *O4'-exo* is never observed in nucleic acid structures. The furanose ring conformation is expressed in terms of five endo-cyclic torsion angles ν_0, ν_1, ν_2, ν_3 and ν_4 **(Table 1.1.1)**. The sugar geometry is also represented by pseudo-rotation phase angle *P* and the amplitude of puckering ν_{max}, where *P* and ν_{max} are given by the following equations (Altona and Sundaralingam, 1972; Schlick, 2010):

$$\boldsymbol{\nu_{max} = \frac{\nu_2}{cosP}} \qquad \text{… Eqn. 1.1}$$

and, $$\boldsymbol{tan\ P = \frac{(\nu_4+\nu_1)-(\nu_3+\nu_0)}{2*\nu_2*(sin36°+sin72°)}} \qquad \text{…Eqn. 1.2}$$

All the endocyclic torsion angles are related to *P* and ν_{max} by (Schlick, 2010):

$$\boldsymbol{\nu_i = \nu_{max} cos\,[P + 144°(i-2)]} \qquad \text{…Eqn. 1.3}$$

The puckering is important as it governs the relative orientation of phosphate group to the sugar ring and, hence, the direction of the sugar-phosphate backbone. According to the description by Altona and Sundaralingam (Altona and Sundaralingam, 1972) , the value of phase angle, *P*, can be divided into ten major classes, depending on the orientation of non-planar atom(s). These classes are *C1'-endo* [P = 306º], *C1'-exo* [P = 126º], *C2'-endo* [P = 152º], *C2'-exo* [P = 342º], *C3'-endo* [P = 18º], *C3'-exo* [P = 198º], *C4'-endo* [P = 234º], *C4'-exo* [P = 54º], *O4'-endo* [P = 90º] and *O4'-exo* [P = 270º].

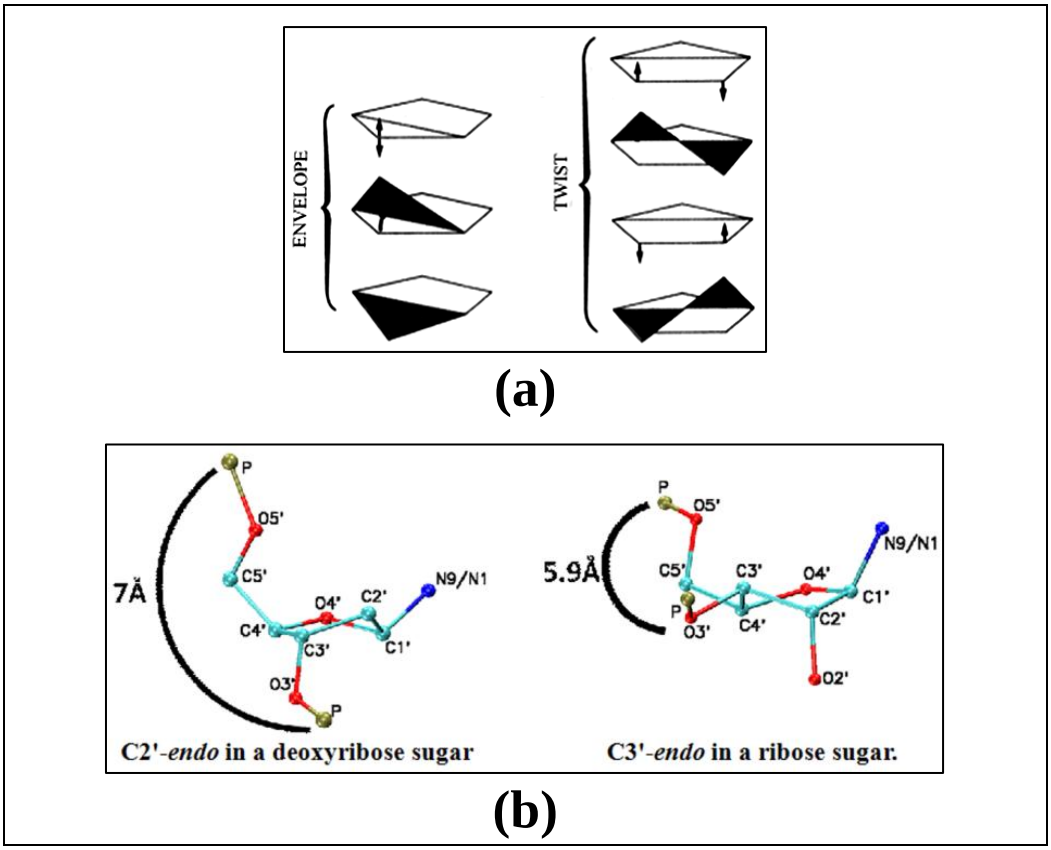


Figure 1.1.6: (a) The envelope (E) or twist (T) forms of the five-membered furanose ring systems in nucleic acids. (b) The C2'-endo and C3'-endo sugar puckers (Left to Right) in DNA and RNA respectively.

1.1.3 Base Pairing Information

1.1.3.1 Structural Characterization of Base Pairs

Base pairs formed by potential hydrogen bond (H-bond) donor and acceptor atoms present at the edges of nucleobases are the building blocks of nucleic acid structures. In case of DNA, the H-bonding between the bases takes place mainly through the formation of standard Watson–Crick (WC) base pairs, *i.e.* Watson-Crick:Watson-Crick (W:W) base pair. But, there may occur base pairs involving H-bonding through highly versatile non-WC base pairing patterns. Thus to further analyze DNA and RNA structures it is important to understand in details the H-bonding between the bases. Purine and pyrimidine bases present three edges for H-bonding: the WC edge, the Hoogsteen edge (purines) or the equivalent "CH" edge (pyrimidines), and the Sugar edge containing the 2'-hydroxyl group of RNA **(Figure 1.1.7)** (Leontis and Westhof, 2001). A given edge of one base can potentially interact

in a plane with any one of the three edges of a second base, and can do so in either the *syn* or *anti* orientation of the glycosidic bonds resulting is *cis* or *trans* base pairs respectively. Thus, twelve distinct edge-to-edge interactions are possible which have been illustrated using the triangle representation for the bases by Leontis and Westhof (Leontis *et al.*, 2002) **(Figure 1.1.8)**. The base pairs are stabilized either through the polar binding region involving N-H….O/N H-bonding interactions (denoted by capital letters, *e.g.*, W, H and S) or through the C-H…O/N type of interactions (non-polar bonding region) which are denoted by small letters (*e.g.*, w, h and s). Thus the base pairs may be classified into the following broad categories:

(a) The canonical base pairs **(Figure 1.1.9a)** that are the WC base pairs where the interacting edges are WC/WC and the glycosidic bond orientation can be *cis* or *trans*. The main WC base pairs are A:T and G:C. The so-called wobble base pairs which are characterized geometrically by a shift of one base relative to the other also belong to this group (Leontis and Westhof, 2001). The four main wobble base pairs are G:U, I:U, I:A and I:C **(Figure 1.1.9b)**. Unlike, in DNA these are found in plenty in RNA. Moreover, the thermodynamic stability of a wobble base pair is comparable to that of a WC base pair. The specific arrangement of WC base pairs forms an isomorphous set with an overall structural similarity. The C1'-C1' distance is similar for both the base pairs as well as their orientation with respect to the sugar phosphate backbones (Saenger, 1984), which allows the DNA/RNA double helices to maintain an overall equivalent conformation irrespective of the base sequence. Nevertheless, they can still adopt different local structures.

(b) The non-canonical base pairs **(Figure 1.1.9c)** that consist of non-WC base pairs involving Hoogsteen and sugar edges of the nucleobases. In addition to the non-WC base pairs formed involving polar binding region, structures of many other base pairs which involve non-polar binding sites, *e.g.*, w:s, w:h, s:s, h:s, *etc.* (Roy *et al.*, 2008) have been optimized. Moreover, some bases can also be protonated from local environmental stress and can form non-canonical pairs with other regular bases (Chawla *et al.*, 2011; Das *et al.*, 2006; Leontis et al., 2002). Various characterization techniques and nomenclatures have been proposed to define the non-canonical base pairing pattern for RNA motifs (Duarte and Pyle, 1998; Gendron *et al.*, 2001; Klosterman *et al.*, 2002; Lee and Gutell, 2004; Lemieux and Major, 2002; Leontis and Westhof, 2001; Leontis *et al.*, 2002; Nagaswamy *et al.*, 2001; Sykes and Levitt, 2005; Walberer *et al.*, 2003).

Structure prediction and modeling of natural or synthetic RNAs have revealed the existence of a number of different non-canonical base pairing arrangements occurring as single, tandem or consecutive base pairs within RNA duplexes (Baeyens *et al.*, 1995; Baeyens *et al.*, 1996; Battiste *et al.*, 1996; Cate *et al.*, 1996a; Lietzke *et al.*, 1996; Pley *et al.*, 1994) Also, the tertiary structures of RNA motifs are often stabilized by such non-canonical interactions resulting in tertiary contacts between different parts of the RNA chain (Butcher and Pyle, 2011; Leontis *et al.*, 2006). Perturbations in regular RNA helices by such non-canonical base pairs are functionally important in adopting unusual structures. Similar non-canonical base pairs are also found in DNA such as in case of G-quadruplex structures.

(c) The bifurcated base pairs involve formally three-centered H-bonds in which two H-atoms point to a single acceptor atom. Thus, they have been observed between the WC edge of one base and one functional group of the second base **(Figure 1.1.9d)**. Water inserted base pairs have also been observed in several high-resolution structures.

In addition to formation of base pairs by interactions between two bases, quite often three or more bases also appear in coplanar orientation within nucleic acid structures forming base triples or quadruples **(Figure 1.1.10a)**. These base triples stack on top of each other giving rise to three-stranded helix while the base quartets give rise to quadruple stranded helix, presumably in the telomere regions of chromosomal DNA. Base triplets are found in several types of RNA structural motifs, *e.g.*, kink-turn motifs, sarcin-ricin loops, tetra loops, A-minor motifs *etc.* At present a considerable research is directed towards finding and classifying the base triples in RNA structures (Abu Almakarem *et al.*, 2012; Lee and Gutell, 2004; Olson *et al.*, 2009; Xin and Olson, 2009). Base quartets can be of *open form* or *close form* **(Figure 1.1.10b)**. The *open forms* are found more often in RNA while the *close form* quartets are dominant in telomeric DNA.

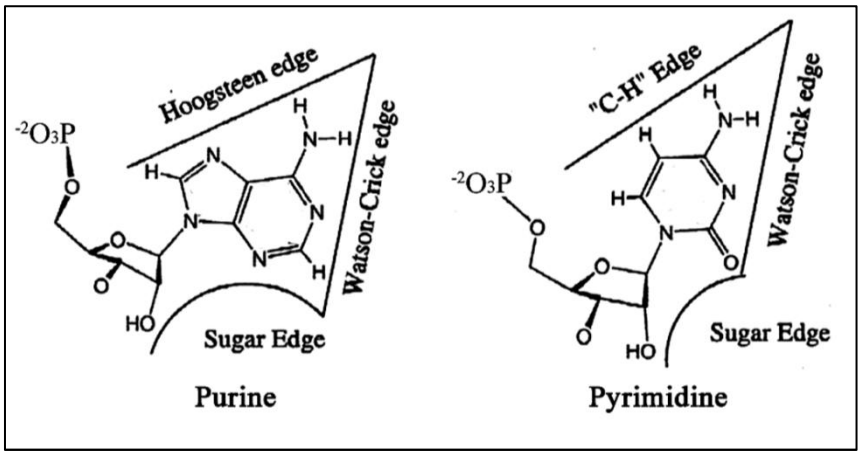


Figure 1.1.7: The three edges for H-bonding: the WC edge, the Hoogsteen edge (in R) or the equivalent "CH" edge (in Y), and the Sugar edge containing the 2'-OH group (in RNA)

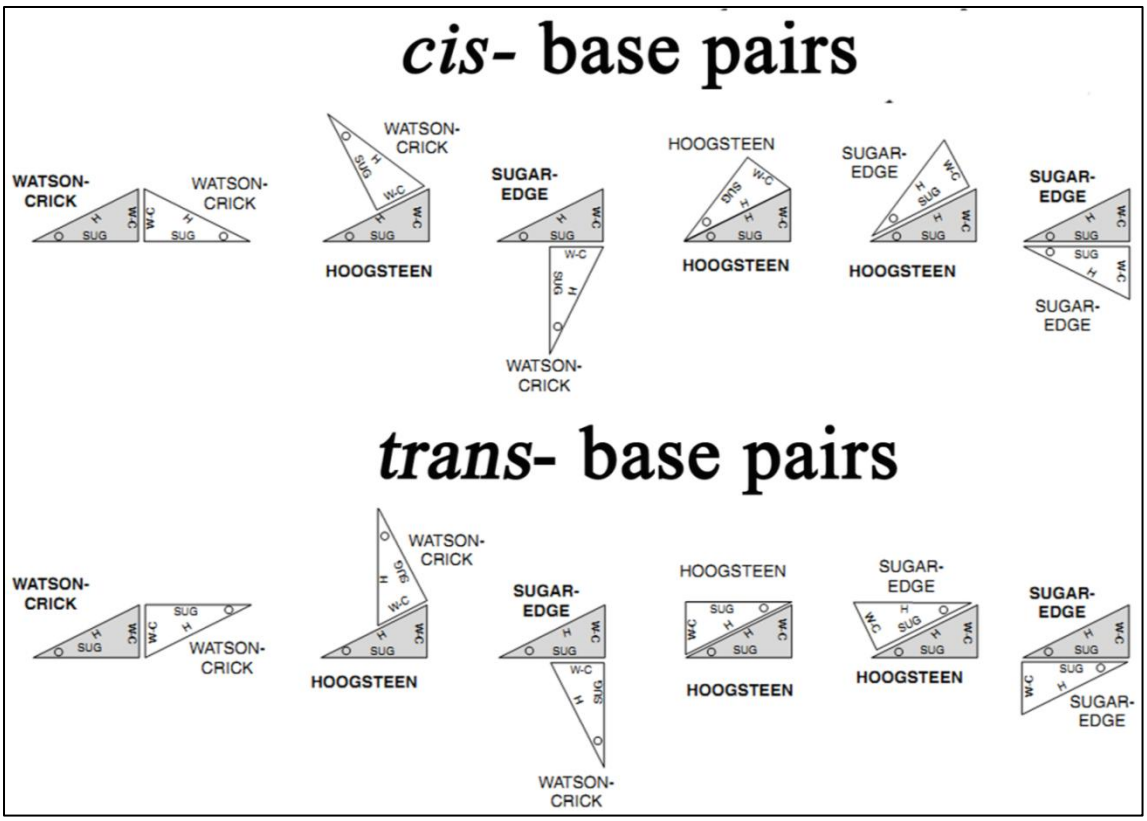


Figure 1.1.8: The twelve distinct edge-to-edge interactions are possible, as taken from (Leontis *et al.*, 2002)

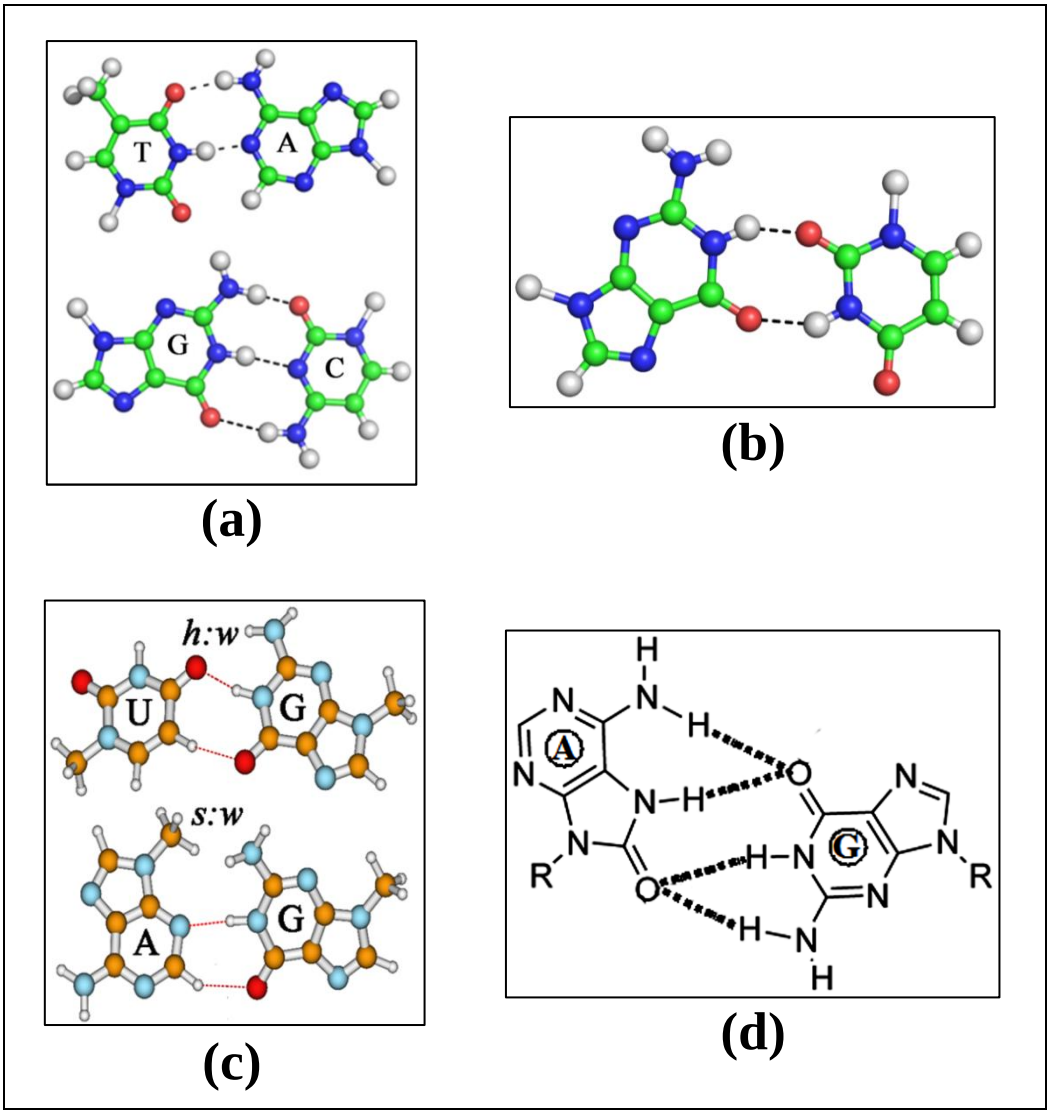


Figure 1.1.9: (a) A:T and G:C WC base pairs. (b) G:U wobble base pair. (c) U:G and A:G non-canonical base pairs. (d) Bifurcated base pair involving three-centered H-bond.

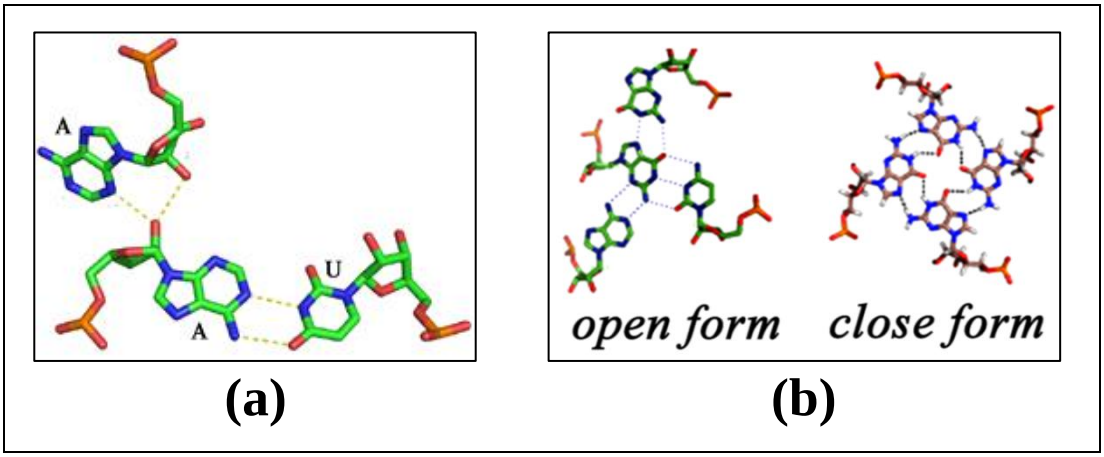


Figure 1.1.10: (a) Base-triplet. (b) Base quartet in open (Left) and close (Right) forms.

1.1.3.2 Base Pair Orientation Parameters

The structural variations in nucleic acids are mostly environ-dependent, but are often sequence-specific. Unlike globular protein structures, no long range interactions are found to prevail between remote base pairs in double helical DNA. Thus, the local structures are defined by the relative displacement and orientation of the bases within a base pair or between two successive base pairs in 5'→3' direction (Hunter and Lu, 1997a, 1997b). On the other hand, RNA can have three-dimensional (3D) architecture like proteins and can have long range interactions through base paring between different bases of the same chain. The secondary structural motifs of RNA are primarily double-stranded, formed by the RNA chain folding back on itself, or unpaired loop regions. Thus, in both the cases, the relative orientation of the bases within a base pair or between two successive base pairs is crucial to understand the structural organizations (Dickerson, 1989; Olson *et al.*, 2001).

The detail understanding of base pairs is extremely important for correct sequence-structure prediction in genome. A combination of four bases in two adjacent positions in same strand maintaining due complementarity on the other can produce ten possible dinucleotide steps (base pair doublets). These are three purine-pyrimidine steps: AT/AT, AC/GT, GC/GC; three pyrimidine-purine: TA/TA, TG/CA, CG/CG; and four purine-purine steps: AA/TT, CT/AG, TC/GA, CC/GG (T is replaced by U in RNA). In a double helical DNA/RNA the phosphate backbone and sugar residues lie on the surface and hence are more dynamic as compared to the base atoms, which are also confirmed from the experimentally reported thermal parameters of the atoms. Hence, structural analysis of DNA/RNA duplexes in terms of the base pair orientation parameters was justified and necessary.

According to the EMBO workshop, 1989 (Dickerson, 1989), geometries of base pairs or base pair doublets are defined with the help of three translational and three rotational degrees of freedom along the three mutually perpendicular axes fixed on the base or base pair planes, respectively (Bansal *et al.*, 1995; Calladine *et al.*, 2004; Dickerson, 1989; Dickerson, 1998; Olson *et al.*, 2001). The geometric sense of these degrees of freedom and the terminology associated with them are shown in **Figure 1.1.11**. Within a base pair, the spatial arrangement of one base with respect to the other can be quantitatively defined with the help of intra-base pair parameters: buckle, propeller, open-angle, shear, stagger and stretch **(Figure 1.1.11a)**.The relative

orientation of two base pairs within a double helical stack can be described by a set of six inter-base pair parameters: tilt, roll, twist, shift, slide and rise **(Figure 1.1.11b)**.

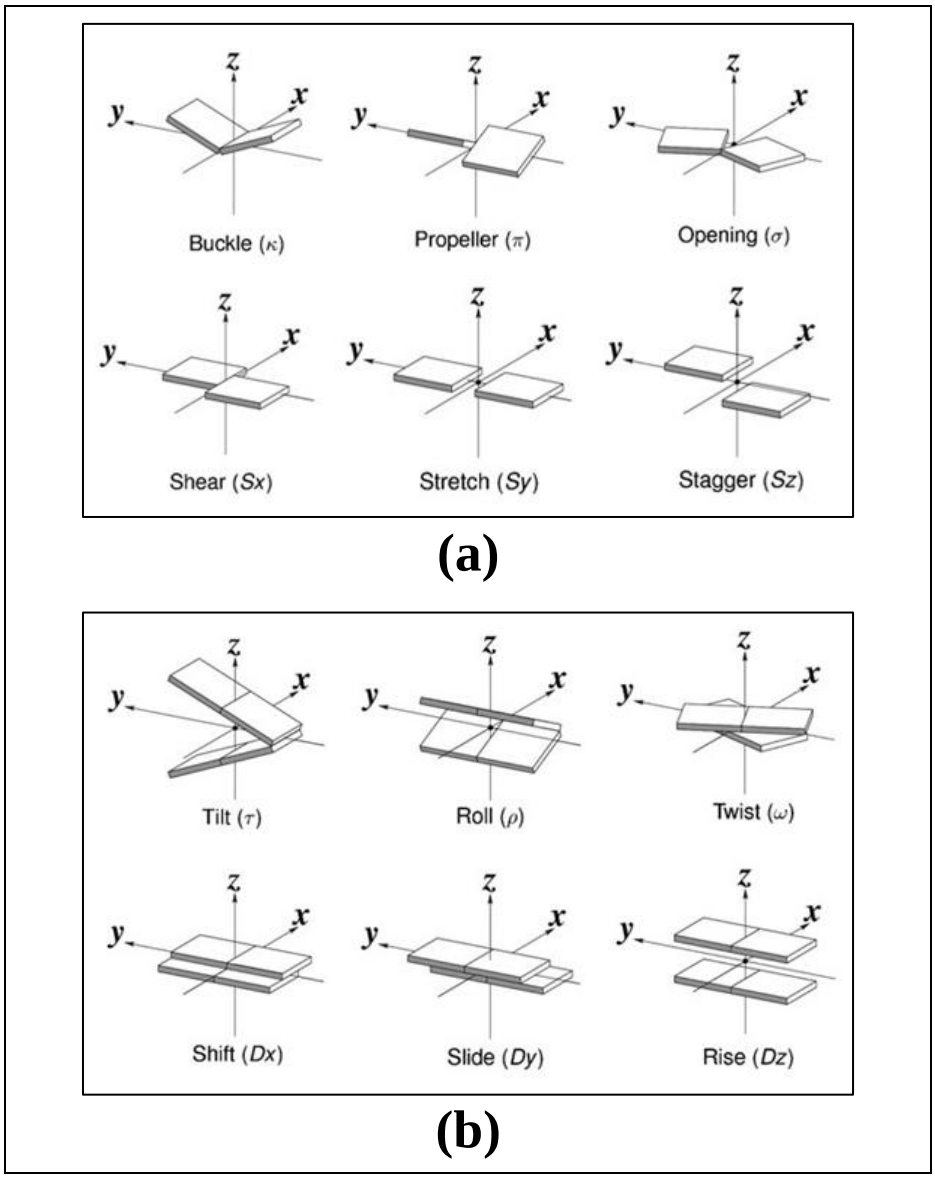


Figure 1.1.11: The translational and rotational (a) base pair parameters and (b) base pair step parameters along x-, y- and z-axis, as taken from (Bansal *et al.*, 1995).

Among the six base pair parameters, shear, stretch and open-angle directly relate to the H-bonding pattern and proximity, while buckle, propeller and stagger describe the overall non-planarity of a base pair compared to the ideal coplanar geometry. Similarly, the relative orientation of two base pairs with respect to one another, within a double helical stack has been described by inter-base pair parameters or dinucleotide step parameters. Since, DNA has a continuous double helical structure with A:T or G:C base pairing, these geometrical parameters display

regular behaviour for a stable structure. Analysis of DNA crystal structures show an invariant sharp peak for tilt, shift and rise at ~0º, ~0Å and ~3.4Å, respectively, whereas the values of roll, twist and slide range within -9º to 5º, 32º to 45º, and -0.6Å to 0.8Å, respectively for the dodecamer d(CGCGAATTCGCG) (Drew and Dickerson, 1981; El Hassan and Calladine, 1997). Crystal structure data base analysis of several DNA and RNA molecules in regular double helical form indicates that some of the base pair parameters (roll, twist and slide) are sequence and environment dependent (El Hassan and Calladine, 1997; Mukherjee *et al.*, 2014, 2015). The sequence dependence in twist, propeller, roll and slide were initially rationalized through steric clashes between substituent atoms on individual bases as depicted by Calladine's Rules on the basis of steric clashes between opposite strand purine bases (Calladine and Drew, 1984; Calladine, 1982) . The steric clash mainly arise between purine bases mainly due to propeller twist in each base pair which is necessary to allow greater overlap of bases within the same strand and to reduce the area of contact between the bases and water. In general this parameter tends to be higher than average in regions containing A:T base pairs, but lower for regions with G:C base pairs. These steric clash rules of Calladine predict reasonably well the structure of B-DNA. However it has limited applicability to other helical types. A number of studies have been carried out to understand sequence dependent structure of DNA and the correlations between different structural parameters (Babcock and Olson, 1994; Bhattacharyya and Bansal, 1990; Calladine and Drew, 1984; Gorin *et al.*, 1995; Subirana and Faria, 1997; Suzuki *et al.*, 1997).

A number of software packages have been developed by different groups for determination of these parameters: CURVES (Lavery and Sklenar, 1988), NGEOM (Tung *et al.*, 1994), RNA (Babcock and Olson, 1994; Babcock *et al.*, 1994), NUPARM (Bansal et al., 1995; Mukherjee *et al.*, 2006), CompDNA (Gorin *et al.*, 1995), SCHNAaP (Lu *et al.*, 1997), 3DNA (Lu and Olson, 2003), FREEHELIX (Dickerson, 1998), *etc.* These programs mainly differ in their choice of reference frames and mathematical definitions of the parameters (Lu et al., 1999). There also have been efforts to obtain a standard reference frame for the description of nucleic acid base pair geometry (Olson *et al.*, 2001).

Crystallographic analyses of a number of diverse sequences have provided further evidence for the variations in sequence dependent structural features (Goodsell *et al.*, 1995; Heinemann *et al.*, 1992; Lipanov *et al.*, 1993; Quintana *et al.*, 1992).

These show that twist, roll and slide have several correlations between them. For example, rise is linearly related to twist. Dependence of DNA structure on sequence was further elucidated by theoretical calculations (De Santis *et al.*, 1990; Hunter and Lu, 1997b; Mohanty and Bansal, 1991; Srinivasan and Olson, 1987). Some of the backbone conformational angles also show considerable variation in B-DNA structures depending on the base sequence (Schneider *et al.*, 1997; Subirana and Faria, 1997). In addition to effect of constituent base pairs, there have been studies to find out effect of flanking residues on base pair steps of the structure (Bandyopadhyay and Bhattacharyya, 2000; Beveridge *et al.*, 2004; Fujii *et al.*, 2007). Both experimental and theoretical studies reveal that base pair parameters are inherently backbone independent (Packer & Hunter, 1998). Studies of experimentally derived data show that C/G containing steps are more context dependent than A/T containing ones (Packer *et al.*, 2000). Steps like CA/TG and CG/CG depend largely on flanking sequences for their structures (Dixit *et al.*, 2005; Subirana and Faria, 1997). Moreover, several studies (Bertrand *et al.*, 1998; Winger *et al.*, 1999; Djuranovic and Hartmann, 2004) have shown that significant variations were observed for slide, roll and twist, while much less changes were found for tilt, shift and rise values.

1.1.3.3 Hydrogen Bonding and Stacking Interaction between Base Pairs

As discussed earlier, the classic WC base pairs that are found in the usual double-stranded nucleic acids are A:T, G:C and A:U (for RNA) where ":" represents base pairing. The distances between C1' atoms of sugars on opposite strands are essentially the same for A:T and G:C base pairs (Donohue and Trueblood, 1960). Theoretical studies of Šponer, Jurečka, and Hobza have reported relevant H-bonding energies in a wide variety of nucleic acid base pairs (Sponer *et al.*, 2004). Studies using density-functional theory combined with symmetry adapted perturbation theory (DFT-SAPT), show that the interaction energy of A:T and G:C base pairs are -15.7 kcal/mole and -30.5 kcal/mole, respectively (Hesselmann *et al.*, 2006). In an effort to systematically characterize H-bonding stability of all the base pairs observed in RNA, Panigrahi *et al.* (Roy *et al.*, 2008) carried out optimizations and determined H-bond strengths of thirty-three different frequently occurring base pairs. A complete study of all possible base pairs stabilized by at least two H-bonds (RNABP COGEST) has also been done far recently (Bhattacharya *et al.*, 2015).

Base pair stacking refers to the favourable interactions between neighboring base pairs which arise from van der Waals and hydrophobic contacts that optimize the water-insoluble areas of contact. In recent years stacking interactions has generated huge interest, where the stacked base pairs undergo π-π interactions due to overlap between the aromatic rings of the bases. Experimental studies reveal that stacking interaction has large effect on the structure and electronic properties of DNA (Hagerman and Hagerman, 1996; Yakovchuk *et al.*, 2006). Stacking energy shows high sequence sensitivity and plays a key role in the stabilization of nucleic acid secondary structures.

1.1.4 DNA- *One molecule, many different helices*

DNA displays a wide range of structures. A partial list of various types of structures depicting polymorphism of DNA is given in **Table 1.1.2**. In many cases the different types of structures result from relatively minor alterations in the helix structures, for example, the difference between the A and B conformations. However, there can be changes that are much more dramatic and differ radically from the double helix structure proposed by Watson and Crick. The two best examples of this class of structures are Z-DNA and parallel-stranded DNA. DNAs with random sequence are found only in A, B and C forms and materials with strictly repetitive oligonucleotide sequence can adopt in addition D, E and Z forms. At first the non-standard conformations were considered to be uncommon with little or no biological relevance. However, it now appears that most, if not all of the structures listed in **Table 1.1.2** can and do exist in cells (Macgregor and Poon, 2003).

Table 1.1.2: Different forms of DNA.

Structure	Description
A-DNA	Duplex isoform found in dehydrated environments. First fiber diffraction by Roselind Franklin was of A-DNA. Only isoform accessible to RNA and RNA/DNA duplexes.
B-DNA	Duplex isoform found in most biological environments. Original structure proposed by Watson and Crick.
Z-DNA	Left-handed duplex favoured by GC-rich sequence under high salt. Discovered in first single-molecule crystallographic structure of DNA (Wang *et al.*, 1979)
M-DNA	Putative duplex isoform in which WC base pairs coordinate divalent ions (Zn^{2+}, Ni^{2+} and Co^{2+}) at alkaline pH. (Aich *et al.*, 1999). Noted for electrical conductive properties.
P-DNA	Once postulated structure of DNA by Linus Pauling, with a interwound phosphate backbone and exposed bases (Pauling and Corey, 1953). Observed in positively supercoiled B-DNA duplex under a stretching force (Allemand *et al.*, 1998).
Triplex	Three-stranded isoform formed by binding of a third strand to the major groove of a B-DNA duplex. Formed by two homopyramidine and one homopurine strands and favoured by low pH or high ionic strength.
H-DNA	Intramolecular triplex form composed of a hairpin duplex by one strand and Hoogsteen base pairing by another. Sequences exhibit mirror symmetry along the strands.
G-quartet	Four-stranded isoform supported by Hoogsteen base pairing among four guanine residues, coordinating a monovalent ion (*e.g.*, K^+, Na^+, NH^{4+})
Frayed wires	Multistranded superstructures formed by oligonucleotides with long consecutive guanine runs (Protozanova and Macgregor, 1998). Non-guanine nucleotides branch out and may be functionalized.

1.1.4.1 Polymorphism in DNA Double Helix:

The most important feature of DNA is that it is usually composed of two polynucleotide chains that are held together by weak, non-covalent bonds between base pairs and twisted around each other in the form of a double helix **(Figure 1.1.12a)**. The backbone of each strand of the helix is composed of alternating sugar and phosphate residues; the bases project inwards. The two strands have the same helical geometry but base pairing holds them together with opposite polarity. These two backbones run in anti-parallel direction along the outer surface of a virtual cylinder containing the helix, in a spiraling motion and divide it into two unequal clefts, through which the base atoms are directly accessible to the external agents like solvent, ions, ligands and proteins. These helical clefts are termed as grooves, the larger one as major and the smaller one as minor groove, in the most stable polymorph of DNA occurring in the cellular environment; the B-DNA (Neidle, 2002). The edges of each base pair are exposed in the major and minor grooves, creating a pattern of H-bond donors and acceptors **(Figure 1.1.12b)** that are characteristic of the type of groove **(Table 1.1.3)**. The minor groove is not as rich in chemical information because the A:T and T:A base pairs and G:C and C:G base pairs look similar to one another in the minor groove ("The Structures of DNA and RNA," 2002). Moreover due to its small size minor groove is less able to accommodate ligands or amino acid side-chains. Groove width is defined as the shortest distance between phosphate groups on opposite strands minus the van der Waals diameter of a phosphate group, while groove depths are normally defined in terms of the differences in cylindrical polar radii between phosphorous and guanine N2 or adenine N6 atoms for minor and major grooves, respectively (Neidle, 2007). Environmental effects, such as relative humidity, salt concentration, nature of counter-ions, *etc.* direct the nucleic acid double helices to adopt different allomorphic conformations, out of which B-, A- and Z-forms have been found naturally (Saenger, 1984) **(Figure 1.1.13)**.

Table 1.1.3: H-bond donor and acceptors that are exposed in the grooves of DNA.

	Adenine	Guanine	Cytosine	Thymine	Uracil
major groove	N6, N7	C6, N7	N4	C4, C5	C4
minor groove	N3	N2, N3	C2	C2	C2

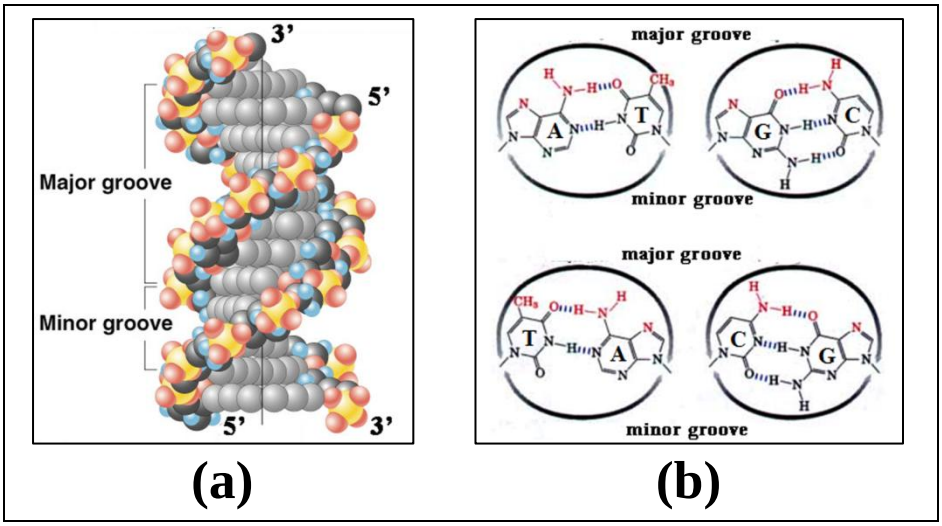


Figure 1.1.12: (a) DNA double helix with major and minor grooves. (b) The atoms of base pairs that are exposed in the major and minor grooves.

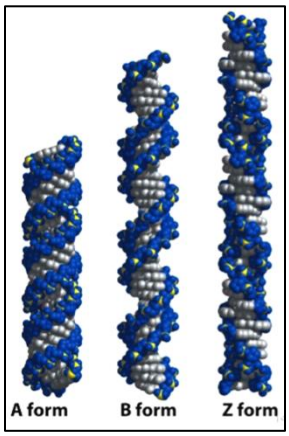


Figure 1.1.13: From Left to Right, right-handed A-DNA, B-DNA and left-handed Z-DNA.

The DNA in cell under physiological conditions generally with more than 92% relative humidity and in presence of alkali metal ions such as Na^+, K^+ adopts the right-handed B-form. It contains two polynucleotide strands that wind about a common axis with a right-handed twist. The helix diameter is 20Å and extends 34Å per 10 base pairs of sequence. The double helix makes one complete turn of 360º about its axis every 10.4-10.5 base pairs in solution. The planes of the bases are nearly perpendicular to the helix axis. B-DNA has two exterior grooves that run along with the sugar-phosphate backbones and the helix axis passing through the approximate centre of each base pair. The major groove is wider and the minor groove is narrow. Correlated variations of ε and ζ dihedral angles sometimes lead to different polymorphic forms of B-DNA, which are called B_I and B_{II} conformations (Drew *et al.*, 1981). In B_I like backbone conformations, the ε-ζ values lie near -90º, while for

B_{II} form it is around 90º (Becker and Wang, 1989; Clark *et al.*, 2000; Gupta *et al.*, 1980; Schneider *et al.*, 1997). It has been seen that B_{II} conformation also moves the bases giving rise to larger slide (2.5Å) and twist (45º). Several studies suggest that BII conformation may play an important role in several biological processes (Djuranovic *et al.*, 2004; Reddy *et al.*, 2003; van Dam *et al.*, 2002; Wellenzohn *et al.*, 2001).

When the relative humidity of the system is reduced to 75%, B-DNA undergoes a reversible conformational transition to the A-form. Conformational pathway for transition from B- to A-form is also clear from X-ray crystallography (Vargason *et al.*, 2000). It was shown by fiber diffraction studies that A-DNA has a wider and flatter right-handed helical form with 11 base pairs per helical turn and a pitch of 28.2Å. The planes of the base pairs are inclined with respect to the helix axis by an angle of ~20º and moved away from the helix axis by ~4Å, which generates an axial hole in the conformation. The inclination associated with helical twist gives rise to a large positive roll angle between successive base pairs. This facilitates the opening of base pairs towards minor groove. Since its helix axis does not pass through the base pairs, A-DNA has a deep and narrow major groove and wider and shallow minor groove.

The crystal structure determination of d(CGCGCG) at multimolar NaCl concentration by Wang and co-workers revealed a left-handed double helix, which was termed as Z-DNA (Wang *et al.*, 1979). Z-DNA has 12 WC base pairs per turn of the helix and a pitch of 45Å. In contrast to A-DNA, there is a narrow and deep minor groove and the major groove bulges out. The base pairs are flipped 180º relative to those in B-form, giving rise to inverse stacking, and the repeating unit in Z-DNA consists of two base pairs with alternating purines (Guanine) and pyrimidines (Cytosine) rather than a single base pair. The alternating Purine/Pyrimidine bases have different sugar pucker and glycosidic χ torsion angle. The lines joining the successive phosphate groups go after a zig-zag pathway around the helix axis. Formation of this structure is generally unfavourable, although certain conditions can promote it, such as, alternating Purine-Pyrimidine sequence (especially poly-d$(GC)_2$), negative DNA supercoiling or high salt concentration (all at physiological temperature, 37°C, and pH 7.3-7.4). A high salt concentration stabilizes Z-DNA by reducing the otherwise increased electrostatic repulsions between closest approaching phosphate groups on opposite strands (8Å in Z-DNA, 12Å in B-DNA). The overall structural features in B-DNA, A-DNA and Z-DNA are summarized in **Table 1.1.4**.

Table 1.1.4: A comparison of A-DNA, B-DNA and Z-DNA.

	A-DNA	B-DNA	Z-DNA
Overall proportions	Short and broad	Longer and thinner	Elongated and slim
Rise per base pair	2.3Å	3.32Å	3.8Å
Helix packing diameter	25.5Å	23.7Å	18.4Å
Helix rotation sense	Right-handed	Right-handed	Left-handed
Base pairs per helix repeat	1	1	2
Base pairs per turn of helix	~11	~10	12
Rotation per base pair	33.6°	35.9°	-60° per 2 base pairs
Pitch per turn of helix	24.6Å	33.2Å	45.6Å
Tilt of base normals to helix axis	+19°	-1.2°	-9°
Base pair mean propeller twist	+18°	+16°	~0°
Helix axis location	Major groove	Through base pairs	Minor groove
Major-groove proportions	Extremely narrow but very deep	Wide and of intermediate depth	Flattened but out on helix surface
Minor-groove proportions	Very broad but shallow	Narrow and of intermediate depth	Flattened out on helix surface
Glycosyl-bond conformation	*anti*	*anti*	*anti* at C, *syn* at G
Sugar pucker	$C_{3'}$- endo	$C_{2'}$- endo	C: $C_{2'}$- endo G: $C_{2'}$- exo
experimental conditions	low humidity	high humidity, most closely corresponds to the average structure of DNA under physiological conditions	alternate purine-pyrimidine residues assume the left-handed conformation only in the presence of high concentrations of positively charged ions.

1.1.4.2 DNA Supercoiling

Physiological DNA can neither maintain the ideal B-form double helix nor remain as a straight rod within a cell of few microns in dimension. Moreover, in cells, the movement and conformation of the duplex DNA is constrained by its non-covalent interactions with proteins or, as in case of bacterial genomes and plasmid DNAs, direct covalent joining to form a continuous circle. Contortions, known as supercoiling, helps the rigid DNA to wrap around proteins, such as histone octamers in nucleosomes, or to form non-standard structures required for several biological processes. The most important aspect of DNA supercoiling is the packaging of DNA double helices to form chromosomes in eukaryotic cells. This compaction allows a long DNA of ~2m length (approximately $6x10^9$ base pairs) to be fitted within the cell nucleus having dimensions 5-10 micrometers. The essential unit of DNA packaging is the nucleosome, where ~147 base pairs of DNA are wrapped around histone heterooctamer units in a left-handed manner (Kornberg, 1974; Luger *et al.,* 1997). These nucleosomes are compacted by H1 histones to give rise to 30nm wide chromatin fibers (Thoma, Koller, & Klug, 1979; Zhou *et al.,* 1998) **(Figure 1.1.14a)**. The polynucleotide backbone is, of course, constrained by rigid bond lengths and angles and it can be twisted to only a limited extent. Thus, double-stranded DNA usually takes up over or under-twisted B-form in physiological environment due to bending and supercoiling, giving rise to strain in the backbones. Further coiling and folding of the chromatin fibres results in the highly condensed form of metaphase chromosomes with a DNA-to-chromosome packing ratio of approximately 10000:1.

DNA supercoils can be of two major types – (i) Plectoneme, a two-start right-handed helix with terminal loops and (ii) Toroid, a one-start left-handed helix **(Figure 1.1.14b)**. In eukaryotes, DNA supercoiling exists on many levels of both plectonemic and toroidal supercoils, with the toroidal supercoiling proving most effective in compaction of DNA into nucleosomes (Bloomfield *et al.*, 2000).

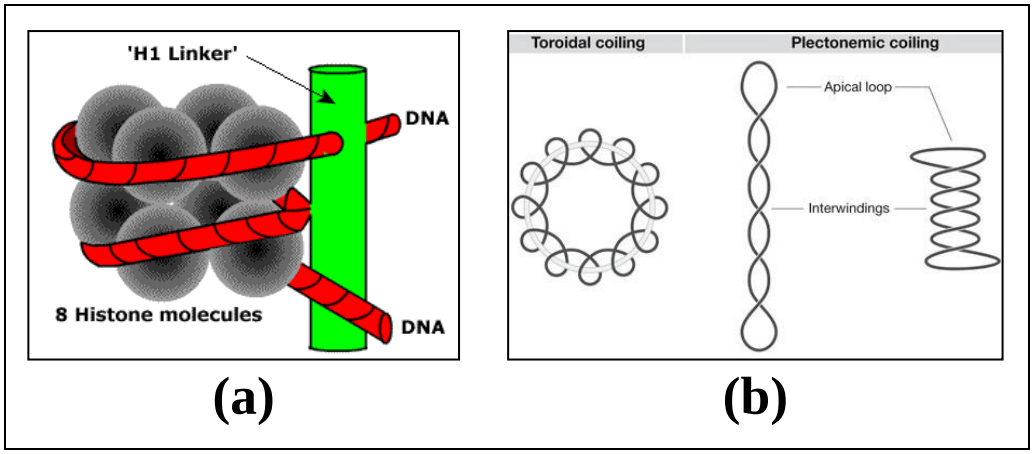


Figure 1.1.14: (a) Structure of nucleosome particle – histone octamer in gray, H1 histone in green and DNA as red liquorice. (b) From Left to Right, Toroidal coiling and Plectonemic coiling.

1.1.4.3 Triple Helical DNA

The first triple-stranded model for nucleic acid was proposed by Pauling and Corey in 1953 on the basis of X-ray crystallography data **(Figure 1.1.15)**. In this model, three polynucleotide strands make up a helix with a seven nucleotide pitch; presumably the uncharged phosphates are localized close to the helix axis, while the bases are oriented outside the molecules (Soĭfer, 1996).

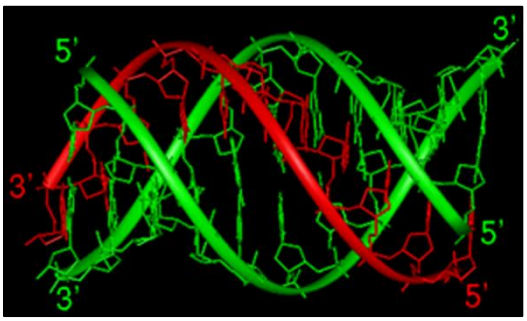


Figure 1.1.15: A triple-helix DNA.

Since the original discovery of triple helical nucleic acids, a number of triplex DNA structures that form under various conditions *in vitro* and/or *in vivo* have been identified. Studies (Frank-Kamenetskii and Mirkin, 1995) show that structure of triplexes may vary substantially as-

(i) Triplexes may consist of two pyrimidine and one purine strands (YR*Y) or of two purine and one pyrimidine strands (YR*R).

(ii) Triplexes can be built from RNA or DNA chains or their combinations.

(iii) Triplexes can be formed within a single polymer molecule (intramolecular triplexes) or by different polynucleotides (intermolecular triplexes).
(iv) Triplex formation may occur by a strand-switch mechanism (alternate strand triplexes) for special DNA sequences consisting of clustered purines and pyrimidines in the same strand.

The building blocks of YR*Y triplexes are the canonical CG*C and TA*T triads. To form such triads, the third strand must be located in the major groove of the double helix that is forming Hoogsteen H-bonds with the purine strand of the duplex. An important feature of the YR*Y triplexes is that formation of the CG*C triad requires the protonation of the N3 of cytosine in the third strand. Thus, such triplexes are favourable under acidic conditions (Frank-Kamenetskii and Mirkin, 1995).

YR*R triplexes are more versatile than YR*Y triplexes. Originally it was believed that they must be built from CG*G and TA*A triads. Later work however showed that TA*T triad may also be incorporated into the otherwise YR*R triplex. Moreover, the stability of triplexes consisting of alternating CG*G and TA*T triads is higher than that of triplexes built of CG*G and TA*A triads (Beal and Dervan, 1991). Thus, the term YR*R triplex, though routinely used in literature, is misleading with regard to the chemical nature of the third strand. Another notable difference between two triplex types is that reverse Hoogsteen base pairs are needed to form reasonable stacking interactions among CG*G, TA*A, and TA*T triads (Beal and Dervan, 1991). Another novel feature of YR*R triplexes is that their stability depends dramatically on the presence of bivalent metal cations. Unlike the case of YR*Y triplexes, where the requirement for H^+ ions has an obvious reason, the metal dependence of YR*R triplexes is an obscure function of the particular metal ion and the triplex sequence (Malkov *et al.*, 1993).

Despite these differences, the YR*R triplexes are similar to YR*Y triplexes in their most fundamental features: (a) the duplex involved in triplex formation must have a homopurine sequence in one strand, and (b) the orientation of the two chemically homologous strands is anti-parallel.

In an intramolecular triplex or H-DNA structure, the third strand is provided by one of the strands of the same duplex DNA molecule at a mirror repeat sequence. Four isomers of intramolecular triplexes can exist depending on the strand that serves as the third strand. Intramolecular triplexes are also known as H-DNA or *H-DNA,

depending on whether the third strand of the triplex is pyrimidine- or purine-rich, respectively (Jain *et al.*, 2008).

1.1.4.4 Quadruple Helix DNA

Poly-guanine sequences fold into G-quadruplexes **(Figure 1.1.16a)** in presence of monovalent cations like Na^+/K^+ *in vitro*. These G-quadruplex structures can be attained by G-rich sequences of immunoglobulin switch regions, mutational hot spots, regulatory elements with oncogene promoters and telomeric regions. The G-rich 3'-overhangs of DNA can fold back and then dimerize to form hairpin loops that are stabilized by G-quartets (Sundquist and Klug, 1989), with four guanine bases involved in a planar four-stranded arrangement with G:G Hoogsteen base pairing (Hoogsteen, 1963).

In vitro, telomeric DNA has the ability to spontaneously undergo an intramolecular rearrangement and form a higher-ordered DNA structure called quadruplex DNA, and they have also been shown to form *in vivo* in some cases (Paeschke *et al.*, 2005). There are various factors such as sequence, concentration of ions, *etc.* that determines how the quadruplex folds. Several structures of G-quadruplex DNA have been solved by X-ray crystallography and NMR spectroscopic methods till date. Depending on the lengths of the oligomers and number of intervening thymine residues, quadruplex can be intramolecular, bimolecular or tetramolecular (Simonsson, 2001). Depending on how the individual runs of guanine bases are arranged in a bimolecular or intramolecular quadruplex, a quadruplex can adopt one of a number of topologies with varying loop configurations (Burge *et al.*, 2006). An in-depth structural study of G-quadruplexes formed by the G-rich region of human telomere has been performed in this thesis and the different topologies adopted by these quadruplexes have been compared using Molecular Dynamics Simulation techniques. A single non-telomeric G-quadruplex has also been studied. All these are discussed in details in Chapter II of this thesis.

1.1.4.5 i-Motif

An i-motif **(Figure 1.1.16a)** is a four-stranded nucleic acid structure formed from C-rich sequences at acidic pH and consists of two parallel stranded duplexes

formed from hemi-protonated cytosines that intercalate. The basic unit of i-motif is the C:C^+ base pair **(Figure 1.1.16b)**, with one of the two cytosines protonated at N3 positions, such that three H-bonds can be found between them (Gehring *et al.*, 1993; Langridge and Rich, 1963). There are different possible I-motif structures with different intercalation and looping topologies (Leroy *et al.*, 1994; Phan & Leroy, 2000).

1.1.4.6 Holliday Junction

A Holliday junction **(Figure 1.1.16a)** is a mobile junction between four strands of DNA. The structure is named after Robin Holliday, who proposed it in 1964 (Holliday, 1964) to account for a particular type of genetic information exchange, he observed in Yeast known as homologous recombination (Stahl, 1994). Holliday junctions are highly conserved structures, from prokaryotes to mammals (Constantinou *et al.*, 2001). They are an intermediate in genetic recombination and also of importance in maintaining genomic integrity (Fu *et al.*, 1994). In the absence of any metal ions and at low ionic strength the four-way junction assumes an extended square planar conformation, whereas in presence of magnesium and other polyvalent ions the junction adopts a folded X structure with tetrahedral geometry, in which the four duplex arms extending from the junction form two quasi-continuous DNA helices with base stacking retained through the cross-over point.

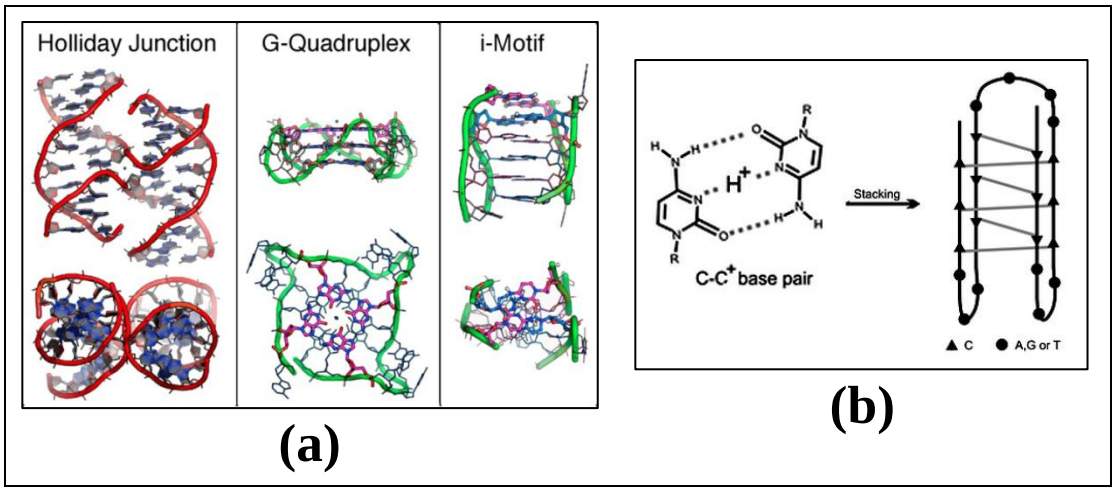


Figure 1.1.16: (a) From Left to Right, Holiday Junction, quadruple helix DNA and i-motif. (b) C-C^+ base pair in an i-motif.

1.1.5 RNA

RNA molecules have a long history and according to the RNA world hypothesis, origin of life is based on RNA molecules that store both the genetic information and catalyze their own replication. Numerous studies have demonstrated that the true catalogue of RNAs encoded within the genome (the "Transcriptome") is more extensive and complex than previously thought (Frith *et al.*, 2005; Kapranov *et al.*, 2007; Mattick and Makunin, 2006). RNA molecules are considered to play an intermediary role in protein coding, carrying the genetic information from DNA to proteins. However, these coding RNAs or messenger RNAs account for only ~2.3% of the human genome, and therefore the vast majority of the transcriptome appears to be non-protein–coding (Frith *et al.*, 2005). Apart from translation of proteins in ribosomes, such non-coding RNAs (ncRNAs) have been found to carry out diverse functions – from mRNA splicing and RNA modification to translational regulation. It is also observed that the proportion of transcribed non-coding sequences increases with developmental complexity of the organism (Frith *et al.*, 2005). A few of the major types of coding and non-coding RNAs found within cells are:

(a) Messenger RNA (mRNA)

(b) Transfer RNA (tRNA)

(c) Ribosomal RNA (rRNA)

(d) Transfer Messenger RNA (tm RNA or SsrA)

(e) Signal Recongition Particle RNA (SRP RNA)

(f) Ribozyme

(g) Ribonuclease P (RNase P)

(h) Riboswitch

(i) Small Nuclear RNA (snRNA)

(j) Small Nucleolar RNA (snoRNA)

(k) MicroRNA (miRNA)

(l) Small Interfering RNA (siRNA)

Chemically, RNA is a close cousin of DNA. However, RNA is involved in a wide range of cellular activities that often require the molecule to fold into a specific structure in order to perform its targeted function. Thus, structurally and functionally it is more closely related to proteins than to DNA.

1.1.5.1 Structural Organization of RNA

RNA structure was once envisioned as a collection of relatively rigid stems comprised of WC base pairs and the single-stranded loops connecting these stems. However, at present it is established that RNA adopts structures that Harry Noller described in a 2005 Science review article as "breathtakingly intricate and graceful" (Noller, 2005). A comprehensive definition of an RNA structural motif should be based on and consist of not only base-pairing or secondary structure constraints, but a complete 3D description, including backbone conformation, all H-bonding and base-stacking interactions, and sequence preferences. RNA structures are described at the sequence or primary structural level, followed by secondary, tertiary and quaternary levels.

RNA structures are described by the sequence at primary structural level. The next level of RNA structure (secondary structures) is the base pairing, which identifies both the base paired regions (helices) and non-paired regions. Tertiary structure describes the overall 3D conformation of a single molecule as determined by crystallography, NMR spectroscopy or modeling methods. The 3D structures are stabilized by long-range intramolecular interactions between basic secondary structural elements – helices and loops – to yield complex motifs, such as pseudoknots, ribose zippers, kissing hairpin loops, tetraloop-tetraloop receptor interactions, co-axial or pseudo-continuous helices.

In addition, a RNA structural motif may have co-factors such as metabolites, bound waters, metals or other ions, to support its conformation. The overall 3D structure of such a recurrent motif is largely independent of the context in which it is found. Moreover, RNA structural motifs are truly structural, and there may be several sequences that obtain the same structure (Hendrix *et al.*, 2005).

1.1.5.2 Secondary Structural Elements

Single-stranded RNA folds into a variety of secondary structural motifs and stabilized by both WC and unconventional base pairing. The common secondary structures that form the building blocks of RNA architecture include: bulges, base-paired helices or "stems", single-stranded hairpin or internal loops, and junctions **(Figure 1.1.17)**.

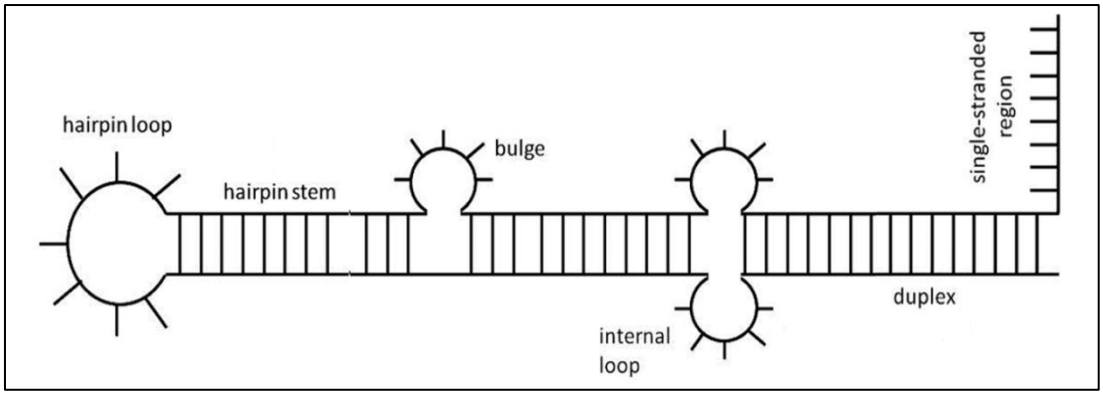


Figure 1.1.17: RNA architecture including the bulges, hairpin stem, hairpin or internal loops, junctions and single-stranded region.

1.1.5.2.1 Double Helix Stem

RNA forms double helix stem which consist of both WC and non-WC base pairs. The isosteric A:U and G:C base pairs give rise to uniform helical regions. However, insertion of non-WC base pairs that are non-isosteric cause disruption of normal helical conformation imparting several structural variations. The conformation of the RNA duplex is rigid, and its behaviour remarkably different in comparison to the DNA duplex.

The conformation of the RNA double helix has many features in common with A-DNA, and accordingly is known as A-RNA **(Figure 1.1.18)**. The A-RNA is 11-fold helix, with a narrow and deep major groove, and a wide and shallow minor groove. The base pairs are inclined to and displaced from the helix axis. The helix shows an average helical twist of about 33° and an axial rise of about 2.8Å. The base pair step parameters of A-RNA differ significantly from that of DNA is clearly indicated by database analysis performed for both independent RNA and protein-bound RNA (Kailasam et al., 2014; Mukherjee *et al.*, 2015). At higher ionic strength a slightly different 12-fold helix known as the A′-RNA is formed, having similar helical features (Arnott, 1999). The extra 2′-OH group in the sugar moiety in RNA nucleotides has a profound effect on the sugar pucker modes of the RNA double helix. Both A-RNA and A′-RNA show C3′-endo sugar pucker, since in other puckering modes 2′-OH group would clash with C8 (for purine) or C6 (for pyrimidine) of the attached bases. A double-helix RNA containing 16 base pairs has been simulated and its structural features including base pair and base pair step

parameters, torsion and pseudo-torsion angles, and grove widths have been reported in Chapter III of this thesis.

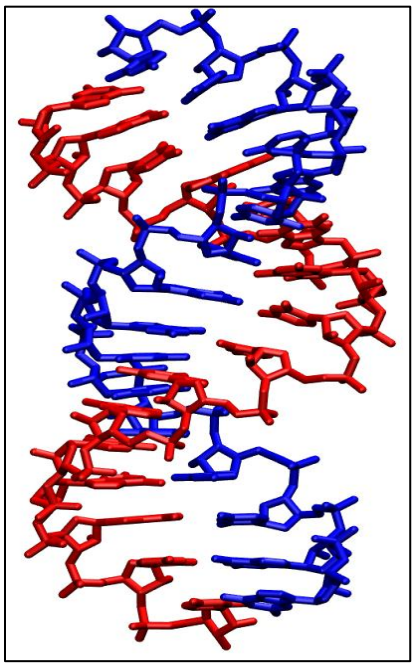


Figure 1.1.18: A-RNA helix, similar to A-DNA.

1.1.5.2.2 Hairpin Loops

Hairpins loops, or stem-loops, have unpaired nucleotides at the apex of a double-stranded stem (Tian *et al.*, 2004) **(Figure 1.1.19)**. They link the 5'- and 3'-ends of a double helix and vary in length from 2 to 14 nucleotides. The most common and well-studied of these are the tetraloops (length=4) **(Figure 1.1.19a)**. There are at least four types of tetraloops that are characterized by their sequence and conserved structures: GNRA (Correll and Swinger, 2003; Jucker *et al.*, 1996; Jucker and Pardi, 1995; Leontis and Westhof, 2002), UNCG (Cheong *et al.*, 1990), ANYA (Convery *et al.*, 1998; Klosterman *et al.*, 2004; Rowsell *et al.*, 1998) and (U/A)GNN (Butcher *et al.*, 1997) (N represents any residue). In these tetraloops, a non-canonical base pair between the first and fourth nucleotides stabilize the stem-loop structure. In ribosomal RNAs, about 70% of the tetraloops belong to either GNRA or UNCG type. Other hairpin loop motifs include T-loop and D-loop motifs of tRNA. The D-loop **(Figure 1.1.19b)** in tRNA is composed of 7 to 11 bases closed by a WC base pair. It contains the modified nucleotide dihydrouridine. The TψC-loop or T-loop **(Figure 1.1.19b)** contains thymine, a base usually found in DNA and pseudouracil (ψ) (Holbrook *et al.*,

1978). There also exists lonepair triloops (LPTLs**) (Figure 1.1.19c).** These motifs contain a single ("lone") base-pair capped by a hairpin loop containing three nucleotides. The two nucleotides immediately outside of this motif (5' and 3' to the lonepair) are not base-paired to one another, restricting the length of this helix to a single base-pair. In ten out of twenty-four LPTLs, U:A occurs as the lonepair, while the remaining fourteen LPTLs contain seven different base-pair types. Only a few of these lonepairs adopt the standard WC base pairing, while the majority of the base pairs have non-standard conformations (Lee *et al.*, 2003).

1.1.5.2.3 Internal Loops or Bulges

Internal-loops in RNA are found where two double stranded RNA segments are separated by a few unpaired nucleotides or a few non WC base pairs. If the numbers of unpaired nucleotides on the two strands are equal, the loop is symmetrical, else it is asymmetrical. Asymmetrical internal-loops formed due to insertions on only strand are also known as bulges **(Figure 1.1.20)**. The bulge can either stack into or out of the helix depending on the sequence context. Fully paired and stacked internal loops of up to eight non-canonical pairs have been structurally observed (Vallurupalli and Moore, 2003). Certain asymmetric internal loop motifs have been identified and characterized as resulting in sharp turns important for tertiary structure formation. These include the kink-turn, reverse kink-turn and hook-turn motifs. Other important structural motifs composed of internal loops are: C-loop, docking-elbow, right-angle, sarcin/ricin loops (also called bulged-G motifs), twist-up motif and the UAA/GAN internal loop motif.

The Chapter III of this thesis includes studies involving three different single residue bugle RNA systems: one protruding out from the helix, another intercalated within the duplex and the last one involves in formation of a base-triple inside the helix.

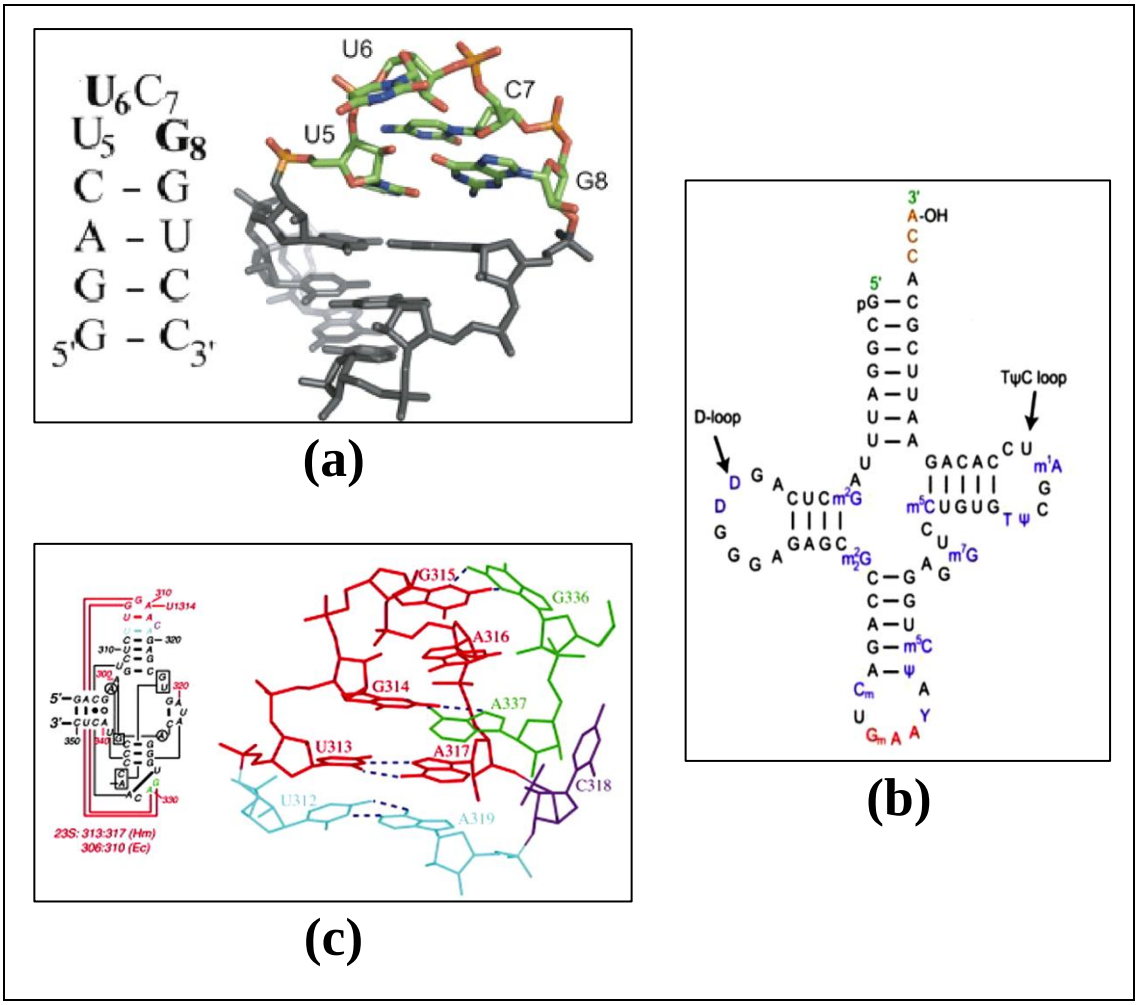


Figure 1.1.19: (a) UNCG tetraloop. (b) D-loop and T-loop. (c) Lonepair triloop.

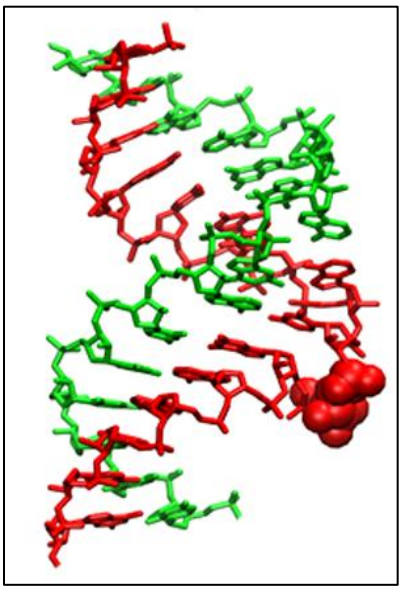


Figure 1.1.20: Bulge formed by a single unpaired residue.

1.1.5.2.4 Junctions

Junctions **(Figure 1.1.21)** are formed by intersection of three or more double helices. The double helices are separated by single-stranded sequences of zero or more residues known as linkers. Junctions have not yet been extensively studied; however, some generalizations have been made for the more common three-way and four-way junctions (Lilley, 1998, 2000). Although Common examples of junction loops are found in tRNAs and hammerhead ribozymes. The tendency for pairwise coaxial stacking of helical arms, the importance of metal ion interaction in the induction of tertiary folding, and the importance of hairpin or internal loop-loop interactions in the stabilization of tertiary structures are prominent features of junction loop architecture (Lescoute and Westhof, 2005; Penedo *et al.*, 2004).

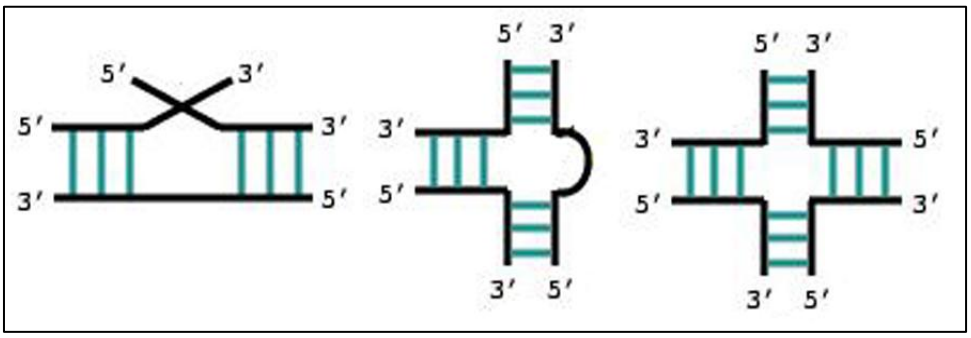


Figure 1.1.21: From Left to Right, 2-way 3-way and 4-way junctions.

1.1.5.3 Tertiary Structural Elements

Francis Crick wrote "tRNA looks like Nature's attempt to make RNA do the job of a protein" in his 1966 paper in the Cold Spring Harbor Symposium on Quantitative Biology. RNA chains fold into unique 3D structures that act similarly to globular proteins. The tertiary structure of RNA is its precise 3D structure, as defined by the atomic coordinates. It is dominated by limited interactions of motifs formed by the secondary structural elements.

1.1.5.3.1 Pseudo-continuous Helix

Bulge size can range from a single unpaired residue to several unpaired nucleotides. When the bulge size is several residues long, so that it appears to be

formed by two different parts of the RNA chain, it is considered to be a pseudo-continuous helix or coaxial stack. Thus, a pseudo-continuous helix is formed by stacking of two helical stretches aligned along the same axis with one continuous strand. Coaxial stacking or inter-helical stacking may occur via a single base or a base pair bridge at the interface between two helices (Kim *et al.*, 1974). They are highly stabilizing and dominant in several RNA structures: tRNAs, group I and group II introns, ribosomal RNAs *etc.* (Murphy *et al.*, 1994; Quigley and Rich, 1976; Toor *et al.*, 2008). The helical stretches taking part in the formation of coaxial stack often have one of the strands in common. In addition to bulges, the Chapter III of this thesis also contains structural analysis of few pseudo-continuous helices containing WC and non-WC base pairing at the coaxially stacked junction region.

1.1.5.3.2 Pseudoknots

A pseudoknot motif **(Figure 1.1.22a)** forms when a single-stranded loop base, pairs with a complementary sequence outside of this loop and folds into 3D structure by coaxial stacking (Echols, 2001). One example of a pseudoknot motif is the highly stable Hepatitis Delta virus ribozyme, in which the backbone shows an overall double pseudoknot topology (Ferré-D'Amaré *et al.*, 1998). Mutations in the pseudoknot region are involved in the human genetic disease, dyskeratosis congenital. Several important biological processes depend on RNA molecules that form pseudoknots. For example, the telomerase RNA component contains a pseudoknot that is critical for its activity (Chen and Greider, 2005). The pseudoknot region of RNase P is one of the most conserved elements in all of evolution.

1.1.5.3.3 Kissing Loop/Kissing-Stem Loop Interactions

RNA kissing stem-loop interactions, also called loop-loop pseudoknots are the single-stranded loop regions of two hairpins that interact through base pairing, forming a composite, coaxially stacked helix (Chang and Tinoco, 1994; Ennifar *et al.*, 2001). This structure allows all of the nucleotides in each loop to participate in base-pairing and stacking interactions. This motif was visualized and studied using NMR analysis by Lee and Crothers (Lee and Crothers, 1998).

1.1.5.3.4 Tetraloop-Tetraloop Receptor Interaction

Tetraloop-receptor interactions combine base pairing and stacking interactions between the loop nucleotides of a tetraloop motif and a receptor motif located within an RNA duplex, creating a loop-helix type of tertiary contact that stabilizes the global tertiary fold of an RNA molecule. Such interactions are critical for proper helical packing within the molecule (Major and Griffey, 2001). The main function of tetraloop-tetraloop receptor seems to be the creation of long-range tertiary interactions with certain types of stem-loop structures (Young and Silverman, 2002). The sequence of the tetraloop and its receptor often co-vary so that the same type of tertiary contact can be made with different isoforms of the tetraloop and its cognate receptor (Michel and Westhof, 1990).

1.1.5.3.5 A-Minor Motif

The A-minor motif is one of the most abundant long-range interactions in large RNA molecules. The host duplex is often a G-C base pair (Ban *et al.*, 2000). In this motif, single-stranded adenines make tertiary contacts with the minor grooves of RNA double helices by H-bonding and van der Waals interactions. The motif is stabilized by both base-base interactions and nucleoside-nucleoside interactions.

There are four subtypes **(Figure 1.1.22b)** depending on the position of the O2' and N3 of the donor adenine with respect to the O2' of the interacting WC base pair: (i) in the rarest form of motif, the N3 of the Ade (or other) residue is at the opposite side of the O2' of far strand of the receptor helix (Type 0), (ii) the O2' and N3 atoms of the A residue are inside the minor groove of the receptor helix (Type I). This is the most abundant A-minor motif type and the inserted base for this interaction must be an adenine, (iii) The N3 atom of the A residue is inside the minor groove and the O2’ is directed away from the O2’ of near strand (Type II). The inserted base for this interaction must be an adenine, and (iv) both the O2' and N3 of the A (or other) residue are positioned away from the near strand O2' of the receptor helix (Type III) (Nissen *et al.*, 2001).

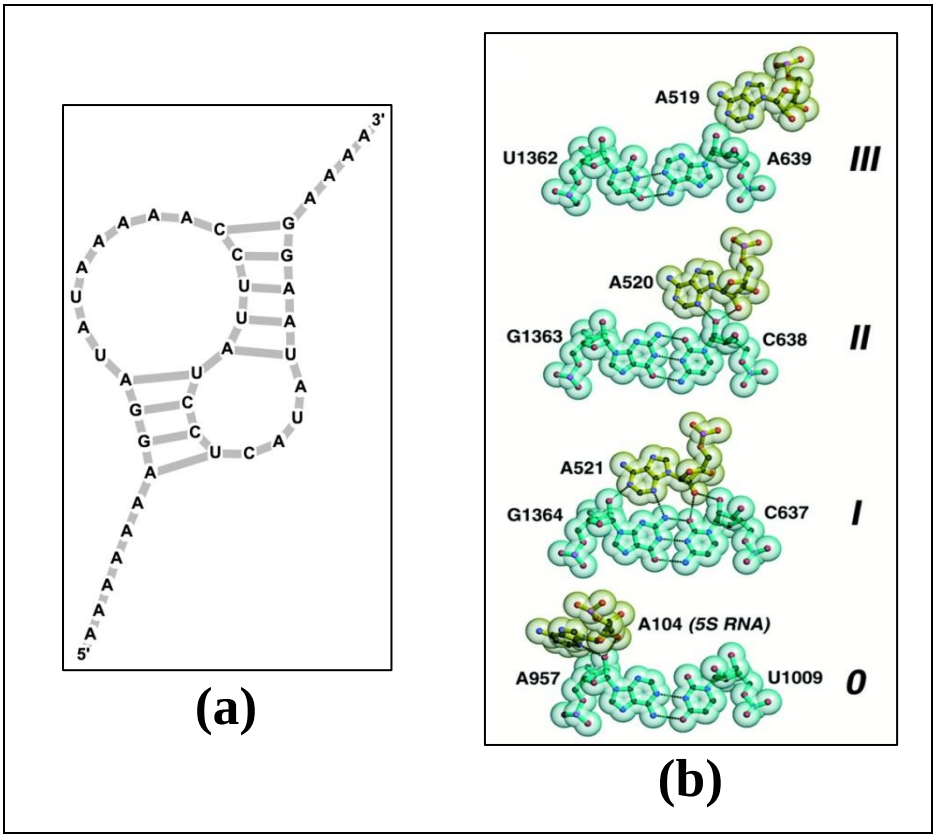


Figure 1.1.22: (a) A pseudoknot. (b) Four subtypes of A-minor motif as taken from (Nissen *et al.*, 2001).

1.1.5.3.6 Ribose zipper motif

Helix-helix interactions are often formed by "ribose zippers" involving H-bonding between the 2'-OH of a ribose in one helix and the O2 of a pyrimidine base (or the N3 of a purine base) of the other helix between their respective minor groove surfaces (Carell, 2004). In practically all ribose zippers, the 2'-OH of adenines interact with the nitrogenous bases of cytosines and/or guanines and their respective 2'-OH (Pley *et al.*, 1994). Ribose zippers allow a bridging connection between two distinct RNA chain segments within the same domain or between different domains (Tamura and Holbrook, 2002). These interactions involve the linking of looped chain segments with stem or stem-like structures. They are found in several functional types of RNA–ribozymes, riboswitches, tRNAs, ribosomal subunits and signal recognition particles.

Chapter I
Section II

Protein
Structure, Biological Role and Mutation

1.2.1 Protein

Like DNA and RNA, proteins too are polymeric biomolecules, but consist of repetitive monomeric units of amino acids. Proteins are more dynamic than nucleic acids because of the larger types of residues involved, increased structural flexibility and lower charge density of the polypeptide backbone. Proteins are structurally complex but functionally by far the most sophisticated. Proteins are the chief actors within a cell, acting as enzymes, structural components, membrane components, templates, substrates and products of many reactions. Protein function expands into a wide range including, including catalyzing metabolic reactions, DNA replication, responding to stimuli, enzymatic catalysis, providing rigidity and mechanical support to nuclei, transmission of nerve impulses, and transporting molecules from one location to another, *etc.* The proteins we observe in nature have evolved through selective pressure to perform specific functions. Proteins occur in a large variety; thousands of different types, ranging from relatively small peptides to huge polymers with molecular weights in millions. The structure and chemistry of each protein has been developed and fine-tuned over billions of years of evolutionary history (Nelson, 2004). Long stretch of specific amino acids that form a protein determines its 3D shape which then determines its function. A comparative structural study of the wild-type and mutant lamin A protein involving its C-terminal immunoglobulin-fold (Ig-fold) domain is reported in Chapter IV of this thesis.

1.2.1.1 Basic Structure of Protein

Proteins are polypeptides formed from specific sequence of closely placed amino acids that are linked to each other through covalent peptide bonds, hence the name polypeptide. The amino acid sequence in a polypeptide chain directs the structure of a protein by generating compact domains with specific 3D pattern.

An amino acid residue is defined by an amino group and a carboxyl group connected to an alpha carbon to which is attached a hydrogen atom and a side-chain group R **(Figure 1.2.1)**. The smallest amino acid, glycine, has hydrogen atom in place of the R-group, whereas the others have distinctive R groups. In a protein, amino acids are connected by peptide bonds **(Figure 1.2.2)**, which are formed by the

condensation of carboxyl (–COOH) group of one amino acid and the amino group (–NH_2) of the next. A peptide consists of a small number of amino acids and a longer chain of amino acids joined in this manner is called a polypeptide. The term protein usually is used to describe the functional unit, which may consist of one or more peptide chains. The direction of the polypeptide chain is defined according to the orientation of the peptide bonds. The amino acid at one end of the chain having a free –NH_2 group, forms the amino- or N-terminal. The amino acid at the other end has a free –COOH group, which forms the carboxyl- or C-terminal **(Figure 1.2.2)**. The protein sequences are conventionally written from N-terminus to C-terminus.

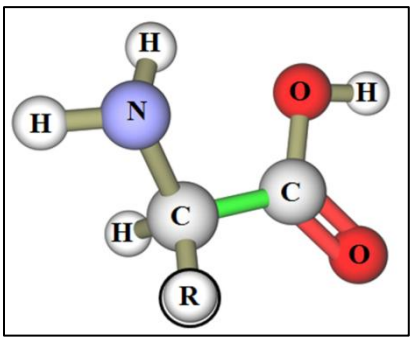


Figure 1.2.1: An amino acid residue. Glycine has H in place of R-group.

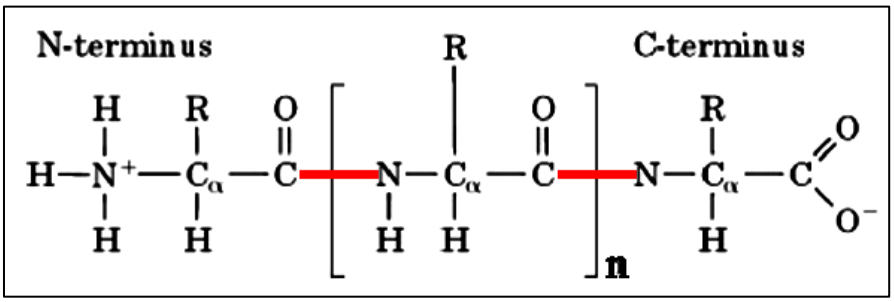


Figure 1.2.2: A polypeptide chain containing amino acids connected by peptide linkages. The N- and C-termini have free –NH_2 and –COOH groups, respectively.

There exist twenty different types of amino acids which are classified into three different classes on the basis of chemical nature of the side-chain: the first class comprises of amino acids with strictly hydrophobic side-chain: Ala, Val, Leu, Ile, Phe, Trp, Pro and Met; the second class consist of the four amino acids having charged residues Asp, Glu, Lys and Arg; and the third class comprises those with polar side-chain, *i.e.* Ser, Thr, Cys, Asn, Gln, His and Tyr. The simplest amino acid glycine (Gln) has special properties and is usually considered either to form a fourth class or to belong to the first class. All amino acids with exception of glycine are

chiral molecules and can exist in two different forms: the L and D forms. Biological systems can recognize chirality, hence all the amino acids forming in protein are in L-form. Moreover, these amino acids can also be classified into five classes depending on the polarity of the –R group **(Figure 1.2.3)** (Nelson, 2004). The distinct property of these side-chains (–R groups) play significant role in biomolecular interaction and recognition.

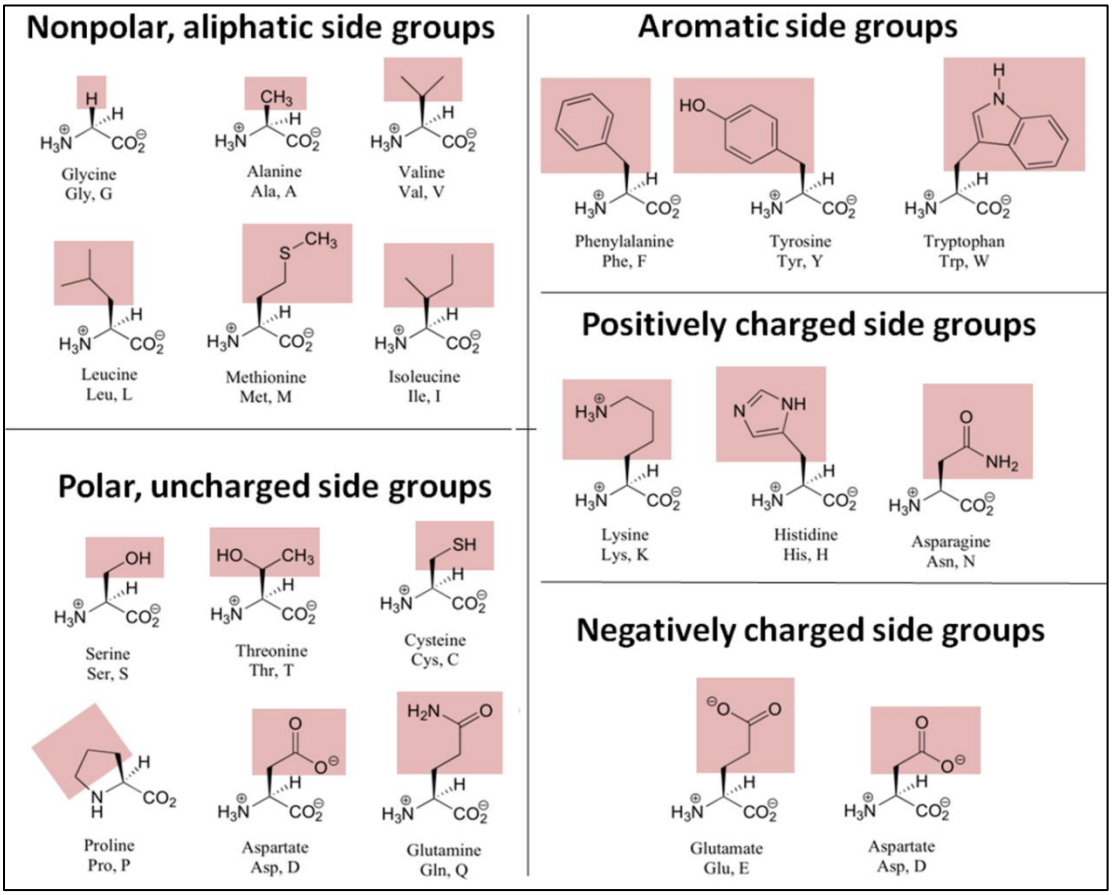


Figure 1.2.3: The 20 amino acids have been classified into five categories as described in (Nelson, 2004).

The Danish biochemist Kai Linderstrom-Lang coined the terms "primary", "secondary" and "tertiary" structure to emphasize the structural hierarchy in proteins. In fact, for monomeric proteins having one chain, there are three hierarchical levels of structure, while for multimeric proteins comprising of many chains, there is yet another level of structure known as the "quaternary" level. A brief description of the structural hierarchy in proteins is provided below (Branden and Tooze, 1992).

1.2.1.2 Primary Structure

Primary structure of protein is basically the arrangement of amino acid residues along a linear polypeptide chain from N-terminal to C-terminal. The primary structure of a protein is determined by the gene corresponding to the protein. The formation of successive peptide bonds during protein synthesis generates a "main-chain" or "backbone". The differences amino acids in the sequence profoundly influence protein stability and function.

The protein main-chain can be divided into peptide repeating units, each ranging from one C_α atom to the next C_α atom **(Figure 1.2.4a)**. Since the peptide units are effectively rigid and linked into a chain by covalent bonds formed by the C_α atoms, the only degrees of freedom they have are rotations around the C_α-C and N-C_α bonds. By convention, the angles of rotation around the N-C_α and C_α-C bonds are called phi *(ϕ)* and psi *(ψ),* respectively **(Figure 1.2.4a)**. In this way each amino acid residue is associated with two conformational angles, ϕ and ψ. A third possible torsion angle within the protein backbone (called omega, ω) is essentially flat and fixed to 180° (the typical *trans* case). This is due to the partial double-bond character of the peptide bond, which restricts rotation around the C-N bond containing C, O, N and H atoms in the peptide plane between two successive C_α-atoms. In case of X-Proline (X is any amino acid), the peptide is often in *cis* orientation and ω is restricted 0°. Since these are the only degrees of freedom, conformation of the main-chain of the polypeptide is completely determined when the torsion angles: ϕ, ψ and ω are determined with high accuracy for each amino acid.

With the objective of finding stable conformations G. N. Ramachandran (Ramachandran *et al.,* 1963) used computer models of small polypeptides of two linked peptide units and systematically varied the ϕ and ψ angles. Atoms were treated as hard spheres with dimensions corresponding to their respective van der Waals radii. For each conformation, the structures were examined for close contacts and other steric interactions. This severely restricted the variety of 3D arrangements of atoms (or conformations) possible. The Ramachandran plot provides an overview of allowed and disallowed regions of torsion angles. Torsion angles are the most important local structural parameters that control protein folding, and if we would have a way to predict the Ramachandran angles for a particular protein, we would be

able to predict its 3D folding. In the Ramachandran plot **(Figure 1.2.4b)** the white area corresponds to sterically disallowed conformations for all amino acids except glycine, which lacks a side-chain. The red region corresponds to conformations that are allowed namely: the alpha-helix and beta-sheet. The yellow area depicts the allowed regions if slightly shorter van der Waals radii are used in the calculations, *i.e.* the atoms are allowed to come a little closer. This brings out an additional region, which corresponds to the left-handed alpha-helix. L-amino acids cannot form extended regions of left-handed helix but occasionally individual residues adopt this conformation.

The side-chain dihedral angles of proteins are denoted as χ_1-χ_5, depending on the length of the side-chain. The χ_1 dihedral angle is defined by atoms N-C^{α}-C^{β}-C^{γ}, the χ_2 dihedral angle is defined by atoms C^{α}-C^{β}-C^{γ}-C^{δ}, and so on. The side-chain dihedral angles tend to cluster near 180°, 60°, and -60°, which are called the *trans*, *gauche*$^{+}$, and *gauche*$^{-}$ conformations. The choice of side-chain dihedral angles is also affected by the associated backbone and neighbouring side-chain dihedrals.

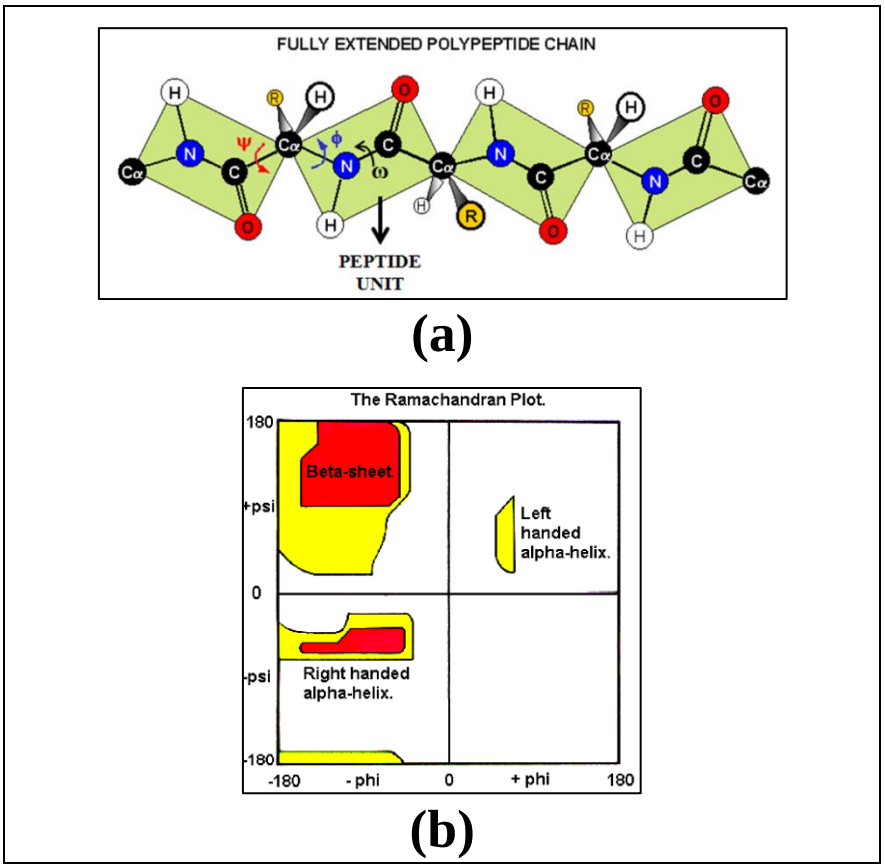


Figure 1.2.4: (a) Peptide repeating units, each ranging from one Cα-atom to the next Cα-atom. The *ω, ϕ and ψ* torsion angles are shown. (b) Ramachandran Plot.

1.2.1.3 Secondary Structure

Secondary structure of a protein refers to the regular recurring arrangement of adjacent amino acids of a polypeptide chain in space, stabilized by H-bonds formed between amine hydrogen (N-H) and carbonyl oxygen (C=O) atoms contained in the backbone of the protein without involving the side-chains of the amino acids. The secondary structure to some extent is determined by the amino acid primary sequence. Polypeptide chains are not free to take up any random 3D structure. Steric constraints and many weak interactions demand that only some arrangements are energetically allowed. The classic secondary structural elements: α-helix and β-sheet are the two regular folding patterns often found in proteins. Certain amino acids favour α-helices and some β-sheets, while others favour formation of loop (coil/unstructured region).

1.2.1.3.1 α-Helix

The α-helix is an important element of protein secondary structure. Linus Pauling initially predicted from the crystallographic analyses of the structures of a range of small molecules, that it would be stable and energetically favourable in proteins. The α-helix is generated when a single polypeptide chain twists around on itself to form a rigid cylinder **(Figure 1.2.5a)** and all the ϕ, ψ angles are approximately -60^{o} and -50^{o} respectively, corresponding to the allowed region in the bottom left quadrant of the Ramachandran plot. The stability of α-helix is affected by total dipole moment of the entire helix due to individual dipoles of the C=O groups involved in H-bonding. Stable α-helices typically end with a charged amino acid to neutralize the dipole moment. The α-helix has 3.6 residues per turn along with a rise of 1.5Å along the helical axis, with H-bond between C=O of residue "n" and N-H of residue "n+4" for several consecutive values of n. Thus, all N-H and C=O groups are engaged in H-bonds except the first three N-H and the last three C=O groups of the α-helix **(Figure 1.2.5a)**. As a result, the ends of polar α-helices are at the surface of protein molecules almost always.

The amino acids that make up a particular helix can be plotted on a helical wheel, a representation that illustrates the helix structure by projection of the Cα backbone. Often in globular proteins, an alpha helix will exhibit two "faces": one

containing predominantly hydrophobic amino acids oriented toward the interior of the protein, in the hydrophobic core, and one containing predominantly polar amino acids oriented toward the solvent-exposed surface of the protein.

The α-helices which are loosely or tightly coiled with respect to the standard helix, involving H-bonding of residue "n" to residues "n+5" or "n+3" instead of "n+4" are called the π helix and 3_{10} helix, respectively. The 3_{10} helix has three residues per turn and contains ten atoms between the H-bond donor and acceptor, hence its name. Both the π and 3_{10} helices occur rarely and usually only at the ends of α-helices or as single turn helices. These are not energetically as favourable as the α-helix, since the backbone atoms are either tightly or loosely packed compared to the regular helix and the H-bonds are not linear.

In proteins, the α-helix is generally right-handed, since a left-handed helix with L-amino acids will result in close contact of the side-chains and C=O groups. However, short stretches of left-handed α-helices (3-5 residues) can occur occasionally with glycin residues which do not have bulky side-chain. Under certain cellular environment, the arrangement of protein structure is such that, the hydrophilic polypeptide backbone is H-bonded to itself in the α-helix form and thus, shielded from the hydrophobic lipid environment of the membrane by its protruding non-polar side-chains. In others, the helices wrap around each other to form a particular stable structure, known as a coiled-coil. This structure can form when the two (or in some cases three) α-helices have most of their hydrophobic side-chains facing another set of non-polar residues of the nearby helix, so that they can twist around each other with these side-chains facing inward. In globular proteins the length of the α-helices vary considerably, ranging from 4-5 amino acids to over 40 residues.

1.2.1.3.2 β-Sheet and β-Bulge

β-sheet: In contrast to α-helix, which involves a continuous region of polypeptide chain, the β-sheets are assembled from a combination of several regions of the polypeptide chain. A β-sheet is formed from several β-strands which are usually 5-10 residues long and are "pleated" with C_α atoms successively a little above and below the plane of the β-sheet. The polypeptide chains involved are in an almost fully extended conformation with the ϕ,ψ angles lying within the broad structurally allowed

region in the upper left quadrant of the Ramachandran plot ($\phi, \psi \approx -80^{o},+150^{o}$). These β-strands are aligned in adjacent positions to form H-bonds between C=O group of one β-strand and N-H group of an adjacent, or *vice versa*. The β-strands can interact in two ways to form two different pleated β-sheets having distinctive H-bonding pattern: (i) all the amino acids in the aligned β-strands run in the same biochemical direction, N-terminal to C-terminal, in which case the sheet is described as "parallel" or (ii) the amino acids in successive strands have alternating directions as a polypeptide chain can folds back and forth upon itself, in which case the sheet is called "anti-parallel" **(Figure 1.2.5b)**. Parallel β-structure almost never occurs in sheets of less than five total strands, whereas anti-parallel β structure often occurs as a twisted ribbon of just two strands. Both the parallel and anti-parallel β-sheets form very rigid structures, held together by H-bonds that connect the peptide bonds in neighbouring chain. In both these forms, the side groups along each strand alternate above and below the sheet, while side groups opposite one another on neighboring strands extend to the same side of the sheet and are quite close together. These close side-chain pairs on neighboring strands show preferences for having hydrophobic groups together, unlike charges together, and branched β-carbons next to unbranched β-carbons (in anti-parallel sheet).

The β-strands can also combine into mixed β-sheets with a mixture of parallel and anti-parallel strands **(Figure 1.2.5b).** Almost all β-sheets, whether they are parallel, anti-parallel or mixed, have the same right-handed twisted strands.

The β-sheets in general show a tendency toward greater hydrophobicity for the central than for the edge strands of the sheet (Sternberg and Thornton, 1977). Parallel β-sheets and the parallel portions of mixed sheets are always thoroughly buried, with other main chain (often α-helices) protecting them on both sides. Anti-parallel sheets, on the other hand, typically have one side exposed to solvent and the other side buried, so that they often show an alternation of side-chain hydrophobicity in the amino acid sequence (Richardson, 1981) . Moreover, the parallel β-structure is less stable because the geometry of the individual amino acid molecules forces the H-bonds to occur at an angle, making them longer and thus weaker. Contrarily, in the anti-parallel arrangement the H-bonds are aligned directly opposite each other, making for stronger and more stable bonds (Richardson, 1977).

β-bulge: The β-bulge (Richardson *et al.*, 1978) is a small piece of non repetitive structure that most often occurs, and is most easily visualized, as an irregularity in anti-parallel β-sheet. It is defined as a region between two consecutive β-type H-bonds that includes two residues on one strand opposite a single residue on the other strand. Bulges can slightly affect the directionality of the polypeptide chain. β-bulges can be classified into several different types, but the most common is the "classic" β-bulge, which occurs between a narrow pair of H-bonds on anti-parallel strands and has the side-chains of all three amino acid residues on the same side of the β sheet.

A globular protein containing anti-parallel β-sheets and β-bulges has been studied by simulations and reported in Chapter IV of this thesis.

1.2.1.3.3 Loop Regions

Most protein structures consist of a combination of secondary structure elements (α-helices and β-strands) connected by loop regions (unstructured region) of various lengths and irregular shapes. A combination of secondary structure elements forms the stable hydrophobic core of the protein molecule. The main chain C=O and N-H groups of the loop regions, which in general are not involved in H-bonding, are exposed to the solvent and can form H-bonds to water molecules. Usually, the exposed loop regions are rich in charged and polar hydrophilic residues. When homologous amino acid sequences from different species were compared, it was found that insertions and deletions of a few residues occur almost exclusively in the loop regions. Analysis of known 3D structures of loops has shown that the loop region falls into a limited set of structures and are not at all random. Loop regions that connect two adjacent anti-parallel β-strands are called hairpin loops **(Figure 1.2.6a)**. Short hairpin loops are generally called reverse turns or simply turns. Type I and Type II turns **(Figure 1.2.6b)** are two of the most frequently occurring turns. The Type II turn usually has a glycine residue as the second of the two residues in the turn. Long loop regions are often flexible and can frequently adopt several conformations, making it difficult to be characterized by X-ray structure and NMR studies. Such loops are frequently involved in protein function and can switch from an "open"

conformation, which allows access to the active site, to a "closed" conformation, which shields reactive groups in the active site from water.

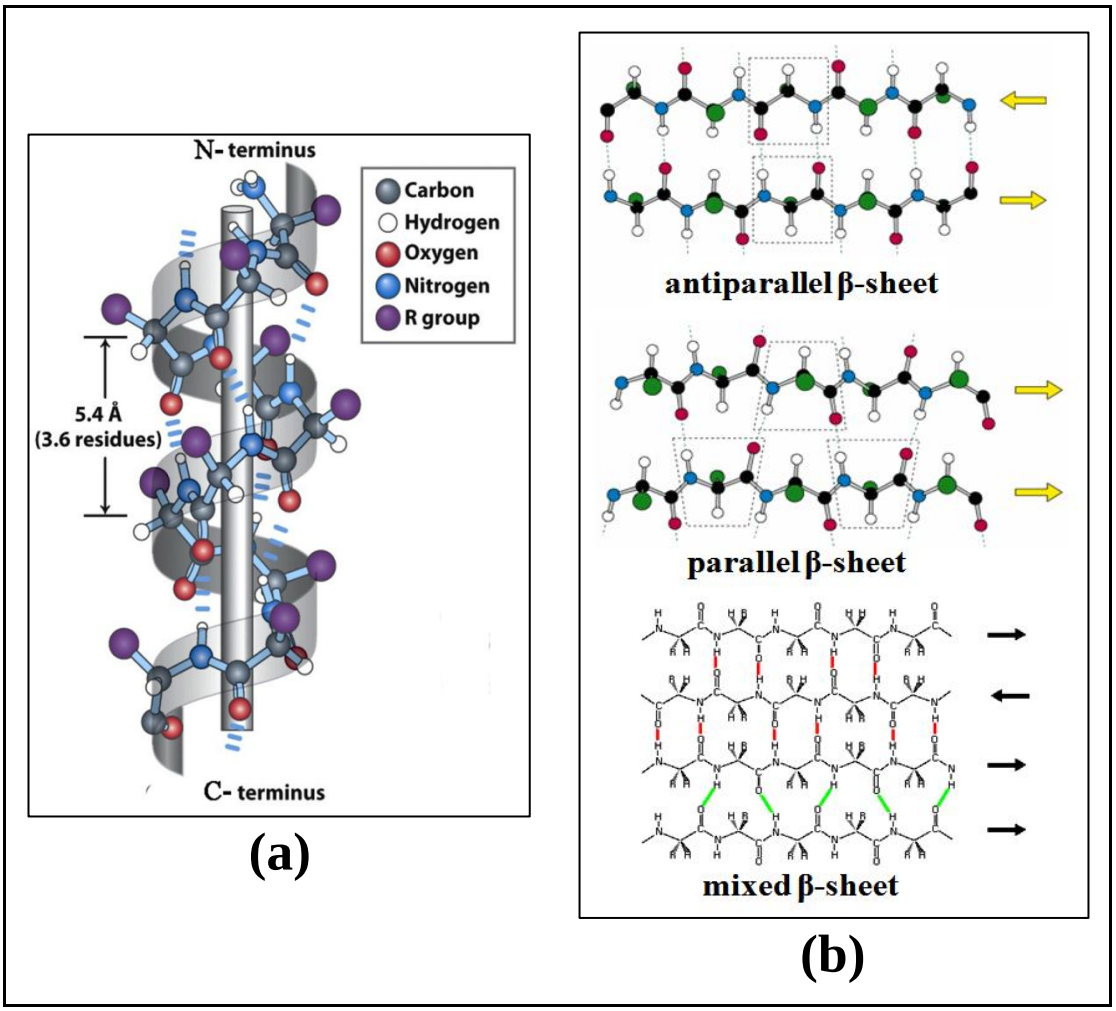


Figure 1.2.5: (a) The α-helix. (b) The anti-parallel, parallel and mixed β-sheets.

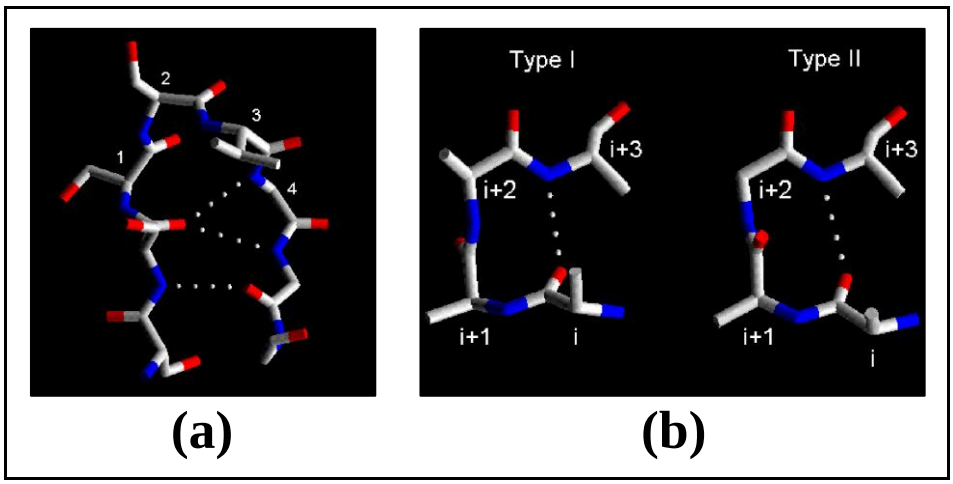


Figure 1.2.6: (a) Hairpin loop. (b) From Left to Right, Type I and Type II hairpin loop.

1.2.1.4 Supersecondary Structures or Motifs

Simple combinations of few secondary structure elements with a specific geometric arrangement occur frequently in protein structures. These units are called either supersecondary structures or motifs. Some of these motifs are associated with a particular function such as DNA binding, while others have no specific biological function alone but are a part of larger structural and functional assemblies (Branden and Tooze, 1992).

The simplest motif with a specific function consists of two α-helices joined by a loop region. Two such motifs, each with its own characteristic geometry and amino acid sequence, have been observed as parts of many protein structures. One of these motifs, called the helix-turn-helix (HTH) motif **(Figure 1.2.7a)**, is specific for DNA binding. Another motif called the EF hand motif **(Figure 1.2.7a)** is specific for calcium binding and is present in parvalbumin, calmodulin, troponin-C and thereby regulate cellular activities.

The simplest motif involving β-strands is adjacent anti-parallel strands joined by a loop, called the hairpin β-motif. These occur quite frequently in protein structure and are present in most anti-parallel β-structures both as isolated ribbons and part of the more complex β-sheets. Another motif involving β-strands is the Greek key motif, which is found in anti-parallel β-sheets where four adjacent β-strands are arranged in a pattern resembling the topology of a Greek key. This motif has not yet been associated with any specific function, but occurs frequently in protein structures. The polypeptide chain connecting two parallel β-strands should cross from one end to the other and such crossovers are frequently made by α-helices. The polypeptide chain must turn twice using loop regions and the motif that is formed is thus, a β-strand followed by a loop, α-helix, another loop, and finally, the second β-strand. This motif is known as the β-α-β motif and is found as part of almost every protein structure that has a parallel β-sheet. The Rossmann fold is one such β-α-β motif found in proteins that bind nucleotides, such as cofactor FAD, NAD and NADP (Hanukoglu, 2015). The core of many proteins contains extensive regions of β sheet. A large β-sheet sometimes twists and coils to form a closed structure in which the first strand is H-bonded to the last. This type of structure is known as β-barrel and the β-strands are typically arranged in an anti-parallel fashion. Another supersecondary structure is β-

meander motif. Diagrammatic representations of almost all possible supersecondary structures are given in **(Figure 1.2.7b)**.

1.2.1.5 Tertiary Structure

The tertiary structure of a protein refers to its geometric shape given by full 3D organization of a polypeptide chain. Unlike the secondary structure of polypeptide chain, which is determined by short-range interactions of backbone amino acid residues, the tertiary structure is conferred by longer range aspects and encompasses the efficient packing of several different side-chains along the polypeptide backbone. Cross-linkage between multiple linear polypeptide chains is mostly observed via disulfide bonds (S-S bonds). Only the cysteine amino acid is capable of forming a disulfide bond. These bonds are responsible for stabilizing the globular structure of a protein and are the strongest intramolecular interaction that a protein can possess. These are one of the major forces responsible for holding proteins in their respective conformations and therefore play an important role in protein folding and stability. Thus, the protein tertiary structure is formed by packing several secondary structural elements and motifs into one or several compact globular units, called domains. A domain is a fundamental unit of organization. A structural domain is a unit of the tertiary structure and the functional domains are the ones acting as functional units. In many cases they also act as functional units and are called as functional domains.

The complete folded 3D structure of a protein may comprise of a single domain or as many as several dozen domains. On the basis of simple considerations of connected motifs, Levitt and Chothia derived taxonomy of protein structures and classified domain structures into three main groups: (i) α-domains, (ii) β-domains and (iii) α/β domains **(Figure 1.2.8)** (Levitt and Chothia, 1976).

In α-domains the core is built up exclusively of α-helices; in β-domains the core comprises of β-sheets and are usually two β-sheets packed against each other. On the other hand α/β domains are made from predominantly parallel β-sheets surrounded by α-helices. Some proteins are built up from a combination of discrete α and β motifs and usually form one small anti-parallel β-sheet in one part of the domain packed against a number of α-helices. These structures can be considered as a fourth category called α+β domain. In addition to these four groups, there are proteins that are strongly influenced by the presence of metal ions and disulfides and often look like

distorted versions of more regular proteins (Branden and Tooze, 1992). Each different combination of domains is known as a protein fold. So far, about thousand different protein folds have been identified from the ten thousand proteins whose detailed conformations are known.

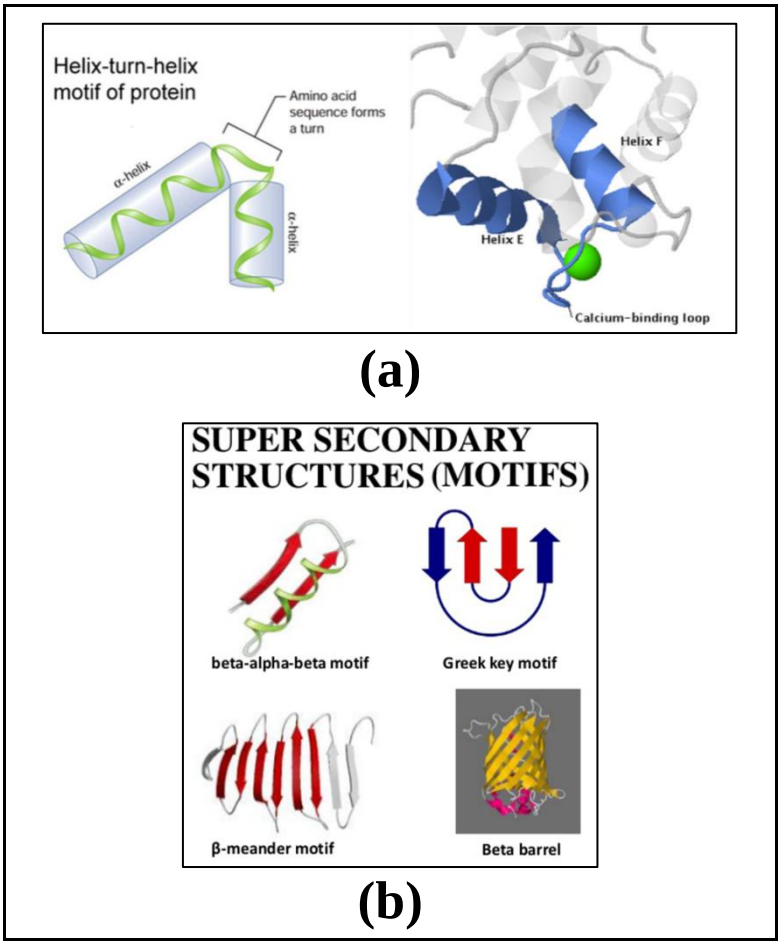


Figure 1.2.7: (a) From Left to Right, HTH motif and EF motif. (b) β-α-β, Greek key, β-meander and β-barrel motifs.

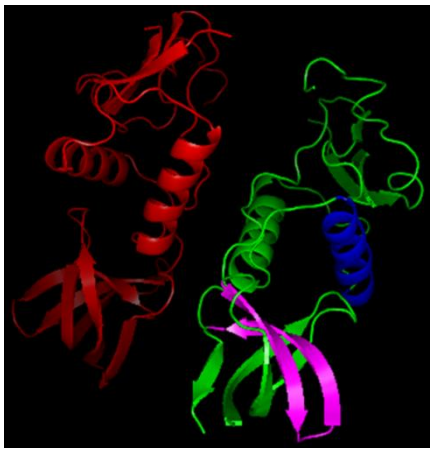


Figure 1.2.8: An α/β-motif.

1.2.1.6 Quaternary Structure

The proteins containing more than one polypeptide chain exhibit a fourth level of structural organization. Each polypeptide chain in such a protein is called a subunit. Quaternary structure refers to the spatial arrangement of these subunits and the nature of their interactions. There are two major categories of proteins based on quaternary structure - fibrous and globular. The simplest quaternary structure is a dimer, consisting of two identical subunits. This organization is present in the DNA-binding CRO-repressor protein found in a bacterial virus called λ. More than one type of subunits, often in variable numbers can be present in a quarternary structure. For example, human haemoglobin consists of two subunits of same type (designated as α) and two subunits of another type (designated as β). The subunits can function either independently of each other, or cooperatively such that the function of one subunit is dependent on the functional state of others. A variety of bonding interactions including H-bonding, salt bridges, and disulfide bonds hold the various subunits into a particular 3D conformation.

In case of globular proteins, the polypeptide chain folds up into a compact shape like a ball with an irregular surface. Enzymes tend to be globular, even though many are large and complicated, with multiple subunits, most have an overall rounded shape. In contrast, there are proteins that play such roles which require each individual protein molecule to span a large distance. Thus, these proteins generally have a relatively simple, elongated 3D structure and are commonly referred to as fibrous proteins. Fibrous proteins are especially abundant outside the cell, where they are the main component of the gel-like extracellular matrix that helps to bind collections of cells together to form tissues. As for example, collagen, the main structural protein of the various connective tissues in animals has elongated fibril form and is mostly found in fibrous tissues such as tendons, ligaments and skin.

The same principles that enable a protein molecule to associate with itself to form rings or filaments operate to generate much larger structures in the cell - supramolecular structures such as enzyme complexes, ribosomes, protein filaments, viruses, and membranes. The information required for formation of the complex assemblies of macromolecules in cells must be contained in the subunits themselves, because purified subunits can spontaneously assemble into the final structure under

the appropriate conditions. The formation of protein 3D conformation from primary amino acid sequence to a quaternary structure is depicted pictorially in **Figure 1.2.9.**

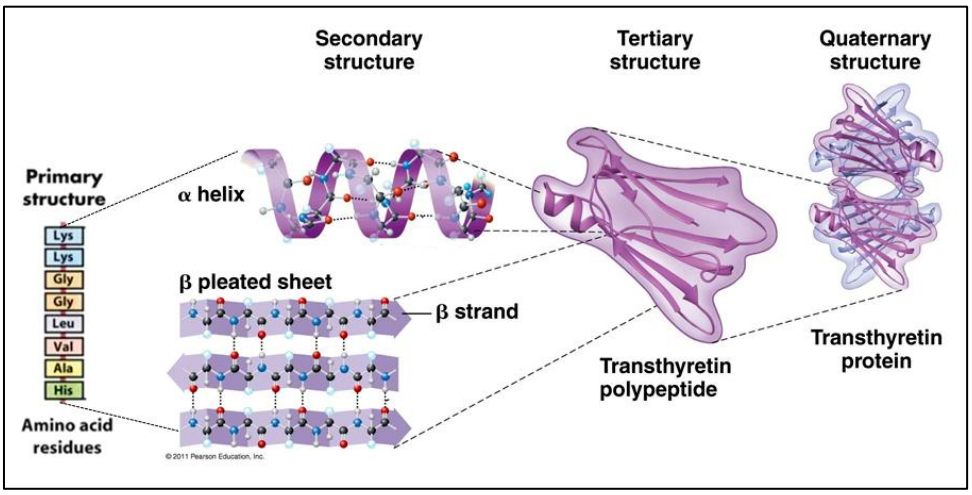


Figure 1.2.9: The journey of protein quaternary structure formation starting from primary amino acid sequence.

1.2.2 Globular Protein

On the basis of solubility, function and 3D shape, proteins may be classified as fibrous, membrane or globular. Globular proteins or spheroproteins are spherical and somewhat water-soluble (forming colloids in water), unlike the fibrous or membrane proteins. The spherical structure of a globular protein is induced by its tertiary structure such that hydrophobic amino acids are placed towards the protein's interior whereas hydrophilic or polar amino acids protrude outwards allowing dipole-dipole interactions with the solvent, accounting for the protein's solubility. Furthermore, the polar side-chains of amino acids tend to exert strong attractive forces toward other polar groups within the protein molecule, or toward polar molecules in the protein's surroundings. Similarly, non-polar side-chains exert attractive toward other non-polar side-chains within the protein. The shape assumed by a globular protein molecule tends to maximize both types of attractive forces, whereby non-polar side-chains "point" inward and interact with one another and polar side-chains are oriented outward such that they are exposed to adjacent polar water molecules.

Globular proteins play many biological roles acting as enzymes, hormones, transport molecules and immunoglobulins. However, the shape of globular proteins is critical to their function and changes in environmental conditions, such as pH or

temperature, may affect the bonds that contribute to the shape of globular proteins, in turn lead to denaturation in which the shape of the protein is so altered that it is no longer functional. Albumins are the globular proteins that are completely soluble in water. Other common globular proteins are hemoglobin (a member of the globin protein family), alpha, beta and gamma globulins, and the immunoglobulins (IgA, IgD, IgE, IgG and IgM). Structural studies involving IgA domain of human lamin protein have been performed extensively and reported in Chapter IV of this thesis.

1.2.3 Protein Mutation

Mutations are both the raw material and the driving force of evolution. The selection for mutation at the protein level is in general much more stringent than at the DNA level. The very consequences of mutations on protein structure depend on the location of the affected nucleotides in the gene those codes for the polypeptide chain. Mutations of parts of a gene other than the protein-coding region can affect production of and even the structure of its protein. The first genetic disease (caused due to mutation) to be characterized at the molecular level, is sickle-cell hemoglobin. In this case, the normal Glu at residue 6 of the hemoglobin beta chain is replaced by Val as a result of mutation. In some cases, a single mutation is sufficient to alter the activity of the protein whereas in others multiple mutations are necessary. More substantial and complex alterations of genes have correspondingly greater effects on protein structure (Creighton, 1993).

Protein evolution involves changes of single residues, insertions and deletions of several residues, gene doubling, gene fusion and splice variants. The acceptance of amino acid changes critically depends on the biological role of the respective amino acids. However, detailed protein biographies are necessary to assess the biological importance of a given residue. It is difficult to assess these biological importances because it depends not only on the function of a protein but also on all the interactions of a protein with other parts of the organism over the protein's entire lifespan. Nevertheless, some general observations have been made (Schulz and Schirmer, 2013):

(i) Residues at the protein surface change more frequently than do the internal ones.

(ii) The effect of an amino acid change in the protein interior is often compensated for by other changes.

(iii) Changes between residues with similar chemical properties occur more frequently than others.

Strictly speaking, probabilities of accepted mutations vary with residue type and are referred to as mutation probabilities. The amino acid Ser, which is usually found at the protein surface, has the largest mutation probability. On the other hand, Trp has the smallest mutation probability, reasonable since Trp is generally an internal residue and cannot be replaced by a side-chain of equal length (Schulz and Schirmer, 2013).

A mutation has its worst impact on the folding dynamics of the polypeptide chain that forms the protein. All accepted mutations in proteins can be taken as natural experiments, which show us variations that affect slightly protein stability and folding dynamics. In addition, random and probably unacceptable mutations represent variations that decrease protein stability by a measurable amount. Study of both these types of mutations along with their respective native protein can be used to improve our knowledge of non-covalent forces acting in proteins.

For this purpose, energy minimization procedures are executed for the wild-type and the mutant proteins on the basis of the known 3D structures. The resulting energy differences and geometric deviations have are compared to experimental data from thermodynamic measurements and high resolution X-ray analyses, respectively. Keeping on the same line, the wild-type and mutant lamin A Ig-domain has been simulated by both Molecular Dynamics and Steered Molecular Dynamics resulting in a better understanding of the effect of mutation, which has been discussed in Chapter IV of this thesis.

1.2.4 Protein Folding and Unfolding Under Force

Cellular functions such as chromosomal segregation, transcription, translation, protein and nucleic acid folding and unfolding, and cell locomotion all involve mechanical forces (Bustamante *et al.*, 2004; Oberhauser and Carrión-Vázquez, 2008). Protein unfolding and refolding are fundamental biological events, yet they are not completely understood. The advent of single molecule force spectroscopy has enabled the application of mechanical force to unfold protein molecules, which in turn has opened doors for characterizing the energy landscape of protein folding and its response to mechanical stress, a biologically important perturbant (Jagannathan and

Marqusee, 2013). In spite of high accuracy, data from these experiments are still lacking a clear molecular picture of the effect of force on the disruption of the folded state. Thus, a merging of novel computational, theoretical and experimental approaches is required.

Steered Molecular Dynamics (SMD) simulations have been applied to investigate the response of protein domains to stretching apart of their terminals. The simulations mimic atomic force microscopy (AFM) and optical tweezers experiments, but proceed on much shorter time scales (Lu and Schulten, 1999). Such SMD simulations have been performed on the wild-type and mutant lamin A Ig-domain and discussed in Chapter IV of this thesis.

Chapter I
Section III

Graphene and Other Biologically Important Molecules
Structure, Properties and Biological Significance

1.3.1 Graphene

Carbonaceous materials have come out as the most promising material with possible application in the interface between biology and material science. One such material, graphene is a rapidly rising star (Geim and Novoselov, 2007) on the horizon of various fields starting from material science to condensed matter physics to even nano-biotechnology. Graphene represents a conceptually new class of materials that are only one atom thick, thus offering new advances into low-dimensional physics that continues to provide a fertile ground for applications not only in the field of physics but more importantly biology. Andre Geim and Konstantin Novoselov won the 2010 Nobel Prize in Physics for their ground breaking work on fabricating graphene sheets by mechanical exfoliation of highly oriented pyrolytic graphite (HOPG) (Novoselov *et al.*, 2004).

1.3.1.1 Unique properties

Graphene has evoked great interest throughout the scientific community since its discovery as a novel nanomaterial due to its exceptional physical, mechanical and chemical properties (Bunch *et al.*, 2008; Castro Neto *et al.*, 2009; Ivanovskii, 2012; Khare *et al.*, 2007; Lee *et al.*, 2008; Loh *et al.*, 2010; Nair *et al.*, 2008) such as:

(i) Thinnest material with greatest strength (Young's modulus of 1000Gpa) (Geim and Novoselov, 2007) and good flexibility.

(ii) Almost transparent (Castro Neto *et al.*, 2009; Nair *et al.*, 2008).

(iii) Most stretchable crystal (20% elasticity) (Geim and Novoselov, 2007).

(iv) High thermal conductivity (above $3000 WmK^{-1}$) (Geim & Novoselov, 2007).

(v) Highest current density at room temperature (Castro Neto *et al.*, 2009).

(vi) Completely impermeable to any gases (Bunch *et al.*, 2008).

(vii) Highest intrinsic mobility (100 times more than in Si) (Castro Neto *et al.*, 2009).

(viii) Conducting electricity in the limit of no electrons (Castro Neto *et al.*, 2009).

(ix) Large surface area ($\sim 2600 m^2 g^{-1}$) (L. Zhang *et al.*, 2013).

(x) Optical absorption of exactly $\pi\alpha \sim 2.3\%$ (in infrared limit, where α is the fine structure constant (R R Nair *et al.*, 2008).

Apart from these, another property of graphene, that makes it readily usable in various biomedical applications such as bio-sensing, drug delivery vehicle, non-viral

gene delivery vectors, *etc.* (Bao *et al.*, 2012), is that it can be chemically functionalized (Elias *et al.*, 2009; Loh *et al.*, 2010; Nair *et al.*, 2010).

The progress in graphene nanomaterials has provided a fascinating opportunity for biotechnological developments because of their exceptional properties. In comparison with its precursor, carbon nanotube (CNT), graphene exhibits merits like low cost, two external surfaces, facile fabrication and modification, and absence of toxic metal particles. Moreover, flat graphene sheets can be easily complexed with functional nanoparticles for potential multimodal imaging and therapeutic applications (Bao *et al.*, 2012).

1.3.1.2 Graphene Synthesis

Graphene was first ever isolated in 2004 by Geim and Novoselov (Novoselov *et al.*, 2004) using the now famous Scotch tape method based on the pioneering work of Ruoff *et al.* (Lu *et al.*, 1999). This method of mechanical exfoliation is considered to be the best for producing high-quality graphene flakes. However, it is a time consuming process limited to small-scale production. There are several methods of mass-production of graphene such as: Thermal Decomposition of SiC, Chemical Vapour Decomposition (CVD), Molecular Beam Decomposition, Liquid-phase exfoliation, Unzipping Carbon Nanotubes, Sodium-Ethanol Pyrolysis, Graphene Oxide (GO), *etc.* (Cooper et al., 2012). Thus, there exist a wide choice in terms of size, quality and price for any particular application **(Figure 1.3.1)**.

In order to achieve bulk scale production, graphene can be epitaxially grown on silicon wafers under ultrahigh vacuum or prepared by chemical vapour decomposition. Apart from these another economical and mass production method to obtain graphene is chemical or thermal (or sonication) reduction of GO. The chemical methods for GO production were all developed before 1960. However, the most recent and most common is Hummers process (Hummers and Offeman, 1958). In this process, graphite is treated for several hours in an anhydrous mixture of sulfuric acid, sodium nitrate and potassium permanganate, followed by addition of water. The resulting material is hydrated graphite oxide. Subsequent NMR and X-ray diffraction studies have revealed that a GO sheet contains a combination of hydroxide, carbonyl, carboxyl, and epoxide groups are covalently bonded to the graphene lattice.

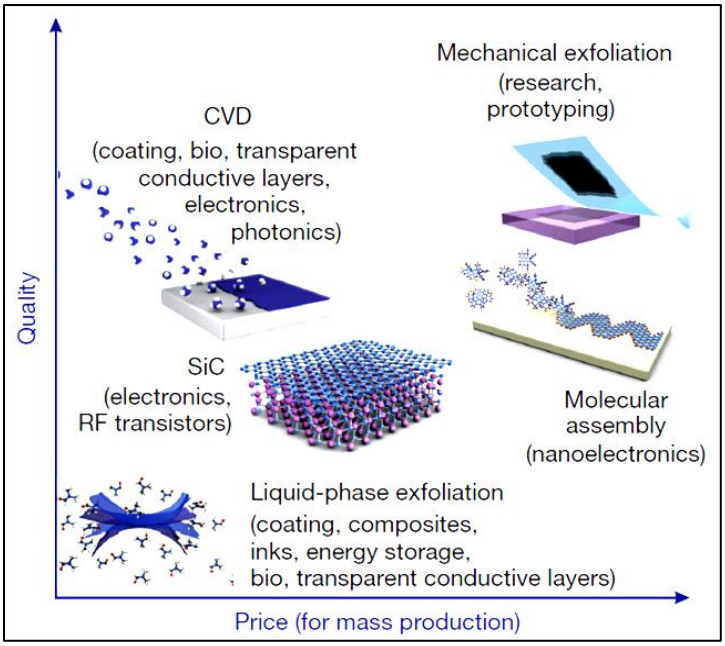


Figure 1.3.1: Graphene synthesis methods in terms of Quality vs. Price of the product, as taken from (Novoselov *et al.*, 2012).

The polar –O– and –OH groups formed during the oxidation of graphitic material makes graphite oxide hydrophilic, and it can be chemically exfoliated in several solvents (Zhu *et al.*, 2010). The graphite oxide solution can then be sonicated to form GO nano-platelets (nanosize sheets). The oxygen groups are then removed by chemical reduction of GO using several reducing reagents. A schematic representation of production of GO from graphite is given in **Figure 1.3.2**. This method was used by Stankovich *et al.* with hydrazine as the reducing agent, however the reduction process could not remove all the oxygen groups (Stankovich *et al.*, 2007). The presence of such polar groups on the basal planes and edges of a GO sheet enables the GO to be functionalized through covalent and non-covalent approaches, hence making it a building block for synthesizing versatile functional nanomaterials (Dreyer *et al.*, 2010; Loh *et al.*, 2010). Scientists and engineers prefer reduction of GO to obtain large quantities of mono- bi- or multi- layer graphene-like materials for large-scale applications. The reduction GO can be achieved by chemical, thermal or electrochemical pathways. One of the proposed schemes for chemical reduction of GO is given in **Figure 1.3.3**. Although, it is the most commonly employed method, but methods involving thermally-mediated reduction and electrochemical reduction can also be applied. Production of graphene by GO treatment can also be attained by other non-chemical methods such as: plasma functionalization, Radio frequency plasma and photoluminescence.

Tearing of GO sheets leading to formation of nanosize graphene or functionalized graphene sheet has been studied by quantum chemical approach and reported in Chapter V of this thesis.

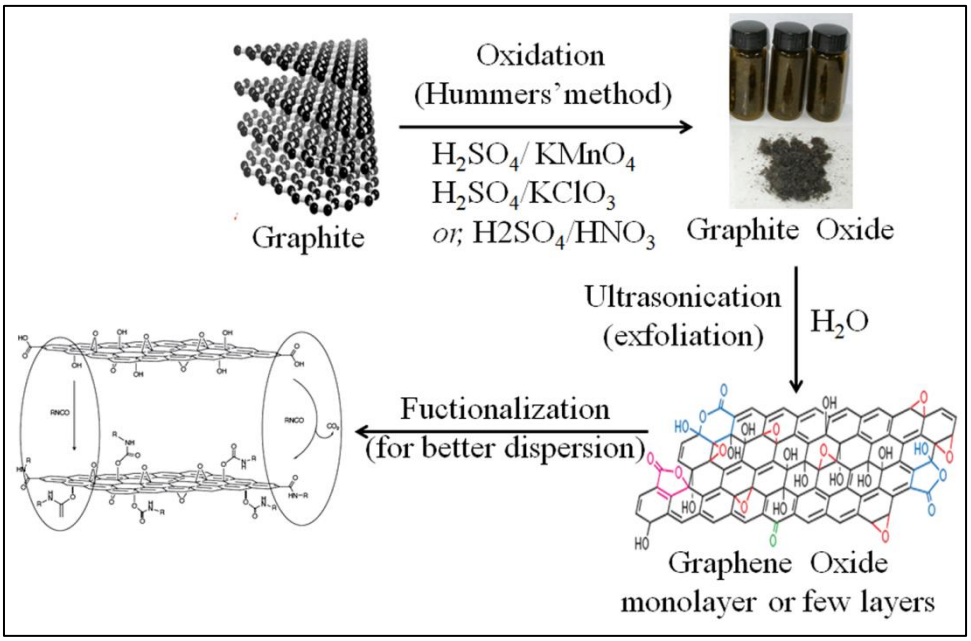


Figure 1.3.2: A schematic representation of production of GO from graphite.

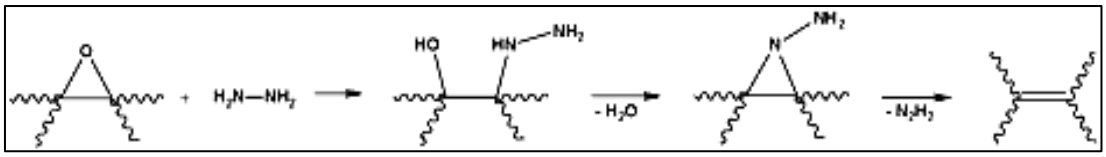


Figure 1.3.3: A proposed scheme for chemical reduction of GO to Graphene.

1.3.1.3 Graphene/GO structure and functionalization

Graphene is the first two dimensional (2D) atomic crystal available to us. It is a single atomic layer of sp^2-bonded carbon atoms tightly packed in a 2D honeycomb lattice **(Figure 1.3.4)**. These graphene sheets can act as building blocks for higher order carbon materials: 2D graphene sheets can be periodically stacked to form 3D graphite; it can be rolled to form 1D nanotubes, wrapped into 0D fullerenes **(Figure 1.3.4)**. In an arbitrarily cut graphene sheet, if one of the edges is arm-chair (*cis*-edge) the other 30° bend cut will be zig-zag (*trans*-edge) **(Figure 1.3.5)**, or vice versa (Banerjee *et al.*, 2005, 2006; Kobayashi *et al.*, 2005). A partial zig-zag-arm-chair edge can be obtained if the cut is made at 60° angle (Kobayashi *et al.*, 2006).The two edges have distinctly different electronic properties, the zig-zag edge accumulates more

electron density as compared to the arm-chair edge (Banerjee *et al.*, 2005, 2006; Banerjee and Bhattacharyya, 2008; Sarkar *et al.*, 2010).

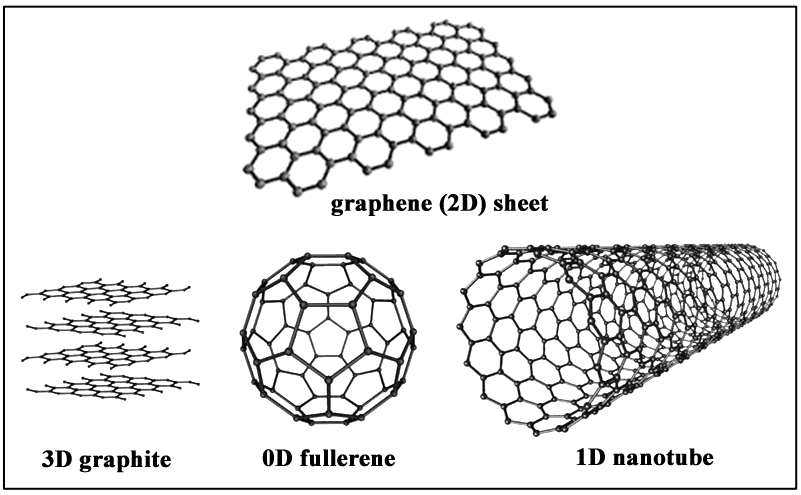


Figure 1.3.4: A 2D graphene sheet which can be periodically stacked to form 3D graphite, rolled to form 1D nanotubes and wrapped into 0D fullerenes.

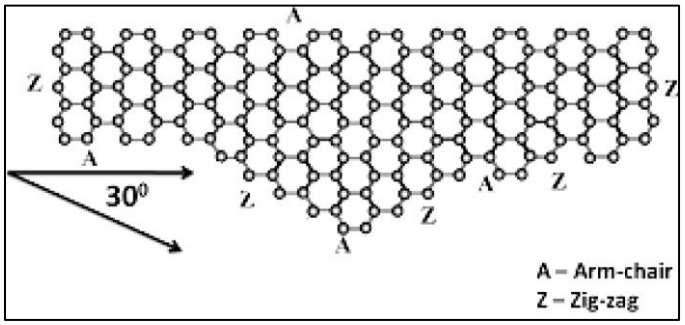


Figure 1.3.5: An arbitrarily cut graphene sheet having both arm-chair and zig-zag edges, taken from (Panigrahi *et al.*, 2011).

As discussed earlier, a promising solution based route for mass production of graphene is chemical reduction of GO. Hence, it becomes necessary to have an understanding of the GO structure. The precise chemical structure of GO has been the subject of considerable debate over the years due to the complexity of the material and lack of precise analytical techniques for characterizing such materials. Many models have been proposed till date (Dreyer *et al.*, 2010) including Hofmann and Holst's structure, Ruess's model, Scholz and Boehm's model, Nakajima-Matsuo's model and the latest proposed by Lerf and Klinowshi (He *et al.*, 1996; Lerf *et al.*, 1998). In 2008 finally, a high-resolution solid state 13C-NMR measurement (Cai *et al.*, 2008) confirmed the existence of C–OH (hydroxyl), C–O–C (epoxide), and sp^2 C-atoms on these layers **(Figure 1.3.6a)**. The data further indicated that a large fraction

of sp^2 C-atoms are bonded directly to C-atoms in the hydroxyl and epoxide groups, and that a large fraction of C-atoms in the hydroxyl and epoxide units are bonded to each other. It has been observed to be energetically favourable for the hydroxyl and epoxide groups to aggregate together and to form specific types of strips with sp^2-C regions in between (Yan *et al.*, 2009) **(Figure 1.3.6b)**.

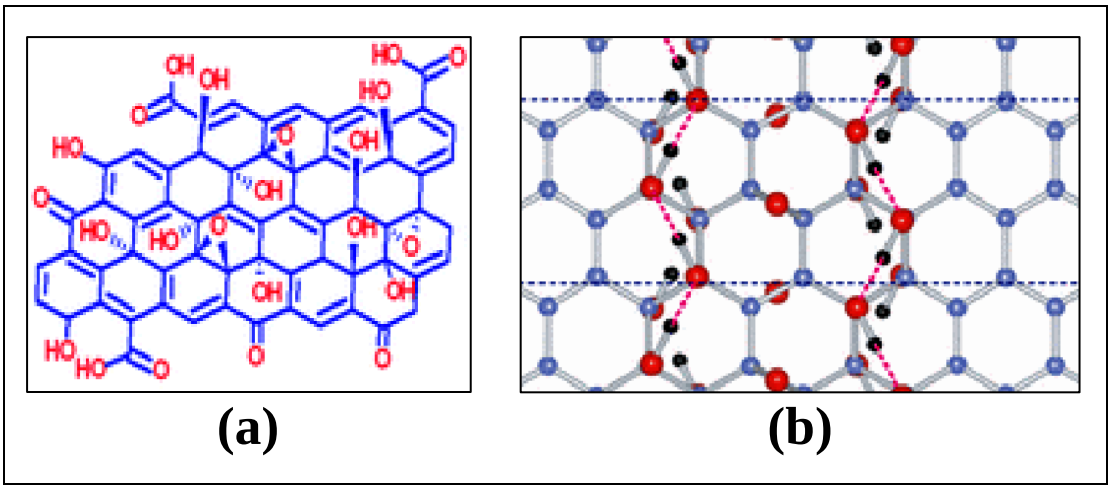


Figure 1.3.6: (a) A high-resolution 13C-NMR of GO. (b) the –OH and –O– groups aggregate together and form specific types of strips with sp^2-C regions in between, as taken from (Yan *et al.*, 2009).

An addition of chemical groups to GO sheets using various chemical reactions that provide for either covalent or non-covalent attachment to the resulting chemically modified graphenes (CMGs). Such approaches which add functionality to groups already present on GO sheets makes the GO a more versatile precursor for a wide range of applications. Reduced GOs too have been frequently modified by non-covalent physisorption of both polymers and small molecules onto their basal planes via π-π stacking or van der Waals interactions, thus, adding functionality to the resultant nanosize graphene sheets. A few structural representation of functionalized graphene/GO sheets are given in **Figure 1.3.7**. A few functionalized graphene systems with functionalization by –NH2, –COOH and –OH at the edges is given in Chapter V of this thesis.

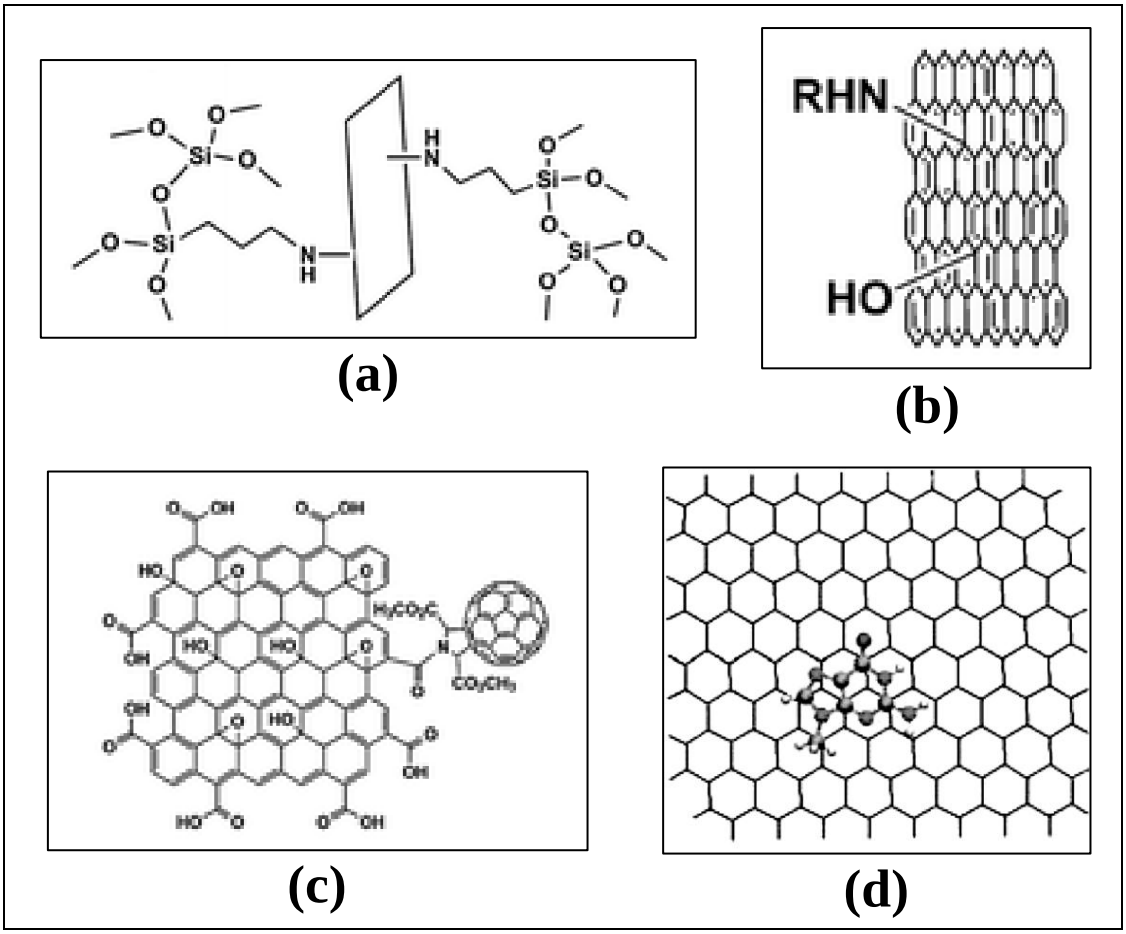


Figure 1.3.7: Funtionalization of GO/graphene sheet by (a) covalent attachment of silane groups, (b) an ionic liquid (R = 3-(3-methylimidazolium)propane), (c) covalent attachment of fullerenes (Dreyer *et al.*, 2010), and (d) non-covalent interactions with nucleobase**.**

1.3.1.4 Biological Significance of Nanosize Graphene/GO sheets

Carbonaceous graphene based nanomaterials has been progressively applied in the biomedical field since its discovery. Graphene-based sensors and biosensors; graphene as nanocarriers for drug delivery, gene delivery and nanomedicine for potential cancer therapies; graphene-based matrices, nanoscfolds, and composites; are the broad categories or bioapplication of graphene. Graphene exhibits excellent capability in fluorescent resonanace energy transfer (FRET) (Bao *et al.*, 2012). Graphene and GO have shown superior quenching efficiency and biomolecular sensitivity with high signal-to-noise (S/N) ratio. Lu *et al.* reported the first graphene-based bio-sensor (Lu *et al.*, 2009). GO exhibits preferential binding to single-strand DNA (ssDNA) over double-strand DNA (dsDNA) (Lu *et al.*, 2009). It interacts strongly with nucleotides through π-stacking interaction between the rings of the

nucleobases and the hexagonal cells of GO (Jang *et al.*, 2010). GO-FRET biosensors are crucial in detecting proteins, heavy metal ions and small molecules too. FRET aptasensors for thrombin detection via dye-labeled aptamer assembled on GO or surfactant-stabilized graphene have been developed (Chang *et al.*, 2010; Lu *et al.*, 2009; Zhang *et al.*, 2011). Wen *et al.* developed FRET sensor for the sensitive detection of Ag^{+} ions (Wen *et al.*, 2010). Zhang *et al.* designed a versatile probe for the multiplex sensing targets such as sequence-specific DNA, protein (thrombin), metal ions (Ag^{+} and Hg^{2+}) and a small molecule (cysteine), and the limit of detection was 1, 5, 20, 5.7 and 60nM, respectively (Zhang *et al.*, 2011). Graphene can be used for in vitro targeting of ATP, in situ localization of mRNA, real-time monitoring and cell imaging. Dai's group has used polyethanol glycol (PEG) functionalized nanoscale GO (NGO-PEG) for living cell imaging in near-infrared because the NGO were found to be photoluminescent in visible and infrared regions (Sun *et al.*, 2008). Jung et al. designed a GO-based immune-biosensor for detecting a rotavirus as a pathogen model (Jung *et al.*, 2010). GO-FRET biosensor has several advantages: (i) GO is cost-effective for manufacturing and the assay can be completed within minutes, (ii) the large planar surfaces of GO make it possible to detect multiple molecular targets in the same solution and (iii) GO-based DNA detection improves the sensitivity by at least an order of magnitude as compared to conventional probes.

A graphene/GO sheet has both hydrophilic and hydrophobic properties due to electron-rich edges and aromatic regions of the plane. Thus, in the recent years, there has been a growing interest in developing graphene for drug loading and delivery because of strong interactions between hydrophobic drugs and aromatic regions of the graphene sheets and the feasibility of crossing a cell-membrane due to hydrophilic edges. Dai's group initially developed NGO-PEG as a nanocarrier to load variant anticancer drugs and evaluated its in vitro cellular uptake capacity and photoluminiscent property (Liu *et al.*, 2008; Lu *et al.*, 2009). Bao and co-workers have synthesized a series of GO-polymeric nanocarriers for water-insoluble anticancer drug delivery. The incorporation of hyrdrophillic and biocompatible polymers onto GO sheets resulted in nanocomposites with high solubility and stability in physiological conditions (Bao *et al.*, 2011; Pan *et al.*, 2011; Sahoo *et al.*, 2011). Along with drug delivery, gene delivery is also influenced by graphene-like materials. Both Liu and Zhang's groups fabricated polyethylenimine-modified GO (GO-PEI)

nanocarriers for pDNA transfection using non-covalent and covalent coupling methods, respectively (Chen *et al.*, 2011; Feng *et al.*, 2011).

One of the critical issues is the intrinsic toxicity of graphene and its derivatives. Although, this is being deliberately explored by several groups, but the conclusion seems to be controversial (Liu *et al.*, 2008; Ryoo *et al.*, 2010; Wang *et al.*, 2011; Zhang *et al.*, 2010). However, it has been indicated that the toxicity of graphene is closely associated with its biocompatible functionalization (Bao *et al.*, 2012). The future development of graphene-based nanomaterials/devices shall be based on better understanding of role of defects and oxygen containing groups at the edges of graphene sheets, the interaction mechanism of biomolecules with graphene surface and the role of doping heteroatoms in graphene.

Considering the biological importance of graphene, several groups tried to understand its different properties by quantum chemical methods. Similarly, we have also carried out DFT-based studies. The possible mechanism for formation of GO and how it may lead to synthesis of biologically important nanosize graphene sheets are discussed in Chapter V of this thesis.

1.3.1.5 Significance of Graphene/GO in Desalination

Access to steady supplies of clean water is difficult in the developing world. The present reverse osmosis plants rely on complicated, expensive and energy-intensive processes. However, a graphene-based salt filter could reduce desalination energy cost by 99%. The single atom-thick sheets of graphene are so thin that water flows through them far more easily than through a conventional thin-film composite.

In January 2012, MIT scientists showed (in simulations) that nanoporous graphene can filter salt water at a rate that is 2-3 orders of magnitude faster than current commercial desalination technologies, reverse osmosis (Phys.org, 2012). This opens the door to smaller and more efficient desalination facilities. In March 2013, Lockheed Martin announced the development of a new graphene-based water desalination technology, with hopes to commercialize it by 2014-2015 (Reuters, 2013). In June 2015, researchers at the University of Aveiro in Portugal designed unique "tea bags" using porous GO foam, which they say can help purify water by removing dissolved mercury (RSC, 2015). In September 2013, researchers from China’s Nanjing University of Aeronautics announced graphyne, an allotrope of

graphene, a promising material for water desalination that may even outperform graphene (Desalination.com, 2013).

Currently, researchers are exploring two different types of graphene-based filters. One is made out of multilayered GO "flakes," or membranes and the second class of filter is made up of a single layer of graphene that's perforated with tiny holes. However, the two approaches have their own benefits and drawbacks (Bennigton-Castro, 2014). It is believed that development of graphene-based filters is important and well carried out but more research is needed before "graphene filters" become a reality. Some theoretical predictions involving interactions Na^+, Cl^- and H_2O with different edges of graphene sheet is discussed in Chapter V of this thesis.

1.3.2 H-bonding Fluorescent Probe: 11-benzoyl-dibenzo[a,c]phenazine (BDBPZ)

The H-bond is a unique phenomenon in structural chemistry and biology. Its fundamental importance lies in its role in molecular association. In all biological systems the H-bonds are of great importance from the perspective of structure and reactivity involving water medium (Desiraju and Steiner, 2001). These H-bonds often significantly modifies the spectroscopic character of organic molecules. Especially when H-bond occurs in the electronically excited state, the changes in spectroscopic properties of a fluorophore are even more prominent (Han and Zhao, 2011). The dynamic behaviour of intermolecular H-bonds in the electronically excited state helps in understanding of microscopic structure and function in many biological systems. Thus, investigation of H-bond dynamics of photoexcited chromophores helps in understanding the effect of environment on their photophysical and photochemical behaviour (Bhattacharyya, 2008; Pal and Zewail, 2004).

The benzoyl derivative of a polarity insensitive H-bond acceptor fluorophore probe dibenzo[a,c]phenazine (DBPZ) is 11-benzoyl-dibenzo[a,c]phenazine (DBPZ) (Dey *et al.*, 2007a) **(Figure 1.3.8)**. The derivative BDBPZ has similar or even better H-bond accepting character like its parent DBPZ. However, in addition it can sense the polarity of the solvent medium. Chapter VI of this thesis includes a look into the different H-bonds formed by BDBPZ in different solvents in both the ground and excited states from quantum chemical approach. Further the interactions of BDBPZ with organic amine such as N,N-dimethyl aniline (DMA) has been studied.

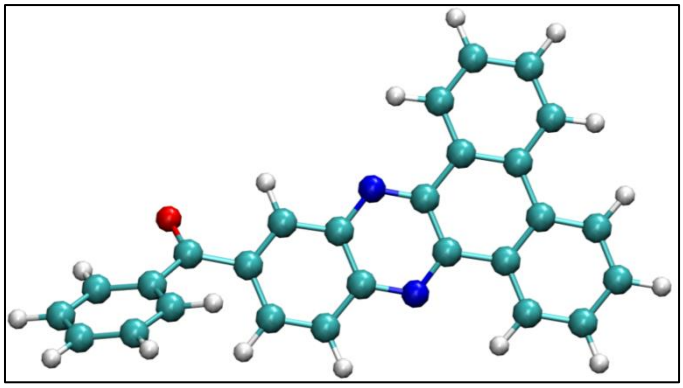


Figure 1.3.8: 11-benzoyl-dibenzo[a,c]phenazine (BDBPZ).

1.3.3 Nicotinamide Adenine Dinucleotide (NAD)

Nicotinamide Adenine Dinucleotide **(Figure 1.3.9)** is an important pyridine nucleotide that functions as an oxidative cofactor in eukaryotic cells. It **(Figure 1.3.9)** serves as a cofactor for dehydrogenases, reductases and hydroxylases, making it major carrier of H^+ and e^- in major metabolic pathways such as glycolysis, the tricarboxylic acid cycle, fatty acid synthesis, *etc*.

Structural and spectral studies involving NAD/NADH has been initiated here and reported in Chapter VI of this thesis.

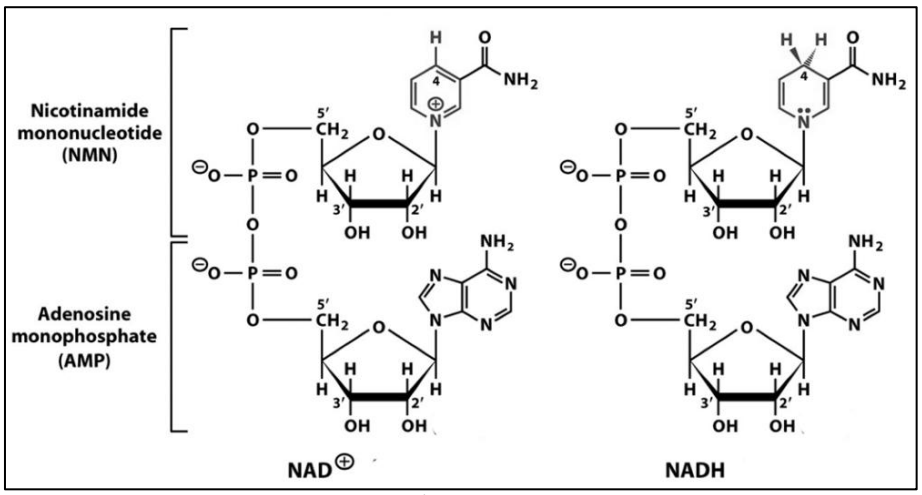


Figure 1.3.9: From Left to Right, NAD^+ and NADH, also shown are the nicotinamide and the adenosine parts.

Chapter I
Section IV

Computational Tools for Understanding the Structure and Dynamics of Biologically Important Macromolecules

The present day research based on experimental and computational techniques has provided an extensive knowledge of nucleic acids, protein and other macromolecules important in the field of biology. The advances in nucleic acid structure prediction and interaction studies have been largely due to the increased power and sophistication of both experimental and computational approach. Along with X-ray crystallography and NMR spectroscopy, molecular modeling or simulation techniques play important roles in providing information on structure, dynamics, flexibility and interactions of nucleic acids and proteins with other organic or inorganic macromolecules. Depending on the system, different computational methodologies have been developed based on classical force filed and *ab initio* quantum mechanics (Cramer, 2013; Jensen, 2006; Leach, 2001; Schlick, 2010). The most commonly used methodologies for understanding structure, dynamics and interactions of nucleic acids and proteins are:

(i) Molecular Dynamics (MD) and Steered Molecular Dynamics (SMD) Simulation

(ii) Umbrella Sampling

(iii) Monte Carlo Simulations (MC)

(iv) Molecular Docking

(v) Density Functional Theory (DFT) based methods

(vi) Hartree-Fock (HF) based mehod

(vii) Møller–Plesset perturbation theory (MP2) (*ab initio* Electron Correlation) based methods.

(viii) Quantum Mechanics/Molecular Mechanics (QM/MM)

In this thesis, MD and SMD techniques have been utilized extensively for generating ensemble of conformations on atomic level. These are discussed in APPENDIX I. The major empirical nucleic acid and protein force fields have been incorporated into algorithms that are used to minimize the conformation of a molecule with respect to its potential energy. This is Molecular Mechanics (MM). Much more extensive explorations of conformations can be made by MD, which applies Newton's equations of motion to an empirical force field, for all the atoms of a molecule. Unlike MM, MD can enable barriers between local energy minima to be traversed. Methods based on MD are especially valuable as they provide a link between the macroscopic properties of the system and the microscopic behaviour of individual molecules. The main limitations of MD stem from force field inaccuracies and the timescale of the simulations. Though typical length of MD simulations performed on systems

containing thousands of atoms range from picoseconds to hundreds of nanoseconds, many biologically interesting conformational motions of bio-macromolecules happen on the microsecond timescale or even at lower scales. The use of MM and MD have greatly increased in recent years, not only due to the availability of high-performance computing facilities, but also due to the widespread availability of well-validated academic and commercially derived computer programs with graphical user interface. A number of modeling and simulation programs are commonly used which have their force fields specifically parameterized for nucleic acids and their components. The most widely used and tested are: AMBER, CHARMM, GROMOS, GROMACS, JUMNA, NAMD, DESMOND, *etc.* Every year development is made in the existing programs leading to improvement in the methodology as well as research is directed towards increasing the computer power required for the same. Moreover, the increasing scope of MD is making it possible to the study systems of increasing complexity with more accurate methods (*e.g.*, improved potentials) on ever increasing time scales (Karplus, 2003). There are many recent examples of the use of MD to obtain functionally important information that is difficult or impossible to obtain experimentally.

The most precise calculations can be performed by the Quantum Mechanics (QM) based methods as these do not require any force field derived from experimental data. The QM methods are usually limited to a few hundred atoms, since; these are very complex and computationally intensive calculations where electrons are explicitly represented. Among the different *ab initio* quantum chemical methods, DFT is quite less expensive and it can be used to study the relatively large systems. DFT allows calculating both ground-state (GS) and excited-states (ES) properties from the electron density of an N-electron system. The DFT-based method has been extensively used in this thesis for studying graphene/GO, BDBPZ and other systems. An overview of the method is given in APPENDIX II.

Hybrid QM/MM techniques have become very fashionable in the past few years (Gogonea *et al.*, 2000; Warshel, 2003). In these schemes the biological system is divided into two parts. The region of biological interest, typically the enzymatic reaction center of a protein or the binding site in drug-DNA complexes, is treated at the *ab initio* level (QM-part) while the remaining protein or DNA residues as well as the solvent and counter ions are treated classically (MM-part). A wide variety of schemes have been developed based on Hartree-Fock (HF), Density functional theory

(DFT) or the empirical valence bond method (EVB) (Grochowski *et al.*, 1996). Considering DFT and Car-Parrinello MD, the CPMD program (Car and Parrinello, 1985) have been developed based on this technique.

Limitations due to the length and time scale in atomistic simulations and quantum chemical methods prohibit studies of many important processes involving biological macromolecules. For examples, phase behaviour in lipid bilayers, vesicle fusion, and dynamics of proteins and their aggregates. Thus, in the past couple of years a new technique called coarse-grain (CG) MD has developed. In this method interactions between single atoms are replaced by effective interactions between the CG units. This process dramatically decreases the degrees of freedom of the model, resulting in speed-up of the simulations by several orders of magnitude. Since, in the coarse-graining procedure some chemical details of the underlying atomistic system is lost, for meaningful application of CG models it is crucial to understand the construct of the CG methods and their range of applicability. MARTINI force field is one of the several force fileds corresponding to these CG model for biomolecular simulation (Marrink *et al.*, 2007).

The primary limitation of all these methods is that they are still approximate. It is here that experiment plays an essential role in validating the simulation methods. Thus, comparisons with experimental data serve to test the accuracy of the calculated results and provide criteria for improving the methodology. When experimental comparisons indicate that the simulations are meaningful, their capacity for providing detailed results often makes it possible to examine specific aspects of the atomic motions far more easily than by making measurements. The work presented in this thesis has been performed by theoretical approaches and compared to experimental data from time to time.

Aim and Scope of the Present Study

Aim and Scope of the Present Study

The 3D structures of biologically important macromolecules (DNA, RNA, proteins, graphene; and molecules like BDBPZ and NAD/NADH) are significant in understanding their biological role. The 3D structures of nucleic acids and proteins are relatively complex and maintained by different covalent and non-covalent interactions between their respective building blocks and surrounding environment. Thus, the nucleic acid and protein structures have been extensively studied. Earlier, modeling consisted of putting together balls and sticks by hand to visualize the 3D structures of biomolecules. However, today, the vast advancement of computational power has allowed not only visualizing static compounds but estimating their conformational changes with time. One of the techniques highly utilized to explore these aspects is MD simulation. The NMR and X-ray techniques provide time-average static pictures of a biomolecule, MD simulations on the other hand, can generate an ensemble of conformations from greater sampling of the system though longer simulation-run. However, there are areas which need treatments beyond classical mechanics. The non-bonded interactions in biomolecules are one where the classical force field methods often fail to generate physiologically relevant results. The much needed *ab initio* quantum chemical calculations until recent days were beyond the scope of biomolecules. These techniques are now evolving rapidly. The DFT based methods have further provided a faster calculation scheme for biolomecular fragments. This thesis emphasizes on the structure and dynamics of some of the biologically important macromolecules utilizing state of the art methods ranging from classical MD simulations to *ab initio* QM methodologies.

The nucleus of a cell contains the usual double helix DNA having canonical WC base pairs (A:T, G:C) and single stranded RNA. Apart from these, there also exist unusual four stranded DNA helices consisting mainly of HW base pairing and RNA bulges and pseudo-continuous helices containing unpaired bases. Over the past decade such unusual nucleic acids have gained huge importance for their functional role in various biological processes. The B-DNA and A-form RNA has already been studied both experimentally and theoretically in the past. We have thus, extended the study of nucleic acid structures to the unusual helices. The Chapter II and Chapter III of this thesis contain an elaborate report of such unusual nucleic acid structures. An in depth study of G-quadruplexes by both MD and *ab initio* methods are given in

Chapter II and Chapter III contained structural studies ofRNA bulges and pseudo-continuous helices based on MD simulations.

The nucleus of a cell not only contains nucleic acids but also has a variety of proteins and cofactors/coenzymes. The nuclear lamina underneath the nuclear envelope consists of two components, lamins and nuclear lamin-associated membrane proteins. Lamin proteins are related diversely to basic nuclear processes like replication, DNA damage repair and transcription. Various mutations in lamin A protein have been reported to produce disease conditions jointly termed as laminopathies. We have studied one such mutation present in the C-terminal domain of lamin A immunoglobulin fold. Chapter IV of this thesis in general focuses on the structural aspect of this mutant and its wild-type protein; however in specific it contains a comparative unfolding study of the proteins based on MD and SMD techniques.

Such and many more widespread disease conditions have prompted researchers to indulge more into biotechnology and biomedical research. Drug discovery is a major branch of this field. Drug design is also associated with developing appropriate drug-carrying vehicles, since it is difficult for the drug molecule to dross the cell-membrane; a drug is always needed to be transported into the cells using a vehicle designed particularly to be able to cross the cell-membrane. Cyclodextrines are often used as such vehicles for transporting small globular drug molecules. Another significant possible drug-carrying vehicle is graphene as it has both hydrophobic and hydrophilic properties at its faces and edges respectively. It is a single layer 2D nanomaterial which has received increasing attention in physical, chemical and biomedical fields for its unique physicochemical properties. Graphene based biosensors for small biomolecules (glucose, dopamine *etc.*), proteins and DNA detection; graphene based bioimaging and photothermal therapy applications are also rapidly developing areas. Tearing of GO sheets to generate smaller graphene sheets with higher functionality along with functionalization of graphene/GO sheets increasing its applicability as a vehicle, has been studied by appropriate QM methods and reported in details in Chapter V of this thesis.

Biomedical research also involves microscopy, imaging, spectroscopy and various other techniques which probe into the details of the biological systems. The time-resolved fluorescence spectroscopy of the molecule 11-benzoyl-dibenzo[a,c]phenazine (BDBPZ), which is a very important H-bonding probe aids in

making a progress in this biomedical field. An enzyme often requires a non-protein chemical component for its biological activity. Nicotinamide adenine dinucleotide phosphate is one such coenzymne found in all living cells. It consists of two nucleotides: adenine and nicotinamide joined through their phosphate linkage. NADPH dependent fluorescent proteins are potential candidate in FRET-based biosensor techniques (PDB ID: 3P19) related to biomedical research. Chapter VI of this thesis is based on the studies of these two molecules: BDBPZ (Section I) and NAD/NADH (Section II). The methodology utilized here is strictly *ab initio* QM.

Chapter II

Structural Studies of Quadruplex DNA

2.1 Introduction

There has been a recent explosion of interest in 3D structure of four-stranded DNAs known as G-quadruplexes or DNA tetraplexes. The structure involved in formation of G-quadruplex was first proposed by Gellert *et al.* in 1962 (Gellert *et al.*, 1962). A G-quartet or a G-tetrad **(Figure 2.1)** may be considered as the basic building block of G-quadruplex DNA (G-DNA). It is a square co-planar array of four guanine bases, in which each base is both donor and acceptor of two hydrogen bonds (H-bonds). Hoogsteen edge of each guanine base forms H-bonds with the Watson-Crick (WC) edge of adjacent guanine base in *cis* orientation, followed by Hoogsteen edge of the second guanine base forming two H-bonds with WC edge of the third guanine base, and in a cyclic way giving rise to four G:G H:W C (Gua:Gua Hoogstein-edge:WC-edge and *Cis* orientation) base pairs in a tetrad **(Figure 2.1)**. Two or more such tetrads stack upon each other forming a four-stranded helix with guanine-tetrad cores (Simonsson, 2001). The sugar phosphate backbones run along the outside four edges of the structure, and monovalent metal ions are generally present in the central G-tetrad (G4) core (Huppert, 2007) with G4-ion-G4-ion-G4 arrangement. The G-quadruplexes exhibit extensive structural diversity and polymorphism, as opposed to double helical DNA (Lee *et al.*, 2005; Neidle and Parkinson, 2003; Phan *et al.*, 2006; Ying *et al.*, 2003). The major structural polymorphism in quadruplex-DNA arises from the various possible combinations of strand direction (Burge *et al.*, 2006). The second type of polymorphism arises due to variations in the nature of the loops, *i.e.* loop size, sequence and their locations with respect to the helix (Burge *et al.*, 2006; Hazel *et al.*, 2004; Smargiasso *et al.*, 2008; Todd *et al.*, 2005). The solution environment, such as the type of metal ions, ligands or molecular crowding conditions, may also influence the topology of quadruplex (Haider *et al.*, 2011; He *et al.*, 2004; Ou *et al.*, 2008).

G-quadruplexes are either formed from an intramolecularly folded single G-rich sequence or by the intermolecular association of two (dimeric) or four (tetrameric) separate strands (Qin and Hurley, 2008). DNA concentration often determines the type of association adopted by the G-rich sequences (Huppert and Balasubramanian, 2005; Palacký *et al.*, 2013; Simonsson and Sjoback, 1999). The backbone strands of the G-quadruplexes are connected by loops in dimeric and monomeric G-quadruplexes. These loops can be classified into four major families:

edgewise loops connecting two adjacent anti-parallel strands, diagonal loops connecting two opposing anti-parallel strands, double-chain-reversal loops connecting adjacent parallel strands and V-shaped loops connecting two corners of a G-tetrad core in which one supporting column is lacking (Phan *et al.*, 2006). The preferred G-quadruplex structures adopted by a G-rich sequence also depend on the nature of cations (Smargiasso *et al.*, 2008). The cavity between two stacked G-quartet planes, which in this study is referred as the G-tetrad core is lined by eight carbonyl oxygen atoms (O6), all of these participate in the precise coordination of cations.

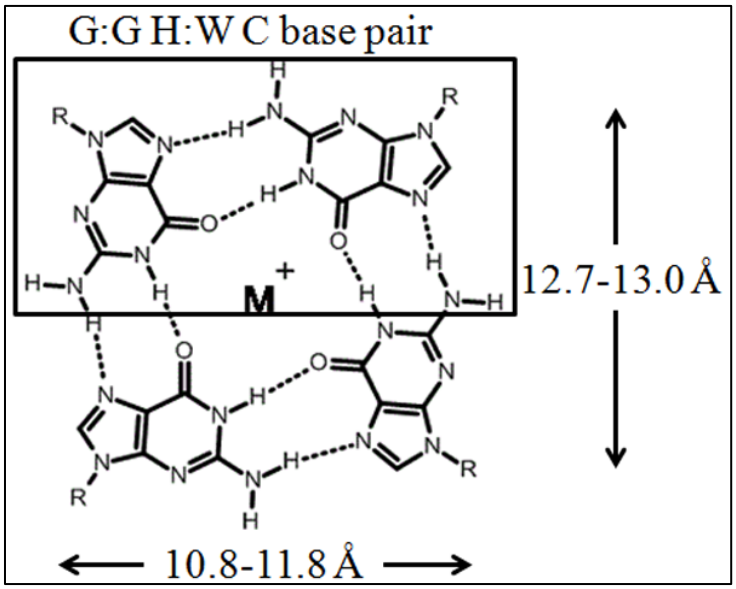
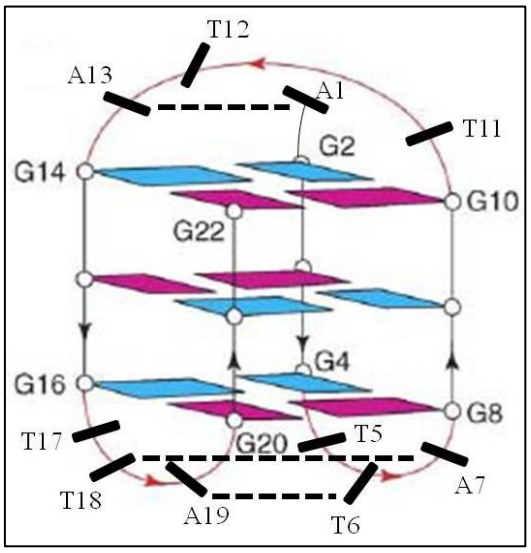


Figure 2.1: A G-quartet involving G:G H:W C base pair and stabilized by a central metal cation (M^+).

Numerous studies on human telomeric G-quadruplexes have been carried out by various spectroscopic techniques like ultraviolet, circular dichroism, fluorescence, raman, electron paramagnetic resonance and nuclear magnetic resonance spectroscopies (Adrian *et al.*, 2012; Dao *et al.*, 2011; Palacký *et al.*, 2013; Renčiuk *et al.*, 2009; Singh *et al.*, 2009; Tran *et al.*, 2011). Experiments based on X-ray crystallography (Parkinson *et al.*, 2002), mass-spectrometry (Smargiasso *et al.*, 2008), calorimetry (Petraccone *et al.*, 2011) and lazer tweezer based methods (De Messieres *et al.*, 2012; Dhakal *et al.*, 2013) have also contributed largely to G-quadruplex research. Several structures are available in RCSB-PDB (Berman *et al.*, 2000) for the human telomeric sequences which differ in strand orientation and loop-topology. The first solution structure of a four-repeat human telomeric sequence, d[AGGG(TTAGGG)$_3$] (in Na^+ solution), was characterized in 1993 by Wang and

Patel (Wang and Patel, 1993) using NMR spectroscopy, distance geometry and molecular dynamics approach. This sequence folds intramolecularly back on itself into an anti-parallel type G-quadruplex **(Figure 2.2.1)** involving three stacked G-tetrads with *anti•anti•syn•syn* glycosidic torsions of the guanine residues. The three connecting TTA loops adopt edgewise, diagonal and edgewise alignments successively such that each strand has both parallel and anti-parallel adjacent strands, giving the complete structure a basket-type look. A decade later the same sequence was reported to form a completely different G-quadruplex in presence of K^+ by X-ray crystallography (Parkinson *et al.*, 2002). In this 'propeller type' structure **(Figure 2.2.2a)**, all four strands are parallel, the connecting TTA loops are double-chain reversal types and all guanines adopt *anti* glycosidic conformation. The same parallel stranded propeller type structure was also obtained by NMR experiment for a human telomeric G-quadruplex under molecular crowding conditions and water depletion (Heddi and Phan, 2011) **(Figure 2.2.2b)**. A near similar observation was reported by Smargiasso *et al.* (Smargiasso *et al.*, 2008), that the human telomeric sequences preferably adopt anti-parallel topology in presence of Na^+ while the same sequence may adopt mutimeric parallel form in presence of K^+. Solution NMR structures in K^+ environment, for the same sequence shows two other possibilities which differ from each other only by the order of loop arrangements (Luu *et al.*, 2006; Phan *et al.*, 2007; Phan *et al.*, 2006). Both structures have one *anti•syn•syn•syn* and two *syn•anti•anti•anti* G-tetrads with three G-tracts oriented in one direction and the fourth tract in the opposite direction. There is one double chain reversal loop and two edgewise loops. In one form the double chain reversal loop is formed by the first linker **(Figure 2.2.3a)** and in the other by the third linker **(Figure 2.2.3b)**. Apart from these topologies there also exists two more possibilities: (i) basket type G-quadruplex (PDB ID: 2KF8) with only two G-tetrad layers (Lim *et al.*, 2013) **(Figure 2.2.4a)** which is identical to the earlier reported anti-parallel topology with three G-tetrads in the sense that it has the diagonal G-stands in anti-parallel orientation forming a similar anti-parallel core and (ii) anti-parallel (2+2) G-quadruplex (PDB ID: 2MBJ) which is obtained with ^{Br}G-substituted oligonucleotides (Lim *et al.*, 2013) although, it differs in loop orientations but complements the anti-parallel human telomeric G-quadruplex in core-type **(Figure 2.2.4b)**.



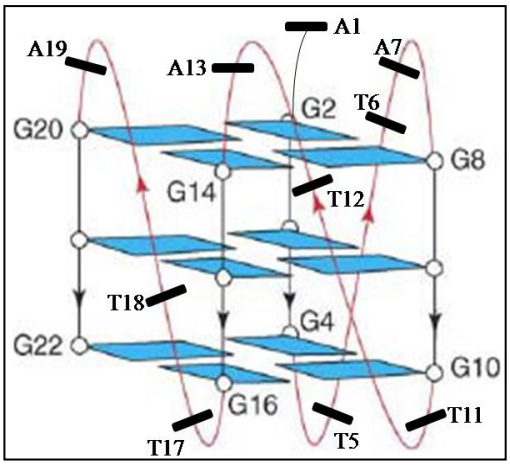
BASE PAIRING INFORMATION

1	ADE	13	ADE	h:h	C
2	GUA	10	GUA	W:H	C
3	GUA	9	GUA	H:W	C
4	GUA	8	GUA	W:H	C
5	THY				
6	THY	19	ADE	W:W	C
7	ADE	18	THY	H:W	T
8	GUA	20	GUA	W:H	C
9	GUA	21	GUA	H:W	C
10	GUA	22	GUA	W:H	C
11	THY				
12	THY				
13	ADE	1	ADE	h:h	C
14	GUA	2	GUA	W:H	C
15	GUA	3	GUA	H:W	C
16	GUA	4	GUA	W:H	C
17	THY				
18	THY	7	ADE	W:H	T
19	ADE	6	THY	W:W	C
20	GUA	16	GUA	W:H	C
21	GUA	15	GUA	H:W	C
22	GUA	14	GUA	W:H	C

Figure 2.2.1: Anti-parallel type G-quadruplex topology (PDB ID: 143D) and its base pairing information as obtained using BPFIND.

Extensive research on G-quadruplex molecules by several different techniques has been done in the past two decades. Nanosecond MD simulations of four-stranded G-quadruplexes formed by d(G_4) and d($G_4T_4G_4$) sequences indicated that the quadruplex stem is rigid and very stable (Špačková *et al.*, 1999). Simulations performed on monomeric quadruplexes containing both the stem and loop-regions have confirmed that the G-quartet stem is highly stable however there exists some flexibility in the loop region (Fadrná *et al.*, 2009; Haider *et al.*, 2008). The loop residues themselves are sometimes involved in stacking with the closest guanine tetrad along with H-bonding interactions with the other loop residues or guanine bases for further stabilizing the G-quadruplex folds (Zhu *et al.*, 2013). Microsecond timescale MD simulation of propeller-type topology in which four parallel G-strands are connected by double-chain-reversal loop provided accounts of interactions of TTA loops with G-quartets, formation of Ade:Ade base pairs, triad, pentad and hexad by loop residues (Islam *et al.*, 2013). It has also been experimentally observed that the base stacking between outer G-tetrads and loop thymine bases slow down the ion

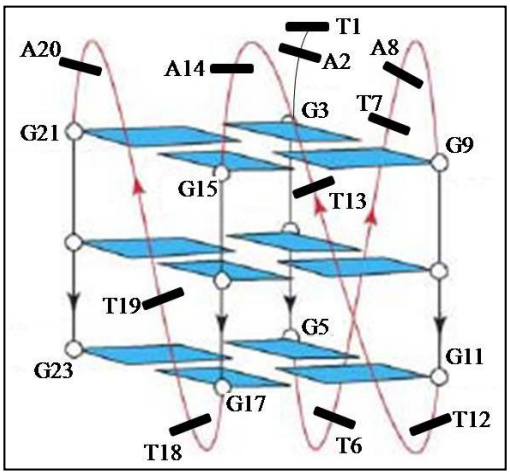
exiting rate, adding to stability of the quadruplexes (Podbevšek *et al.*, 2008). It has been reported by Suhnel and co-workers (Meyer *et al.*, 2001) that cations like Be^{2+} which have small radius and high charge density incorporate a non-planarity to the G-quartet, however, Na^+ and K^+ act differently. Since, the central cavity is too small for K^+ ion, this favours a non-planar geometry. Moreover, it is now well established that coordination of potassium (Simonsson and Sjoback, 1999), and more rarely of sodium (Laughlan *et al.*, 1994) adds to the stability of G-quadruplexes (Lane *et al.*, 2008; Pilch *et al.*, 1995). Stabilization of G-quadruplex by NH_4^+ and $^{23}Na^+$ ion has also been studied extensively by NMR spectroscopy (Guschlbauer *et al.*, 1990; Podbevšek et *al.*, 2007, 2008) and indicated ion-exchange on a large millisecond to second time scale (Guschlbauer *et al.*, 1990; Ida and Wu, 2008; Podbevšek *et al.*, 2007; Sket and Plavec, 2007; Snoussi and Halle, 2008). Along with equilibrium MD simulations, recently some SMD simulations have been performed on G-quadruplexes indicating that the unfolding path of a G-quadruplex is topology-dependant (Bergues-Pupo *et al.*, 2015).



(a)

BASE PAIRING INFORMATION

1	ADE			
2	GUA	8	GUA	W:H C
3	GUA	9	GUA	W:H C
4	GUA	10	GUA	W:H C
5	THY			
6	THY			
7	ADE			
8	GUA	14	GUA	W:H C
9	GUA	15	GUA	W:H C
10	GUA	16	GUA	W:H C
11	THY			
12	THY			
13	ADE			
14	GUA	20	GUA	W:H C
15	GUA	21	GUA	W:H C
16	GUA	22	GUA	W:H C
17	THY			
18	THY			
19	ADE			
20	GUA	2	GUA	W:H C
21	GUA	3	GUA	W:H C
22	GUA	4	GUA	W:H C

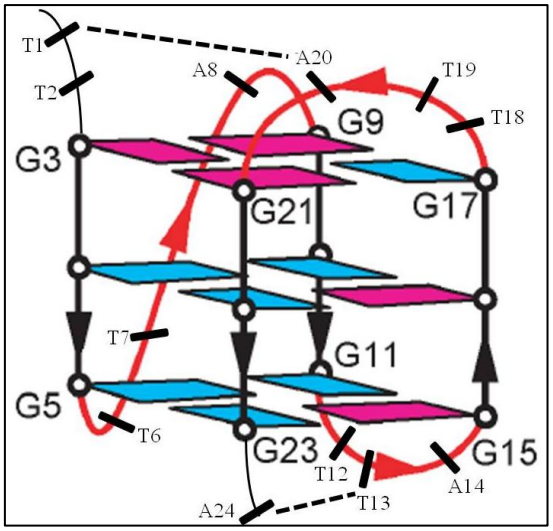


(b)

BASE PAIRING INFORMATION

1	THY			
2	ADE			
3	GUA	9	GUA	W:H C
4	GUA	10	GUA	W:H C
5	GUA	11	GUA	W:H C
6	THY			
7	THY			
8	ADE			
9	GUA	15	GUA	W:H C
10	GUA	16	GUA	W:H C
11	GUA	17	GUA	W:H C
12	THY			
13	THY			
14	ADE			
15	GUA	21	GUA	W:H C
16	GUA	22	GUA	W:H C
17	GUA	23	GUA	W:H C
18	THY			
19	THY			
20	ADE			
21	GUA	3	GUA	W:H C
22	GUA	4	GUA	W:H C
23	GUA	5	GUA	W:H C

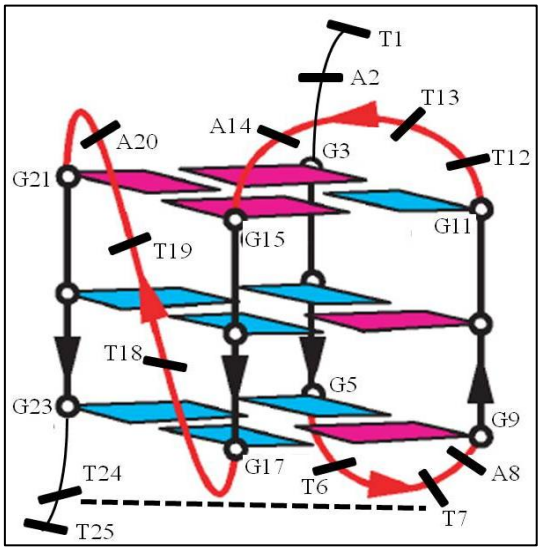
Figure 2.2.2: Parallel type G-quadruplex topology of (a) crystal structure corresponding to PDB ID: 1KF1 and (b) NMR reported structure (PDB ID: 2LD8). The corresponding base pairing information are given in box.



(a)

BASE PAIRING INFORMATION

1	THY	20	ADE	W:W C
2	THY			
3	GUA	9	GUA	H:W C
4	GUA	10	GUA	W:H C
5	GUA	11	GUA	W:H C
6	THY			
7	THY	5	GUA	W:S C
8	ADE			
9	GUA	17	GUA	H:W C
10	GUA	16	GUA	W:H C
11	GUA	15	GUA	W:H C
12	THY			
13	THY	24	ADE	W:W T
14	ADE			
15	GUA	23	GUA	W:H C
16	GUA	22	GUA	W:H C
17	GUA	21	GUA	H:W C
18	THY			
19	THY			
20	ADE	1	THY	W:W C
21	GUA	3	GUA	H:W C
22	GUA	4	GUA	W:H C
23	GUA	5	GUA	W:H C
24	ADE	13	THY	W:W T



(b)

BASE PAIRING INFORMATION

1	THY			
2	ADE			
3	GUA	11	GUA	H:W C
4	GUA	10	GUA	W:H C
5	GUA	9	GUA	W:H C
6	THY			
7	THY	24	THY	W:W T
8	ADE			
9	GUA	17	GUA	W:H C
10	GUA	16	GUA	W:H C
11	GUA	15	GUA	H:W C
12	THY			
13	THY			
14	ADE			
15	GUA	21	GUA	H:W C
16	GUA	22	GUA	W:H C
17	GUA	23	GUA	W:H C
18	THY			
19	THY			
20	ADE			
21	GUA	3	GUA	H:W C
22	GUA	4	GUA	W:H C
23	GUA	5	GUA	W:H C
24	THY	7	THY	W:W T
25	THY			

Figure 2.2.3: (a) Mixed-(3+1)-form1 (PDB ID: 2GKU) and (b) mixed-(3+1)-form2 (PDB ID: 2JSL) G-quadruplex topology types with corresponding base pairing information in the box.

G-quadruplex DNA motifs are widely dispersed in the eukaryotic genome and are abundant in regions of biological significance, for example, at immunoglobulin switch regions (Sen and Gilbert, 1988), gene promoter regions (Evans *et al.*, 1984; Kilpatrick *et al.*, 1986) sequences associated with human disease (Fry & Loeb, 1994) and telomeres (Blackburn, 1990). However, G-quadruplex DNA has probably attracted the greatest attention of researchers because of the natural existence of a G-rich single-strand overhang at 3′ ends of chromosomes (telomere) and its relevance to a potential anticancer drug targets (Lech *et al.*, 2013). Although, an order of mechanical stability of uni-molecular G-quadruplexes at the telomeres has been predicted as anti-parallel > hybrid (mixed-forms) > parallel (Bergues-Pupo *et al.*, 2015), but as indicated earlier its strand orientation and loop topology is not yet confirmed. Furthermore, several experiments performed in physiological K^+ solutions indicate that the same telomeric sequence can adopt different topologies even under identical solution environment (Burge *et al.*, 2006; Dhakal *et al.*, 2013; Hazel *et al.*, 2004; Lee *et al.*, 2005) due to the alteration of micro-environment but obviously there exist some unique structural features inside the eukaryotic cells, which account for the stability, recognition by specific proteins and hence biological role of the telomeres. Thus, it becomes necessary to ascertain the conformational preference, *i.e.* the form which will represent the majority in the cellular environment. Moreover, conversion of one form to the other appears highly unlikely as the process demands rupture of at least 24 H-bonds. We have therefore, tried to narrow down to one amongst the four possible core types attained by the major conformation. There exist numerous reports on MD simulations of G-quadruplexes by AMBER force field with various corrections (Chowdhury and Bansal, 2000; Fadrná *et al.*, 2009; Hazel *et al.*, 2006; Šponer *et al.*, 2012; Stefl, Spackova *et al.*, 2001). We have found only a single paper based on human telomeric uni-molecular G-quadruplex simulations by charmm27 force field (Fadrná *et al.*, 2009), but it has criticized the force-field on the basis of ion-expulsion and loop flexibility. However, since our recent MD studies on non-canonical nucleic acid structures by charmm27 force field (Foloppe and MacKerell, 2000) have explained many experimental observations (Halder and Bhattacharyya, 2012), we considered this appropriate for all our simulations. We have also attempted to understand these structures having different core-types, from the perspective of base pair parameters using both quantum chemical and MD simulation studies.

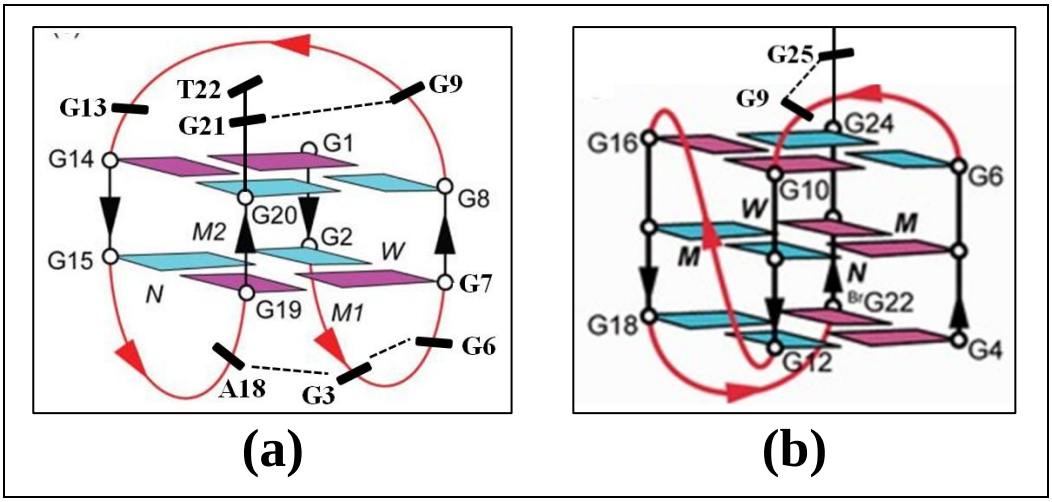


Figure 2.2.4: (a) Basket-type G-quadruplex (PDB ID: 2KF8) and (b) 2+2 anti-parallel G-quadruplex (PDB ID: 2MBJ).

2.2 Methodology

2.2.1 Quantum Chemical Method

Atomic coordinates of the guanine quartet selected for optimization were extracted from the PDB file (PDB ID: 156D) (Schultze *et al.*, 1994). This quartet was initially stacked between two other quartets in the reported NMR structure but for our quantum chemical calculations we selected only the four guanine bases forming the quartet. The sugar phosphate moieties connected to N9 positions of the purines were replaced by methyl groups to remove the unnatural dipole moments or H-bonding. Modeling of all the hydrogen atoms were done with the help of MOLDEN (Schaftenaar and Noordik, 2000). Geometry optimization of the quartets were performed without any constrains or symmetry restrain by several standard *ab initio* quantum chemical methods namely MP2/6-31G(2d,2p) (Second order Møller–Plesset perturbation) (Møller and Plesset, 1934), HF/cc-pVDZ (Dunning correlation consistent basis set) (Dunning Jr, 1989), B3LYP/6-31G(2d,2p) (Becke, 1993; Lee *et al.*, 1988) and GGA:PW91/DZP (Perdew *et al.*, 1996) using Gaussian09 (g09) (Frisch *et al.*, 2009) and ADF (te Velde et al., 2001) softwares and under different ionic conditions: (i) in presence of single Na^+ or K^+ ions, (ii) in presence of two Na^+ or K^+ ions and (iii) in absence of any ion. The initial geometries of G-quartet containing single Na^+ or K^+ ion had the ions placed inside the G-tetrad core at the centre of the plane formed by the four bases whereas, for the optimization of G-tetrad containing

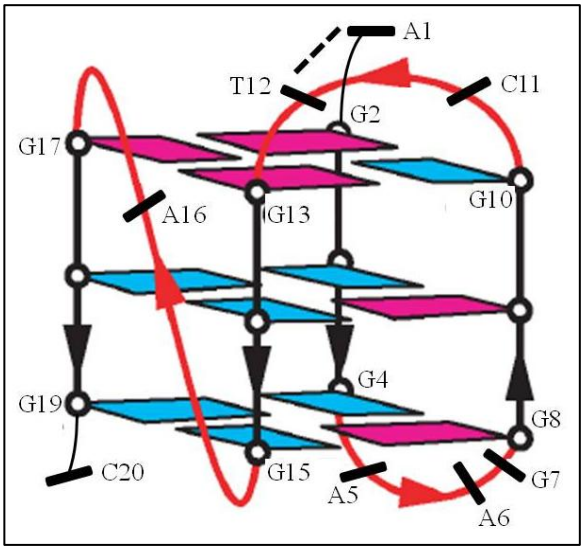
two ions the initial geometries were modeled by placing an entire tetrad between the two ions such that the ions lie on opposite sides of the G-tetrad core. Any solvent effect was not considered as the quartet is supposed to be stable in the hydrophobic core of the somewhat globular helix. Furthermore, empirical dispersion interaction was neglected as such interactions are more dominant in stacking interactions between two quartets.

2.2.2 Molecular Dynamics Simulation Method

Molecular dynamics (MD) simulations were carried out for four different uni-molecular human telomeric G-quadruplex topologies and one non-telomeric G-quadruplex motif, whose structures are available. The first models of the respective NMR derived structures corresponding to the topologies with parallel core (2LD8) (Heddi and Phan, 2011), anti-parallel core (143D) (Wang and Patel, 1993), mixed-(3+1)-form1core (2GKU) (Luu *et al.*, 2006) and mixed-(3+1)-form2 core (2JSL) (Phan *et al.*, 2007) along with the G-rich non-telomeric form (2KZD) (Lim *et al.*, 2010) were taken from RCSB-PDB. None of these structures reported any ion inside the quadruplex. A combination of telomeric and non-telomeric sequences was selected to compare the structural features of G-rich sequences of different regions. Along with the NMR structures, two X-ray crystal structures of similar telomeric sequences with parallel topology (1KF1 and 3R6R) (Bazzicalupi *et al.*, 2013; Parkinson *et al.*, 2002) were also selected for our simulation study. The PDB-files corresponding to these, unlike the NMR structures, reported K^+ ions inside the G-tetrad core, which made simulation of these structures significant. The crystal structure with PDB ID: 3R6R is a complex of G-quadruplex with drug-like ligand (natural alkaloid berberine). Apart from presence of the ligand, the overall DNA topology was similar to that of 1KF1. Since, the presence of ligand may cause some structural differences with the native; MD simulation was performed considering the initial coordinates from the PDB (3R6R) after removing the ligand molecules. The details of all the sequences and corresponding base pairing information, as obtained from BPFIND software (Das *et al.*, 2006), are already given in **Figure 2.2.1-2.2.4**.

The non-telomeric hTERT promoter G-rich sequence with topology similar to the mixed-(3+1)-form2 does not follow the general human telomeric sequence pattern -TAGGG- and contains an ionosine base in the loop region **(Figure 2.2.5)**. In order to

make this system comparable to the human telomeric sequences, which have only adenine and thymine bases forming the loops, we have replaced the ionosine base by adenine in the initial coordinate file. The minimized structure obtained had three G-quartet stacks connected by three loops: the third is the double-chain-reversal loop having a single adenine residue while the first and the second are edgewise loops.



BASE PAIRING INFORMATION

1	ADE	12	THY	H:W T
2	GUA	10	GUA	H:W C
3	GUA	9	GUA	W:H C
4	GUA	8	GUA	W:H C
5	ADE			
6	ADE			
7	GUA			
8	GUA	15	GUA	W:H C
9	GUA	14	GUA	W:H C
10	GUA	13	GUA	H:W C
11	CYT			
12	THY	1	ADE	W:H T
13	GUA	17	GUA	H:W C
14	GUA	18	GUA	W:H C
15	GUA	19	GUA	W:H C
16	ADE			
17	GUA	2	GUA	H:W C
18	GUA	3	GUA	W:H C
19	GUA	4	GUA	W:H C
20	CYT			

Figure 2.2.5: topology of a non-telomeric G-rich hTERT sequence (2KZD)

The MD simulations were carried out using NAMD software (Phillips *et al.*, 2005) with charmm27 force field. In each simulation, the nucleic acid molecules were explicitly solvated with TIP3P water molecules in a rectangular periodic box whose dimensions were at least 30Å larger than the size of the corresponding solute molecules. For carrying out MD simulation, Na^+ is usually considered as the charge neutralizing cation as it is a good representation of point charge. However, inside the cell, under physiological conditions K^+ ion concentration is higher than that of Na^+. Hence, the simulations of the molecules obtained from NMR data were performed with both Na^+ and K^+ counter-ions. The details of the MD simulations carried out are tabulated in **Table 2.1**. A point to be mentioned here is that the crystal structures with parallel topology and containing multiple K^+ ions in their G-tetrad cores were

simulated only with additional charge neutralizing Na^+ counter-ions, since simulation of these in presence of only K^+ ions would represent a situation exactly similar to the simulation of parallel topology of simulation1b. The positions of the counter-ions were generated by Monte Carlo simulation in absence of water, considering only the interactions between the ions and the nucleic acid concerned (Bandyopadhyay, 2005). The initial round of equilibration with explicit solvent and ions involved 100 steps of steepest descent, followed by 1000 steps of conjugate gradient and 20000 steps of the adopted basis Newton–Raphson energy minimization using CHARMM (Brooks *et al.*, 1983). The particle mesh Ewald (PME) summation method (Darden *et al.*, 1993), with width of the Gaussian distribution in the PME as 0.35, was used to treat the long-range electrostatic interactions. A force-switch method was applied for non-bonded interactions (van der Waals) with a 12Å cutoff. The systems were gently heated from 0K to 300K in 30ps with a time step of 1.0fs after energy minimization followed by the final production run of 200ns by Constant Pressure Temperature dynamics algorithm (Feller *et al.*, 1995) at 1atm pressure, with 1fs time step. Some of the simulations **(Table 2.1)** were carried out only up to 100ns as the structure appeared to destabilize during the MD-run. The trajectories were made up by frames collected every 1.0ps and analyzed using CHARMM. Different structural and base pair parameters of DNA were calculated using NUPARM software (Bansal *et al.*, 1995; Mukherjee *et al.*, 2006).

Table 2.1: Details of MD simulations of G-quadruplexes.

Sim no.	PDB ID	Topology	Expt. method	sequence	Loop details	Na+/ K+	Length of MD-run (ns)	All-atom average RMSD (Å)	Average RMSD for G-quartet stacks (Å)
1a	2LD8	parallel	K^+ sol^n NMR in crowded medium	$d[TAG_3(T_2AG_3)_3]$	3 double chain reversal loops	22/0	200	3.74 (0.28)	1.53 (0.13)
1b						0/22	200	4.46 (0.34)	1.51 (0.17)
2a	143D	anti-parallel	Na^+ sol^n NMR	$d[AG_3(T_2AG_3)_3]$	2 lateral loops and 1 central diagonal loop	21/0	200	2.46 (0.22)	1.66 (0.12)
2b						0/21	200	3.32 (0.20)	1.85 (0.09)
3a	2GKU	mixed-(3+1) form1	K^+ sol^n NMR	$d[TTG_3(T_2AG_3)_3A]$	1 double chain reversal loop and 2 edgewise/lateral loops	23/0	200	3.12 (0.22)	1.02 (0.09)
3b						0/23	200	3.36 (0.15)	1.23 (0.14)
4a	2JSL	mixed-(3+1) form2	K^+ sol^n NMR	$d[TAG_3(T_2AG_3)_3TT]$	1 double chain reversal loop and 2 edgewise/lateral loops	24/0	200	3.87 (0.61)	1.23 (0.13)
4b						0/24	200	4.60 (0.30)	1.57 (0.19)
5a	2KZD	G-rich sequence in hTERT promoter region	K^+ sol^n NMR	d[AGGGIAGGGGC TGGGAGGGC]	1 double chain reversal loop and 2 edgewise/lateral loops	19/0	100	3.11 (0.31)	2.14 (0.16)
5b						0/19	100	4.07 (0.70)	3.59 (0.67)
6	1KF1	parallel	X-ray crystal data	$d[AG_3(T_2AG_3)_3]$	3 double chain reversal loops	18/3	200	3.66 (0.32)	1.30 (0.12)
7	3R6R	parallel	X-ray crystal data	$d[AG_3(T_2AG_3)_3]$	3 double chain reversal loops	19/2	100	4.68 (0.27)	1.33 (0.11)

2.2.3 Steered Molecular Dynamics Simulation Method

Steered Molecular Dynamics (SMD) simulations (Balsera *et al.*, 1997; Evans and Ritchie, 1997; Izrailev *et al.*, 1997) were carried out using NAMD software with charmm27 force field and Generalized Born Implicit Solvent (GBIS) model for water (Feig & Brooks, 2004; Qiu, Shenkin, Hollinger, & Still, 1997) on the four human telomeric structures. The simulations performed by employing implicit solvent models are computationally less expensive compared to simulations in explicit solvent, but absence of water molecules may give lower force peaks than otherwise (Gao *et al.*, 2002; Xia *et al.*, 2002). However, in our work the qualitative force peaks are of sufficient significance as we have employed SMD simulations in a comparative study of the four different topologies. The molecules of the explicitly solvated models, heated to 300K, were taken and implicitly solvated using the GBIS model. SMD simulations were then performed with a time step of 1fs, solvent dielectric constant of 78.5, and cutoff of Coulomb forces with a switching function starting at a distance of 15Å and reaching zero at 16Å. The SMD simulations with constant velocity stretching (SMD-CV protocol) were carried out by fixing one of the terminal residues and applying external forces to the dummy atoms attached to another terminal residue (SMD residue) with a virtual spring of spring constant 0.6Kcal/mol/Å^2 and pulling velocity of 0.0000025Å/fs similar to our earlier study (Bera *et al.*, 2014). This spring constant and pulling velocity is much less than that considered recently by Burgues-Pupo *et al.* (Bergues-Pupo *et al.*, 2015) for their SMD simulations in explicit solvent. Two different pulling vectors were considered: i) parallel and ii) perpendicular to the plane containing H-bonds of the G:G H:W C base pairs. The direction of pulling was such that the end-to-end distance always increased, *i.e.* the SMD residue was pulled away from the fixed residues. Each simulation was repeated twice with different initial seeds to check the reproducibility. The details of fixed residue, SMD residue and direction of pulling are given in **Table 2.2**. The force experienced by the pulled terminal residue, F is $F = k(vt - x)$, where, x is the displacement of the pulled atom from its original position, v is the pulling velocity, and k is the spring constant. The simulation trajectory was made up by frames collected every 1.0ps and then analyzed using CHARMM, BPFIND and NUPARM.

Table 2.2: Details of SMD simulations of G-quadruplexes.

Topology type	PDB-ID	Pulling vector parallel to plane of H-bonds			Pulling vector perpendicular to plane of H-bond		
		SMD residue	Fixed residue	direction of force	SMD residue	Fixed residue	direction of force
parallel	2LD8	GUA (residue 23)	THY (residue 1)	$1/\sqrt{2}\hat{j}+1/\sqrt{2}\hat{k}$	GUA (residue 23)	THY (residue 1)	$1/\sqrt{2}\hat{\imath}+1/\sqrt{2}\hat{j}$
anti-parallel	143D	ADE (residue 1)	GUA (residue 22)	$\hat{k}$	ADE (residue 1)	GUA (residue 22)	$\hat{j}$
mixed-(3+1)-form1	2GKU	THY (residue 1)	ADE (residue 24)	$\hat{\imath}$	THY (residue 1)	ADE (residue 24)	$\hat{j}$
mixed-(3+1)-form2	2JSL	THY (residue 1)	THY (residue 25)	$\hat{k}$	THY (residue 1)	ADE (residue 24)	$\hat{j}$

2.3 Results

2.3.1 Quantum Chemical Analysis

The G:G H:W C base pair is an integral part of the G-quartet, which has been studied in the context of non-canonical base pairing in RNA (Bhattacharya *et al.*, 2015; Panigrahi *et al.*, 2011). In G-quadruplex topology, however, four such base pairs are arranged in a cyclic orientation, mostly along with a cation of appropriate size. We initially optimized a single G-quartet structure in absence of any ion using the popular *ab initio* quantum chemical methods. The N-H…N and N-H…O H-bonds and bond angles **(Table 2.3)** are similar to those reported earlier (Meyer *et al.*, 2001). We have also calculated the base pair orientation parameters for all the optimized structures **(Table 2.4)**, which were not available in literature. It was clearly evident from B3LYP/6-31G(2d,2p) optimized structures that in absence of any ion the base pairs of the quartet have comparatively high propeller values indicating that the base pairs have a tendency to acquire non-planar geometry **(Figure 2.3a)**. Similar was the situation when the same G-quartet was optimized by MP2/6-31G(2d,2p), HF/cc-pVDZ and GGA:PW91/DZP **(Table 2.4)** methods. Hoogsteen base pairing mostly require large shear similar to G:U Wobble base pairs (Mukherjee *et al.*, 2006) but we further noticed an enhancement of shear for the base pairs after optimization of these G-quartet in absence of any ion, by all the methods. This indicated deformation of the H-bond geometry holding the four guanine bases as a quartet possibly due to electrostatic repulsion between the four O6 atoms. All these facts pointed to the necessity of presence of a cation, a ligand, or some topological strain from loop or bead formation in the G-quartets for their stability and planarity. Although, K^+ and Na^+ are both biologically important, relatively small Na^+ ion can be coordinated in the plane of a G-quartet but large K^+ ions are coordinated only inside G-tetrad stacks (Haider *et al.*, 2002). We have therefore, also optimized this quartet in presence of single Na^+ and K^+ ions separately with the same quantum chemical methods. Optimizations in presence of these ions led to a comparatively planar base quartet orientation **(Figure 2.3b and 2.3c)**. The base pairs have high buckle value when the G-quartet was optimized in presence of a single K^+ ion by B3LYP/6-31G(2d,2p) method **(Table 2.4)** inducing a non-planarity in the quartet plane, whereas, the Na^+ ion containing structures showed lower buckle value **(Table 2.4)**. It may be noted that

almost all the WC base pairs in DNA double helix show high negative propeller (Halder and Bhattacharyya, 2013), which might arise due to pyramidal amino groups. Large buckle, on the other hand, distorts H-bonds. Similar, non-planar structures of G-quartets in presence of K^+ ion have also been observed by high level DFT calculations (Gkionis *et al.*, 2014; Meyer *et al.*, 2001; Yurenko *et al.*, 2014). It was also indicated earlier that a G-quartet formation is favoured by K^+ over Na^+ due to higher dehydration energy of Na^+ (Deng and Braunlin, 1996; Hud *et al.*, 1996). When the base quartet was optimized using HF/cc-pVDZ method without any ion, the shear values for the base pairs reached 4.11Å as opposed to ~2.5Å for the base pairs of Na^+ ion containing quartets. Similarly, when a single K^+ ion was present in the optimized structure instead of Na^+ and MP2/6-31G(2d,2p) was used as method, the base pairs of ion containing quartet were seen to attain small shear value of 1.7Å. In both these situations low shear values were accompanied by low propeller values of about 7° **(Table 2.4)**.

A G-quadruplex consists of two or more stacked G-quartets. We have therefore, calculated inter- and intra- base pair parameters for the four NMR structures and two crystal structures of telomeric G-quadruplex **(Table 2.5 and 2.6)**. Furthermore, from experimental observations it is known that inside a quadruplex a G-quartet is mostly stabilized between two ions (Špačková *et al.*, 1999). It is difficult to evaluate the true nature of a G-tetrad plane from the available structural information as these indicate non-planarity of various extents. While the NMR structures indicate non-planar geometry with large buckle and propeller values, the X-ray crystal structures mostly report planar G-quartets **(Table 2.5)**. Thus, we have also optimized the G-quartet in presence of two Na^+ and two K^+ ions separately with B3LYP/6-31G(2d,2p) method. The buckle, propeller and stagger values are found lower when two K^+ ions are present inducing planarity to the G-quartet **(Table 2.4)**. We have observed G-quartet stability relative to ion-coordination **(Table 2.3)** as presence of two K^+ ions on either side of a G-quartet gives it the more planarity as compared to Na^+ situation **(Figure 2.3d and 2.3e)**. On the other hand, one Na^+ inside the G-tetrad core gives similar planarity to the G-quartets **(Figure 2.3b)**.

We have calculated partial charges of the atoms using NBO renormalization (Glendening *et al.*, 2012; Reed *et al.*, 1988) and found slight charge reduction of the ions (ca 0.8 from 1.0). Similar charge redistribution was also observed in buried charge systems (Halder *et al.*, 2014; Mukherjee and Bhattacharyya, 2013). However,

high charge redistribution was indicated in an earlier report (Yurenko *et al.*, 2014) considering Mulliken charges. Nevertheless a single G-quartet does not encompass the ion from all sides, for which the full G-quadruplex fold is required and in this case charge reduction of the ion is quite possible.

Table 2.3: Details of H-bond parameters (bond length and bond angle) and O6–(ion) –O6 angle obtained from optimized geometries of G-tetrad in presence and absence of ions.

parameters	1-sod-1G4	1-pot-1G4	no-ion-1G4	2-sod-1G4	2-pot-1G4
H-bond length (N-H…O)	1.853 (0.00)	1.899 (0.00)	1.751 (0.01)	1.909 (0.00)	1.928 (0.00)
H-bond length (N-H…N)	1.909 (0.00)	1.915 (0.00)	1.990 (0.01)	1.889 (0.01)	1.883 (0.01)
H-bond angle (N-H…O)	164.095 (0.04)	168.652 (0.03)	167.907 (9.59)	154.976 (0.15)	160.007 (0.15)
H-bond angle (N-H…N)	175.426 (0.02)	171.247 (0.30)	166.568 (1.78)	176.369 (0.93)	178.222 (0.07)
O6– (ion)–O6 (diagonal O6)	175.648 (0.21)	134.028 (0.33)		115.784 (12.81)	100.442 (1.12)
O6–(ion)–O6 (adjacent O6)	90.083 (0.03)	81.229 (0.12)		74.147 (0.16)	65.839 (0.14)

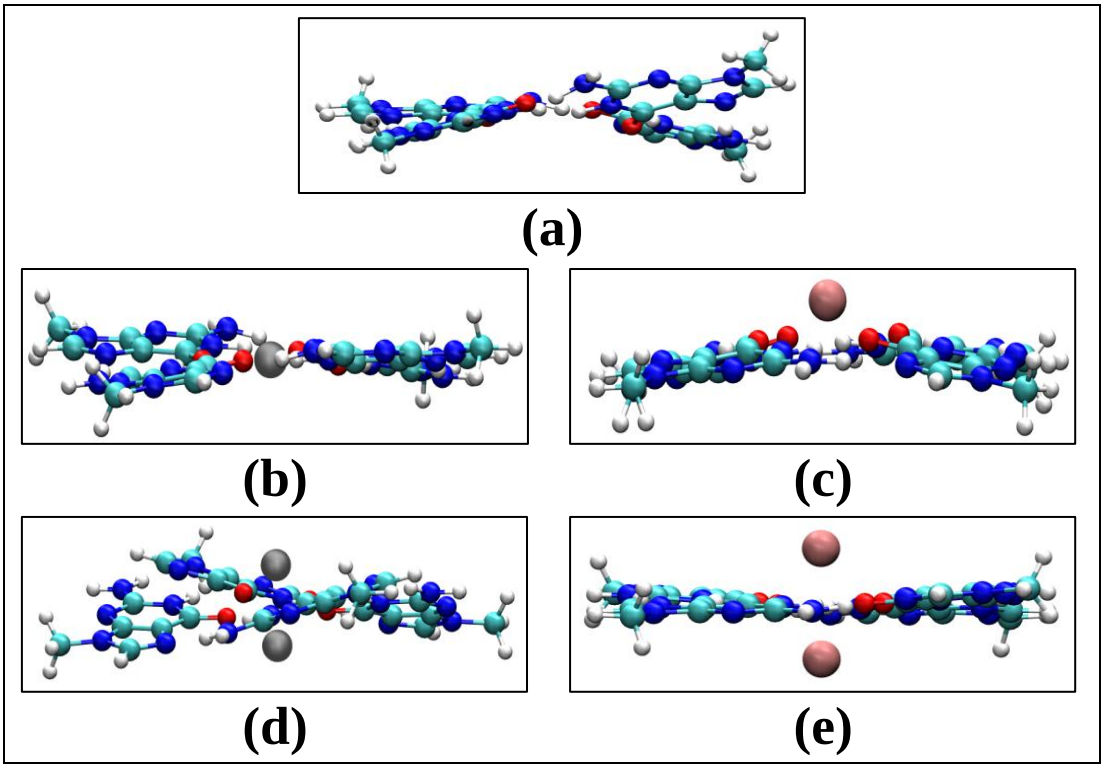


Figure 2.3: Optimized structure of a G-quartet (a) in absence of any ions, (b) containing a single Na^+ ion (steel), (c) in presence of one K^+ ion (pink) in its core, (d) containing two Na^+ ions (steel) on either side of its core and (e) containing two K^+ ion (pink) on either side of its core.

Table 2.4: Base pair orientation parameters of crystal structure of a G-quartet (156D) and its optimized geometry obtained by quantum chemical calculations in presence and absence of ions (Na^+/K^+).

method	base pair	buckle	open	propeller	stagger	shear	stretch	ion
GGA:PW91/ DZP	G1:G2	0.37	-5.55	-19.05	-0.12	2.83	2.86	none
		-3.09	-4.53	-3.68	-0.16	2.72	2.92	K^+
		-1.26	-5.47	-8.18	-0.05	2.67	2.92	Na^+
	G2:G3	10.82	-5.40	2.42	0.04	2.83	2.88	none
		-11.17	-4.99	10.14	-0.09	2.72	2.90	K^+
		-8.58	-5.30	10.90	-0.01	2.65	2.92	Na^+
	G3:G4	11.93	-6.09	-16.05	-0.13	2.90	2.92	none
		-16.53	-3.48	2.14	-0.07	2.70	2.98	K^+
		4.99	-5.68	-11.01	-0.11	2.55	2.96	Na^+
	G4:G1	-0.41	-8.92	-25.09	0.12	3.01	2.84	none
		-15.08	-3.63	-0.45	-0.09	2.71	2.98	K^+
		4.82	-5.69	-10.83	-0.11	2.55	2.96	Na^+
B3LYP/ 6-31G(2d,2p)	G1:G2	-0.41	-8.92	-25.09	0.12	3.01	2.84	none
		-15.08	-3.63	-0.45	-0.09	2.71	2.98	K^+
		4.82	-5.69	-10.83	-0.11	2.55	2.96	Na^+
		-20.79	-7.48	-3.91	-0.33	2.28	2.93	2 Na^+
		-1.04	-5.61	6.18	0.08	2.39	2.97	2 K^+
	G2:G3	-0.79	-5.62	16.25	0.19	2.95	2.91	none
		-12.88	-3.59	5.55	-0.02	2.69	2.98	K^+
		-5.02	-5.73	10.83	0.11	2.55	2.96	Na^+
		21.63	-7.46	4.15	0.33	2.28	2.93	2 Na^+
		1.14	-5.61	-6.03	-0.08	2.39	2.97	2 K^+
	G3:G4	11.93	-6.09	-16.05	-0.13	2.90	2.92	none
		-16.53	-3.48	2.14	-0.07	2.70	2.98	K^+
		4.99	-5.68	-11.01	-0.11	2.55	2.96	Na^+
		-21.15	-7.45	-4.03	-0.32	2.28	2.93	2 Na^+
		-1.11	-5.59	6.38	0.09	2.39	2.97	2 K^+
	G4:G1	-0.41	-8.92	-25.09	0.12	3.01	2.84	none
		-15.08	-3.63	-0.45	-0.09	2.71	2.98	K^+
		4.82	-5.69	-10.83	-0.11	2.55	2.96	Na^+
		21.75	-7.48	3.79	0.34	2.28	2.93	2 Na^+
		0.95	-5.62	-6.10	-0.08	2.39	2.97	2 K^+
$^{\$}$MP2/ 6-31G**	G1:G2	23.73	-9.45	-26.08	-0.06	2.70	2.93	none
		-13.22	2.58	-38.97	-0.75	1.74	2.54	K^+
	G2:G3	-4.34	-7.92	26.83	0.15	2.78	2.91	none
		31.59	-2.92	-8.39	0.08	2.17	2.54	K^+
	G3:G4	20.60	-10.27	-29.45	-0.04	2.69	2.92	none
		9.03	-9.35	6.36	-0.27	1.73	2.92	K^+
	G4:G1	-0.79	-7.28	23.68	0.21	2.77	2.93	none
		-13.70	-6.93	12.13	-1.00	2.65	2.79	K^+
*HF/cc-pVDZ	G1:G2	7.30	-3.41	-15.19	-0.33	3.48	2.97	none
		6.04	-4.90	5.18	0.21	2.48	3.07	Na^+
	G2:G3	7.30	-3.41	-15.19	-0.33	3.48	2.97	none

		6.04	-4.90	5.18	0.21	2.48	3.07	Na^+
	G3:G4	20.60	-10.27	-29.45	-0.04	2.69	2.92	none
		9.03	-9.35	6.36	-0.27	1.73	2.92	Na^+
	G4:G1	-16.33	-6.79	1.40	0.01	4.11	2.94	none
		-0.90	-5.12	-0.09	-0.13	2.48	3.07	Na^+
crystal structure	G1:G2	31.12	-1.40	-10.04	0.11	2.25	2.91	none
	G2:G3	1.26	-4.73	7.04	0.31	1.86	2.82	
	G3:G4	-9.61	-3.20	0.89	-0.33	2.16	2.91	
	G4:G1	-4.92	-4.77	-1.38	-0.16	2.23	2.96	

[$] optimizations of G-quartet with Na^+ ions using MP2/6-31G(2d,2p) did not converge.
[*] parameters for K^+ were not available for this basis set.

Table 2.5: Intra-base pair parameters of initial geometry of the four NMR structures and two crystal structures of telomeric G-quadruplexes.

buckle	open	propeller	stagger	shear	stretch	base pair (bp)	quartet & bp type
NMR structure with parallel topology (PDB ID: 2LD8) (initial geometry for simulations 1a & 1b)							
4.19	-4.47	3.2	0.82	3.14	2.80	G3:G9	
-0.10	-8.37	3.94	-0.27	3.10	2.95	G9:G15	Q1
-8.79	1.50	9.20	0.01	3.15	2.77	G15:G21	W:H
-5.94	-6.78	-9.28	-0.29	2.67	2.95	G21:G3	
16.54	-4.81	-0.79	0.39	2.80	2.82	G4:G10	
5.98	0.0	-6.58	-0.44	2.05	2.82	G10:G16	Q2
8.63	-10.74	8.58	-0.11	2.69	2.99	G16:G22	W:H
-3.69	-2.73	-14.49	-0.41	2.21	2.78	G22:G4	
20.85	-2.56	-9.45	-0.54	2.57	2.78	G5:G11	
-3.27	-6.97	-13.60	-0.39	2.22	2.82	G11:G17	Q3
31.73	-3.96	-0.57	-0.17	2.70	2.84	G17:G23	W:H
-12.47	-6.89	-2.56	-0.33	2.31	2.87	G23:G5	
NMR structure with anti-parallel topology (PDB ID: 143D) (initial geometry for simulations 2a & 2b)							
-0.95	-2.66	1.47	0.67	2.60	2.76	G2:G10	
-19.99	-4.79	12.89	0.08	2.64	2.87	G10:G22	Q1
-25.57	-0.41	-6.20	-0.29	2.89	2.85	G14:G2	W:H
-21.06	-5.69	7.69	-0.48	2.75	2.86	G22:G14	
-13.95	-4.87	-3.73	0.36	-2.80	2.88	G3:G9	
-8.92	-1.90	8.17	0.37	-2.50	2.86	G9:G21	Q2
-17.78	-1.80	-2.78	0.26	-3.06	2.82	G21:G15	H:W
-10.74	-6.65	3.91	0.28	-2.78	2.95	G15:G3	
16.32	7.39	19.13	1.22	2.53	2.53	G4:G8	
9.87	-8.38	12.42	0.46	2.07	2.83	G8:G20	Q3
-4.85	-7.22	4.03	0.01	3.01	2.96	G20:G16	W:H
25.03	-7.67	-10.86	0.78	2.51	2.83	G16:G4	
NMR structure with mixed-(3+1)-form1 topology (PDB ID: 2GKU) (initial geometry for simulations 3a & 3b)							
-4.51	-3.28	5.78	0.20	-2.30	2.96	G3:G9	
-14.58	-6.06	-2.30	-0.25	-2.18	2.92	G9:G17	Q1
1.09	-2.58	-10.39	-0.05	-2.16	2.88	G17:G21	H:W
1.60	-5.38	3.25	0.03	-2.15	2.96	G21:G3	

-4.63	-4.44	-8.28	-0.28	2.61	3.00	G4:G10	
9.71	-9.75	-0.27	0.22	2.45	2.97	G10:G16	Q2
-4.38	2.09	5.92	0.13	2.73	2.93	G16:G22	W:H
-4.33	174.9	-0.19	0.29	0.50	-7.94	G22:G4	
10.54	-5.18	-5.78	-0.07	2.73	2.91	G5:G11	
12.00	0.59	-4.67	-0.06	2.53	2.94	G11:G15	Q3
13.68	-6.58	0.88	0.11	2.59	3.03	G15:G23	W:H
5.73	-4.02	-1.74	-0.14	2.60	3.05	G23:G5	
NMR structure with mixed-(3+1)-form2 topology (PDB ID: 2JSL) (initial geometry for simulations 4a & 4b)							
-14.43	-4.96	3.50	-0.24	-2.61	3.05	G3:G11	
-3.01	-4.32	-12.12	0.06	-2.32	3.05	G11:G15	Q1
1.23	-4.37	-1.19	0.05	-2.55	3.07	G15:G21	H:W
1.76	-4.10	3.52	-0.14	-2.45	3.05	G21:G3	
-1.10	-5.71	4.77	-0.13	2.81	3.08	G4:G10	
0.10	-1.23	-7.22	0.07	2.87	3.04	G10:G16	Q2
7.33	-5.23	-1.94	0.04	2.50	3.08	G16:G22	W:H
6.94	-4.62	1.27	0.02	2.92	3.02	G22:G4	
-9.89	-0.35	4.88	-0.20	2.77	3.01	G5:G9	
7.90	-6.13	-10.84	-0.09	2.73	3.05	G9:G17	Q3
10.58	-7.81	-4.78	0.39	2.99	3.00	G17:G23	W:H
17.49	-2.82	8.62	0.22	3.31	2.93	G23:G5	
Crystal structure (parallel topology) (PDB ID: 1KF1) (initial geometry for simulation6)							
1.75	-5.95	-3.60	0.20	2.52	2.97	G2:G8	
1.86	-4.27	1.38	0.01	2.50	2.87	G8:G14	Q1
3.98	-4.38	-1.65	-0.03	2.34	2.99	G14:G20	W:H
-0.04	-5.66	3.62	-0.09	2.52	2.84	G20:G2	
10.19	-1.60	-2.45	-0.15	2.40	2.91	G3:G9	
11.32	-7.96	-3.10	0.02	2.43	2.91	G9:G15	Q2
6.17	-2.10	2.61	-0.21	2.60	2.81	G15:G21	W:H
6.09	-5.78	-7.18	0.12	2.47	2.92	G21:G3	
13.54	-1.59	-1.87	0.12	2.38	2.82	G4:G10	
15.66	-4.77	-4.09	-0.15	2.43	2.90	G10:G16	Q3
15.93	-2.56	-2.85	0.20	2.51	2.86	G16:G22	W:H
16.51	-4.51	-1.88	-0.04	2.48	2.99	G22:G4	
Crystal structure (parallel topology) (PDB ID: 3R6R) (initial geometry for simulation7)							
-1.17	-5.44	-4.33	-0.15	2.79	2.83	G2:G8	
2.28	-3.22	-3.84	-0.10	2.73	2.85	G8:G14	Q1
3.33	-4.41	-3.24	-0.30	2.70	2.88	G14:G20	W:H
-0.04	-5.66	3.62	-0.09	2.52	2.84	G20:G2	
3.34	-0.93	-5.65	-0.18	2.53	2.88	G3:G9	
8.98	-5.17	-1.20	-0.15	2.48	2.84	G9:G15	Q2
1.60	-2.67	-2.72	-0.16	2.56	2.98	G15:G21	W:H
6.10	-6.98	-3.14	-0.12	2.50	2.98	G21:G3	
5.73	-0.43	0.42	-0.06	2.82	2.87	G4:G10	
14.10	-5.95	-3.72	0.05	2.66	2.95	G10:G16	Q3
4.63	-1.58	-0.24	0.01	2.79	2.90	G16:G22	W:H
14.78	-6.71	-4.93	0.05	2.71	2.98	G22:G4	

Table 2.6: Intra-base pair parameters of initial geometry of the four NMR and two crystal structures of telomeric G-quadruplexes.

tilt	roll	twist	shift	slide	rise	base pair (bp)	strand orient[n]	stack	bp type
parallel topology (PDB ID: 2LD8) (initial geometry for simulations 1a & 1b)									
1.77	6.53	22.17	0.12	-1.90	3.55	G3:G9::G4:G10	***p****		W:H::W:H
-3.14	4.78	22.33	-0.33	-1.84	3.34	G9:G15::G10:G16	***p***	Q1-	W:H::W:H
2.13	4.59	21.11	0.54	-1.45	3.13	G15:G21::G16:G22	***p***	Q2	W:H::W:H
-3.43	5.60	25.94	0.22	-2.40	3.25	G21:G3::G22:G4	***p***		W:H::W:H
-2.13	-2.48	30.96	0.03	-2.22	3.57	G4:G10::G5:G11	***p***		W:H::W:H
-5.18	2.24	30.58	-0.72	-2.80	3.41	G10:G16::G11:G17	***p***	Q2-	W:H::W:H
-1.02	-0.74	30.92	0.17	-1.93	3.08	G16:G22::G17:G23	***p***	Q3	W:H::W:H
-0.94	6.95	29.31	-0.38	-2.96	3.09	G22:G4::G23:G5	***p***		W:H::W:H
anti-parallel topology (PDB ID: 143D) (initial geometry for simulations 2a & 2b)									
-0.41	10.46	37.99	0.92	-1.99	4.35	G2:G10::G3:G9	***ap****		W:H::H:W
9.28	3.59	36.39	1.48	3.41	4.24	G9:G21::G10:G22	***p***	Q1-	H:W::W:H
0.75	-5.09	35.03	-0.19	3.97	3.18	G21:G15::G22:G14	***ap***	Q2	H:W::W:H
9.36	-0.43	34.76	-0.37	-2.61	3.23	G14:G2::G15:G3	***p***		W:H::H:W
6.77	16.99	8.07	0.42	-1.66	3.10	G3:G9::G4:G8	***ap***		H:W::W:H
0.84	-11.81	10.99	-0.28	1.13	3.72	G8:G20::G9:G21	***p***	Q2-	W:H::H:W
1.30	-6.67	11.47	-0.53	0.79	3.76	G20:G16::G21:G15	***ap***	Q3	W:H::H:W
-7.63	6.62	16.99	0.93	-0.39	2.98	G15:G3::G16:G4	***p***		H:W::W:H
mixed-(3+1)-form1 topology (2GKU) (initial geometry for simulations 3a & 3b)									
-1.32	1.31	14.95	-0.49	-1.72	3.64	G3:G9::G4:G10	***p***		H:W::W:H
0.08	-3.20	18.52	-0.14	-0.68	2.99	G9:G17::G10:G16	***ap***	Q1-	H:W::W:H
1.14	-6.63	16.95	0.51	1.50	3.63	G16:G22::G17:G21	***ap***	Q2	W:H::H:W
1.61	2.19	17.63	0.13	-1.72	3.85	G21:G3::G22:G4	***p***		H:W::W:H
-5.53	0.91	27.58	-0.84	-2.67	3.00	G4:G10::G5:G11	***p***		W:H::W:H
-6.27	16.04	23.98	-0.76	-1.51	4.46	G10:G16::G11:G15	***ap***	Q2-	W:H::W:H
7.38	-5.83	25.17	0.43	1.44	4.46	G15:G23::G16:G22	***ap***	Q3	W:H::W:H
-0.26	0.48	26.14	0.21	-2.89	3.03	G22:G4::G23:G5	***p***		W:H::W:H
mixed-(3+1)-form2 topology (2JSL) (initial geometry for simulations 4a & 4b)									
-0.57	5.74	19.83	0.46	-1.29	3.42	G3:G11::G4:G10	***ap***		H:W::W:H
2.63	-2.24	19.93	0.25	2.42	3.73	G10:G16::G11:G15	***ap***	Q1-	W:H::H:W
-1.90	1.91	19.51	-0.75	-1.78	3.48	G15:G21::G16:G22	***p***	Q2	H:W::W:H
0.05	4.67	19.34	-0.30	-1.20	3.52	G21:G3::G22:G4	***p***		H:W::W:H
-3.85	8.16	24.01	-0.01	-1.87	4.19	G4:G10::G5:G9	***ap***		W:H::W:H
5.53	3.85	23.42	0.30	2.47	3.64	G9:G17::G10:G16	***ap***	Q2-	W:H::W:H
-5.33	-0.58	24.69	-0.66	-2.37	3.21	G16:G22::G17:G23	***p***	Q3	W:H::W:H
-0.51	1.87	25.92	-0.33	-1.77	3.31	G22:G4::G23:G5	***p***		W:H::W:H
Crystal structure (parallel topology) (PDB ID: 1KF1) (initial geometry for simulation6)									
0.80	5.98	27.88	-0.21	-2.44	3.42	G2:G8::G3:G9	***p***		W:H::W:H
-2.29	1.79	28.12	-0.30	-2.25	3.32	G8:G14::G9:G15	***p***	Q1-	W:H::W:H
0.55	3.74	28.43	-0.13	-2.21	3.43	G14:G20::G15:G21	***p***	Q2	W:H::W:H
-4.36	2.50	27.94	0.05	-2.41	3.47	G20:G2::G21:G3	***p***		W:H::W:H
-3.59	3.49	29.08	-0.49	-2.49	3.40	G3:G9::G4:G10	***p***		W:H::W:H
-3.50	2.86	28.40	-0.43	-2.41	3.44	G9:G15::G10:G16	***p***	Q2-	W:H::W:H
-3.54	3.81	28.23	-0.38	-2.25	3.34	G15:G21::G16:G22	***p***	Q3	W:H::W:H
-1.97	3.44	29.41	-0.24	-2.45	3.19	G21:G3::G22:G4	***p***		W:H::W:H
Crystal structure (parallel topology) (PDB ID: 3R6R) (initial geometry for simulation7)									
-1.38	3.00	31.09	-0.07	-2.59	3.31	G2:G8::G3:G9	***p***	Q1-Q2	W:H::W:H
0.55	0.92	29.20	-0.13	-2.48	3.23	G8:G14::G9:G15	***p***		W:H::W:H
-1.71	1.36	29.40	-0.15	-2.44	3.30	G14:G20::G15:G21	***p***		W:H::W:H

-1.75	1.80	29.85	0.00	-2.43	3.36	G20:G2::G21:G3	***p***		W:H::W:H
-1.47	2.66	26.69	-0.48	-2.41	3.21	G3:G9::G4:G10	***p***		W:H::W:H
-2.94	0.97	27.69	-0.24	-2.19	3.32	G9:G15::G10:G16	***p***	Q2-	W:H::W:H
-1.37	3.96	27.57	-0.37	-2.29	3.32	G15:G21::G16:G22	***p***	Q3	W:H::W:H
-1.87	0.80	27.38	-0.13	-2.28	3.18	G21:G3::G22:G4	***p***		W:H::W:H

***p**: parallel
****ap**: anti-parallel

2.3.2 Molecular Dynamics Simulation Analysis

2.3.2.1 Dynamics and Fluctuation

The root-mean-square deviations (RMSD) of the trajectories with respect to the energy-minimized structures of the corresponding quadruplexes were calculated with the aim of studying overall behaviour of the simulations. The RMSD values were calculated for both the all-atom structure and the G-quartet stacks **(Table 2.1)**. The structural integrity of all the human telomeric G-quadruplexes with Na^+ or K^+ counter-ions persisted during the complete course of 200ns MD-run as the RMSD values rarely reached 2Å for the guanine residues **(Figure 2.4)**. The all-atom RMSD vs. time plots indicated that only after a time period of about 20ns the systems settle down. Therefore, we treated this as equilibration time and carried out all analyses on the trajectory portions beyond this point, except for simulation2b, which took 25ns to settle down and hence its equilibration time was considered to be 25ns. The average RMSDs of the all-atom structures were found to be slightly greater than those of the G-quartet stacks, which can be considered as a result of loop flexibility. The standard deviation in RMSD for G-tetrad stacks are lowest for the anti-parallel topology **(Figure 2.4)**, indicating most stable structures. The mixed-(3+1)-form2 shows slight instability with higher fluctuations of RMSD. The average RMSD along with their standard deviations for G-quartet stacks in simulations 5a and 5b for the non-telomeric form was found to be the highest **(Table 2.1)**.

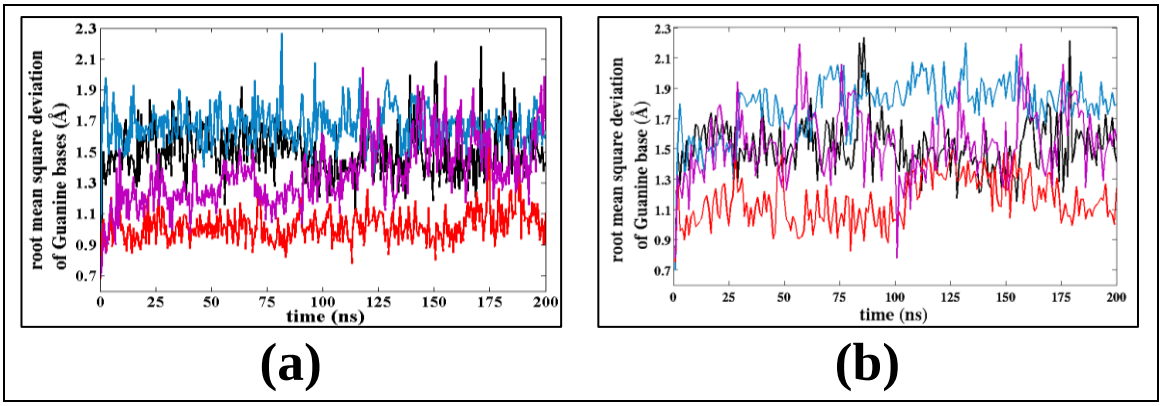


Figure 2.4: RMSD of guanine stacks of the four topology-types simulated with (a) Na^+ counter-ions and (b) K^+ counter-ions. The black, blue, red and magenta lines represent parallel, anti-parallel, mixed-(3+1)-form1 and mixed-(3+1)-form2 topologies respectively.

It is already established that G-quadruplex formation requires presence of monovalent cations such as Na^+, K^+, or NH_4^+ (Neidle and Parkinson, 2003) in its electronegative central channel created by the O6 carbonyl oxygen atoms of guanine residues. Our Monte Carlo ion-placement method positioned two ions (Na^+/K^+) inside this channel. These ions were retained in their positions during the energy minimization steps. The MD simulations may have overcome the local energy barrier and moved the ions from inside the G-tetrad core into the solvated state. We have therefore plotted time development of ion-position inside the quadruplex for the all simulations **(Figure 2.5.1 and 2.5.2)**. We found that the Na^+ ion which was expelled into the solvent re-entered the G-tetrad core both in case of anti-parallel and mixed-(3+1)-form1 topologies. The simulations with K^+ ions showed a similar trend of ion-movement, although to a lower extent. However, in all the simulations, one of the counter-ions retained its position inside the central-ion channel. Thus, in spite of such ion movements, the G-quadruplex structures do not become completely unstable and a clear idea about their topologies can still be obtained. It may further be noted that such ions expulsion, possibly due to mutual electrostatic repulsion was also indicated from high resolution crystallography (Laughlan *et al.*, 1994). Movement of ions from the G-tetrad core into the solvent have also been reported from theoretical calculations and attributed to force field errors (Chowdhury and Bansal, 2000; Fadrná *et al.*, 2009; Hazel *et al.*, 2006; Šponer *et al.*, 2012). As indicated earlier, significant charge diffusion from the cation to the guanine bases is possible, especially within a G-tetrad core (Halder *et al.*, 2014; Yurenko *et al.*, 2014). This force-field, however, does not consider such polarization effect and may overestimate ionic repulsions.

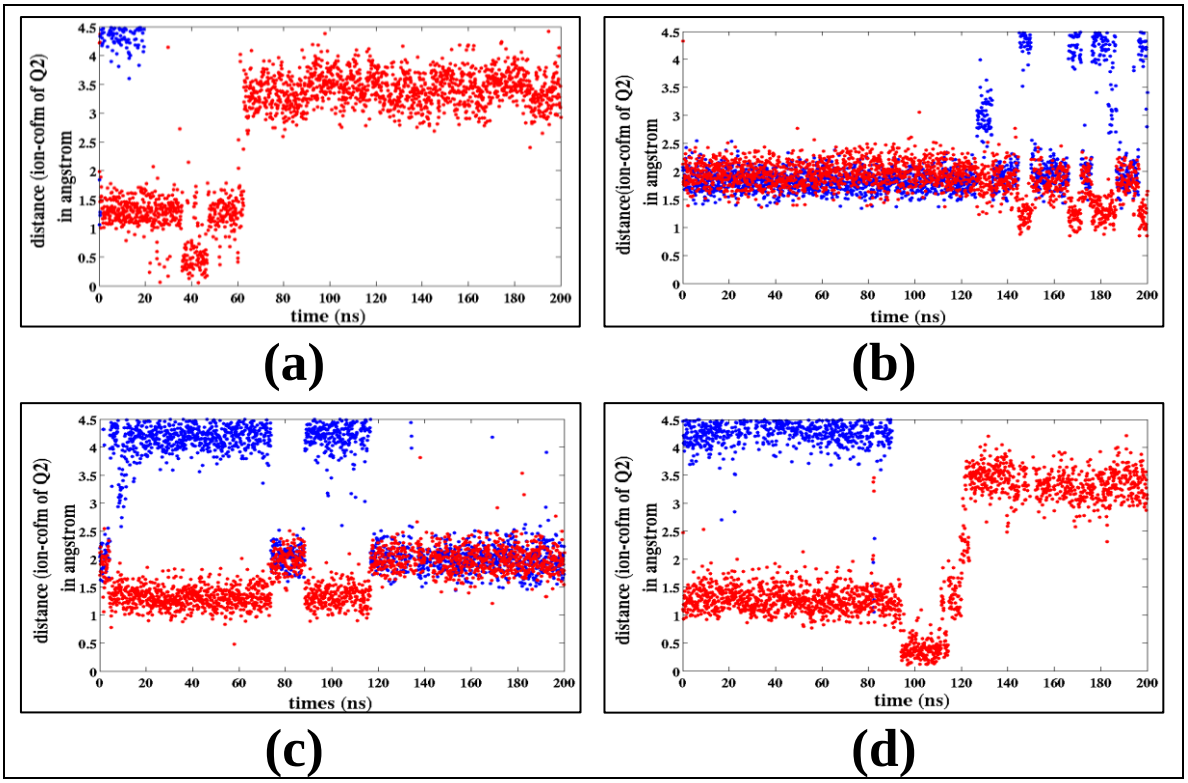


Figure 2.5.1: Time evolution plot of two Na+ ions (red and blue) initially present inside G-tetrad core of (a) parallel, (b) anti-parallel, (c) mixed-(3+1)-form1 and (d) mixed-(3+1)-form2 topologies.

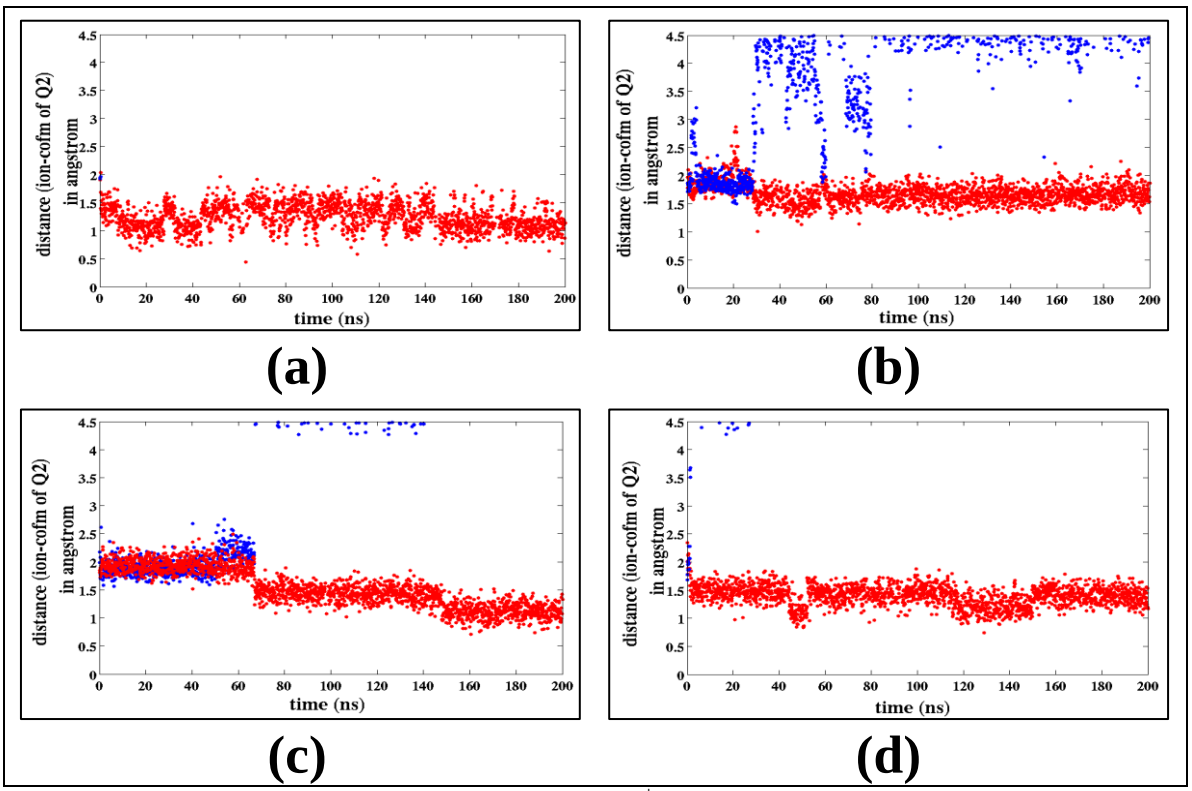


Figure 2.5.2: Time evolution plot of two K^+ ions (red and blue) initially present inside G-tetrad core of (a) parallel, (b) anti-parallel, (c) mixed-(3+1)-form1 and (d) mixed-(3+1)-form2 topologies.

The Root-mean-square fluctuations (RMSF) of the base atoms for all the simulations were calculated using CHARMM **(Figure 2.6)** software. The basket-type anti-parallel quadruplex (simulations 2a and 2b) has low RMSF for most of the bases (less than 1.5Å). Although, the guanine bases in propeller-type parallel-stranded quadruplex were found to be stable with less than 1.5Å RMSF, some of its loop bases were seen to have comparatively high RMSF of about 3.5Å. The guanine bases of mixed-(3+1)-form1 topology were seen to have lowest fluctuations in presence of both Na^+ (simulation3a) and K^+ (simulation3b) ions. The mixed-(3+1)-form2 topology followed a similar trend with slightly higher RMSF.

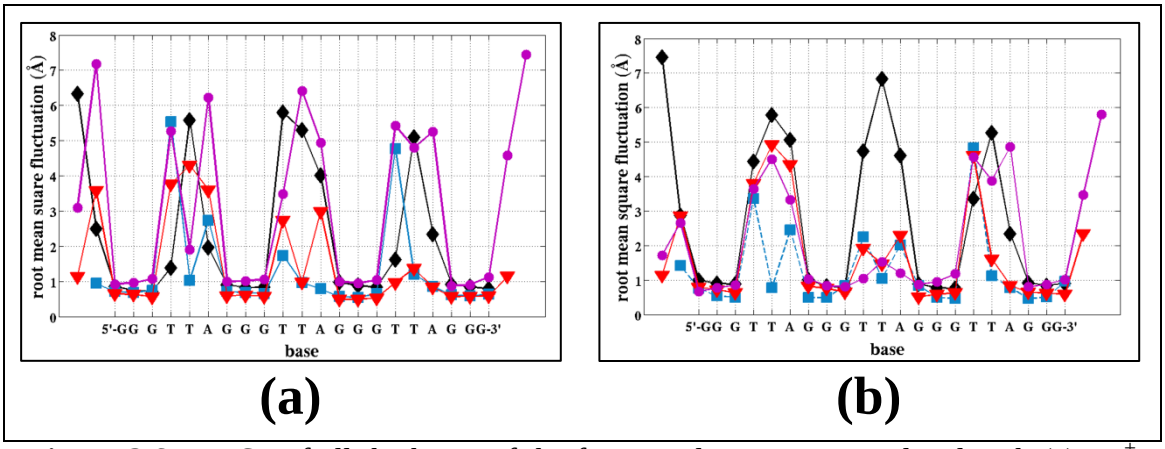


Figure 2.6: RMSF of all the bases of the four topology-types simulated with (a) Na^+ counter-ions and (b) K^+ counter-ions. The black, blue, red and magenta lines represent parallel, anti-parallel, mixed-(3+1)-form1 and mixed-(3+1)-form2 topologies respectively.

The residue average temperature factor (B-factors) for the initial crystal structures were calculated from the multiple models reported in NMR structure by GROMACS (Hess *et al.*, 2008) software to compare with our simulation data **(Table 2.7.1)**. Similarly, the B-factors for these structures were also obtained from the respective simulated snapshots **(Table 2.7.2)**. We found that the B-factors are always higher for the loop residues of all the experimentally reported structures as well as our simulations. We observed that the difference in B-factors for the loop residues and the G-stem residues were amplified when calculated from simulations as compared to those from crystal. The relatively smaller B-factors in crystal structures are probably due to contact between symmetry related molecules restricting the movement of the loops. The guanine bases in propeller-type parallel-stranded quadruplex were found to have B-factors greater than $15Å^2$, whereas, for the anti-parallel and mixed-(3+1)-

form1 quadruplexes the B-factors attain values less than 10Å^2. It is noted that the some of the adenine and thymine residues of the loop-region of anti-parallel and mixed-(3+1)-form1 have B-factors quiet comparable to those of guanine residues, which is in correspondence to the fact that there exists Ade:Thy base pairs in the loops of these topologies **(Figure 2.2.3a)**. Interestingly, such base pairing between the loop residues is completely absent in the parallel topology while they appear as a capping elements in the anti-parallel and mixed-forms.

Table 2.7.1: Average temperature-factor (B-factor in Å^2) of each residue for the initial structure as obtained from PDB. Number of models (conformers submitted) in the NMR structures are given below the PDB-IDs.

base no.	seq	1KF1	3R6R	parallel (2LD8) (10)	anti-parallel (143D) (6)	mixed-(3+1)-form1(2GKU) (12)	mixed-(3+1)-form2 (2JSL) (10)
		X-ray crystal structures (parallel)		NMR reported structures			
1				30.96		14.44	19.41
2		58.58	41.79	18.29	36.08	9.98	29.96
3	GUA	40.51	27.50	4.86	20.48	4.12	2.07
4	GUA	37.25	25.74	3.84	21.54	2.53	4.36
5	GUA	40.36	23.78	5.29	20.27	2.8	3.89
6	THY	52.89	37.36	11.4	298.78	52.79	25.33
7	THY	55.18	30.61	19.8	33	17.08	6.15
8	ADE	41.54	29.09	11.03	23.93	119.73	5.35
9	GUA	36.85	20.53	5.32	7.39	12.04	3.37
10	GUA	38.30	17.75	2.75	12.73	3.16	2.09
11	GUA	36.30	19.56	3.6	31.76	3.57	3.89
12	THY	52.90	31.52	8.49	348.94	328.69	10.36
13	THY	47.51	28.83	12.99	144.86	8.09	8.71
14	ADE	47.30	28.29	4.75	65.29	10.39	3.32
15	GUA	40.36	21.85	4.09	13.94	8.15	1.77
16	GUA	35.77	21.64	3.04	17.27	4.76	3.11
17	GUA	35.25	22.62	2.58	34.59	2.45	5.64
18	THY	50.02	37.41	9.32	763.63	37.34	41.41
19	THY	47.78	29.86	14.48	673.32	19.72	98.47
20	ADE	48.74	31.96	5.98	17.81	4.73	223.15
21	GUA	34.62	20.24	2.77	26.14	10.11	10.09
22	GUA	34.78	21.41	4.16	18.55	2.12	8.67
23	GUA	41.60	23.92	6.51	9.08	2.14	5.2
24						18.31	5.04
25							9.52

Table 2.7.2: Average temperature-factor (B-factor in Å^2) of each residue calculated from simulation trajectories.

		B-factor (temperature factor) calculated based on snapshots from the last 180ns simulation trajectory								
base no.	**seq**	**Simulation1a**	**Simulation1b**	**Simulation2a**	**Simulation2b**	**Simulation3a**	**Simulation3b**	**Simulation4a**	**Simulation4b**	**Simulation6**
		parallel topology		**anti-parallel topology**		**mixed-(3+1)-form1 topology**		**mixed-(3+1)-form2 topology**		**parallel topology**
1		1097.27	1512.63			34.17	34.33	265.44	84.431	
2		167.25	204.88	23.80	56.40	352.63	221.61	1387.33	206.329	175.31
3	GUA	18.26	24.76	12.72	14.78	10.29	15.05	20.43	11.755	14.23
4	GUA	14.71	20.93	12.11	7.11	10.01	12.73	22.43	14.686	10.04
5	GUA	12.90	19.98	14.77	6.24	8.18	9.98	29.03	19.128	9.86
6	THY	53.01	527.05	863.52	311.11	387.77	412.31	759.58	356.157	88.37
7	THY	842.66	867.83	**26.33**	**15.70**	499.64	647.47	92.07	561.152	659.10
8	ADE	106.19	744.42	204.53	166.16	396.84	509.55	1050.75	301.817	109.80
9	GUA	20.85	26.59	12.48	6.21	58.49	16.90	23.6	24.745	16.32
10	GUA	17.20	17.85	11.32	6.08	9.54	14.79	24.9	16.558	10.94
11	GUA	16.83	14.74	9.86	18.54	8.81	10.91	26.15	15.583	10.98
12	THY	908.44	636.70	79.01	142.70	205.45	100.05	339.49	27.606	68.63
13	THY	771.11	1171.11	**24.87**	**28.92**	**25.48**	57.01	1055.64	62.336	723.20
14	ADE	451.47	595.76	15.96	109.12	241.20	142.70	679.29	37.670	457.27
15	GUA	23.54	20.18	8.14	18.15	5.96	7.04	24.05	17.977	18.82
16	GUA	18.80	14.25	7.89	6.31	6.03	9.32	22.35	22.292	12.19
17	GUA	16.23	13.16	11.96	5.26	6.82	10.82	26.57	34.515	11.23
18	THY	66.59	304.40	606.51	635.71	23.72	568.64	832.78	570.049	229.99
19	THY	713.88	797.78	38.13	32.13	48.02	68.17	616.55	413.273	897.50
20	ADE	146.03	144.55	**18.40**	**15.36**	**17.95**	**18.41**	768.56	654.486	562.40
21	GUA	20.19	20.46	10.40	5.59	7.95	10.22	19.62	17.804	16.74
22	GUA	17.43	17.09	9.02	6.78	8.46	9.83	20.57	19.616	11.49
23	GUA	15.49	20.10	10.07	24.28	8.66	8.70	31.53	26.792	11.91
24						35.68	145.58	568.73	323.187	
25								14821.4	908.678	

2.3.2.2 Structural Variability

An idea about the initial geometry of the structures reported by NMR was obtained from their base pair and base pair step parameters **(Table 2.5 and 2.6)** calculated using NUPARM software. The intra-base pair parameters viz., propeller, buckle, open, shear, stagger, stretch have been calculated for all the simulated systems **(Table 2.8.1, 2.8.2, 2.9.1, 2.9.2 and 2.10)**. As the G-quadruplex structures are formed by guanine containing sequences only, there is no question of sequence dependant structural variations, as is observed in DNA or RNA double helices. The four stranded helix, however has two G-quartet stacks formed by stacking of base pairs on one another. The possible base pair stacking types are: (i) W:H::W:H, (ii) H:W::W:H and (iii) W:H::H:W, where '::' indicates stacking and ':' indicates base pairing. The roll, twist and slide values were seen to depend on the base pair stacking types **(Table 2.9.1 and 2.9.2)**. For obvious reasons the slide values change sign between H:W::W:H and W:H::H:W stacks of G:G base pairs.

The parallel topology is composed of only W:H C base pairs and follows W:H::W:H type of base pair step pattern. The buckle, open, stagger and stretch parameters of all the W:H C base pairs forming the parallel topology were found to be low and close to B-DNA like in simulation1a *i.e.* in presence of Na^+ counter-ions **(Table 2.8.1)**. However, when the simulation was performed with K^+ counter-ions, the buckle and stagger parameters for the base pairs of Quartet3 were found to be higher **(Table 2.8.2)**. Such difference in values is probably due to the size difference of the counter-ions, larger-ions inducing a non-planarity in the G-quartet (Gkionis *et al.*, 2014; Yurenko *et al.*, 2014). The shear value is always seen to be greater than 2.5Å and independent of the type of counter-ion present, except for the situation where Quartet3 of simulation2b has values less than 2Å but is accompanied by high buckle. The initial geometry from NMR had base pairs steps with twist values around 20^o and 30° along with roll values within 6° **(Table 2.6)**. The twist values for all the steps all of parallel topology (simulations 1a, 1b, 6 and 7) were close to 30°, similar to the values in crystal structures **(Table 2.6, 2.9.1, 2.9.2 and 2.10)**.

The intra-base pair parameter, **(Table 2.5)** calculated for the initial crystal structures of the parallel topology quadruplex (simulations 6 and 7), indicated a slight difference between their initial geometries, which might be attributed to the presence of a ligand in contact with the quadruplex of simulation7. The same intra-base pair

parameters calculated over the trajectories of these simulations **(Table 2.10)** showed high standard deviations for the propeller, buckle, shear and stagger parameters of the base pairs forming the external G-quartets.

The 3D topology of anti-parallel form contains a diagonal loop encompassing Quartet1 and one Ade:Ade (A:A h:h C) base pair acting as a capping for the same. The Quartet3 is capped by two Ade:Thy (A:T H:W T) base pairs. It was shown from analysis of non-coding RNA that hairpin loops of different forms give additional stability to the double helical structures (Lee and Gutell, 2014). The shear values for all the G:G W:H C base pairs of anti-parallel topology (simulation2a) were seen to be less that 2Å which is also lower than that seen for the base pairs of parallel-topology. The capping provided to the outer quartets in case of anti-parallel topology probably gives more firmness to the structure as compared to that of the parallel topology. Since, this topology consists of a mixture of W:H C and H:W C base pairs, there forms two types of base pair steps viz, W:H::H:W and H:W::W:H. The twist values are found to be around 20° accompanied by rather high roll (> 6°). The situation is almost similar for simulation2a and simulation2b, although the variations in roll values are higher when the simulation is performed with larger K^+ counter-ions.

Similar to the anti-parallel topology, the mixed-(3+1)-form1 3-D structure also contains two base pairs involving the loop residues. The capping effect is evident from the low buckle value (< 1°) of G21:G3 H:W C base pair of Quartet1 which is restricted by T1:A20 W:W C base pair, in both simulation3a (Na^+ counter-ion) and simulation3b (K^+ counter-ion). A similar low buckle value is also seen for some base pairs of Quartet3 capped by T13:A24 W:W T base pair. During simulation3a, one of the Na^+ ions attempted to move into the solvent leaving the G-tetrad core, but after about 120ns we observed that both the Na^+ ions remain inside the G-tetrad channel and close to Quartet2 **(Figure 2.5.1)**, similar to our G-quartet and 2-ions model system studied by quantum chemical approach. It is to be noted that, similar to the anti-parallel form, in case of mixed-(3+1)-form1 also there exist restrains imposed on both the outer quartets by two loop-residue base pairs. Unlike this topology, the mixed-(3+1)-form2 contains only one base pair involving the loops residues. The T24:T7 W:W T base pair formed by the distant bases T24 and T7 bases is incapable of providing a capping to Quartet3 which is as strong as the restrain provided by loop region base pairs of other topologies. Thus, the overall shear value is slightly higher than 2Å for both simualtion4a and simulation4b. The G5:G9 W:H C base pair of

Quartet3 has high buckle value compared to the other base pairs of this topology. It might be due to the involvement of T7 in base pair formation which exerts a straining effect on G9, in turn slightly deforming the G9:G17 W:H C base pair.

The mixed topologies have identical 3-D structure and similar base pair step pattern. The Quartet1-Quartet2 stack has one W:H::H:W (G9:G17::G10:G16) and three H:W::W:H C type of base pair steps. The backbone strands may run either in the same direction or in opposite directions giving rise to parallel or anti-parallel strand orientations respectively. The Quartet2-Quartet3 stack has the same W:H::W:H base pair step pattern throughout. The twist parameter in both the topologies clearly attains values in two different ranges: nearly 18° for base pair steps of Quartet1-Quartet2 and ~30° for Quartet2-Quartet3 (W:H::W:H) stack.

The large standard deviations in structural parameters of some of the quartets and G4-stacks can be attributed to the movement of ions. It can thus be stated that the anti-parallel topology and mixed-(3+1)-form1 are the topologies which experienced least structural variability during simulations making them comparatively more rigid and thus, may exist as major conformations. This is in agreement with the experimental finding that the basket-type structure becomes the most stable in 0.1M NaCl solutions (Maruyama *et al.,* 2011). The primary reason may be the presence of restrictions imposed by loops and base pairs in the loop region. Also, the base pair stacking pattern of W:H::H:W or H:W::W:H in these systems add to the stability, since such stacks have base pairs steps with twist ~20° as compared to W:H::W:H stacks with 30° twist.

The structure of the non-telomeric G-rich promoter sequence resembles the mixed-(3+1)-form2 closely in terms of base pairing pattern and loop orientation. In simulation5b most of the base pairs of Quartet1 attained abnormal average buckle of 131°, average open angle of -73°, average stagger of -9Å, *etc.* **(Table 2.8.1 and 2.8.2)** indicating totally unstructured form with broken H-bonds. The average twist angle of this non-telomeric sequence for Quartet1-Quartet2 and Quartet2-Quartet3 stacks are ~20° and ~30°, respectively, in Na^+ environment. In K^+ environment the Quartet1-Quartet2 step parameters adopt absurd values. All these features clearly indicate un-stacking of the Quartet1-Quartet2 stack in presence of K^+ ion in its core, differentiating this from the structure adopted by mixed-(3+1)-form2. Hence, considering such deformity we did not extend these simulations to 200ns.

Table 2.8.1: Average values along with standard deviations (within parenthesis) of the six base pair parameters for the base pairs of the four telomeric topologies (NMR structure) and one non-telomeric structure. This data set has been generated from the trajectories of simulations 1a, 2a, 3a and 4a and 5a where the simulation has been performed with Na^+ counter-ions. Q1, Q2 and Q3 represent Quartet1, Quartet2 and Quartet3 respectively.

buckle	open	propeller	stagger	shear	stretch	base pair (bp)	Quartet	bp type
parallel topology (PDB ID: 2LD8) (simulation1a)								
4.96 (11.26)	-3.30 (5.69)	-8.87 (9.24)	-0.21 (0.55)	3.27 (0.69)	2.87 (0.19)	G3:G9	Q1	W:H
6.17 (10.95)	-3.40 (5.66)	-5.74 (9.76)	-0.23 (0.54)	2.82 (0.73)	2.93 (0.22)	G9:G15		
8.57 (10.14)	-3.50 (5.36)	-4.37 (10.10)	-0.12 (0.54)	3.16 (0.70)	2.89 (0.20)	G15:G21		
1.69 (9.98)	-4.75 (5.06)	-2.28 (9.50)	-0.34 (0.48)	2.84 (0.58)	2.89 (0.15)	G21:G3		
2.97 (9.88)	-3.47 (4.90)	-7.10 (8.69)	-0.25 (0.41)	2.95 (0.73)	2.89 (0.14)	G4:G10	Q2	W:H
2.55 (10.09)	-3.79 (4.81)	-4.44 (8.20)	-0.28 (0.42)	2.95 (0.73)	2.88 (0.15)	G10:G16		
5.51 (9.51)	-3.52 (4.95)	-7.36 (8.93)	-0.25 (0.42)	2.97 (0.74)	2.89 (0.15)	G16:G22		
2.79 (10.05)	-3.60 (5.07)	-1.89 (8.04)	-0.26 (0.41)	2.95 (0.69)	2.88 (0.14)	G22:G4		
7.39 (10.15)	-3.57 (3.77)	-1.69 (8.62)	0.08 (0.44)	2.09 (0.42)	2.98 (0.14)	G5:G11	Q3	W:H
10.59 (11.21)	-3.59 (3.85)	0.65 (8.76)	0.08 (0.46)	2.14 (0.40)	2.98 (0.14)	G11:G17		
8.55 (10.67)	-3.12 (3.88)	-2.66 (9.16)	0.05 (0.46)	2.09 (0.43)	2.98 (0.14)	G17:G23		
11.56 (11.07)	-3.46 (3.82)	1.84 (8.79)	0.10 (0.46)	2.11 (0.40)	2.97 (0.14)	G23:G5		
anti-parallel topology (PDB ID: 143D) (simulation2a)								
-14.39 (8.78)	-2.06 (4.62)	16.12 (12.05)	-0.36 (0.59)	1.17 (0.36)	2.84 (0.15)	G2:G10	Q1	W:H
-37.46 (9.55)	-0.48 (4.01)	-3.57 (9.29)	-0.63 (0.52)	1.64 (0.40)	2.82 (0.16)	G10:G22		
-7.09 (10.00)	-4.18 (4.24)	-13.12 (8.21)	-0.29 (0.49)	1.87 (0.56)	2.88 (0.14)	G14:G2		
-24.17 (9.86)	-1.91 (3.80)	-14.63 (8.47)	-0.28 (0.41)	1.36 (0.30)	2.87 (0.13)	G22:G14		
-6.08 (8.12)	-5.19 (3.89)	5.69 (10.42)	-0.36 (0.44)	-1.37 (0.47)	2.91 (0.14)	G3:G9	Q2	H:W
-15.30 (9.16)	-3.14 (3.59)	-7.73 (8.11)	-0.51 (0.38)	-1.29 (0.47)	2.91 (0.14)	G9:G21		
-3.87 (8.42)	-4.82 (3.45)	-12.42 (8.77)	-0.13 (0.39)	-1.30 (0.43)	2.89 (0.12)	G21:G15		
2.91 (9.70)	-2.35 (3.83)	-5.47 (7.50)	-0.17 (0.49)	-1.18 (0.43)	2.87 (0.13)	G15:G3		

16.79 (10.21)	-2.23 (3.95)	14.15 (10.49)	0.39 (0.50)	1.78 (0.32)	2.82 (0.15)	G4:G8		
-1.71 (8.82)	-2.84 (3.65)	8.27 (8.06)	0.54 (0.44)	1.72 (0.34)	2.87 (0.16)	G8:G20	Q3	W:H
7.43 (8.77)	-4.72 (3.40)	-7.35 (9.57)	0.28 (0.40)	1.67 (0.34)	2.90 (0.13)	G20:G16		
20.19 (17.17)	-3.88 (3.51)	-1.27 (8.38)	0.19 (0.40)	1.78 (0.32)	2.93 (0.13)	G16:G4		
mixed-(3+1)-form1 topology (PDB ID: 2GKU) (simulation3a)								
-4.82 (9.18)	-3.88 (3.60)	8.76 (7.66)	0.43 (0.44)	-1.83 (0.31)	2.87 (0.14)	G3:G9		
-13.09 (12.33)	-3.07 (3.41)	7.29 (8.09)	0.34 (0.36)	-1.75 (0.29)	2.94 (0.14)	G9:G17	Q1	H:W
-4.44 (8.07)	-6.35 (3.24)	-3.79 (10.18)	0.06 (0.43)	-1.77 (0.32)	2.91 (0.12)	G17:G21		
-0.23 (8.47)	-1.05 (3.75)	7.72 (8.49)	0.75 (0.44)	-1.62 (0.32)	2.85 (0.17)	G21:G3		
2.03 (7.04)	-4.80 (3.56)	-7.40 (6.79)	-0.64 (0.36)	1.98 (0.41)	2.86 (0.13)	G4:G10		
5.62 (7.73)	-2.82 (3.48)	-4.11 (6.40)	-0.19 (0.38)	1.70 (0.47)	2.91 (0.12)	G10:G16	Q2	W:H
-2.91 (7.93)	-4.55 (3.31)	-9.22 (7.58)	-0.30 (0.43)	1.86 (0.40)	2.85 (0.11)	G16:G22		
7.47 (8.05)	-3.48 (3.54)	-9.11 (7.99)	-0.64 (0.38)	1.74 (0.45)	2.86 (0.14)	G22:G4		
7.82 (8.99)	-4.34 (3.91)	4.07 (7.86)	-0.01 (0.38)	2.07 (0.55)	2.92 (0.12)	G5:G11		
17.34 (9.54)	-0.32 (3.71)	-1.55 (8.01)	0.21 (0.35)	1.33 (0.29)	2.92 (0.13)	G11:G15	Q3	W:H
12.91 (8.48)	-5.03 (3.36)	3.12 (7.69)	0.64 (0.35)	1.83 (0.40)	2.87 (0.14)	G15:G23		
18.50 (10.07)	-1.80 (3.36)	3.38 (8.17)	0.26 (0.42)	1.50 (0.31)	2.90 (0.13)	G23:G5		
mixed-(3+1)-form2 topology (PDB ID: 2JSL) (simulation4a)								
-16.39 (10.36)	-5.22 (4.28)	9.44 (8.41)	0.24 (0.47)	-2.13 (0.37)	2.95 (0.15)	G3:G11		
1.18 (10.20)	-5.16 (3.61)	-6.83 (9.36)	0.08 (0.44)	-2.20 (0.51)	2.93 (0.13)	G11:G15	Q1	H:W
-6.59 (10.42)	-2.17 (3.93)	8.80 (8.31)	0.58 (0.50)	-2.01 (0.38)	2.88 (0.17)	G15:G21		
2.00 (10.27)	-3.75 (4.12)	2.93 (8.61)	0.35 (0.53)	-2.32 (0.50)	2.92 (0.18)	G21:G3		
13.54 (9.32)	-3.76 (5.08)	4.49 (9.16)	-0.09 (0.54)	2.65 (0.80)	2.89 (0.15)	G4:G10		
-7.84 (8.86)	-4.33 (4.20)	-9.27 (7.79)	-0.44 (0.40)	2.56 (0.73)	2.87 (0.14)	G10:G16	Q2	W:H
9.58 (10.21)	-4.09 (4.88)	-9.31 (8.57)	-0.33 (0.47)	2.69 (0.83)	2.88 (0.14)	G16:G22		
5.82 (11.66)	-2.98 (4.93)	-13.25 (8.14)	-0.49 (0.44)	2.63 (0.75)	2.86 (0.15)	G22:G4		
30.55 (11.26)	1.01 (7.33)	-3.36 (10.04)	0.17 (0.51)	1.93 (0.87)	2.91 (0.16)	G5:G9	Q3	W:H

3.47 (12.55)	-6.27 (5.03)	5.25 (9.21)	0.32 (0.51)	2.48 (0.78)	2.92 (0.21)	G9:G17		
18.97 (14.81)	-4.40 (5.73)	-4.52 (12.60)	0.15 (0.53)	2.15 (0.82)	2.94 (0.21)	G17:G23		
4.11 (13.58)	-4.14 (7.91)	-6.35 (14.05)	-0.27 (0.60)	2.83 (0.94)	2.89 (0.21)	G23:G5		
Non-telomeric G-rich sequence (PDB ID: 2KZD) (simulation5a)								
-17.29 (9.62)	-8.83 (8.48)	9.36 (8.56)	0.05 (0.50)	-2.86 (0.67)	3.03 (0.38)	G2:G10		
-10.60 (8.68)	-3.33 (4.48)	-12.94 (8.41)	0.24 (0.46)	-3.17 (0.54)	2.84 (0.14)	G10:G13	Q1	H:W
2.44 (9.86)	-3.45 (4.58)	1.08 (7.31)	0.27 (0.50)	-2.45 (0.49)	2.94 (0.15)	G13:G17		
-1.82 (8.81)	0.24 (7.08)	6.32 (7.84)	0.21 (0.44)	-3.46 (0.68)	2.85 (0.16)	G17:G2		
7.11 (8.72)	-2.72 (4.01)	-2.07 (8.29)	-0.39 (0.42)	2.42 (0.58)	2.89 (0.15)	G3:G9		
-6.11 (8.88)	-5.30 (4.07)	-1.90 (7.38)	-0.26 (0.40)	2.50 (0.69)	2.92 (0.13)	G9:G14	Q2	W:H
-1.20 (9.29)	-4.80 (4.64)	-10.91 (8.37)	-0.38 (0.40)	2.69 (0.71)	2.93 (0.14)	G14:G18		
2.55 (11.06)	-1.86 (4.79)	-9.99 (8.44)	-0.50 (0.40)	2.57 (0.71)	2.90 (0.16)	G18:G3		
17.42 (10.55)	-2.03 (3.98)	1.29 (8.02)	-0.14 (0.38)	2.04 (0.67)	2.93 (0.13)	G4:G8		
1.62 (9.98)	-6.65 (5.27)	8.75 (7.75)	0.29 (0.46)	2.25 (0.65)	2.98 (0.18)	G8:G15	Q3	W:H
1.75 (11.82)	-2.86 (4.94)	-7.49 (10.55)	0.16 (0.48)	2.23 (0.83)	2.98 (0.18)	G15:G19		
10.11 (13.20)	-2.92 (4.37)	-6.80 (11.84)	-0.24 (0.49)	2.18 (0.55)	2.93 (0.14)	G19:G4		

Table 2.8.2: Average values along with standard deviations (within parenthesis) of the six base pair parameters for the base pairs of the four telomeric topologies (NMR structure) and one non-telomeric structure. This data set has been generated from the trajectories of simulations 1b, 2b, 3b and 4b and 5b where the simulation has been performed with K^+ counter-ions. Q1, Q2 and Q3 represent Quartet1, Quartet2 and Quartet3 respectively.

buckle	open	propeller	stagger	shear	stretch	base pair (bp)	quartet	bp type
parallel topology (PDB ID: 2LD8) (simulation1b)								
3.52 (11.71)	-3.12 (5.98)	-8.29 (9.48)	-0.20 (0.53)	2.91 (0.73)	2.88 (0.16)	G3:G9	Q1	W:H
3.44 (10.53)	-3.89 (5.07)	-3.63 (10.01)	-0.34 (0.52)	2.82 (0.81)	2.90 (0.17)	G9:G15		
4.07 (10.64)	-4.65 (5.41)	-4.75 (10.15)	-0.25 (0.49)	2.90 (0.73)	2.92 (0.18)	G15:G21		
-0.19 (10.77)	-4.00 (4.68)	-3.17 (9.25)	-0.26 (0.45)	2.86 (0.60)	2.90 (0.14)	G21:G3		
-0.16 (8.85)	-4.17 (3.93)	-7.01 (7.91)	-0.41 (0.37)	2.64 (0.41)	2.94 (0.13)	G4:G10	Q2	W:H
0.49 (8.48)	-3.48 (4.08)	-5.38 (7.52)	-0.41 (0.37)	2.64 (0.40)	2.94 (0.14)	G10:G16		
1.12 (8.63)	-3.22 (4.11)	-7.64 (8.12)	-0.36 (0.37)	2.62 (0.40)	2.94 (0.13)	G16:G22		
1.11 (8.64)	-3.56 (3.94)	-4.09 (8.00)	-0.32 (0.37)	2.57 (0.40)	2.94 (0.13)	G22:G4		
12.19 (9.48)	-3.28 (3.76)	1.67 (7.95)	0.28 (0.41)	1.91 (0.35)	2.97 (0.14)	G5:G11	Q3	W:H
14.52 (9.88)	-2.73 (3.85)	2.50 (7.72)	0.26 (0.42)	1.91 (0.36)	2.96 (0.14)	G11:G17		
12.27 (9.65)	-2.81 (3.84)	1.01 (7.85)	0.25 (0.41)	1.89 (0.35)	2.96 (0.14)	G17:G23		
15.19 (10.22)	-2.82 (3.78)	2.67 (8.01)	0.27 (0.42)	1.88 (0.35)	2.95 (0.14)	G23:G5		
anti-parallel topology (PDB ID: 143D) (simulation2b)								
-14.85 (10.88)	-20.58 (12.03)	39.86 (8.54)	-0.63 (0.45)	1.16 (1.07)	2.73 (0.19)	G2:G10	Q1	W:H
-23.14 (11.15)	36.61 (16.27)	-27.72 (9.41)	-0.39 (0.65)	1.29 (1.08)	3.80 (0.63)	G10:G22		
-8.37 (9.07)	-6.08 (7.28)	3.41 (9.24)	0.23 (0.70)	-3.84 (1.42)	2.94 (0.88)	G22:G14		
8.15 (8.64)	-32.05 (7.60)	-9.66 (8.75)	1.73 (0.58)	-0.68 (0.69)	2.99 (0.33)	G14:G2		
-23.84 (7.75)	-5.55 (3.63)	12.09 (7.02)	0.12 (0.34)	-2.34 (0.39)	2.96 (0.13)	G3:G9	Q2	H:W
-6.32 (10.99)	-1.34 (4.01)	-23.21 (7.29)	-0.63 (0.44)	-2.00 (0.37)	2.86 (0.15)	G9:G21		
6.13 (6.83)	-6.67 (3.75)	3.16 (8.94)	0.22 (0.42)	-2.37 (0.40)	3.00 (0.14)	G21:G15		
2.65 (7.72)	-2.82 (3.80)	6.75 (6.22)	0.54 (0.36)	-1.99 (0.36)	2.95 (0.14)	G15:G3		

21.59 (7.92)	-2.43 (3.79)	15.85 (6.68)	0.35 (0.40)	2.35 (0.30)	2.89 (0.13)	G4:G8		
-10.28 (7.41)	-2.47 (3.90)	2.27 (9.00)	0.73 (0.46)	2.17 (0.32)	2.91 (0.17)	G8:G20	Q3	W:H
11.43 (8.97)	-6.16 (3.49)	-6.24 (6.96)	-0.04 (0.35)	2.25 (0.29)	2.97 (0.12)	G20:G16		
12.21 (8.32)	-4.54 (3.63)	-7.77 (6.93)	-0.26 (0.34)	2.52 (0.29)	2.97 (0.13)	G16:G4		
mixed-(3+1)-form1 topology (PDB ID: 2GKU) (simulation3b)								
-2.03 (8.24)	-3.42 (3.89)	4.00 (7.89)	0.42 (0.41)	-2.51 (0.42)	2.89 (0.13)	G3:G9		
-5.41 (11.89)	-3.66 (3.83)	5.03 (7.26)	0.11 (0.35)	-2.56 (0.45)	2.97 (0.13)	G9:G17	Q1	H:W
-0.30 (7.56)	-5.50 (3.81)	-1.33 (9.77)	0.10 (0.42)	-2.50 (0.41)	2.94 (0.13)	G17:G21		
0.08 (9.01)	-2.43 (4.26)	6.81 (7.55)	0.51 (0.44)	-2.45 (0.48)	2.93 (0.17)	G21:G3		
5.53 (8.46)	-3.54 (3.63)	-7.46 (7.03)	-0.45 (0.35)	2.37 (0.70)	2.95 (0.14)	G4:G10		
7.06 (8.11)	-3.24 (3.69)	-3.16 (6.99)	-0.31 (0.34)	2.26 (0.63)	2.97 (0.13)	G10:G16	Q2	W:H
-2.72 (7.96)	-4.15 (3.61)	-6.69 (7.39)	-0.41 (0.36)	2.32 (0.68)	2.90 (0.13)	G16:G22		
4.94 (9.50)	126.30 (101.25)	9.04 (10.14)	0.07 (0.41)	0.72 (0.88)	-6.21 (4.93)	G22:G4		
12.57 (9.30)	-3.03 (3.66)	0.46 (7.60)	0.01 (0.39)	2.22 (0.36)	2.96 (0.13)	G5:G11		
17.70 (9.28)	-1.11 (3.71)	1.26 (6.80)	0.07 (0.39)	1.78 (0.32)	2.93 (0.13)	G11:G15	Q3	W:H
10.36 (8.92)	-4.64 (3.75)	2.39 (7.38)	0.55 (0.36)	2.13 (0.34)	2.94 (0.15)	G15:G23		
15.82 (10.52)	-1.96 (3.80)	-1.18 (7.56)	0.21 (0.47)	1.92 (0.32)	2.95 (0.15)	G23:G5		
mixed-(3+1)-form2 topology (PDB ID: 2JSL) (simulation4b)								
-17.03 (9.02)	-1.51 (4.20)	16.89 (8.62)	0.58 (0.44)	-2.05 (0.31)	2.88 (0.16)	G3:G11		
-5.38 (9.33)	-5.02 (3.72)	-7.97 (7.41)	0.47 (0.43)	-2.16 (0.31)	2.89 (0.13)	G11:G15	Q1	H:W
-7.82 (9.21)	-4.81 (3.63)	8.91 (7.00)	0.35 (0.38)	-2.08 (0.33)	2.93 (0.13)	G15:G21		
2.57 (9.19)	-4.44 (3.70)	1.14 (7.82)	0.24 (0.39)	-1.99 (0.34)	2.93 (0.13)	G21:G3		
16.84 (8.19)	-5.72 (3.99)	6.48 (8.35)	-0.19 (0.48)	2.57 (0.38)	2.97 (0.14)	G4:G10		
-3.11 (7.66)	-2.47 (3.57)	-12.99 (6.80)	-0.17 (0.37)	2.52 (0.34)	2.91 (0.12)	G10:G16	Q2	W:H
17.56 (8.86)	-3.36 (3.73)	-3.32 (8.26)	-0.11 (0.43)	2.43 (0.38)	2.94 (0.13)	G16:G22		
7.17 (9.26)	-3.37 (3.82)	-12.23 (7.36)	-0.34 (0.40)	2.50 (0.38)	2.92 (0.13)	G22:G4		
33.43 (14.87)	0.06 (5.70)	-5.96 (9.87)	-0.10 (0.54)	2.37 (0.82)	2.93 (0.18)	G5:G9	Q3	W:H

-0.74 (13.54)	-6.70 (6.17)	1.75 (11.18)	0.02 (0.57)	3.31 (0.85)	2.93 (0.33)	G9:G17		
18.56 (13.85)	-5.24 (5.27)	-8.55 (14.52)	-0.01 (0.49)	2.70 (0.61)	2.97 (0.39)	G17:G23		
0.79 (11.86)	-2.87 (6.15)	-12.69 (13.22)	-0.42 (0.53)	3.56 (0.99)	2.85 (0.34)	G23:G5		
Non-telomeric G-rich sequence (PDB ID: 2KZD) (simulation5b)								
131.00 (40.30)	-19.00 (13.03)	-23.08 (33.52)	-8.98 (2.70)	2.24 (2.17)	-0.90 (1.33)	G2:G10	Q1	H:W
-14.13 (14.44)	12.16 (13.41)	-13.39 (8.98)	0.42 (0.64)	-3.74 (0.57)	2.85 (0.33)	G10:G13		
7.53 (64.18)	-73.34 (75.12)	51.51 (79.69)	-3.49 (3.74)	3.79 (5.76)	-1.52 (2.53)	G13:G17		
7.64 (95.10)	17.35 (84.00)	3.81 (51.46)	-0.59 (4.72)	3.83 (4.24)	-0.43 (5.14)	G17:G2		
7.75 (8.61)	-5.31 (4.16)	1.25 (7.99)	-0.59 (0.47)	2.19 (0.34)	2.88 (0.15)	G3:G9	Q2	W:H
-9.09 (7.35)	-3.68 (3.49)	-9.96 (7.11)	-0.52 (0.33)	2.15 (0.30)	2.90 (0.12)	G9:G14		
4.25 (8.51)	-3.62 (4.22)	-5.69 (9.27)	-0.47 (0.41)	2.27 (0.33)	2.94 (0.16)	G14:G18		
-2.01 (9.51)	-3.06 (4.06)	-13.51 (9.57)	-0.24 (0.43)	1.99 (0.39)	2.94 (0.14)	G18:G3		
23.26 (9.86)	-3.15 (4.22)	5.17 (7.27)	-0.26 (0.41)	1.90 (0.37)	2.99 (0.14)	G4:G8	Q3	W:H
-1.81 (9.05)	-4.62 (3.98)	1.87 (7.59)	0.35 (0.40)	1.97 (0.38)	3.01 (0.16)	G8:G15		
14.83 (9.91)	-3.27 (3.87)	-3.33 (8.61)	0.09 (0.41)	1.91 (0.38)	2.99 (0.14)	G15:G19		
6.56 (11.27)	-3.27 (3.85)	-6.56 (9.08)	0.07 (0.43)	1.97 (0.33)	2.97 (0.14)	G19:G4		

Table 2.9.1: Average values along with standard deviations (within parenthesis) of the six base pair stacking parameters for the base pairs of the four telomeric topologies (NMR structure) and one non-telomeric structure. This data set has been generated from the trajectories of simulations 1a, 2a, 3a, 4a and 5a where the simulation has been performed with Na^+ counter-ions. Q1-Q2 and Q2-Q3 represents Quartet1-Quartet2 and Quartet2-Quartet3 stacks respectively.

tilt	roll	twist	shift	slide	rise	base pair (bp)	stack	bp type
parallel topology (PDB ID: 2LD8) (simualtion1a)								
0.52 (3.64)	0.27 (5.20)	28.51 (3.29)	-0.05 (0.61)	1.14 (1.83)	3.32 (0.41)	G3:G9::G4:G10	Q1-Q2	W:H::W:H
0.12 (3.41)	0.20 (5.11)	26.62 (3.70)	-0.00 (0.69)	0.42 (2.50)	3.42 (0.38)	G9:G15::G10:G16		W:H::W:H
-1.22 (3.85)	-0.74 (5.49)	28.70 (3.30)	0.00 (0.60)	-1.65 (1.73)	3.50 (0.42)	G15:G21::G16:G22		W:H::W:H
0.21 (3.45)	-0.56 (5.14)	26.84 (3.87)	-0.06 (0.58)	-0.68 (2.38)	3.33 (0.34)	G21:G3::G22:G4		W:H::W:H
1.08 (3.92)	-2.83 (5.76)	29.99 (2.68)	-0.02 (0.40)	1.82 (1.85)	3.20 (0.38)	G4:G10::G5:G11	Q2-Q3	W:H::W:H
-0.19 (3.57)	0.00 (5.86)	29.86 (2.74)	-0.02 (0.39)	-0.36 (2.44)	3.18 (0.38)	G10:G16::G11:G17		W:H::W:H
-1.79 (3.92)	2.80 (5.82)	30.03 (2.73)	-0.02 (0.42)	-1.86 (1.74)	3.26 (0.39)	G16:G22-G17:G23		W:H::W:H
0.09 (3.72)	-0.39 (6.08)	29.61 (2.86)	-0.01 (0.38)	0.27 (2.44)	3.22 (0.38)	G22:G4::G23:G5		W:H::W:H
anti-parallel topology (PDB ID: 143D) (simulation2a)								
-4.53 (2.51)	9.27 (4.38)	20.26 (3.43)	0.70 (0.34)	-1.32 (0.38)	3.70 (0.57)	G2:G10::G3:G9	Q1-Q2	W:H::H:W
7.02 (3.50)	-8.74 (4.85)	24.63 (3.01)	0.62 (0.33)	2.25 (0.36)	3.73 (0.42)	G9:G21::G10:G22		H:W::W:H
3.96 (3.01)	-7.36 (4.30)	23.30 (2.84)	-0.01 (0.32)	2.20 (0.33)	3.51 (0.37)	G21:G15::G22:G14		H:W::W:H
4.16 (2.61)	5.42 (4.20)	23.53 (2.54)	0.20 (0.35)	-1.36 (0.38)	2.93 (0.38)	G14:G2::G15:G3		W:H::H:W
1.18 (2.57)	7.47 (7.22)	17.18 (3.63)	-0.22 (0.36)	-0.96 (0.55)	3.10 (0.33)	G3:G9::G4:G8	Q2-Q3	H:W::W:H
0.45 (2.67)	-8.06 (5.38)	17.69 (3.63)	0.21 (0.38)	1.33 (0.45)	3.86 (0.25)	G8:G20::G9:G21		W:H::H:W
-2.08 (3.43)	-6.38 (5.41)	17.67 (2.48)	-0.10 (0.49)	1.78 (0.35)	4.04 (0.32)	G20:G16::G21:G15		W:H::H:W
-0.33 (2.81)	5.78 (5.15)	18.32 (3.55)	-0.60 (0.40)	-1.48 (0.62)	3.44 (0.47)	G15:G3::G16:G4		H:W::W:H
mixed-(3+1)-form1 topology (PDB ID: 2GKU) (simulation3a)								
0.35 (2.11)	1.88 (5.24)	18.37 (2.27)	-0.25 (0.31)	-1.50 (0.37)	3.39 (0.31)	G3:G9::G4:G10	Q1-Q2	H:W::W:H
-2.79 (2.28)	0.52 (5.42)	19.29 (2.54)	0.26 (0.31)	-1.10 (0.43)	3.20 (0.32)	G9:G17::G10:G16		W:H::H:W
2.49 (3.10)	-5.60 (4.77)	18.27 (1.83)	0.17 (0.38)	1.88 (0.30)	3.81 (0.31)	G16:G22::G17:G21		H:W::W:H
0.42 (2.14)	2.77 (6.38)	19.37 (2.35)	-0.17 (0.36)	-1.47 (0.37)	3.43 (0.30)	G21:G3::G22:G4		H:W::W:H

-1.35 (3.14)	6.79 (4.37)	28.79 (2.95)	-0.55 (0.35)	-2.56 (0.36)	3.05 (0.26)	G4:G10::G5:G11	W:H::W:H
-0.95 (2.80)	4.19 (4.55)	30.49 (2.72)	-0.18 (0.30)	-1.80 (0.39)	3.06 (0.26)	G10:G16::G11:G15	W:H::W:H
-1.59 (2.72)	-3.03 (4.41)	28.41 (2.70)	0.19 (0.32)	2.33 (0.27)	3.34 (0.28)	G15:G23::G16:G22	W:H::W:H
-3.34 (2.95)	6.61 (4.70)	30.23 (2.38)	-0.18 (0.32)	-2.62 (0.38)	2.90 (0.31)	G22:G4::G23:G5	W:H::W:H
mixed-(3+1)-form2 topology (PDB ID: 2JSL) (simulation4a)							
-0.70 (2.23)	2.01 (5.74)	17.58 (3.49)	0.25 (0.48)	-0.82 (0.43)	2.96 (0.42)	G3:G11::G4:G10	H:W::W:H
2.25 (3.41)	-7.74 (5.29)	18.66 (2.61)	0.20 (0.44)	2.20 (0.42)	3.66 (0.31)	G10:G16::G11:G15	W:H::H:W
-0.92 (2.73)	0.61 (5.91)	18.68 (3.33)	-0.52 (0.48)	-1.59 (0.42)	3.32 (0.38)	G15:G21::G16:G22	H:W::W:H
-1.03 (2.74)	6.30 (5.66)	19.14 (3.11)	-0.57 (0.40)	-1.45 (0.46)	3.69 (0.41)	G21:G3::G22:G4	H:W::W:H
-2.27 (3.19)	3.03 (6.13)	28.47 (4.48)	0.24 (0.51)	-1.25 (0.66)	3.35 (0.41)	G4:G10::G5:G9	W:H::W:H
0.85 (3.73)	-4.59 (5.04)	24.45 (3.56)	0.17 (0.54)	2.78 (0.47)	3.31 (0.38)	G9:G17::G10:G16	W:H::W:H
-3.63 (4.19)	3.53 (6.46)	28.45 (4.22)	-0.35 (0.60)	-2.37 (0.70)	3.13 (0.45)	G16:G22::G17:G23	W:H::W:H
-2.60 (4.90)	6.00 (6.28)	26.03 (3.94)	-0.82 (0.59)	-2.38 (0.72)	3.34 (0.44)	G22:G4::G23:G5	W:H::W:H
Non-telomeric G-rich sequence (PDB ID: 2KZD) (simulation5a)							
-3.50 (2.21)	6.78 (5.27)	18.96 (3.20)	0.23 (0.39)	-0.66 (0.48)	3.33 (0.38)	G2:G10::G3:G9	H:W::W:H
7.46 (2.67)	-3.64 (5.66)	16.34 (3.08)	0.76 (0.44)	1.68 (0.37)	3.97 (0.27)	G9:G14::G10:G13	W:H::H:W
0.20 (2.89)	0.74 (5.02)	20.81 (2.45)	0.33 (0.41)	-1.98 (0.41)	3.67 (0.39)	G13:G17::G14:G18	H:W::W:H
-2.23 (2.57)	3.06 (5.46)	18.69 (2.97)	-0.52 (0.42)	-1.95 (0.42)	3.56 (0.36)	G17:G2::G18:G3	H:W::W:H
1.42 (2.65)	3.61 (5.17)	27.29 (3.71)	0.35 (0.40)	-1.68 (0.52)	3.23 (0.36)	G3:G9::G4:G8	W:H::W:H
1.56 (3.40)	-2.24 (4.93)	27.26 (2.63)	0.45 (0.37)	2.50 (0.32)	3.31 (0.36)	G8:G15::G9:G14	W:H::W:H
-4.58 (4.46)	2.69 (5.43)	28.08 (3.36)	-0.29 (0.47)	-2.78 (0.49)	3.09 (0.40)	G14:G18::G15:G19	W:H::W:H
-2.17 (4.36)	3.95 (5.71)	27.22 (3.00)	-0.60 (0.62)	-2.02 (0.51)	3.26 (0.41)	G18:G3::G19:G4	W:H::W:H

Table 2.9.2: Average values along with standard deviations (within parenthesis) of the six base pair stacking parameters for the base pairs of the four telomeric topologies (NMR structure) and one non-telomeric structure. This data set has been generated from the trajectories of simulations 1b, 2b, 3b, 4b and 5b where the simulation has been performed with K^+ counter-ions. Q1-Q2 and Q2-Q3 represents Quartet1-Quartet2 and Quartet2-Quartet3 stacks respectively.

tilt	roll	twist	shift	slide	rise	base pairing (bp)	stack	bp type
parallel topology (PDB ID: 2LD8) (simulation1b)								
-0.19 (3.34)	-0.94 (4.95)	26.62 (3.42)	-0.12 (0.55)	-2.17 (0.61)	3.34 (0.33)	G3:G9::G4:G10	Q1-Q2	W:H::W:H
-0.56 (3.19)	-1.97 (5.35)	26.79 (3.43)	0.07 (0.63)	-2.06 (0.60)	3.44 (0.37)	G9:G15::G10:G16		W:H::W:H
-0.85 (3.80)	-0.57 (5.30)	26.73 (3.40)	0.21 (0.59)	-2.37 (0.54)	3.42 (0.37)	G15:G21::G16:G22		W:H::W:H
0.29 (3.97)	-1.43 (5.03)	26.23 (3.82)	-0.14 (0.52)	-2.36 (0.59)	3.45 (0.35)	G21:G3::G22:G4		W:H::W:H
-1.87 (3.65)	6.19 (4.76)	30.52 (2.49)	0.00 (0.31)	-2.53 (0.30)	3.01 (0.31)	G4:G10::G5:G21	Q2-Q3	W:H::W:H
-1.58 (3.24)	6.09 (4.89)	30.36 (2.36)	0.04 (0.31)	-2.47 (0.29)	3.05 (0.31)	G10:G16::G11:G17		W:H::W:H
-1.59 (3.52)	6.52 (4.76)	30.47 (2.45)	0.00 (0.32)	-2.56 (0.29)	2.97 (0.32)	G16:G22::G17:G23		W:H::W:H
-1.81 (3.56)	5.79 (4.84)	30.29 (2.49)	0.00 (0.31)	-2.46 (0.32)	3.03 (0.32)	G22:G4::G23:G5		W:H::W:H
anti-parallel topology (PDB ID: 143D) (simulation2b)								
-4.07 (5.33)	-0.61 (5.14)	12.91 (3.15)	0.28 (0.47)	-0.73 (1.07)	4.80 (0.43)	G2:G10::G3:G9	Q1-Q2	W:H::H:W
5.13 (8.80)	-2.61 (7.24)	23.04 (3.40)	-0.21 (0.67)	1.40 (2.42)	2.86 (0.46)	G9:G21::G10:G22		H:W::W:H
-0.59 (2.47)	-5.35 (8.70)	19.67 (3.28)	-0.83 (1.14)	0.74 (1.21)	3.93 (0.37)	G21:G15::G22:G14		H:W::W:H
2.57 (4.75)	0.74 (5.11)	22.31 (2.46)	0.27 (0.49)	-0.26 (0.71)	3.88 (0.34)	G14:G2::G15:G3		W:H::H:W
-0.09 (1.89)	0.20 (4.95)	16.84 (2.36)	0.04 (0.26)	-0.56 (0.81)	2.45 (0.28)	G3:G9::G4:G8	Q2-Q3	H:W::W:H
1.70 (3.63)	-3.52 (9.04)	19.82 (1.98)	0.11 (0.32)	0.89 (1.65)	3.66 (0.26)	G8:G20::G9:G21		W:H::H:W
-2.07 (3.42)	-3.89 (6.93)	17.37 (2.01)	-0.10 (0.36)	1.01 (1.36)	3.76 (0.24)	G20:G16::G21:G15		W:H::H:W
-0.12 (2.14)	4.12 (8.45)	19.75 (2.07)	-0.22 (0.49)	-0.82 (1.40)	3.42 (0.27)	G15:G3::G16:G4		H:W::W:H
mixed-(3+1)-form1 topology (PDB ID: 2GKU) (simulation3b)								
-0.39 (2.35)	0.70 (4.81)	18.50 (2.21)	-0.08 (0.35)	-0.61 (1.27)	3.39 (0.31)	G3:G9::G4:G10	Q1-Q2	H:W::W:H
-0.64 (2.99)	1.31 (5.37)	18.37 (2.58)	0.17 (0.38)	-0.31 (1.42)	3.34 (0.32)	G9:G17::G10:G16		W:H::H:W
1.07 (3.35)	-2.14 (5.87)	18.72 (1.99)	0.12 (0.40)	0.87 (1.63)	3.70 (0.25)	G16:G22::G17:G21		H:W::W:H

-0.36 (2.27)	1.50 (4.86)	18.43 (2.17)	-0.10 (0.38)	-0.88 (1.32)	3.50 (0.34)	G21:G3::G22:G4	H:W::W:H
-1.04 (3.28)	3.07 (6.87)	28.42 (2.42)	-0.34 (0.56)	-1.09 (2.19)	3.15 (0.30)	G4:G10::G5:G11	W:H::W:H
-0.17 (2.74)	0.60 (5.39)	29.46 (2.38)	-0.06 (0.31)	-0.45 (1.99)	3.10 (0.28)	G10:G16::G11:G15	W:H::W:H
-0.90 (3.47)	-0.78 (5.30)	27.82 (2.26)	-0.02 (0.35)	0.96 (2.23)	3.37 (0.30)	G15:G23::G16:G22	W:H::W:H
-2.55 (4.00)	3.68 (6.55)	29.59 (2.52)	-0.25 (0.44)	-1.56 (2.04)	2.95 (0.34)	G22:G4::G23:G5	W:H::W:H
mixed-(3+1)-form2 topology (PDB ID: 2JSL) (simulation4b)							
0.03 (2.30)	7.89 (6.25)	16.82 (2.93)	0.19 (0.29)	-0.89 (0.35)	2.92 (0.35)	G3:G11::G4G10	H:W::W:H
1.36 (2.97)	-6.08 (4.70)	19.99 (1.80)	0.00 (0.29)	2.19 (0.24)	3.82 (0.29)	G10:G16::G11:G15	W:H::H:W
-0.94 (2.58)	1.95 (5.13)	17.41 (2.75)	-0.51 (0.31)	-1.36 (0.32)	3.12 (0.32)	G15:G21::G16:G22	H:W::W:H
-1.77 (2.78)	10.52 (5.43)	18.13 (2.41)	-0.47 (0.32)	-1.52 (0.31)	3.74 (0.37)	G21:G3::G22:G4	H:W::W:H
-1.45 (2.98)	2.45 (5.54)	27.22 (4.46)	0.20 (0.59)	-0.91 (0.56)	3.49 (0.33)	G4:G10::G5:G9	W:H::W:H
2.17 (4.21)	-3.56 (4.73)	22.42 (3.71)	0.06 (0.54)	3.05 (0.44)	3.13 (0.39)	G9:G17::G10:G16	W:H::W:H
-5.20 (3.87)	0.06 (6.33)	28.10 (4.41)	-0.81 (0.65)	-2.29 (0.88)	3.52 (0.39)	G16:G22::G17:G23	W:H::W:H
-5.20 (3.87)	0.06 (6.33)	28.10 (4.41)	-0.81 (0.65)	-2.29 (0.88)	3.52 (0.39)	G22:G4::G23:G5	W:H::W:H
Non-telomeric G-rich sequence (PDB ID: 2KZD) (simulation5b)							
-2.90 (3.84)	4.80 (11.19)	34.48 (4.88)	-1.95 (1.04)	0.49 (1.20)	13.73 (8.53)	G2:G10::G3:G9	H:W::W:H
5.90 (4.42)	-5.94 (5.11)	11.02 (4.79)	-0.19 (0.48)	2.62 (0.43)	4.12 (0.38)	G9:G14::G10:G13	W:H::H:W
8.46 (20.38)	-8.85 (15.66)	32.61 (31.37)	2.18 (2.49)	-1.46 (1.19)	2.13 (3.38)	G13:G17::G14:G18	H:W::W:H
-8.26 (16.93)	10.89 (28.77)	11.00 (23.83)	-0.51 (1.84)	-5.87 (3.79)	2.43 (10.52)	G17:G2::G18:G3	H:W::W:H
1.06 (2.40)	4.19 (4.43)	25.39 (3.02)	0.03 (0.30)	-1.45 (0.38)	3.16 (0.27)	G3:G9::G4:G8	W:H::W:H
-0.74 (2.76)	-3.69 (4.34)	24.61 (2.33)	0.06 (0.29)	2.23 (0.25)	3.19 (0.30)	G8:G15::G9:G14	W:H::W:H
-3.16 (3.55)	5.15 (5.20)	25.94 (2.94)	0.10 (0.34)	-2.15 (0.34)	3.17 (0.32)	G14:G18::G15:G19	W:H::W:H
-1.60 (3.70)	5.76 (4.65)	24.65 (2.70)	-0.66 (0.35)	-2.31 (0.37)	3.13 (0.30)	G18:G3::G19:G4	W:H::W:H

Table 2.10: Average values along with standard deviations (within parenthesis) of: (a) the six base pair orientation parameters and (b) the six base pair stacking parameters for the base pair of (X-ray crystal) parallel-type topology. Q1, Q2 and Q3 correspond to Quartet1, Quartet2 and Quartet3. These data have been obtained from trajectories of simulation6 and simulation7 where the G-tetrad core contains K^+ ion and Na^+ is the charge neutralizing cation.

buckle	open	propeller	stagger	shear	stretch	base pair (bp)	bp type
Parallel (X-ray crystal, PDB ID: 1KF1) (simulation6)							
4.37 (10.71)	-3.82 (5.11)	-7.98 (8.71)	-0.21 (0.50)	3.04 (0.66)	2.91 (0.21)	G2:G8	Q1 W:H
4.17 (10.02)	-3.98 (5.14)	-4.81 (9.81)	-0.25 (0.50)	2.82 (0.67)	2.95 (0.27)	G8:G14	
7.12 (10.18)	-3.75 (5.04)	-4.51 (9.56)	-0.17 (0.51)	2.99 (0.66)	2.91 (0.19)	G14:G20	
0.66 (9.86)	-4.07 (4.85)	-2.08 (8.88)	-0.28 (0.45)	2.80 (0.57)	2.91 (0.15)	G20:G2	
1.67 (8.19)	-3.25 (3.90)	-7.61 (7.55)	-0.35 (0.36)	2.61 (0.43)	2.94 (0.13)	G3:G9	Q2 W:H
1.02 (8.24)	-3.87 (3.91)	-4.46 (7.60)	-0.32 (0.37)	2.57 (0.43)	2.95 (0.13)	G9:G15	
1.99 (8.10)	-3.43 (3.91)	-7.12 (7.71)	-0.39 (0.36)	2.61 (0.43)	2.94 (0.13)	G15:G21	
0.53 (8.44)	-3.99 (3.96)	-4.15 (7.34)	-0.36 (0.37)	2.59 (0.41)	2.95 (0.13)	G21:G3	
11.39 (9.04)	-3.20 (3.73)	0.10 (7.70)	0.24 (0.40)	1.89 (0.35)	2.96 (0.14)	G4:G10	Q3 W:H
15.27 (9.63)	-2.93 (3.78)	2.51 (7.75)	0.26 (0.41)	1.93 (0.35)	2.96 (0.14)	G10:G16	
12.44 (9.38)	-2.61 (3.78)	1.08 (7.89)	0.25 (0.41)	1.90 (0.35)	2.96 (0.14)	G16:G22	
14.27 (9.82)	-2.93 (3.78)	3.62 (7.48)	0.26 (0.42)	1.89 (0.36)	2.97 (0.14)	G22:G4	
Parallel (X-ray crystal, PDB ID: 3R6R) (simulation7)							
5.88 (10.21)	-4.91 (5.26)	-8.51 (9.44)	-0.22 (0.50)	3.03 (0.61)	2.90 (0.19)	G2:G8	Q1 W:H
0.87 (10.01)	-3.45 (5.46)	-4.32 (9.52)	-0.43 (0.48)	2.66 (0.60)	2.91 (0.21)	G8:G14	
7.22 (9.57)	-4.41 (4.78)	-3.81 (9.09)	-0.25 (0.47)	3.05 (0.61)	2.90 (0.19)	G14:G20	
-3.90 (10.48)	-2.90 (4.49)	-2.95 (8.36)	-0.19 (0.46)	2.58 (0.51)	2.91 (0.13)	G20:G2	
3.20 (8.79)	-4.00 (4.14)	-8.11 (7.53)	-0.41 (0.38)	2.63 (0.47)	2.94 (0.14)	G3:G9	Q2 W:H
-1.19 (8.79)	-3.91 (4.00)	-3.10 (7.29)	-0.31 (0.36)	2.63 (0.43)	2.94 (0.13)	G9:G15	
2.72 (7.80)	-2.13 (4.08)	-8.73 (8.29)	-0.38 (0.37)	2.57 (0.41)	2.93 (0.13)	G15:G21	
-0.54 (8.41)	-4.57 (3.99)	-3.68 (7.67)	-0.33 (0.36)	2.52 (0.37)	2.96 (0.13)	G21:G3	
10.32 (9.59)	-3.31 (3.76)	-1.84 (8.13)	0.15 (0.41)	1.86 (0.35)	2.98 (0.14)	G4:G10	Q3 W:H
16.49 (9.54)	-3.19 (3.75)	3.07 (7.70)	0.25 (0.42)	1.96 (0.35)	2.96 (0.14)	G10:G16	

10.99 (9.75)	-2.17 (3.80)	0.83 (7.79)	0.23 (0.41)	1.86 (0.35)	2.95 (0.14)	G16:G22	
13.75 (9.63)	-3.27 (3.88)	3.91 (7.67)	0.29 (0.42)	1.89 (0.36)	2.97 (0.15)	G22:G4	
tilt	**roll**	**twist**	**shift**	**slide**	**rise**	**base pair (bp)**	**bp step**
Parallel (X-ray crystal, PDB ID: 1KF1) (simulation6)							
-0.10 (3.16)	-1.30 (4.82)	27.92 (3.07)	0.19 (0.52)	-2.19 (0.49)	3.37 (0.38)	G2:G8::G3:G9	Q1-Q2 W:H::W:H (parallel strand orientation)
-0.41 (3.17)	-1.77 (5.12)	26.93 (3.34)	0.07 (0.56)	-2.43 (0.49)	3.41 (0.37)	G8:G14::G9:G15	
-1.34 (3.32)	-0.69 (5.24)	27.66 (3.31)	-0.04 (0.55)	-2.42 (0.54)	3.53 (0.38)	G14:G20::G15:G21	
-0.25 (3.22)	-1.79 (5.03)	27.15 (3.34)	-0.19 (0.49)	-2.21 (0.57)	3.41 (0.33)	G20:G2::G21:G3	
-2.15 (3.24)	6.06 (4.63)	30.62 (2.29)	0.01 (0.30)	-2.57 (0.27)	3.02 (0.30)	G3:G9::G4:G10	Q2-Q3 W:H::W:H (parallel strand orientation)
-1.64 (3.36)	5.24 (4.71)	30.34 (2.41)	-0.03 (0.30)	-2.48 (0.29)	3.01 (0.31)	G9:G15::G10:G16	
-2.08 (3.37)	6.36 (4.74)	30.54 (2.40)	-0.01 (0.32)	-2.52 (0.30)	3.02 (0.31)	G15:G21::G16:G22	
-1.19 (3.08)	5.39 (4.66)	30.19 (2.46)	0.03 (0.30)	-2.45 (0.31)	3.00 (0.31)	G21:G3::G22:G4	
Parallel (X-ray crystal, PDB ID: 3R6R) (simulation7)							
0.35 (3.25)	-1.39 (4.96)	27.78 (3.05)	0.08 (0.51)	-0.95 (1.70)	3.41 (0.37)	G2:G8::G3:G9	Q1-Q2 W:H::W:H (parallel strand orientation)
0.43 (3.30)	1.14 (5.40)	25.80 (3.21)	-0.01 (0.61)	0.83 (2.10)	3.32 (0.35)	G8:G14::G9:G15	
1.68 (4.07)	-0.77 (5.27)	27.45 (3.00)	-0.08 (0.60)	1.18 (2.19)	3.58 (0.41)	G14:G20::G15:G21	
0.11 (4.14)	-1.07 (4.78)	25.40 (3.45)	-0.25 (0.64)	-1.04 (2.28)	3.38 (0.33)	G20:G2::G21:G3	
0.05 (4.75)	-0.22 (8.28)	30.60 (2.57)	-0.03 (0.32)	-0.30 (2.55)	3.13 (0.32)	G3:G9::G4:G10	Q2-Q3 W:H::W:H (parallel strand orientation)
-0.10 (3.82)	-2.69 (5.73)	30.04 (2.52)	-0.04 (0.31)	1.46 (1.98)	2.99 (0.32)	G9:G15::G10:G16	
0.04 (3.70)	-0.55 (8.22)	30.72 (2.30)	0.04 (0.30)	0.33 (2.57)	3.03 (0.32)	G15:G21::G16:G22	
-0.53 (3.44)	2.52 (5.67)	30.14 (2.51)	0.02 (0.30)	-1.58 (1.87)	2.98 (0.31)	G21:G3::G22:G4	

Since, a quadruplex is somewhat globular, it is possible to get an idea about its compactness from its radius of gyration. On analyzing simulation5b we observed an increase in radius of gyration during the timescale 50-80ns **(Figure 2.7a)**. The time-evolution plot for number of base pairs along the simulation trajectory indicated a decrease in base pairing during the same time-scale **(Figure 2.7b)**, indicative of melted-out flexible structures. It should be noted that this stretch of promoter sequence appears in the genomic DNA double stranded region (Gagniuc and Ionescu-Tirgoviste, 2012), thus having an inherent tendency to remain in double helix form. Single stranded DNA can occur in situations like replication bubble where the complimentary strand gets stabilized by single strand binding proteins or in R-loop (Thomas *et al.*, 1976). This sequence may adopt telomere-like topology, as reported, only if it is stabilized by a regulatory protein.

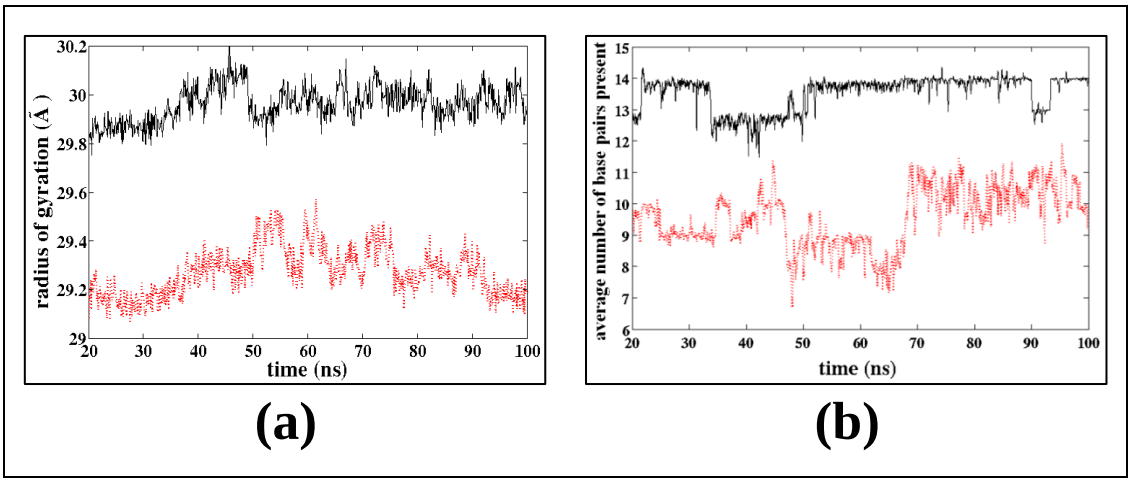


Figure 2.7: (a) radius of gyration vs. time plot for mixed-(3+1)-form1 topology (black and bold line) and hTERT G-rich sequence (red and dotted line); and (b) base pair present vs. time plot for mixed-(3+1)-form1 topology (black and bold line) and hTERT G-rich sequence (red and dotted line), when the simulations have been done with K^+ counterions.

A general idea about the conformational changes introduced in the sugar-phosphate backbone during the MD-run was obtained by detailed analysis of backbone conformation of the four different topology types. Only the trajectories corresponding to simulations performed in Na^+ counter-ions were analyzed **(Table 2.11)** as the larger K^+ ion only introduces unsteadiness in the MD run of the systems. The average values for most of the backbone dihedrals (α, β, γ, δ, ε, ζ) and glycosidic torsion angle (χ) obtained from our simulations lied within the allowed regions of conformational space and close to their initial experimental values. The

pseudorotation phase angle, *P*, for most of the sugars, remain in the C2'-endo regions while there are sugars which also take up O4'-endo conformation.

The parallel-stranded G-quadruplex contained all the guanine bases in *anti* conformation of the glycosidic torsion angle with the corresponding sugars adopting C2'-endo pucker. Except three guanine residues, all others maintained the canonical g-/g+ conformation for the α/γ torsion angles. The α-torsion angle in these three guanine bases adopted unusual g+ and t conformations too, along with the usual g- conformation during the course of 200ns long MD-run of simulation1a accounting for the high standard deviation. It may be noted that double helical B-DNA is known to adopt g- and g+ conformations around α and γ torsion angles but their preferences in other types of structures are not known.

The initial NMR structure reported in PDB, for the anti-parallel topology contained six guanine bases with *syn* conformation of the glycosidic torsion angle. However, during the MD run corresponding to simulation2a, one of the guanine bases adopted *anti* conformation (291.4 (13.0)). The corresponding sugar pucker was not observed in the stable C2'-endo region, instead adopted phase angle in the range (95°-130°). One of the guanine residues had *syn* orientation in the reported NMR structure but after the MD run it showed γ-value of 291.7° (29.2) which is also unusual as compared to double helical DNA. In both the mixed topologies, five guanine bases were seen to be in *syn* conformation with corresponding sugar puckers adopting O4'-endo or C1'-exo conformations.

Table 2.11: Average values along with standard deviations (within parenthesis) of the α/γ torsion angles, sugar pucker (χ) and phase angle for the guanine bases of the four topology types (NMR structure). This data set has been generated from the trajectories of simulations 1a, 2a, 3a and 4a where the simulation has been performed with Na^+ counter-ions.

Residue	**α (°)**	**γ (°)**	**χ (°)**	**phase angle**
parallel topology (PDB ID: 2LD8) (simulation1a)				
G3	290.18 (29.19)	54.94 (15.29)	256.70 (14.49)	155.03 (22.86)
G4	297.52 (26.05)	54.66 (18.49)	236.00 (16.76)	138.42 (33.46)
G5	296.88 (20.85)	45.85 (13.95)	241.06 (11.37)	150.92 (14.22)
G9	145.73 (84.04)	63.54 (35.31)	250.24 (18.90)	143.09 (36.81)
G10	292.01 (32.66)	54.38 (21.53)	234.09 (18.43)	134.61 (40.59)
G11	295.42 (23.98)	46.88 (18.40)	242.04 (13.47)	148.42 (31.32)
G15	146.28 (94.54)	62.60 (28.84)	252.56 (18.68)	146.44 (27.50)
G16	296.08 (27.44)	51.75 (19.35)	237.89 (15.79)	141.18 (32.90)
G17	297.20 (23.60)	45.69 (14.26)	241.42 (11.80)	151.44 (14.54)
G21	165.87 (95.56)	73.04 (56.25)	239.89 (21.27)	136.64 (40.60)

G22	294.74 (32.47)	51.64 (17.18)	237.18 (18.22)	139.66 (40.49)
G23	296.13 (26.29)	44.69 (20.83)	244.43 (12.94)	153.88 (18.41)
anti-parallel topology (PDB ID: 143D) (simulation2a)				
G2	292.83 (38.22)	56.89 (21.68)	257.79 (13.99)	145.78 (22.97)
G3	213.73 (36.82)	291.73 (29.21)	291.45 (13.00)	142.37 (14.64)
G4	148.84 (52.76)	68.66 (24.77)	289.06 (11.73)	127.46 (11.62)
G8	196.05 (84.91)	107.94 (56.72)	66.66 (13.75)	114.42 (24.81)
G9	275.10 (48.43)	53.93 (12.40)	252.20 (20.73)	128.00 (26.78)
G10	277.70 (35.72)	64.93 (28.19)	54.83 (8.85)	103.17 (14.96)
G14	273.69 (42.49)	56.59 (17.43)	276.85 (12.32)	163.05 (14.38)
G15	295.11 (14.59)	56.63 (16.33)	54.60 (19.45)	99.81 (9.06)
G16	299.47 (24.13)	49.69 (12.36)	253.26 (16.34)	143.07 (17.61)
G20	238.63 (87.17)	88.00 (67.18)	57.65 (10.69)	129.64 (27.85)
G21	303.24 (22.83)	46.90 (11.91)	247.77 (17.64)	124.95 (29.79)
G22	290.49 (24.07)	57.47 (23.13)	57.65 (23.87)	104.72 (12.33)
mixed-(3+1)-form1 topology (PDB ID: 2GKU) (simulation3a)				
G3	200.37 (75.14)	96.30 (58.95)	56.95 (13.80)	118.46 (22.57)
G4	302.50 (31.71)	55.83 (15.24)	245.85 (12.54)	139.94 (16.41)
G5	290.29 (29.73)	51.47 (27.13)	236.69 (11.08)	150.73 (16.75)
G9	64.45 (28.72)	62.28 (18.77)	64.64 (10.63)	97.44 (11.76)
G10	230.81 (42.29)	61.58 (27.55)	256.11 (14.67)	135.05 (16.94)
G11	288.86 (56.56)	55.53 (34.06)	238.52 (16.27)	141.83 (40.82)
G15	108.95 (62.67)	180.04 (48.76)	57.64 (19.49)	145.31 (19.75)
G16	304.05 (12.99)	51.22 (10.57)	42.83 (8.13)	106.16 (9.49)
G17	298.40 (19.31)	47.37 (11.87)	254.46 (14.13)	143.44 (13.29)
G21	273.25 (33.12)	56.80 (27.33)	51.85 (9.14)	112.14 (18.33)
G22	310.11 (16.46)	52.64 (11.30)	248.57 (11.96)	140.89 (15.58)
G23	293.72 (29.14)	45.27 (27.09)	240.07 (13.06)	155.89 (18.91)
mixed-(3+1)-form2 topology (PDB ID: 2JSL) (simulation4a)				
G3	237.33 (68.40)	93.92 (58.65)	59.40 (12.62)	115.05 (20.44)
G4	293.44 (38.24)	52.55 (12.24)	258.53 (13.96)	143.74 (30.00)
G5	297.21 (31.79)	54.63 (16.33)	246.84 (17.30)	142.09 (40.78)
G9	186.83 (109.92)	176.57 (100.48)	53.98 (59.75)	114.20 (21.05)
G10	299.28 (18.60)	51.11 (13.42)	46.61 (9.58)	106.59 (11.69)
G11	303.12 (19.43)	49.69 (12.10)	254.75 (11.20)	152.33 (14.22)
G15	298.13 (22.69)	64.23 (21.98)	47.59 (42.16)	107.55 (15.15)
G16	300.43 (32.06)	52.96 (13.04)	251.98 (13.82)	145.13 (21.65)
G17	293.30 (29.47)	50.08 (21.70)	240.87 (15.91)	146.98 (27.38)
G21	124.67 (91.00)	60.23 (13.53)	48.68 (32.32)	107.72 (15.19)
G22	296.46 (36.72)	50.82 (14.64)	254.35 (14.07)	148.65 (21.19)
G23	281.29 (35.16)	52.61 (31.38)	236.95 (25.19)	117.74 (65.46)
non-telomeric G-rich sequence (PDB ID: 2KZD) (simulation5a)				
G2	147.09 (83.93)	116.35 (68.91)	58.39 (10.53)	136.81 (23.43)
G3	302.89 (19.83)	52.59 (11.67)	255.77 (12.48)	153.50 (18.88)
G4	296.12 (26.57)	50.05 (15.45)	253.37 (14.15)	148.56 (14.05)
G8	267.52 (43.76)	126.65 (54.42)	51.60 (29.66)	108.47 (14.96)
G9	281.04 (39.40)	56.07 (13.37)	41.09 (11.57)	103.61 (9.56)
G10	269.95 (42.43)	49.76 (16.40)	263.53 (15.83)	121.99 (57.55)
G13	271.19 (44.37)	67.05 (28.61)	58.54 (8.49)	137.44 (19.08)
G14	308.09 (13.75)	55.16 (11.45)	234.60 (17.67)	128.29 (23.70)
G15	302.10 (30.02)	50.02 (21.24)	230.90 (11.05)	147.52 (15.22)
G17	72.25 (32.35)	62.24 (11.78)	38.62 (29.17)	118.70 (17.31)

G18	298.78 (34.49)	46.41 (13.12)	242.15 (17.63)	134.53 (31.52)
G19	285.74 (32.89)	51.56 (25.16)	246.58 (18.89)	135.82 (44.33)

2.3.2.3 G-tetrad Stack and Ion-coordination

As stated earlier, stacking of three G-quartets form two G-tetrad cores, each lined by eight carbonyl oxygen atoms (O6) **(Figure 2.8)**. As nucleic acids are negatively charged polyion, we added requisite number of counter-ions to neutralize the system. Generally the ions neutralize the phosphate groups staying close to them. In case of G-quadruplex structures additional electronegative pockets are present at the G-tetrad core. An ion at the centre of such a G-tetrad core surrounded by eight O6-atoms would give Coordination Number (CN)=8 for the ion, which in turn suggested three main geometries for the arrangement- i) Cubic type, ii) Square antiprism and iii) Dodecahedron. The angle O6—(ion)—O6 can give an idea about the most probable geometry adopted by the G-quartet stacks and thus were analyzed **(Figure 2.9)**.

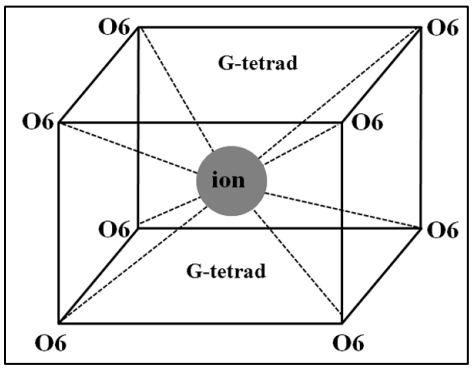


Figure 2.8: Geometric representation of an ion placed at the G-tetrad core (between two G-tetrads) line by eight O6 atoms (CN=8).

The normalized frequency plot of O6—Na^+—O6 angle **(Figure 2.9a)** corresponding to the parallel topology showed two distinct peaks at 88° and 170° and a hump at 142°; the mixed-(3+1)-form1 and mixed-(3+1)-form1 showed a strong peak at 78° and broad hump from 120°-150°. On the other hand, the anti-parallel topology had three distinct peaks at 88°, 120° and 160° for the same angle. Similar to the O6—Na^+—O6 angle normalized frequency plot, the plot for O6—K^+—O6 angle **(Figure 2.9b)** showed three peaks for the parallel topology but the peaks were left shifted. The anti-parallel and mixed-(3+1)-form2 topologies were seen to have similar

distributions for the angles with three strong peaks at 70°, 110° and 170°. In mixed-(3+1)-form1 topology, the distribution of O6—K^+—O6 angle was seen to have three distinct peaks at 75°, 110° and 135° along with a hump at around 155°, unlike the earlier situation. A perfect cubic geometry is expected to have only three possible angles: 70°, 109° and 180°, thus, it could be assumed that the G-tetrad cores adopted geometry close to square antiprism. It is already known that an ion sandwiched between two G-quartets is arranged in a square antiprism coordination (Islam *et al.*, 2013), however, since a varied range of angles were seen for each topology, it would be wise to assume that the G-tetrad geometry is rather a distorted square antiprism as the adjacent G-quartets are arranged in a twisted manner.

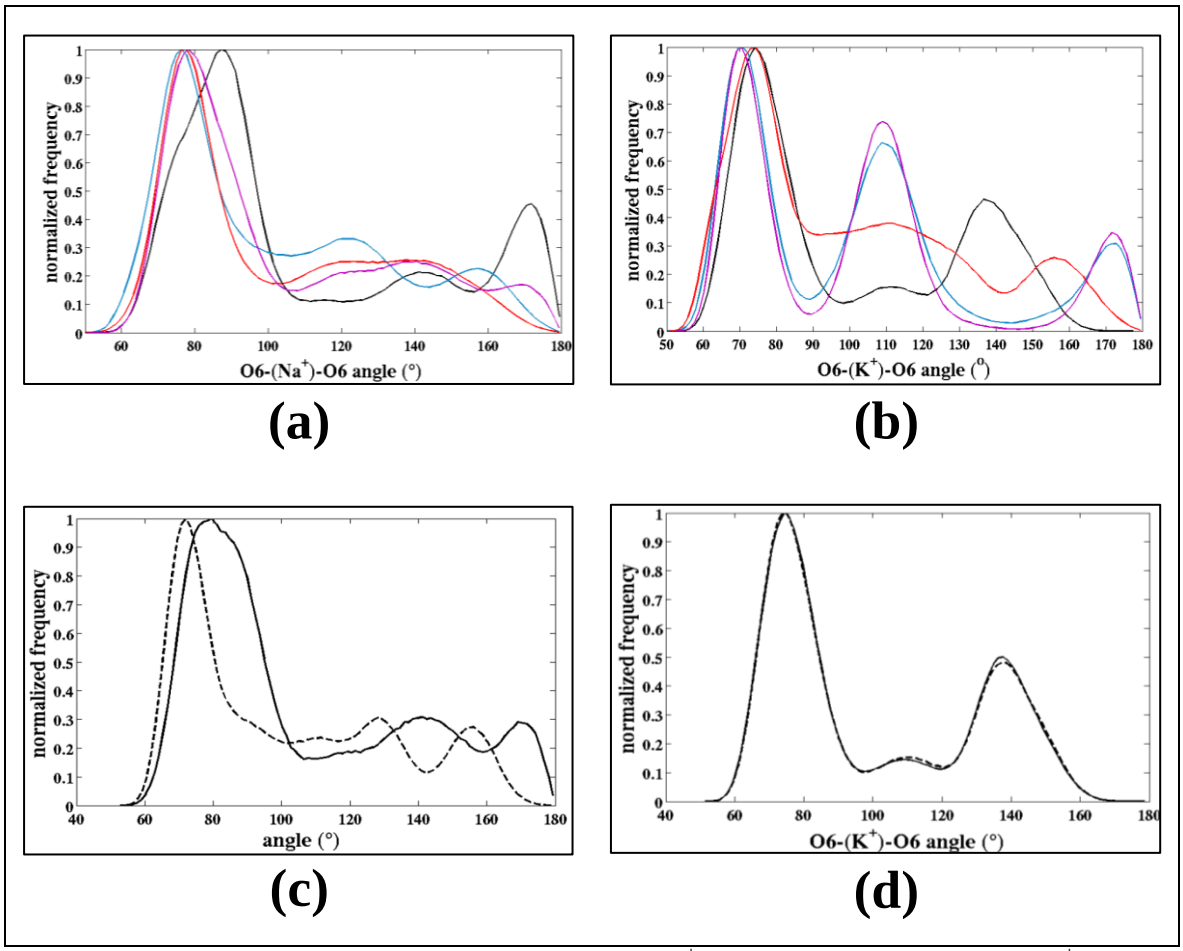


Figure 2.9: Frequency histogram of (a) O6—Na^+—O6 angle and (b) O6—K^+—O6 angle in the four human telomeric topology-types (NMR structure); (c) O6—Na^+—O6 angle (bold line) and O6—K^+—O6 angle (dotted line) for the non-telomeric hTERT sequence; and (d) O6—K^+—O6 angle in the X-ray crystal structures: simulation6 (bold line) and simulation7 (dotted line). The black, blue, red and magenta in (a) and (b) represent parallel, anti-parallel, mixed-(3+1)-form1 and mixed-(3+1)-form2 topologies respectively.

The analysis of simulation5 for G-rich hTERT promoter sequence showed different distributions for O6—Na^+—O6 and O6—K^+—O6 angles **(Figure 2.9c)**. Like the four telomeric topologies, the peak near 80° was maintained. The O6—Na^+—O6 angle distribution showed two more peaks at 140° and 170° whereas the O6—K^+—O6 angles distribution showed only two peaks at 130° and 158°. These distributions indicated that the G-quartet stacks have geometric similarity with the G-quartet stacks of telomeric topologies but more number of peaks point towards higher fluctuations during the simulation.

The crystal structures of simulation6 and simulation7 initially had K^+ in the G-tetrad cores and Na^+ ions in the medium. The frequency histogram of O6—K^+—O6 angles with peaks at approximately 78°, 110° and 138° **(Figure 2.9d)** indicate their similarities with simulation1b. Thus, the impact of presence of ligand on the G-tetrad is not as high as to disrupt the entire geometry of the quadruplex G-tetrad stacks.

2.3.3 Steered Molecular Dynamics Simulation Analysis

The difficulties in studying G-quadruplex unfolding by equilibrium MD simulation can be overcome by SMD simulations, where the ends of the molecule are linked to Hookean springs and pulled with constant velocity (Gräter *et al.*, 2005). This method gives the non-equilibrium force required to attain a particular end-to-end distance, *i.e.* amount of force needed to unfold the quadruplex (Li *et al.*, 2009). Although, force-dependant unfolding of a model human telomeric quadruplex in presence of ions has been analyzed by integrated fluorescence and magnetic tweezers spectroscopy (Long *et al.*, 2013), SMD simulations have not yet been extensively performed for unfolding studies. Furthermore, as stated earlier the experimental conditions often give multiple topologies together, making the analysis very difficult. Here, we have studied comparative unfolding of four different telomeric topologies and the unfolding force we obtained from SMD is comparable to the earlier reported stretching forces (Long *et al.*, 2013). The four NMR-reported human telomeric G-quadruplexes with respective topologies: parallel, anti-parallel, (3+1)-mixed-form1 and (3+1)-mixed-form2, have the common structural feature of quite globular formation. Initially, all the quadruplexes have a minimum of twelve base pairs as there are three G-quartet stacks and each stack has four G:G base pairs. In addition to these G:G base pairs some A:T base pairs are formed in the loop regions of some of

the quadruplexes. In order to understand the relative mechanical stabilities of these four quadruplex-DNA topology types, which show somewhat equivalent conformational stability during equilibrium MD simulation, we carried out SMD simulations for the same four systems.

On application of constant velocity SMD all the quadruplexes were found to elongate gradually with time until the oligonucleotide chains were completely straightened and attained maximum extension of ~125Å. We have plotted Potential of Mean Force (pmf) vs. extension **(Figure 2.10a and 2.10b)** for each system after simulating more than once with different initial seed. The force vs. extension plots **(Figure 2.10c)** indicated that when the pulling vector is in the plane of H-bonds, a high force of ~1.48pN is required to attain the end-to-end extensions of ~22Å and ~35Å by mixed-(3+1)-form1 and mixed-(3+1)-form2 respectively. However, when the force is applied perpendicular to the plane of H-bonds, two peaks at ~28Å and ~88Å extensions were observed for the mixed-(3+1)-form1 topology, both corresponding to 1.75pN force **(Figure 2.10d)**. The peaks at 42Å and 90Å extensions for the mixed-(3+1)-form2 topology correspond to the forces 1.5pN and 2.2pN respectively. Thus, from all the circumstances discussed, it emerges that higher amount of force is required to achieve complete extensions for the mixed-(3+1) topology types. The parallel topology requires the least force to completely stretch and the anti-parallel topology requires nearly 1pN force to attain complete extension, independent of the nature of pulling force vector. These intermediate peaks are possibly similar to those reported by folding studies on G-quadruplexes (Bian *et al.*, 2014). The slight difference in pulling forces between the experiments and our theoretical observation is mainly due to the fact that experimental conditions involve ions, large DNA-segments, solvent viscosity and perhaps signals from a mixture of different topologies, whereas, in our study we have used a single quadruplex moiety in an implicit solvent model.

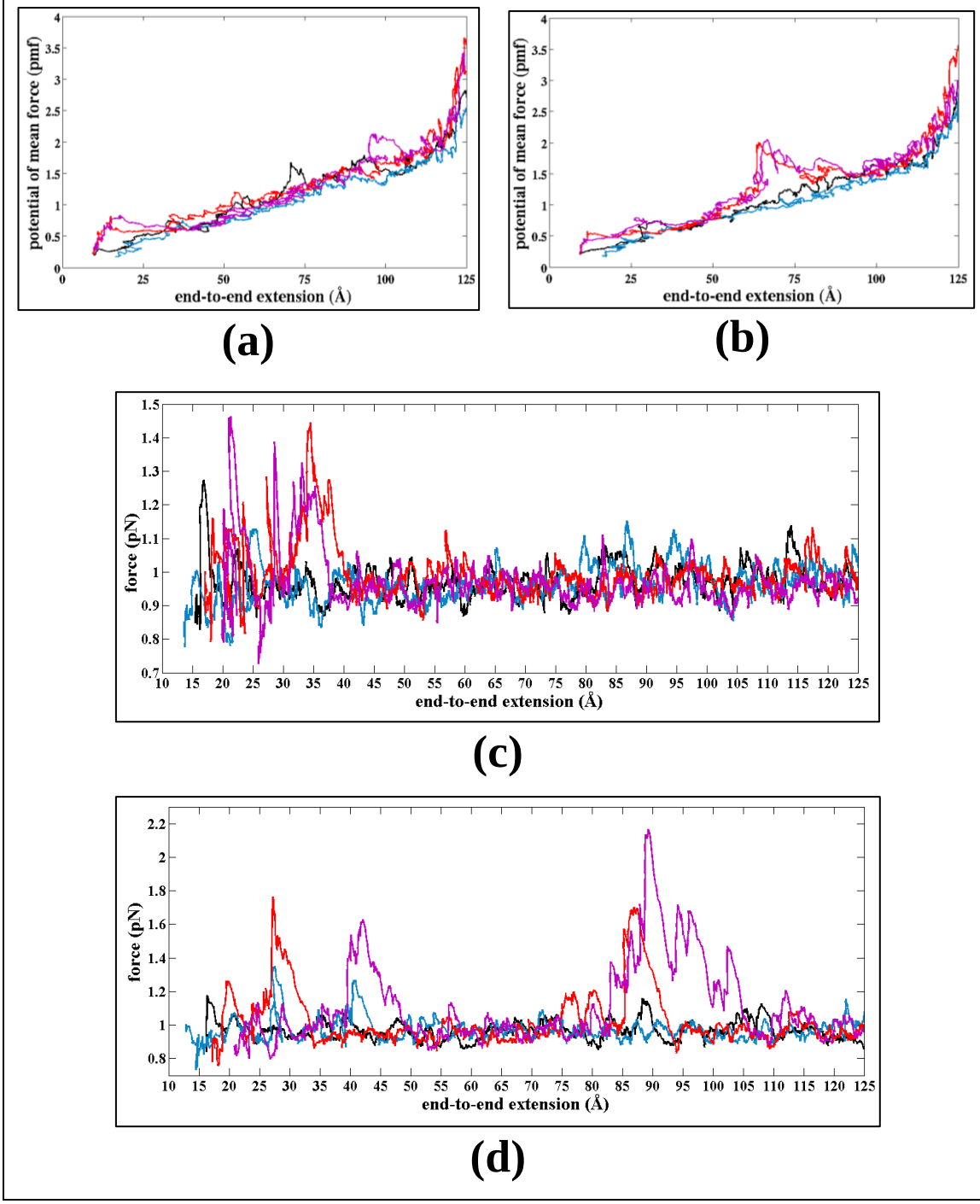


Figure 2.10: potential of mean force (pmf) vs. extension plot for the SMD simulations when the force is applied (a) parallel to the plane of H-bonds and (b) perpendicular to the plane of H-bonds; and force vs. extension plot for the SMD simulations when the force is applied (c) parallel to the plane of H-bonds and (d) perpendicular to the plane of H-bonds. The pulling is performed on the four telomeric quadruplexes in such a way that the 3'-end and 5'-end always move away from each other. The black, blue, red and magenta lines represent parallel, anti-parallel, mixed-(3+1)-form1 and mixed-(3+1)-form2 topologies respectively.

Each of the snapshots obtained during SMD were analyzed for base pairing by BPFIND software. This software detects a base pair only when two H-bonds are present between the involved bases without requiring the positions of hydrogen atoms. During the SMD process all the H-bonds were gradually broken with a decrease in the number of base pairs **(Figure 2.11)**. The figure indicates that when the pulling vector is in the plane of H-bonds the mixed-(3+1)-form1 and mixed-(3+1)-form2 topologies retain maximum number of base pairs even at 40-50Å extension **(Figure 2.11a)**. However, when the pulling vector is perpendicular to the plane of H-bonds, the base pair persistence for these two systems is highest even after 60Å extension **(Figure 2.11b)**. Thus, the mixed-topologies require the maximum force to completely stretch out as compared to others under similar conditions of applied velocity and associated spring constant. However, consideration of implicit solvent model may have imposed some restrictions and indicated slightly higher stability for the mixed-forms as compared to the anti-parallel topology.

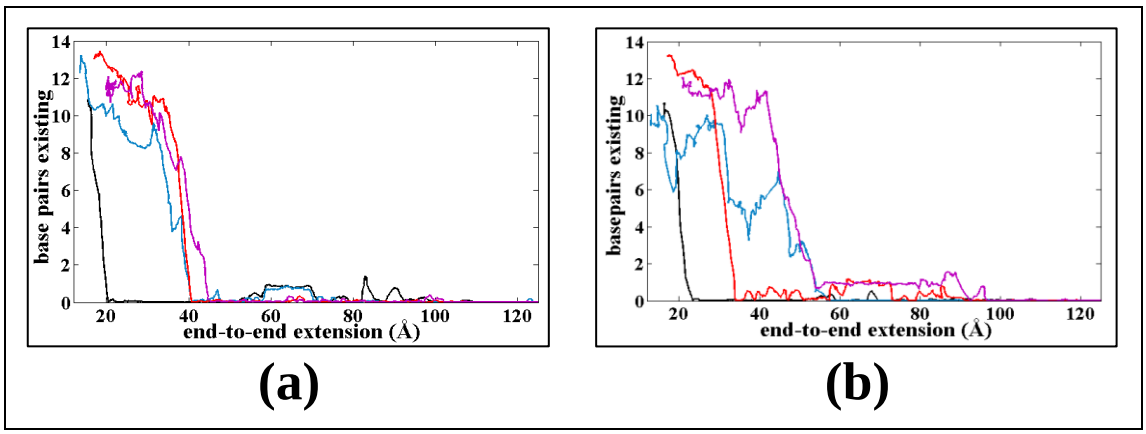


Figure 2.11: Average no. of base pairs existing vs. extension plot for the SMD simulations of the four telomeric quadruplexes when force is applied (a) parallel to the plane of H-bonds and (b) perpendicular to the plane of H-bonds. The black, blue, red and magenta lines represent parallel, anti-parallel, mixed-(3+1)-form1 and mixed-(3+1)-form2 topologies respectively.

2.3 Discussion

The present study provided an insight into the structural features of the four different telomeric quadruplex-DNA topologies, *viz.*, parallel, anti-parallel, mixed-(3+1)-form1 and mixed-(3+1)-form2, from the perspective of base pair and base pair step parameters. The quantum chemical studies indicated that the presence of ions (Na^+ or K^+) near the G-quartets adds stability to it. However, the planarity of the G-

quartet is dependent on the type and position of ion, *i.e.* the G-quartet becomes planar when a single Na^+ ion is inside the core of the G-tetrad whereas the quartet attains planarity when two K^+ ions are present on either side of the G-tetrad. The presence or absence of ions inside the core formed by stacking of two G-quartets affects the geometry of that stack, which thereby affects the overall structure of G-quadruplex-DNA. It is clear that G-tetrads are composed of cyclic G:G H:W C base pairs and a G-quadruplex is formed by stacking of two or more such G-tetrads. The inter- and intra-base pair orientation parameters of these G:G H:W C base pairs are found to be dependent on the topology of the G-quadruplex. Some of the quadruplexes consisted of capping over the extreme G-tetrads formed by A:T or A:A base pairs involving the loop region bases. These were found to have least fluctuation in their RMSDs and retained their compactness for a longer time during SMD simulations. Thus, such topologies can be considered as the most suitable major conformation. The MD simulation of hTERT promoter sequence was found to be inherently unstable with anomalous values of the base pair and base pair step parameters along with high standard deviation. It was clearly reflected from the breaking or opening up of some of the base pairs of this quadruplex during the 100ns equilibrium MD-run. Thus, all G-rich sequences may not adopt a topology similar to the ones adopted by G-rich sequences of the telomeric region. Combining the MD and SMD studies we could conclude that the anti-parallel topology is the most suited as a telomeric topology, although the mixed-(3+1)-form1 should not be completely neglected while considering major conformation. It would be interesting to study unfolding and folding of G-quadruplex structures with different topology in the light of helicoidal parameters and ion coordination.

Chapter III

Structural Studies of Unusual RNA motifs

Bulges and Pseudo-continuous Helices

3.1 Introduction

In recent years, great advances have been made in identifying and understanding non-coding RNAs (ncRNA), since evidences suggest that the majority of the genomes of mammals and other complex organisms is transcribed into ncRNAs, many of which are alternatively spliced and/or processed into smaller products (Mattick and Makunin, 2006). These ncRNA folds exhibit fascinating structural diversity. Double stranded RNA helices alone cannot account for such diversity. It is the occurrence of scattered non-helical motifs along with helical stems that confer such extensive structural diversity to RNA duplexes (Hermann and Patel, 1999; Moore, 1999).

Structured RNAs fold hierarchically from their sequence to native 3D tertiary form by formation of secondary structures, comprising mostly of such duplexes (Boyle *et al.*, 1980; Chai, 2008; Tinoco and Bustamante, 1999). A large number of different secondary structural motifs of RNA have been identified and are categorically listed in the RNA Structure Atlas and RNA 3D Motif Atlas databases (Petrov *et al.*, 2013) such as pseudo-continuous (coaxial) helices, hairpin loops, internal loops (symmetric & asymmetric or bulge loops), junction loops, *etc.* The knowledge about junction-regions of the recurring secondary structural motifs can significantly influence the prediction of long range tertiary interactions and higher order RNA structures which in turn can assist in exploring the RNA folding pattern. Among these motifs, bulges exhibit wide distribution in all types of large structured functional RNAs. Identification of the bulges as key structural elements in a wide range of RNAs underline their importance and versatility in RNA architecture and molecular recognition (Hermann & Patel, 2000). Bulges are either a single unpaired nucleotide (residues) or a stretch of unpaired residues located within a strand of double helical RNA (Woese and Gutell, 1989). These bulge loops are very common features of folded RNA structures, where they can present highly recognizable features for specific protein, ligand or metal-ion binding (Dassonneville, 1997; Ennifar *et al.*, 1999; Grate and Wilson, 1997; Hermann and Westhof, 1999; Mei *et al.*, 1998; Naryshkin *et al.*, 1998; Ye *et al.*, 1999, 1995). An example is the binding of human immunodeficiency virus transactivator protein Tat to the three-pyrimidine bulge in the response element TAR peptide (Ye *et al.*, 1995).

In RNA, along with the standard Watson-Crick (WC) base pairs, several types of non-canonical base pairs (Barciszewski and Clark, 1999; Leontis *et al.*, 2006; Sponer *et al.*, 2005) are found. These are as stable as the canonical base pairs (Bhattacharya *et al.*, 2015; Halder and Bhattacharyya, 2010; Sponer *et al.*, 2005) and play an integral part in RNA structural organization by: (i) spatially connecting the structural motifs formed by distant parts of the RNA chains and (ii) providing rigidity to the linker nucleotides (Butcher and Pyle, 2011; Kim *et al.*, 1996). Non-canonical base pairs are often present surrounding the bulges, as they are expected to introduce flexibility in the RNA backbone through their unusual conformations. Their substantial stability is indicative of their importance as 'nucleation site' for folding or unfolding of functional RNAs. In terminal regions, these non-canonical base pairs play the role of helix-capping, preventing terminal melting which is necessary for stabilization of short double helices (Lee & Gutell, 2014).

The conformations of bulge nucleotides are governed by the competing interactions of both the unpaired residues and the surrounding base pairs. The unpaired residues may participate in continuous stacking of the flanking regions with the helix **(Figure 3.1a)**, or they may be extruded from the duplex with the unpaired bases pointing out into the solvent **(Figure 3.1b)** (Hermann and Patel, 2000). When the unpaired residue is bulged inside and stacked well within the neighbouring base pairs, it can induce a kink in the helix-axis at the bulge site (Varani *et al.*, 1999). On the other hand, when the bulge nucleotides are looped out into the solvent, the overall duplex geometry is expected to remain close to the regular A-form (Portmann *et al.*, 1996). Both these type of bulges can create unique recognition sites in RNA 3D structures by directly acting as molecular handles within helical regions which are otherwise uniform. In internal loops, a special case of non-canonical base pairing called dinucleotide platform occurs. It is defined as two neighboring consecutive nucleotides arranged in a side-by-side planar arrangement with hydrogen bonds (H-bonds) between the residues (Sharma *et al.*, 2010). The best known examples of such platforms include: the ApA or adenosine platform (Cate *et al.*, 1996b) and the GpU platform **(Figure 3.2)** (Wimberly *et al.*, 1999). These dinucleotide platforms, frequently anchor helices against loops and remain conserved as the GpUpA/GpA miniduplex at the core of the loop-E/bulged-G motif (Lu *et al.*, 2010). Such H-bonded three co-planar bases with two of the bases sometimes forming a dinucleotide

platform are known as base triples commonly found in Loop E motifs, and Sarcin–ricin loop (Nagaswamy, 2002) of RNA structures.

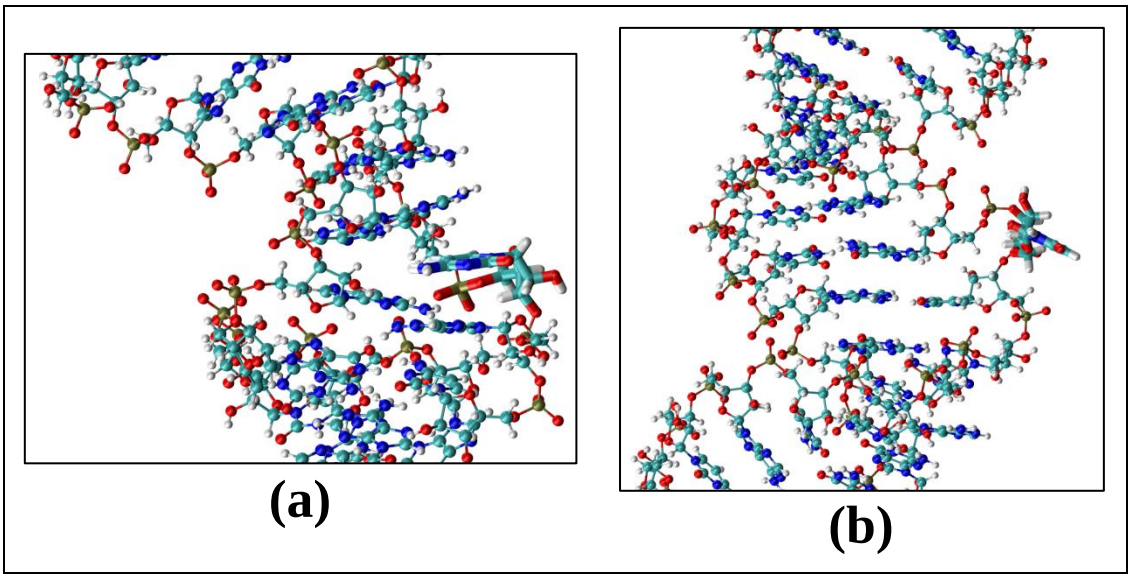


Figure 3.1: An unpaired residue (bonds) having (a) bulged-in conformation and (b) bulged-out flanking conformation.

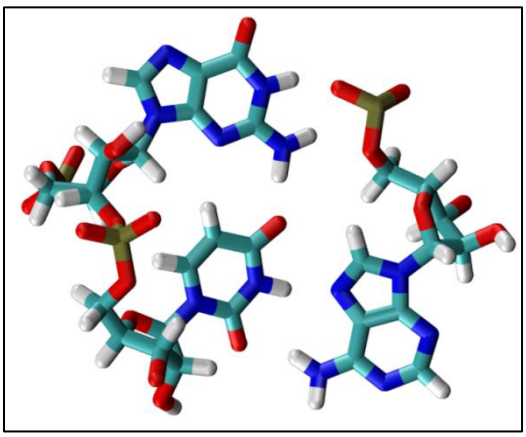


Figure 3.2: A GpU dinucleotide platform.

The bulges can indirectly distort the RNA backbone allowing access to base pairs in a widened deep major groove (Jiang and Patel, 1998; Ye *et al.*, 1995). The dimensions of the major and minor grooves in RNA duplexes can be altered by bulges introducing distortions in the nucleic acid backbone. Widening of the major groove by bulges, which exposes the H-bonding edges of base pairs, is frequently found at interaction sites where RNA domains dock into duplexes, or ligands bind to RNA (Hermann and Patel, 1999; Hermann and Westhof, 1999). Thus, these motifs may act as contact points in the tertiary interactions and folding of RNA (Woese and Gutell,

1989). Moreover, bulges form constitutive elements of peptide- and protein-binding sites in RNA either indirectly, by shaping the architecture of the binding region, or directly, by providing the recognition motifs (Diener and Moore, 1998; Jiang and Patel, 1998; Wimberly *et al.*, 1999). In some RNA-peptides or small molecule complexes, bulge nucleotides participate as 'flap' residues closing over the ligand-binding site, *e.g.*, the tobramycin aptamer RNA complex (Jiang and Patel, 1998).

As discussed earlier, bulge size can range from a single unpaired residue to several unpaired nucleotides. If the bulge size is several residues long, so that it appears to be formed by two different parts of the RNA chain, it is described as a pseudo-continuous helix (also called pseudohelix) or a coaxial stack. Coaxial helices are highly stabilizing and are dominant in several large RNA structures. Inter-helical stacking may occur *via* a single base or base pair bridge between helical stretches aligned along the same axis, resulting in a continuous helical stack spanning multiple helices. The continuous strand, opposite to the bulge containing strand, facilitates the coaxial stacking of the flanking residues forming double helix with another RNA strand.

The biological function of RNA is often directly related to the presence of bulges and loops, which can form recognition sites for ligands or mediate a possible catalytic activity. Some of these elements can also function as isolated structures, without the context of a large folded RNA structure. These non-helical RNA motifs are increasingly being considered as possible drug targets (Varani *et al.*, 1999; Zacharias, 2000). Moreover, recent breakthrough has led to the detection of a class of naturally occurring small non-coding RNAs known as micro-RNAs (miRNA) which function in RNA silencing and post-transcriptional regulation of gene expression **(Figure 3.3)** (He and Hannon, 2004) *via* base pairing with complementary sequences within mRNA. The matured miRNA is formed by enzymes which recognize a helix-loop motif where the helix region mostly has an asymmetric bulge **(Figure 3.3)**. The structural diversity of bulges in RNA architecture, their role in intramolecular interactions and intermolecular recognition, and various other biological processes, motivated us to study these secondary structural motifs focusing on composite double helices comprising of bulge loops or coaxial-stacks.

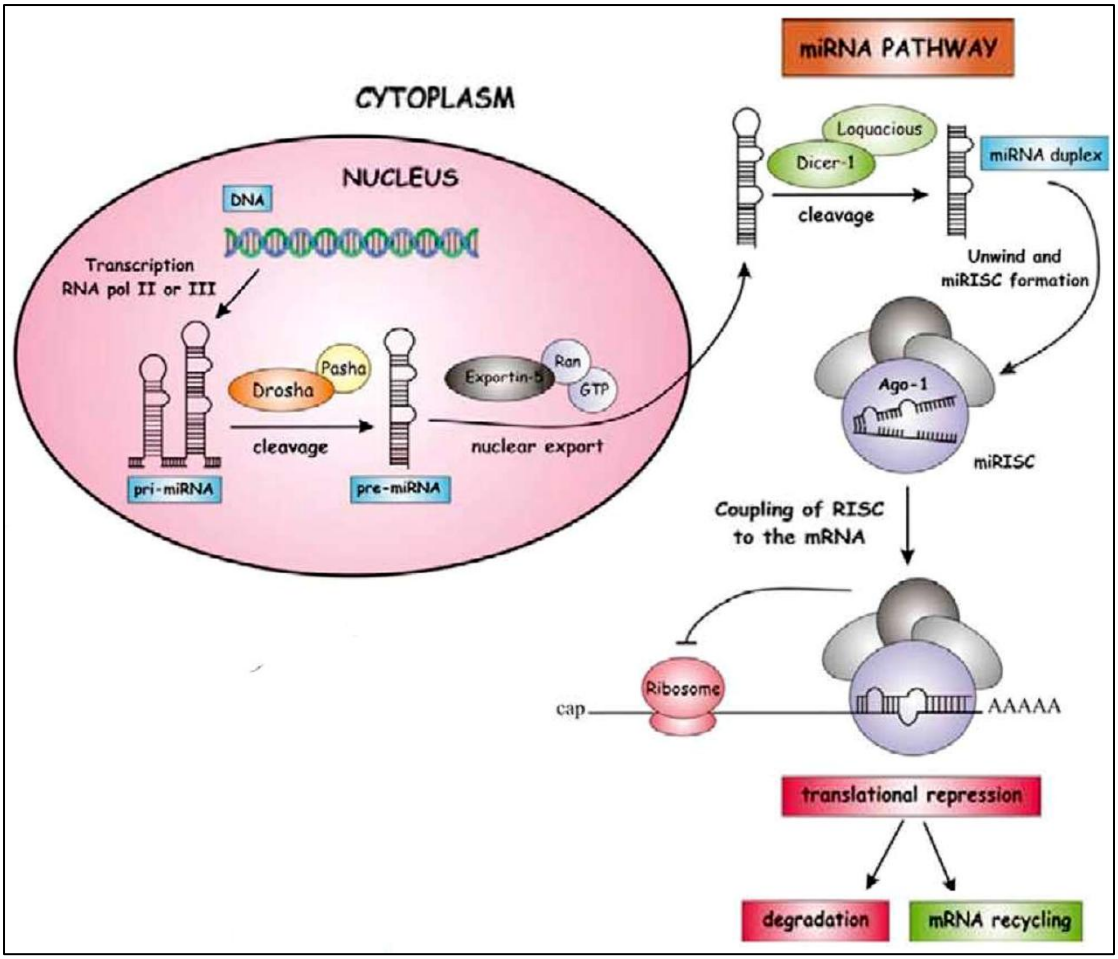


Figure 3.3: Micro-RNA Biogenesis

The sub-micro second time scale MD simulations with explicit solvent model can appropriately mimic the behaviour and dynamics of biomolecules at atomic level and has proved to be an important tool in the structural studies of nucleic acids. We have found plenty of literature based on structure and dynamics of non-canonical base pairs situated at termini of RNA double helices or at the beginning of hairpin loop (Proctor *et al.*, 2004; Villa *et al.*, 2009; Widjajakusuma *et al.*, 2012) and a few on structural variations in RNA duplex containing non-canonical base pairs (Ditzler *et al.*, 2010; Halder and Bhattacharyya, 2010). However, despite the huge biological importance of the bulge motifs, we have rarely found reports after year 2000 regarding the structure and dynamics of bulge containing RNA helices from ribosomal RNA structures.

In this work, we have therefore attempted to understand initially the structure and dynamics of one residue bulges and pseudo-continuous coaxial stacks in RNA structures by MD simulation. The bulge containing RNA fragments behave differently depending on the bulge residue and the surrounding base pairs of the

helical stem region. The RNA stretch behaves similar to that of standard A-form, when the bulge base is extra-helical, whereas, intercalated bulge residue introduces significant change in the overall RNA helix, affecting particularly the junction region. The structural features based on base pair and dinucleotide step parameters, stacking in terms of overlap of the base pairs, bending of two co-axial helices, *etc*. at the junction has been carefully studied and reported in this work.

3.2 Methodology

3.2.1 Model System Description

The bulge or pseudo-continuous stack containing helical fragments revealed the existence of various kinds of bulge containing RNAs (such as 1-residue, 2-residue, 3-residue, and 4-residue bulge motifs) of different functions in different source organisms (Agarwal, 2015). Apart from these, different pseudo-continuous helices also exist with variable base pairing at the junction and in combination with various other structural motifs such as hairpin loops. The base pairs above and below the bulge residue or forming a coaxial stack has been referred to as the junction **(Figure 3.4)**. In general it is expected that small bending angle between two helical fragments of a continuous RNA stretch is accompanied by good overlap between the base pairs of the junction. In case of single residue bulge containing helices, the anti-correlation is not strictly maintained. As for example, in 1VQO (Fragment ID: 9) the overlap at the junction is 48.5Å^2 but the bending angle is 40.9^o. Similarly, in 3U5D (Fragment ID: 3) only 12^o bend in the helix gives a small overlap value of 18Å^2. However, a good overlap in the range of 25Å^2 to 45Å^2 was seen in some pseudo-continuous helices, while in some systems the overlap at the junction was very poor or zero due to absence of stacking of the base pairs. Only those pseudo-continuous helices where selected where base pairs of the two fragments sufficiently overlap giving a composite single helix. Further, as reported earlier, in the crystal structures the canonical and non-canonical base pairs can lead to significant changes of the structural parameters of the overall duplex fragment (Halder and Bhattacharyya, 2010, 2012). Hence, to account for the structural changes introduced due to base pairing pattern at junction (bulge or coaxial stack), both canonical and non-canonical base pair containing systems were selected for simulation. Thus, combination of three single residue bulge

containing helices and three pseudo-continuous stacks containing fragments was selected for the present study on the basis of length of the helical fragments and dinucleotide step parameters at the junction from the non-redundant HDRNAS database (Ray *et al.*, 2012). The systems simulated have been denoted by a certain system ID, for the simplicity of representation throughout this report **(Table 3.1)**.

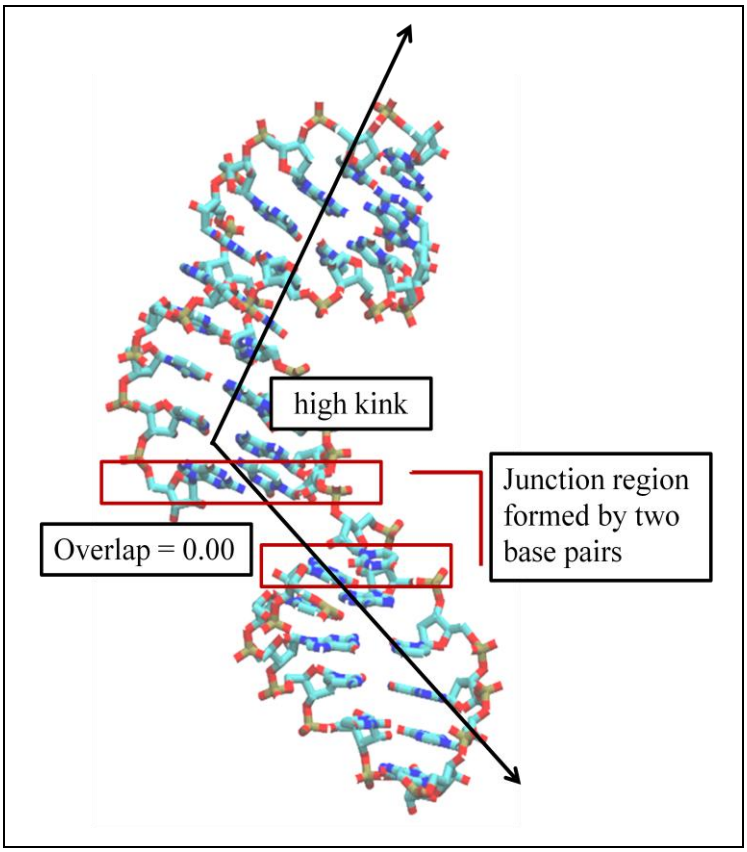


Figure 3.4: A representative RNA stretch taken from (PDB ID: 1VQO(2)) containing a bulge/pseudo helical motif. The junction region is formed by two base pairs that are on either side of the bulge/discontinuous region. The base pairs in red box form the junction region. The helix axis bending angle representation is shown by black axis. This system has a high kink (helix axis bending) with zero overlap at the junction.

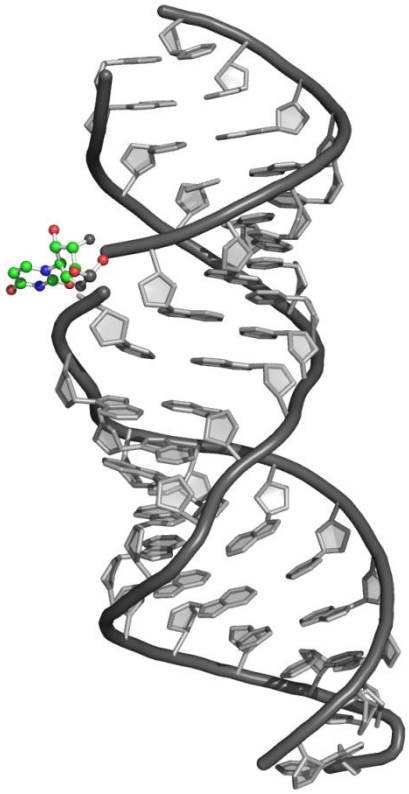
System 1a (PDB ID: 3U5B, Fragment ID: 3) and System 2a (PDB ID: 3U5D, Fragment ID: 3) are single residue bulge containing helices, differing in orientation of the bulge residue whereas, System 3a (PDB ID: 1VQO, Fragment ID: 9) is a dinucleotide platform containing helix but the bases orient to form a single residue bulge. In System 1a the bulge residue, Uracil (1657U) is protruding out from the duplex into the solvent **(Figure 3.5.1a)**, System 2a has the bulge residue Cytosine (743C) intercalated between two non-canonical base pairs **(Figure 3.5.1b)** and System 3a contains a Guanine (3078G) residue at the bulge region intercalated within

the duplex and involved in formation of a dinucleotide platform with adjacent Uracil (3079U) residue of the same strand using its sugar edge and a base triplet with the Uracil:Adenine (3079U:3103A) base pair present adjacent to the bulge region towards the 3'-end **(Figure 3.5.2)**. The systems 1b, 2b and 3b are the modeled counter-parts of the bulge-containing systems developed by removing the unpaired residue. Simulation studies were performed with both the native and model systems as these presented a way of understanding the role of unpaired flanking residues at the junction. The systems 4a (PDB ID: 3U5D, Fragment ID: 2) and 4b are pseudo-continuous helices with pure WC (W:W) base pairing at the junction **(Figure 3.6a)**. The difference between the two systems is that the former (System 4a) contains extra flanking residues near the junction, representative of pseudohelix, whereas System 4b was modeled such that it is terminated at the junction. Similarly, System 5a (PDB ID: 2ZJR, Fragment ID: 2) has flanking residues whereas its counter model System 5b is terminated at the junction, however both these pseudo-continuous helices contain non-canonical A:G W:W C base pair at the junction **(Figure 3.6b)**. Finally the pseudohelix System 6 (PDB ID: 3DIO, Fragment ID: 3) containing a 3-residue hairpin loop was selected **(Figure 3.7)**. A detailed description involving, PDB ID (Fragment ID), sequence, location of the bulge or coaxial-stack in the helix, *etc.*, for each system selected for the present work is given in **Table 3.1**

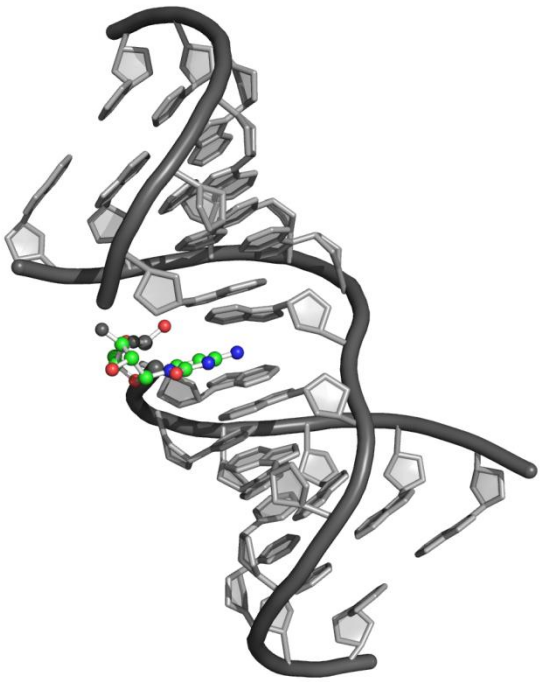
Table 3:1: Description of the systems studied.

System ID	PDB ID (Frag ID)	No. of RNA strands	A	B	C	strand A	strand B	strand C	helix-axis 1	helix-axis 2	stretch of helix 1	stretch of helix 2
System Details			(strand) no. of residues			residue ID			length in terms of base pair		residue ID	
1a	3U5B (3)	2	20	19	_	1646-1665	1736-1754	_	9	6	1647U:1753A—1655A:1745G	1659A:1742U—1664C:1737G
1b		3	11	8	19	1646-1656	1658-1665	1736-1754				
2a	3U5D (3)	2	12	13	_	721-732	737-749	_	4	4	722G:748U—725G:745C	728G:741U—731U:738A
2b		3	12	6	6	721-732	737-742	744-749				
3a	1VQO (9)	2	18	17	_	3069-3086	3096-3112	_	7	6	3070U:3111U—3076G:3105A	3080A:3102G—3085A:3097U
3b		3	9	8	17	3069-3077	3079-3086	3096-3112				
4a	3U5D (2)	3	18	12	10	495-512	579-590	609-618	6	8	496C:617G—501A:612U	504A:587U—511G:580C
4b		3	18	10	8	495-512	579-588	611-618				
5a	2ZJR (2)	3	13	9	8	2647-2659	2623-2631	2707-2714	5	4	2626U:2652G—2630C:2648G	2708U:2658A—2711G:2655C
5b		3	13	7	6	2647-2659	2625-2631	2707-2712				
6	3DIO (2)	2	38	11	_	112-149	154-164	_	10	7	115G:139C—124G:130C	142U:161G—148A:155U

1646	C	:	A	1754	W:W T
1647	U	:	A	1753	W:W C
1648	A	:	U	1752	W:W C
1649	G	:	C	1751	W:W C
1650	U	:	A	1750	W:W C
1651	A	:	A	1749	w:w C
1652	C	:	G	1748	W:W C
1653	C	:	G	1747	W:W C
1654	G	:	A	1746	S:H T
1655	A	:	G	1745	H:S T
1656	U	:	A	1744	W:W C
1657	U	:			
1658	G	:	U	1743	W:W C
1659	A	:	U	1742	W:W C
1660	A	:	U	1741	W:W C
1661	U	:	A	1740	W:W C
1662	G	:	C	1739	W:W C
1663	G	:	U	1738	W:W C
1664	C	:	G	1737	W:W C
1665	U	:	G	1736	W:W C



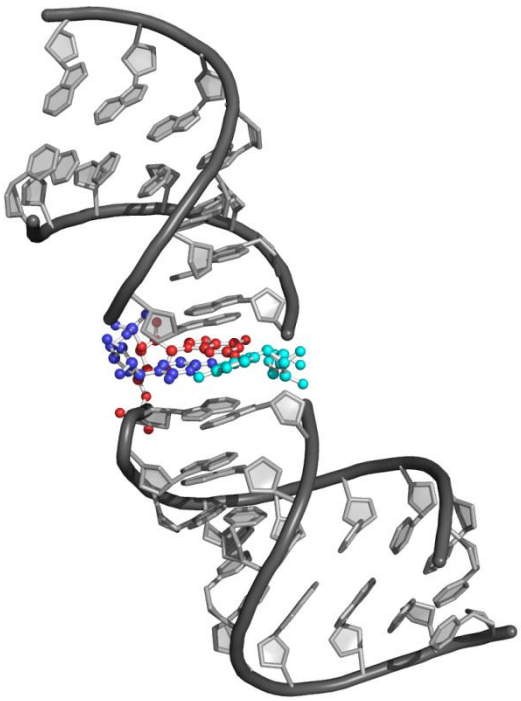
(a)



721	G	:	C	749	W:W C
722	G	:	U	748	W:W C
723	U	:	A	747	W:W C
724	U	:	A	746	W:W C
725	G	:	C	745	W:W C
726	G	:	A	744	S:H T
			C	743	
727	G	:	G	742	H:S T
728	G	:	U	741	W:W C
729	C	:	G	740	W:W C
730	C	:	G	739	W:W C
731	U	:	A	738	W:W C
732	C	:	G	737	W:W C

(b)

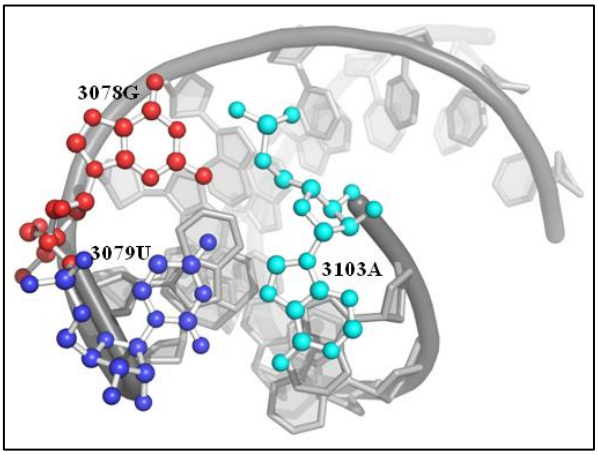
Figure 3.5.1: RNA representation with base pairing pattern (in box) of (a) System 1a (3U5B(3)) and (b) System 2a (3U5D(3)). The unpaired residues are drawn in ball and stick (CPK colour) scheme.



3069	U	:	U	3112	W:W C
3070	U	:	U	3111	W:W C
3071	C	:	G	3110	W:W C
3072	C	:	G	3109	W:W C
3073	G	:	C	3108	W:W C
3074	G	:	C	3107	W:W C
3075	G	:	C	3106	W:W C
3076	G	:	A	3105	S:H T
3077	A	:	A	3104	H:H T
3078	G	:	U	3079	s:h C
3079	U	:	A	3103	W:H T
3080	A	:	G	3102	H:S T
3081	C	:	G	3101	W:W C
3082	U	:	G	3100	W:W C
3083	G	:	U	3099	W:W C
3084	G	:	C	3098	W:W C
3085	A	:	U	3097	W:W C
3086	G	:	C	3096	W:W C

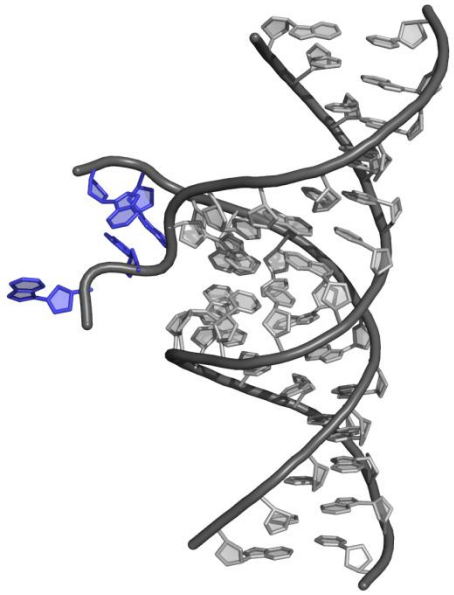
BASE PAIRING AT JUNCTION				
3077	A	...	A	3104
3078	G			
3079	U	...	A	3103

(a)



(b)

Figure 3.5.2: Line and CPK (unpaired residue) representation with base pairing pattern (in box) of System 3a (1VQO(9)).

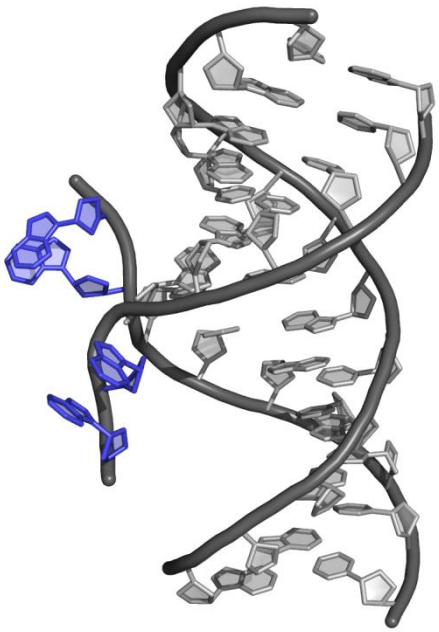


495	G	:	C	618	W:W C
496	C	:	G	617	W:W C
497	C	:	G	616	W:W C
498	A	:	U	615	W:W C
499	G	:	C	614	W:W C
500	C	:	G	613	W:W C
501	A	:	U	612	W:W C
502	U	:	A	611	W:W C
			G	610	
			G	609	
			G	590	
			A	589	
503	C	:	G	588	W:W C
504	A	:	U	587	W:W C
505	G	:	C	586	W:W C
506	U	:	A	585	W:W C
507	U	:	G	584	W:W C
508	U	:	G	583	W:W C
509	U	:	G	582	W:W C
510	G	:	U	581	W:W C
511	G	:	C	580	W:W C
512	U	:	G	579	W:W C

JUNCTION				
502 U	:	A 611	610 G	609 G
503 C	:	G 588	589 A	590 G

(a)

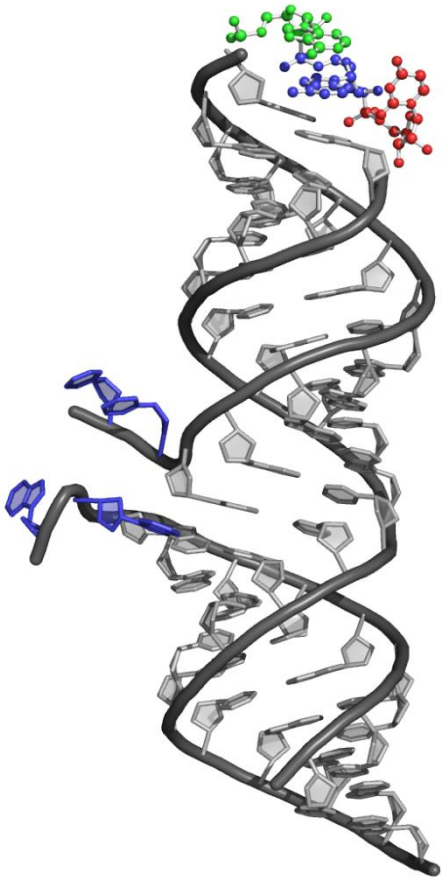
2647	G	:	C	2631	W:W C
2648	G	:	C	2630	W:W C
2649	A	:	U	2629	W:W C
2650	G	:	C	2628	W:W C
2651	U	:	G	2627	W:W C
2652	G	:	U	2626	W:W C
2653	A	:	U	2625	W:W C
			G	2624	
			A	2623	
			A	2714	
			A	2713	
2654	A	:	G	2712	W:W C
2655	C	:	G	2711	W:W C
2656	G	:	C	2710	W:W C
2657	G	:	C	2709	W:W C
2658	A	:	U	2708	W:W C
2659	C	:	G	2707	W:W C



JUNCTION				
2653 A	:	U 2625	2624 G	2623 A
2654 A	:	G 2712	2713 A	2714 A

(b)

Figure 3.6: RNA representation with base pairing pattern (in box) of (a) System 4a (3U5D(2)) and (b) System 5a (2ZJR(2)). The flanking residues (pseudohelix region) at the junction are shown in blue colour.



				127 A			
			126 A		128 U		
		125	G	...	A	129	S:H T
		124	G	...	C	130	W:W C
		123	G	...	C	131	W:W C
		122	G	...	C	132	W:W C
		121	U	...	A	133	W:W C
		120	A	...	U	134	W:W C
		119	U	...	A	135	W:W C
		118	G	...	C	136	W:W C
		117	G	...	C	137	W:W C
		116	G	...	A	138	W:W C
		115	G	...	C	139	W:W C
		114	G	...	U	140	W:W C
	113 U						
112 C		162	A	...	G	141	W:W C
	163 G	161	G	...	U	142	W:W C
164 A		160	G	...	C	143	W:W C
		159	U	...	A	144	W:W C
		158	G	...	C	145	W:W C
		157	C	...	G	146	W:W C
		156	C	...	G	147	W:W C
		155	U	...	A	148	W:W C
		154	G	...	G	149	W:W C

Figure 3.7: RNA representation with base pairing pattern (in box) of System 6 (3DIO(2)). The loop residues are shown in ball and stick model and the terminal flanking residues (pseudohelix region) is shown in blue colour.

3.2.1 Simulation Set-up

The initial coordinates of all the systems were taken from the Protein Data Bank (PDB) (Berman *et al.*, 2000). All the systems were simulated using GROMACS software v.4.5.3 (Hess *et al.*, 2008) with AMBER-99sb force field (Hornak *et al.*, 2006; Pérez et al., 2007). In each simulation, the RNA stretches were explicitly solvated with TIP3P water molecules maintaining periodic boundary conditions and at least 15Å cut-off between the molecule and the walls of the rectangular box. Appropriate numbers of Na^+ charge neutralizing counter-ions were added to maintain electro-neutrality of the whole system. The solvated systems were then subjected to energy minimization involving a maximum of 20000 Steepest Descent steps followed

by 50000 Conjugate Gradient steps. In each case, Verlet cut-off scheme with short-range electrostatics and van der Waals cut-off of 1nm was considered. The long-range electrostatic interactions were treated by Particle-Mesh-Ewald (PME) summation method (Sagui and Darden, 1999) with 1.6Å grid spacing. The systems were then gently heated to 300K followed by NVT ensemble equilibration run employing Leap-Frog integrator algorithm and Berendson thermostat with time-constant of 0.1ps for temperature coupling and time step of 2fs. The constraint algorithm LINCS was used to maintain lengths of bonds involving hydrogen atoms. After equilibration, the final MD run was carried out up to a minimum of 500ns for each system. Some of the simulations however had to be extended further due to their high structural variations even near 500ns. The details of the simulations are given in **Table 3.2**. The trajectories were made up by frames collected every 2.0ps and then analyzed by using GROMACS.

The bulge-containing RNAs (systems 1a, 2a and 3a) were also simulated with charmm36 (Vanommeslaeghe *et al.,* 2010; Yu *et al.,* 2012) force field using NAMD software (Phillips *et al.,* 2005) for the purpose of comparison. In each simulation, the RNA molecules were explicitly solvated with TIP3P water molecules in a rectangular periodic box whose dimensions were at least 30Å larger than the size of the corresponding solute molecules. Na^+ was considered as the charge neutralizing counter-ion. The positions of the counter-ions were generated by Monte Carlo simulation in absence of water, considering only the interactions between the ions and the nucleic acid concerned. The initial round of equilibration with explicit solvent and ions involved 100 steps of steepest descent, followed by 1000 steps of conjugate gradient and 20000 steps of the adopted basis Newton–Raphson energy minimization using CHARMM (Brooks *et al.,* 1983). The particle mesh Ewald (PME) summation method (Darden *et al.,* 1993), with width of the Gaussian distribution in the PME as 0.35, was used to treat the long-range electrostatic interactions. A force-switch method was applied for non-bonded interactions (van der Waals) with a 12Å cutoff. In these simulations, the constraint algorithm SHAKE was used to maintain lengths of bonds involving hydrogen atoms. The systems were gently heated from 0K to 300K in 30ps with a time step of 1.0fs after energy minimization followed by the final production run of 200ns by Constant Pressure Temperature dynamics algorithm at 1atm pressure, with 1fs time step. The trajectories were made up by frames collected every 1.0ps and analyzed using CHARMM and VMD (Humphrey *et al.,* 1996).

The base pair information for all the simulated systems was obtained using BPFIND software (Das *et al.*, 2006) and the base pair/base pair step parameters were calculated using NUPARM software (Bansal *et al.*, 1995; Mukherjee *et al.*, 2006; Pingali *et al.*, 2014). The inherent bent at the junction of the RNAs were estimated using NUPARM in terms of bending angle for which two average helix axes were fitted to the RNA systems by considering C1' atoms of the non-terminal base pairs on either sides of the junction and excluding those present near the junction. The helix-axis determination algorithm follows that adopted by Dickerson in FREEHELIX program (Dickerson, 1998). The details of the helix-axis chosen for the RNA systems are mentioned in **Table 3.1**.

3.3 Results

3.3.1 MD Simulation by AMBER-99sb force field

3.3.1.1 Dynamics and Fluctuation

An initial idea about the dynamics and overall behaviour of the simulation trajectories was obtained from the root-mean-square deviations (RMSD) **(Figure 3.8)** and root-mean-square fluctuations (RMSF) **(Figure 3.9.1 and 3.9.2)** of the simulation snapshots calculated with respect to the corresponding energy-minimized structures**.** MD simulations generate an ensemble of structures that have similar conformations. But, each system requires some time to reach that equilibrated state, referred to as its equilibration time. The stretch of MD-run during which a system remains in equilibrated state is different for individual simulation and is given in **Table 3.2**, The RMSFs per residue were calculated for only the MD-stretches during which the respective systems remain in equilibrated state. This indicated time-average fluctuation of each residue for the simulation snapshots which in turn can be compared to experimental B-factor. Low RMSF indicates higher rigidity of the residues and in turn stable base pairing involving these residues. A very high RMSF is expected for the terminal and near-terminal residues of the RNAs due to possible terminal fraying which keeps on decreasing on moving towards the middle of the stem.

Table 3.2: Details of the MD simulations.

ID	PDB ID (Frag ID)	Sequence	Base pair step at the junction	MD run (ns) (GROMACS and AMBER-99sb)	MD run (ns) (NAMD-charmm36)	Equilibrium state maintained during (ns)	
						GROMACS and AMBER-99sb	NAMD-charmm36
1a	3U5B (3)	5'-CUAGUACCGAUUGAAUGGCU-3' 3'-AAUCAAGGAGA UUUACUGG-5'	(U:G W:W C)/ (A:U W:W C)	500	100	70-500	20-100
1b		5'-CUAGUACCGAUGAAUGGCU-3' 3'-AAUCAAGGAGAUUUACUGG-5'		500	NA	120-500	NA
2a	3U5D (3)	5'-GGUUGG GGCCUC-3' 3'-CUAACACGUGGAG-5'	(G:A S:H T)/ (G:G H:S T)	600	100	400-600	20-60*
2b		5'-GGUUGGGGCCUC-3' 3'-CUAACAGUGGAG-5'		600	NA	400-600	NA
3a	1VQO (9)	5'-UUCCGGGGAGUACUGGAG-3' 3'-UUGGCCCAA AGGGUCUC-5'	(A:U H:W T)/ (A:A H:H T)	500	100	20-500	20-100
3b		5'-UUCCGGGGAUACUGGAG-3' 3'-UUGGCCCAAAGGGUCUC-5'		500	NA	120-500	NA
4a	3U5D (2)	5'-GCCAGCAU CAGUUUUGGU-3' 3'-CGGUCGUAGG-5' 3'-GAGUCAGGGUCG-5'	(U:A W:W C)/ (C:G W:W C)	500		20-500	
4b		5'-GCCAGCAU CAGUUUUGGU-3' 3'-CGGUCGUA-5' 3'-GUCAGGGUCG-5'		500		20-500	
5a	2ZJR (2)	5'-GGAGUGA ACGGAC-3 3'-CCUCGUUGA-5' 3'-AAGGCCUG-5'	(A:U W:W C)/ (A:G W:W C)	500		20-500	
5b		5'-GGAGUGAACGGAC-3' 3'-CCUCGUUGGCCUG-5'		600		120-470	
6	3DIO (2)	5'-CUCCGUGGAGA-3' 5'-CUGGGGGUAUGGGGAAU 3'-GAGGCACUG UCACCAUACCCA	(G:U W:W C)/ (G:A W:W C)	500		20-500	

*after 60ns, the RMSD shows a steep rise indicating a completely different state.

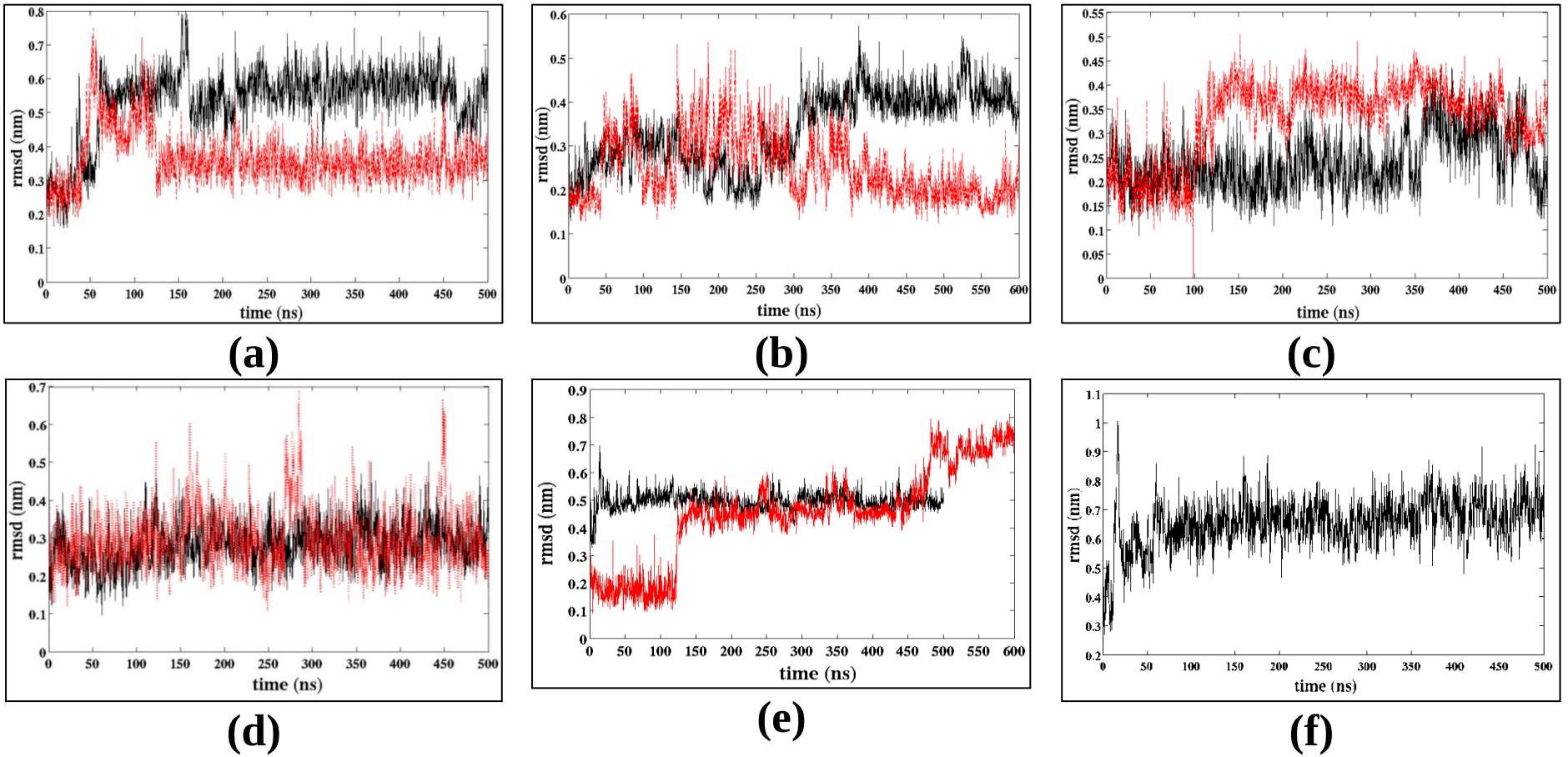


Figure 3.8: RMSD of (a) systems 1a (black and solid) and 1b (red and dotted); (b) systems 2a (black and solid) and 2b (red and dotted); (c) systems 3a (black and solid) and 3b (red and dotted); (d) systems 4a (black and solid) and 4b (red and dotted); (e) systems 5a (black and solid) and 5b (red and dotted); and (f) System 6.

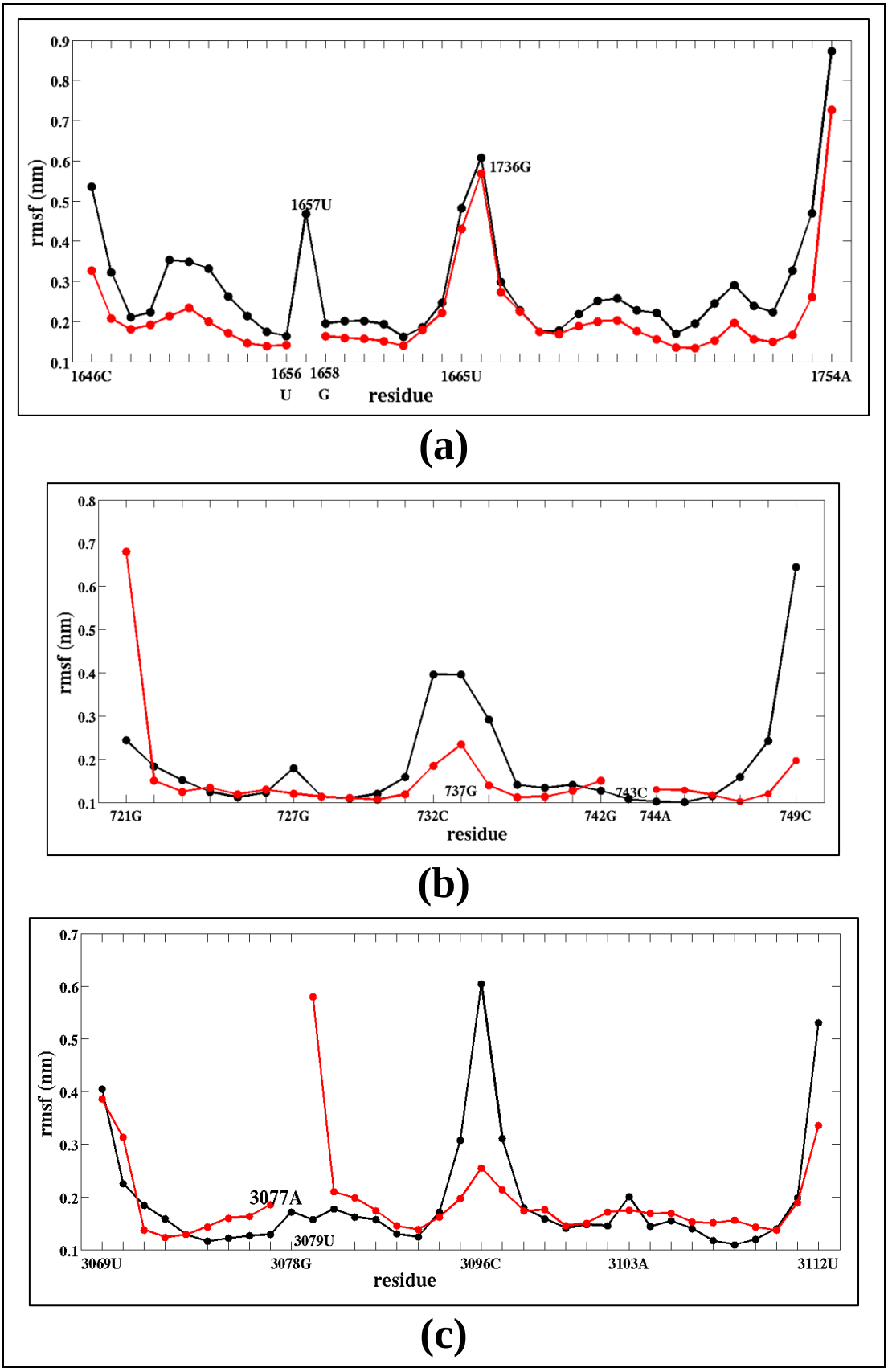


Figure 3.9.1: RMSF of (a) systems 1a (black) and 1b (red); (b) systems 2a (black) and 2b (red); and (c) systems 3a (black) and 3b (red).

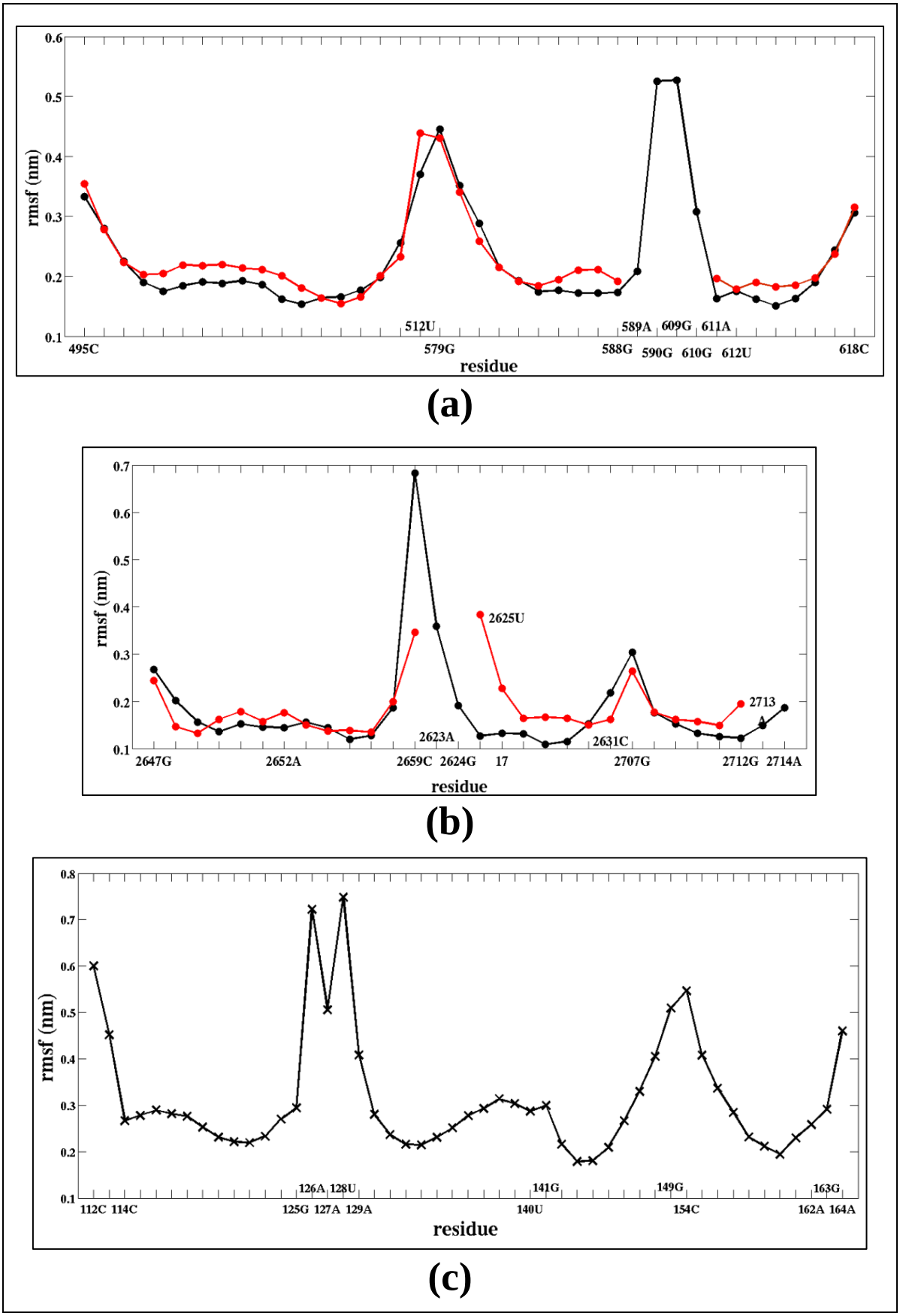


Figure 3.9.2: RMSF of (a) systems 4a (black) and 4b (red); (b) systems 5a (black) and 5b (red); and (c) Systems 6 (black).

3.3.1.1.1 Bulge Systems

The RMSD plot corresponding to systems 1a and 1b **(Figure 3.8a)** clearly indicates that System 1a was well stabilized within 70ns of the 500ns MD-run whereas; System 2b was stabilized almost after 120ns. In general, both these systems show a similar trend in RMSF **(Figure 3.9.1a)** for most of the bases except for a few involved in unusual non canonical base pairing (G:A S:H T, A:A w:w C, C:A W:W T) (Halder and Bhattacharyya, 2010). These bases however, do not form the junction region. The bulged-out Uracil residue (1657U) of System 1a has a higher fluctuation of 0.48nm as compared to bulge residues of systems 3a and 4a due to its extra-helical nature and greater flexibility.

The helical region of System 2a is a short twelve base pair stretch. As a result, it showed tendencies of high structural fluctuations in the first 300ns of the MD-run and after 300ns a jump in the RMSD was observed **(Figure 3.8b)**. Therefore, the simulation was extended to 600ns and it was found that the system retained possibly a second equilibrium state from 400ns-600ns. This might indicate major readjustments of conformation taking place after 300ns of MD-run. In this system the bulge residue is stacked inside the double-helix, intercalated between the base pairs at the junction, giving rise to a kink in the molecule. The bulged-in Cytosine residue has a low fluctuation of 0.18nm resulting from its intercalated nature **(Figure 3.9.1b)**. This residue (743C) participates in cross-strand H-bonding with residues 726G and 727G involving sugar and phosphate groups also **(Figure 3.10a)**. The H-bonds initially present between 742G and 727G was broken after about 400ns of MD-run with simultaneous increase in H-bonding of type O2P…N2 between 727G and 743C **(Figure 3.10b)**. It may be mentioned that such G:G H:S T base pairs, involving sugar 2'-OH group, was rarely found in RNA structures (Bhattacharya *et al.*, 2015) and also found unstable by DFT studies. Thus, the bulged-in residue (743C) which is intercalated into the helix seriously weakens the base pairing 727G:742G H:S T making the 727G residue flexible **(Figure 3.11)**. Absence of the bulge residue in counter-model System 2b may be an additional reason for its longer equilibration time of 400ns **(Figure 3.8b)** as it required more time to adjust and stabilize. Thus, it was also extended to 600ns. The Guanine residue (727G) of System 2a, has RMSF of 0.2nm due to the flexibility it attained during the MD-run, which is decreased to 0.14nm in case of System 2b, *i.e.* in absence of the bulge residue.

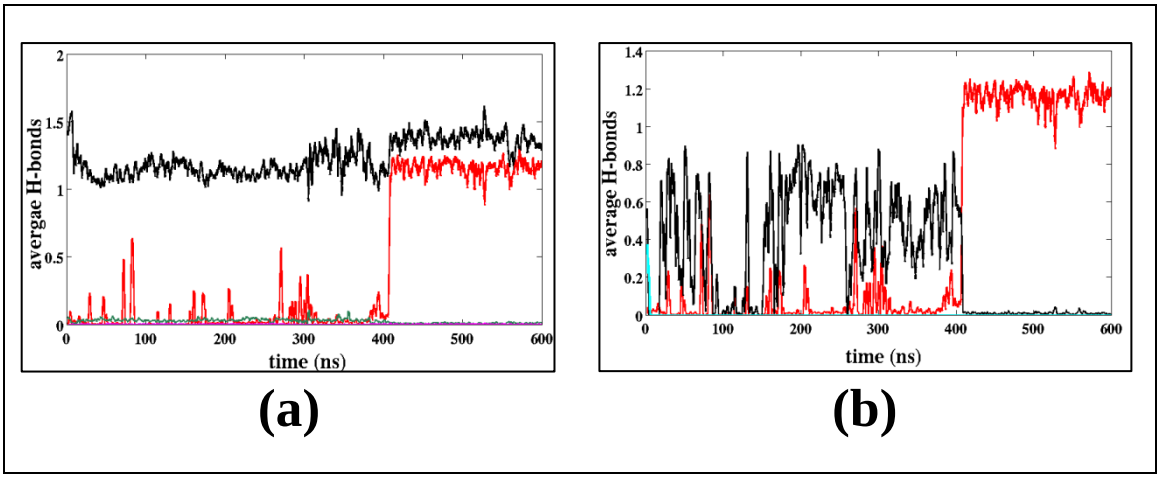


Figure 3.10: (a) Time evolution plot for average H-bonds (running average with window of 1000 snaps) involved in base pairs 743C:726G (black), 743C:727G (red), 743C:742G (green) and 743C:744A (magenta). (b) At 400ns, the 742G:727G base pairing (back line represents H-bonding via O2P of 727G residue and cyan line represents H-bonding via polar atoms) is broken and H-bond between 743C and 727G involving O2P of 727G increases (red).

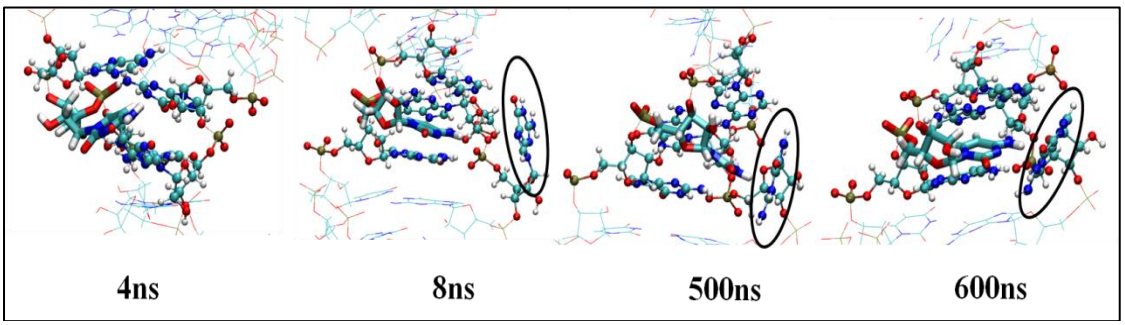


Figure 3.11: Snapshots from simulation of System 3a indicating movement of residue 727G (black oval) from inside the helix towards the solvent breaking the base pairing.

The RMSD values clearly indicate that System 3a was well stabilized within first 20ns of MD-run. However, its modeled counterpart (System 3b) equilibrated only after 120ns **(Figure 3.8c)**. This indicates that the RNA stretch achieves additional stability due to the dinucleotide platform and base-triple. However, this was lost on removal of the bulge residue (System 3b) as it disrupts the dinucleotide platform and base triple formation. In the System 3a the bulge residue 3078G forms s:h C base pairing involving weak C-H…O H-bond with the next residue 3079U of the same strand. This type of base pairing between two successive residues is generally called dinucleotide platform. The 3079U is also involved in W:H T base pairing with 3103A of the adjacent strand. When the dinucleotide platform and base triple formation are disrupted the system takes additional time to adjust to the most suitable double helix form. The Guanine residue (3078G) forming the bulge exhibited

a low RMSF of 0.18nm due to its H-bonded state with adjacent Uracil (3079U) of the same strand and Adenine (3103A) of the complimentary strand in System 3a. Here, the same 3079U like other non-terminal residues had low RMSF of 0.15nm, whereas in the counter-model System 3b in absence of the bulge-G residue, this jumped to 0.6nm **(Figure 3.9.1c)**. Since, the bulge residue is involved in formation of dinucleotide platform with 3079U and a base triple with the adjacent base pair, System 3a was expected to remain in a more stable during the MD-run, whereas, absence of the bulge residue, might affect the 3079U:3103A base pairing, thereby, weakening it and making the bases more flexible, contributing to its unusually high RMSF.

3.3.1.1.2 Pseudo-continuous Helix

The systems 4a, 4b, 5a and 6 stabilize well within 20ns of the MD-run, considered as their equilibration time. However, System 5b shows an unusual pattern in the RMSD plot **(Figure 3.8e)** with sharp jump in the RMSD after 120ns indicating a phase transition. The system then attained a stable conformation, which persisted till 470ns, beyond which again a phase change occurred marked by yet another jump in the RMSD. Thus, three distinct regions: 20-120ns, 120-470ns and 480-600ns, are seen which indicate the existence of three very different conformations this system can attain. However, since the system remains in the same phase for quite a long time of 120-470ns, it was considered as the contributing stable state of the system. The systems 4a and 5a follow a similar trend in fluctuation of most of the bases with their modeled counter-parts, systems 4b and 5b respectively. Few residues in the RNA stem of systems 4b and 5b showed higher RMSF **(Figure 3.9.2a and 3.9.2b)** due to the absence of extra flanking residues at the junction. A common feature found in both these pseudohelices is that the base pairs present at the junction tend to fluctuate more in absence of the extra flanking residues (systems 4b and 5b) as compared to that in presence of such residues (systems 4a and 5a).

3.3.1.2 Structural variability

3.3.1.2.1 Backbone Conformations

The backbone torsion angles: α, γ and χ; pseudorotation phase angle (*P*) and pseudotorsion angles: η and θ, and were calculated for all the bulge residues and bases forming the junction region. These are given in **Table 3.3**.

Table 3.3: The mean (standard deviation) torsion and pseudotorsion angles for the junction and bulge residues of all the simulated systems. The distribution of angles during the course of MD can be unimodal Gaussian like or bimodal or multimodal.

PDB ID	ID	res IDs	α	β	χ	phase (*P*)	η	θ
3U5B (3)	1a	1656U	-62.12 (53.3)	77.91 (44.9)	−148.99 (12.2)	17.03 (23.5)	167.24 (15.9)	-21.49 (18.5)
		1657U (bulge)	26.50 (75.2) bimodal	-60.97 (93.8) multimodal	3.88 (75.27) bimodal	108.16 (64.40) bimodal	-80.26 (28.9)	-1.10 (24.7)
		1658G	-36.44 (80.2) multimodal	177.24 (99.3) multimodal	−176.06 (8.2)	-3.36 (14.1)	-94.47 (26.4)	-155.7 (17.9)
		1743U	80.34 (28.5)	68.06 (24.7)	−155.18 (9.2)	20.29 (13.2)	164.47 (7.4)	-163.51 (16.3)
		1744A	148.55 (65.6)	137.93 (61.4)	−174.7 (9.7)	13.81 (16.4)	-179.05 (10.9)	-147.66 (18.6)
	1b	1656U	-75.93 (15.5)	62.05 (11.6)	-152.61 (16.0)	23.17 (39.7)	NA	NA
		1658G	NA	130.90 (69.7) bimodal	−176.21 (8.2)	-2.04 (16.1)	NA	NA
		1743U	-78.55 (23.3)	66.06 (20.1)	−154.55 (9.6)	20.16 (13.0)	163.62 (7.4)	-165.49 (16.1)
		1744A	142.50 (66.0) bimodal	131.18 (62.8) bimodal	−173.77 (9.2)	12.54 (15.0)	179.23 (11.3)	-148.32 (17.3)
3U5D (3)	2a	726G	-74.01 (7.4)	181.73 (8.9)	-142.30 (8.2)	24.89 (9.53)	-140.80 (8.0)	72.11 (15.3)
		727G	102.95 (19.3)	185.76 (8.1)	50.23 (11.7)	31.60 (15.1)	-151.84 (24.54)	-71.42 (10.4)
		742G	-119.34 (64.9) bimodal	106.01 (59.1) bimodal	-151.39 (18.3)	13.23 (15.16)	174.16 (10.19)	163.38 (18.1)
		743C (bulge)	-72.71 (8.5)	62.37 (8.40)	-146.59 (17.4)	51.68 (21.2)	183.52 (5.23)	-145.51 (7.9)
		744A	-71.23 (13.9)	67.61 (12.54)	-108.73 (12.2)	178.02 (15.9)	-163.22 (48.13)	-55.77 (12.6)

	2b	726G	-76.18 (15.8)	59.11 (13.4)	−154.05 (9.7)	6.71 (11.0)	160.36 (8.8)	176.72 (13.7)
		727G	-82.18 (34.6)	79.30 (39.0)	−163.72 (11.1)	12.25 (16.9)	-177.89 (13.1)	-140.25 (17.4)
		742G	107.94 (33.3)	60.79 (25.3)	48.76 (10.9)	141.08 (18.2)	NA	NA
		744A	NA	105.55 (67.3) bimodal	−168.72 (12.5)	9.41 (17.4)	NA	NA
1VQO (9)	**3a**	3077A	112.89 (41.8)	77.01 (20.0)	-134.94 (20.4)	173.49 (13.1)	55.45 (11.1)	159.42 (9.0)
		3078G (bulge)	58.69 (33.4)	-87.86 (45.6) multimodal	-51.30 (19.3)	159.80 (24.6)	-96.35 (11.9)	71.57 (17.3)
		3079U	151.36 (69.9) multimodal	110.89 (59.5) multimodal	−158.63 (9.3)	8.91 (12.9)	132.89 (13.4)	-150.60 (12.9)
		3103A	-118.68 (34.4)	-176.31 (11.8)	−168.81 (12.9)	-20.76 (65.0) multimodal	-140.62 (27.4)	-174.01 (24.9)
		3104A	-71.74 (13.5)	60.12 (11.9)	-137.52 (48.4) bimodal	151.92 (46.1)	157.94 (17.4)	-147.82 (12.9)
	3b	3077A	142.17 (49.1) bimodal	62.67 (12.1)	-127.66 (14.4)	143.10 (109.4) bimodal	NA	NA
		3079U	NA	108.94 (66.0) bimodal	-9.25 (79.5) multimodal	76.34 (79.8) multimodal	NA	NA
		3103A	-91.59 (40.8)	104.32 (56.1) bimodal	-144.21 (44.6)	22.57 (22.6)	-167.14 (16.0)	166.99 (26.6)
		3104A	-71.65 (11.9)	71.56 (11.6)	-103.47 (21.2)	169.35 (101.7) bimodal	-177.81 (8.8)	-120.65 (19.0)
3U5D (2)	**4a**	502U	-80.88 (32.7)	71.46 (28.9)	-153.77 (20.2)	17.42 (15.8)	170.82 (9.2)	-142.91 (11.6)
		503C	-79.01 (22.5)	64.50 (20.0)	-146.08 (15.4)	17.59 (12.7)	161.84 (9.6)	-152.70 (11.7)
		588G	-73.63 (12.7)	56.88 (9.3)	-114.65 (14.1)	129.24 (61.4) bimodal	-177.98 (9.9)	104.79 (13.6)
		611A	110.22 (67.2) multimodal	-178.96 (45.3)	-97.07 (49.4) bimodal	100.89 (48.5) bimodal	148.49 (57.2) multimodal	126.31 (20.3)
	4b	502U	-81.77 (33.8)	70.23 (28.8)	-150.85 (14.5)	17.72 (15.2)	164.88 (9.2)	-150.2 (16.6)
		503C	-101.71 (53.9) bimodal	90.96 (52.2) bimodal	-151.84 (14.0)	22.22 (15.8)	171.04 (12.4)	-162.07 (20.4)
		588G	-92.57 (47.5)	62.04 (37.7)	-117.73 (29.3)	81.41 (75.8)	NA	NA
		611A	NA	136.64 (64.9) bimodal	-44.07 (26.3)	150.95 (19.8)	NA	NA

2ZJR (2)	**5a**	2653A	-85.68 (43.8)	73.68 (39.0)	-161.85 (9.7)	6.59 (14.4)	177.27 (11.2)	-120.91 (21.9)
		2654A	-110.72 (57.5) bimodal	171.37 (41.7)	-169.55 (10.5)	29.57 (18.1)	-163.92 (15.7)	172.39 (29.5)
		2625U	-58.95 (77.6) multimodal	82.54 (60.0) bimodal	-165.29 (8.9)	5.52 (17.5)	145.97 (42.5)	-130.52 (19.4)
		2712G	-71.29 (11.1)	26.45 (13.7)	-78.37 (10.3)	-132.91 (18.8)	-143.35 (6.5)	85.69 (11.4)
	5b	2653A	-83.13 (39.7)	77.83 (38.2)	-129.15 (15.9)	173.78 (73.8) bimodal	-156.82 (14.7)	-49.96 (64.4) multimodal
		2654A	173.75 (88.6) bimodal	126.79 (72.6) bimodal	-60.64 (15.6)	-38.30 (46.2)	117.76 (39.3)	-172.43 (32.3)
		2625U	NA	117.94 (66.2) bimodal	138.55 (79.9) bimodal	60.53 (71.6) bimodal	NA	NA
		2712G	-65.41 (44.7)	64.06 (50.7)	-57.43 (28.3)	-54.76 (76.5) multimodal	NA	NA
3DIO (3)	**6**	114G	132.91 (86.1) multimodal	175.43 (50.9) bimodal	-168.93 (11.2)	10.45 (17.0)	174.79 (36.5)	-168.84 (24.5)
		125G	-86.87 (43.0)	74.40 (39.9)	-154.39 (13.3)	21.48 (66.7) bimodal	175.56 (19.7)	-147.92 (45.8) multimodal
		126A	59.02 (61.7) bimodal	149.75 (66.5) bimodal	-32.38 (60.1) bimodal	156.09 (48.7) broad	54.74 (66.6) multimodal	-95.74 (33.2)
		127A	-77.12 (33.6)	53.65 (17.9)	-69.10 (20.3)	-167.00 (94.1) bimodal	145.84 (17.9)	-75.95 (39.4)
		128U	-21.72 (78.6) bimodal	132.10 (76.5) multimodal	-13.26 (82.4) multimodal	100.65 (77.8) multimodal	-95.64 (54.6) bimodal	33.43 (53.7) bimodal
		129A	-32.93 (78.6) multimodal	99.96 (58.6) bimodal	-152.88 (25.9)	4.39 (17.8)	174.22 (41.5) multimodal	-93.57 (24.6)
		140C	-85.16 (38.9)	69.99 (35.9)	-139.71 (17.9)	60.53 (60.0) bimodal	167.22 (10.1)	-161.39 (21.4)
		141G	-74.69 (11.8)	61.32 (15.0)	-111.15 (20.6)	157.25 (51.2) multimodal	172.99 (12.2)	-110.54 (20.1)
		162A	-73.91 (22.3)	59.68 (19.6)	-150.71 (19.7)	14.08 (13.0)	164.48 (8.6)	-175.5 (16.3)

3.3.1.2.1.1 Bulge Systems

The residues at the junction of systems 1a and 1b maintain the canonical g^-/g^+ conformation for the α/γ torsion angles. Some residues in System 1a adopt the non-canonical g+/g-, t/g-, and t/g+ conformations, for a short time during the MD-run whereas only the canonical g^-/g^+ conformation is found in System 1b.The α torsion angle for the bulge residue Uracil (1657U) is found to fluctuate between g- and g+ conformations during the MD-run. The corresponding γ torsion angle also adopts unusual g-, and t conformations, along with the usual g+ conformation. This accounts for the high standard deviations in the α/γ torsion angles of the bulge residue. The glycosidic torsion angle χ for the same base retained the less stable *syn* conformation for a long period of time and the *anti* conformation for a short period, indicating its greater flexibility due to extra-helical conformation. Also, the corresponding sugar pucker was not in the usual C3'-endo conformation but was fluctuating between the C2'-endo and C3'-endo/C4'-exo states, as evident from the two peaks in the radial plot for the pseudorotation phase angle near 150° and 45° respectively **(Figure 3.12a)**. However, in System 2a, the bulged-in Cytosine (743C) residue maintained a stable g^-/g^+ conformation for the α/γ torsion angles, and a stable *anti* conformation of the χ torsion angle throughout the MD-run, indicating its greater rigidity due to less freedom of movement resulting from intercalation between two base pairs at the junction. The corresponding sugar pucker adopted a stable C3'-endo/C4'-exo conformation for a sufficiently long time during the simulation **(Figure 3.12b)**. The bases at the junction adopted canonical g^-/g^+ as well as unusual g^-/g^- conformations in presence of the bulge (System 2a) but adopted only the stable g^-/g^+ conformation in its absence (System 2b). This may be corroborated with greater structural stability of the System 2b duplex in absence of the bulged-in residue as was seen from RMSD. In System 3a, the bulge Guanine (3078G) residue adopted the unusual g+/g- conformation for the α/γ torsion angles and the unusual *syn* conformation for the χ torsion angle along with the corresponding sugar pucker in the C2'-endo region for the entire simulation time **(Figure 3.12c)**. The Uracil residue at the junction, forming a dinucleotide platform with the bugled residue, adopted a stable *anti* conformation for the χ angle, however, in System 3b, in absence of the bulge residue, a χ-switching from *anti* to *syn* was observed for the same. The 3079U:3103A base pair involved in base triple formation with the bulge residue thus seems to be more stabilized in

presence of the bulge residue as compared to that in its absence (System 3b), where high standard deviations account for the multiple conformations adopted by the α/γ torsion angles, glycosidic torsion angle and pseudorotation phase angle, indicating a rather more fluctuating structure.

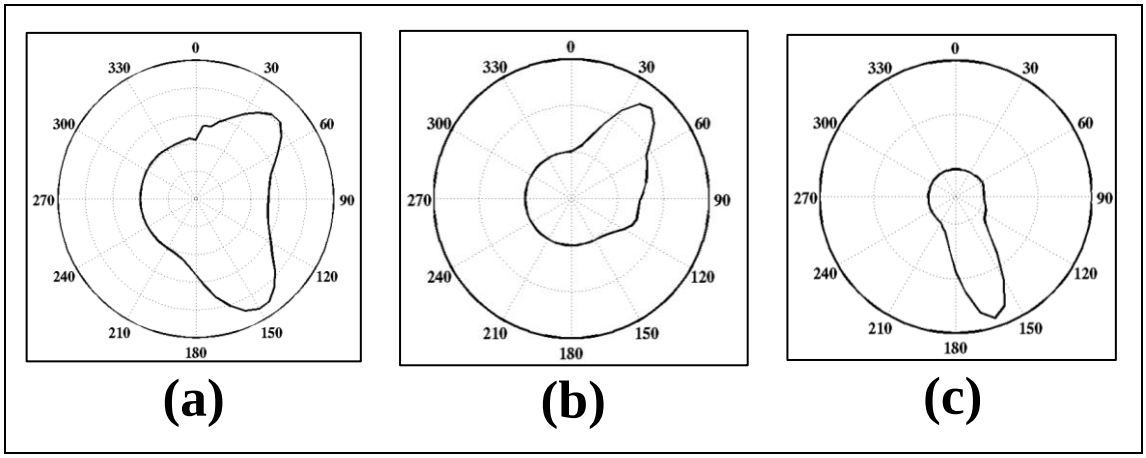


Figure 3.12: The sugar puckering for (a) System 1a (3U5B(3)), (b) System 2a (3U5D(3)), and (c) System 3a (1VQO(9)).

The pseudotorsion angles η and θ calculated for the bulge containing systems provided an idea regarding the variety of conformations adopted by the RNA backbone. The bulged-out 1657U residue of System 1a adopted a wide range of η, θ with average values near -80° and -1°, respectively. The η, θ values in System 1b are near *trans* regions for all the residues including the junction region. The values of the pseudotorsion angles for the bulged-in 743C residue of System 2a adopted a regular conformation close to the normal helical region with the η and θ values clustered around 170° and -135°, respectively **(Figure 3.13)** (Duarte and Pyle, 1998). However, the 727G residue, whose H-bonding with 742G was disrupted during MD simulation show average θ value of -71° in the second equilibrium state. As indicated earlier, this Guanine base forms H-bond with phosphate group of 743C after 400ns. The 744A also show unusual θ value along with C2'-endo puckered sugar to accommodate this structural alteration. The bulge residue 3078G in System 4a shows unusual η and θ values around -96° and 72°, respectively, along with C2'-endo sugar. These abnormal values were to accommodate the strain. All the residues in System 4b show usual helical signatures.

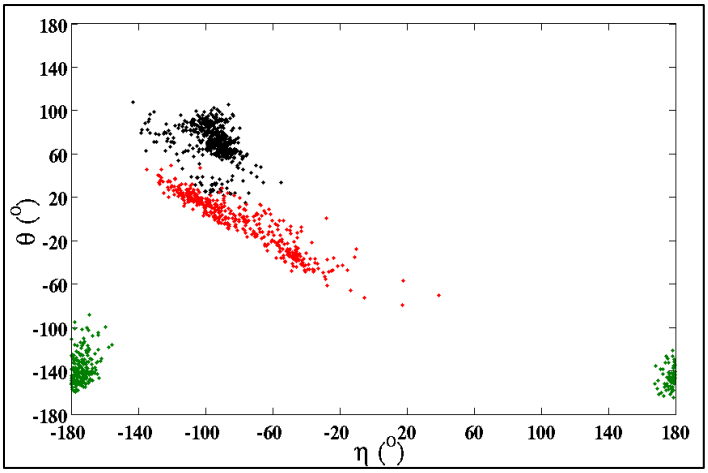


Figure 3.13: The pseudotorsion angle (η vs. θ) plot for (a) Bulge containing RNA duplexes (System 1a (3U5B(3))-Red, System 2a (3U5D(3))-Green and System 3a (1VQO(9))-Black).

3.3.1.2.1.2 Pseudo-continuous Helical Systems

The α/γ torsion angles lied within the allowed regions of the conformational space with canonical g^-/g^+ conformation prevalent for most of the bases at the junction of systems 4a and 4b. The Adenine residue (611A) at the junction of System 4a showed multimodal and bimodal distributions for the α and γ torsion angles respectively. The corresponding angle was seen to be fluctuating between the *syn* and *anti* conformations, whereas the same residue maintained a constant *syn* orientation in System 4b. The 611A residue adopted a variable sugar pucker in System 4a but in System 4b, it maintained the C2'-endo conformation throughout its MD-run. In systems 5a and 5b, the g^-/g^+ conformation of the α/γ torsion angles was prevalent for most bases at the junction, however, t/g^+, g^+/t, g^+/g^-, and g-/t conformations were also adopted as evident from the high standard deviations for the 2654A and 2625U residues. All the residues at the junction of systems 5a and 5b adopted the favourable *anti* conformation of the χ torsion angle except for the 2654A residue which adopted *syn* conformation. Moreover, in System 5b, the 2626U was seen to be fluctuating between both *syn* and *anti* conformations. The pseudorotation phase angles for all the bases showed a bimodal or multimodal distribution, indicating higher structural variability. In the pseudohelix System 6 capped at one end by a tri-loop, the residues forming the loop and the base pair connecting these adopted variable α/γ torsion angles, whereas residues at the junction adopted the usual g^-/g^+ conformation. The junction residues adopted the favourable *anti* conformation, whereby the loop

residues were seen to be fluctuating between the *syn* and *anti* conformations. The sugar puckers fluctuated between variable conformations for the terminal residues, as evident from their high standard deviations as these bases were subjected to terminal fraying.

The residues at junction of systems 4a and 4b show helical signatures in terms of η and θ values. The strain for accommodation of junction was probably absorbed by the sugar-puckers. The 2712G residue in System 5a adopted unusual θ values along with sugar pucker variability to accommodate the junction. On the other hand the 2653A residue of the continuous strand in System 5b shows unusual values of η and θ. The junction residues of System 6 adopts the η and θ values which mostly fall in the defined helical region, (η=175°, θ=-135°) (Duarte and Pyle, 1998), whereas in the other pseudohelical systems (systems 4a and 5a) these junction residues adopt variable η and θ values.

3.3.1.2.2 Groove Width

It has earlier been reported that the occurrence of bulges in RNA double helix can impart variations in the groove width, especially major groove leading to formation of potential interaction sites for specific RNA–RNA and RNA–protein recognition (Hermann and Patel, 1999; Hermann and Westhof, 1999; Woese and Gutell, 1989). Widening of the major groove due to presence of bulge is frequently observed at such interaction sites (Hermann and Westhof, 1999).

The width of major groove of a regular fiber model A-RNA is reported to be 9.97Å (Arnott *et al.*, 1972), which is a measure of shortest distances between the phosphate atoms of two strands across the major groove. The groove width for the initial bulge containing and representative pseudohelical RNA systems is given in **Table 3.4**. The bulged-out (System 1a) residue containing and base-triple forming (System 3a) RNAs were found to have groove width above 14Å in their initial crystal structures. However, the initial crystal structure of System 2a, exhibited a far shorter groove width of ~11Å, probably because of its short-length and insufficient number of phosphate-phosphate distances that can determine correctly the groove width.

From the simulations we observed that the groove widths are not fixed and may vary, whereby, widening of the major groove width is seen. Also, we found that

the groove width at the extremes all the RNA structures (System 1a, 1b, 2a, 2b, 3a and 3b) is increased during the course of simulation, probably due to terminal fraying effect and is independent of presence or absence of bulge residues. The average groove width for systems 1b, 2b and 3b were found to be similar to that of the standard A-RNA. However, the groove width of the central region surrounding the junction differs in presence of the unpaired bulge residue.

Table 3.4: The mean (and standard deviation) groove width of major groove of the RNA stretches obtained from simulation.

System ID	System 1a	System 2a	System 3a	System 4a	System 5a	System 6
PDB ID (Frag ID)	3U5B(3)	3U5D(3)	1VQO(9)	3U5D(2)	2ZJR(2)	3DIO(2)
groove width (Å)	14.43	10.85	14.78	11.76	12.94	9.84
in terms of PP dist	(3.5)	(1.0)	(5.2)	(2.3)	(3.3)	(1.6)

3.3.1.2.3 Base Pair/Base Pair Step Parameters and Helix-axis Bending

The two base pairs surrounding the bulge residue forms the junction region in case of bulge motifs and for the pseudohelices the base pairs at the discontinuous region is the junction. The dinucleotide step (base pair step) and base pair parameters for these junction regions were calculated **(Tables 3.5 and 3.6)**. Moreover, the kink or bend induced in the helices due to bulges has also been estimated and correlated with the structural parameters.

The angles between axes of the two helical fragments and overlap between the base pairs at the junction across the bulge (in case of bulge containing RNAs) and the pseudo-continuous region were analyzed for all the simulated systems including the respective counter-models **(Figure 3.14 and 3.15)**.

Table 3.5: Dinucleotide step parameters of the base pair steps at the junction. "^" represents a bulge residue for the bulge containing RNAs and a discontinuity in strand for the pseudo-continuous helices.

System ID	PDB ID (Frag ID)	Dinucleotide step with residue IDs	tilt	roll	twist	shift	slide	rise	overlap	Base Pairing
1a	3U5B(3)	1656U^1658G/ 1743U1744A	-0.18 (4.4)	6.11 (7.3)	11.03 (6.8)	1.97 (0.6)	-3.95 (0.6)	3.52 (0.3)	32.76 (5.8)	(W:W C)/ (W:W C)
1b		1656U1658G/ 1743U1744A	-0.94 (4.6)	6.44 (7.5)	9.78 (6.7)	2.13 (0.5)	-3.99 (0.6)	3.58 (0.3)	32.08 (5.8)	
2a	3U5D(3)	726G727G/ 742G^744A	13.57 (28.79)	23.18 (13.63)	22.04 (82.43)	0.33 (2.66)	1.39 (2.90)	7.10 (4.94)	23.24 (5.36)	(S:H T)/ (H:S T)
2b		726G727G/ 742G744A	12.04 (10.78)	1.32 (9.55)	80.42 (61.43)	1.75 (1.09)	-0.50 (1.31)	2.62 (0.55)	44.54 (8.07)	
3a	1VQO(9)	3077G^3079U/ 3103A3104A	0.39 (6.8)	-6.65 (7.2)	5.10 (7.9)	1.00 (1.3)	2.71 (0.9)	3.85 (0.3)	34.78 (12.4)	(H:H T)/ (W:H T)
3b		3077G3079U/ 3103A3104A	-1.57 (25.4)	4.47 (19.1)	27.25 (36.8)	0.34 (2.2)	4.59 (2.6)	4.35 (5.8)	33.03 (5.4)	
4a	3U5D(2)	502U503C/ 588G^611A	4.70 (3.7)	3.07 (6.9)	40.61 (7.9)	1.27 (0.8)	-0.19 (1.0)	3.43 (0.3)	38.84 (6.9)	(W:W C)/ (W:W C)
4b			3.27 (5.2)	5.10 (7.1)	24.23 (9.9)	-0.13 (0.8)	-1.76 (0.9)	3.50 (0.4)	40.23 (8.8)	
5a	2ZJR(2)	2653A2654A/ 2712G^2625U	5.86 (4.1)	6.25 (6.9)	8.40 (6.7)	-0.38 (0.7)	-2.63 (0.6)	3.35 (0.3)	44.05 (8.3)	(W:W C)/ (W:W C)
5b			4.19 (9.8)	12.44 (13.2)	34.17 (23.3)	0.55 (1.7)	3.76 (2.3)	3.29 (5.0)	29.93 (7.5)	
6	3DIO(2)	140U141G/ 162A^114G	-38.04 (16.3)	-40.39 (15.5)	68.13 (16.9)	2.79 (1.5)	-2.84 (1.4)	6.02 (0.6)	27.49 (6.6)	(W:W C)/ (W:W C)

Table 3.6: Base pair parameters for the two base pairs at the junction of the RNAs.

System ID	PDB ID (Frag ID)	Base pairs with res IDs	buckle	open	propeller	stagger	shear	stretch	Base Pairing
1a	3U5B(3)	1656U:1744A	10.98 (11.0)	2.89 (6.2)	1.76 (9.3)	-0.20 (0.5)	-0.21 (0.4)	2.88 (0.2)	W:W C
1b			13.13 (9.8)	4.02 (6.0)	-1.00 (8.1)	-0.38 (0.4)	-0.20 (0.3)	2.87 (0.1)	
1a		1658G:1743U	-8.33 (12.0)	-3.54 (6.2)	-7.10 (8.3)	-0.23 (0.5)	-2.50 (0.3)	2.87 (0.1)	W:W C
1b			-9.35 (12.1)	-3.61 (7.0)	-6.15 (8.4)	-0.30 (0.6)	-2.48 (0.3)	2.85 (0.2)	
2a	3U5D(3)	726G:744A	-4.54 (12.33)	13.89 (3.91)	-11.53 (8.07)	0.23 (0.43)	2.23 (0.22)	3.39 (0.16)	S:H T
2b			-3.18 (11.05)	-9.66 (8.15)	3.49 (9.23	-0.07 (0.45)	2.35 (0.33)	3.35 (0.19)	
2a		727G:742G	102.64 (41.8)	38.59 (11.5)	-25.27 (23.4)	9.05 (2.9)	4.11 (0.9)	1.13 (0.9)	H:S T
2b			8.87 (10.13)	-80.36 (16.76)	22.59 (9.88)	-0.03 (0.54)	6.05 (0.92)	0.71 (0.81)	
3a	1VQO(9)	3077A:3104A	-15.68 (13.2)	1.96 (5.5)	6.32 (9.0)	-0.06 (0.5)	2.32 (0.3)	2.83 (0.2)	H:H T
3b			-36.27 (10.9)	2.12 (4.5)	-9.00 (9.6)	-0.05 (0.4)	2.45 (0.3)	2.85 (0.1)	
3a		3079U:3103A	-34.61 (11.4)	11.11 (7.7)	-9.09 (8.7)	-0.17 (0.7)	0.12 (0.3)	2.79 (0.2)	W:H T
3b			-34.57 (112.0)	-29.11 (78.8)	0.50 (58.8)	0.10 (4.2)	0.87 (2.4)	1.31 (6.0)	
3a		*3078G:3079U	-9.40 (16.10)	-4.07 (6.43)	-3.99 (9.18)	-0.63 (0.52)	0.82 (0.52)	3.38 (0.28)	s:h C
4a	3U5D(2)	502U:611A	-10.85 (12.5)	7.74 (10.1)	-19.17 (9.7)	-0.21 (0.5)	-0.21 (0.3)	2.90 (0.2)	W:W C
4b			-5.46 (14.4)	7.01 (9.3)	-10.25 (10.0)	-0.44 (0.9)	-0.10 (0.4)	2.64 (0.9)	
4a		503C:588G	-10.61 (13.5)	-1.51 (3.3)	0.75 (8.3)	-0.12 (0.4)	0.02 (0.3)	2.93 (0.1)	W:W C
4b			1.42 (15.0)	-1.29 (4.5)	-3.31 (10.6)	-0.13 (0.4)	0.00 (0.3)	2.92 (0.1)	
5a	2ZJR(2)	2653A:2625U	-11.12 (11.10)	-5.48 (7.76)	-13.07 (10.52)	-0.16 (0.48)	0.22 (0.31)	3.07 (0.26)	W:W C
5b			1.86 (74.63)	-3.35 (34.02)	7.98 (55.09)	-0.68 (1.69)	0.22 (1.21)	0.95 (3.27)	
5a		2654A:2712G	-8.06 (9.27)	9.39 (10.31)	-25.38 (7.65)	-1.13 (0.36)	0.00 (0.34)	2.73 (0.30)	W:W C
5b			16.26 (14.44)	6.27 (16.09)	27.56 (13.13)	1.17 (0.70)	0.55 (1.20)	2.55 (0.77)	
6	3DIO(2)	140U:114G	9.13 (11.34)	2.55 (13.81)	-8.59 (10.15)	-0.14 (0.51)	2.51 (0.40)	2.94 (0.29)	W:W C
		141G:162A	24.73 (9.28)	-8.92 (12.62)	23.36 (15.03)	-6.50 (0.48)	-4.30 (1.42)	-1.46 (1.14)	W:W C

**in System 3a the base pair is involved in forming dinucleotide platform thus, the parameters are given.*

3.3.1.2.3.1 Bulge Systems

The average helix-axis bending angle for systems 1a and 1b were calculated to be 17.15° (8.3) and 16.15°(8.3) respectively **(Figure 3.14a)**. A good average overlap of ~32Å^2 for the 1656U^1658G/1743U1744A (*i.e.* the step involving 1656U:1744A and 1658G:1743U base pairs with additional bulge residue 1657U, "^" indicates presence of a bulge residue) and 1656U1658G/1743U1744A steps, respectively **(Figure 3.14b)** indicated that no additional kinks were introduced into the helix at the junction region after removal of the bulged-out residue. The base pair parameters along with tilt and roll step parameters at the junction **(Table 3.5 and 3.6)** of systems 1a and 1b are similar to that of A-form RNAs reported earlier. However, the average twist, shift and slide values of this dinucleotide step are unusual. Unusual twist values are possible for G:U base pair containing dinucleotides, but it was shown earlier that a larger twist value is expected for the above UG/UA sequence (Halder and Bhattacharyya, 2010; Mondal *et al.*, 2015; Pingali *et al.*, 2014). The simulation results, however, show small twist around 11°, possibly to accommodate the discontinuity, thereby simultaneously resulting in large shift and slide. Thus, the removal of bugle residue (System 1b) did not have any strong impact in alteration of the overall structure of the double-helix as the residue was already protruding out from the helix and did not involve into any stacking interactions with the base pairs at the junction.

The helix-axis bending angle for System 2a was found to be very high (~50°) whereas for its counter-model System 2b the bending of ~15° was much lower **(Figure 3.14c)**. Presence of an unpaired base intercalated between two base pairs at the junction leads to the high kink in System 2a. A similar trend is also reflected from the overlap of junction base pair steps **(Figure 3.14d)**. The 726G727G/742G^744A step containing the 743C bulge shows poor average overlap of 20.82Å^2 (6.1) in System 2a, whereas in its counter-model System 2b, the average overlap is increased to 42.28Å^2(9.0) for the somewhat continuous 726G727G/742G744A step **(Table 3.5)**. As mentioned earlier in System 2a the 727G:742G H:S T base pair was disrupted during MD run after 400ns. Hence, the overlap and base pair step parameters may not give true indication of (G:A S:H T)/(G:G H:S T) stacking. The disruption of base pair was also reflected in the base pair parameters **(Table 3.6)**. When the bulge residue is removed in System 2b, the helical system behaves more or less like A-form RNA,

except with large twist for stacking between two non WC base pairs **(Table 3.5)**. The G:G H:S T base pair, however, maintained planarity but showed some unusual values for open, shear and stretch, indicating disruption of H-bonds between the two bases. These indicate the role of bulged-in residue in changing the base pair geometry near the bulge and in turn disrupting the overall geometry of the duplex.

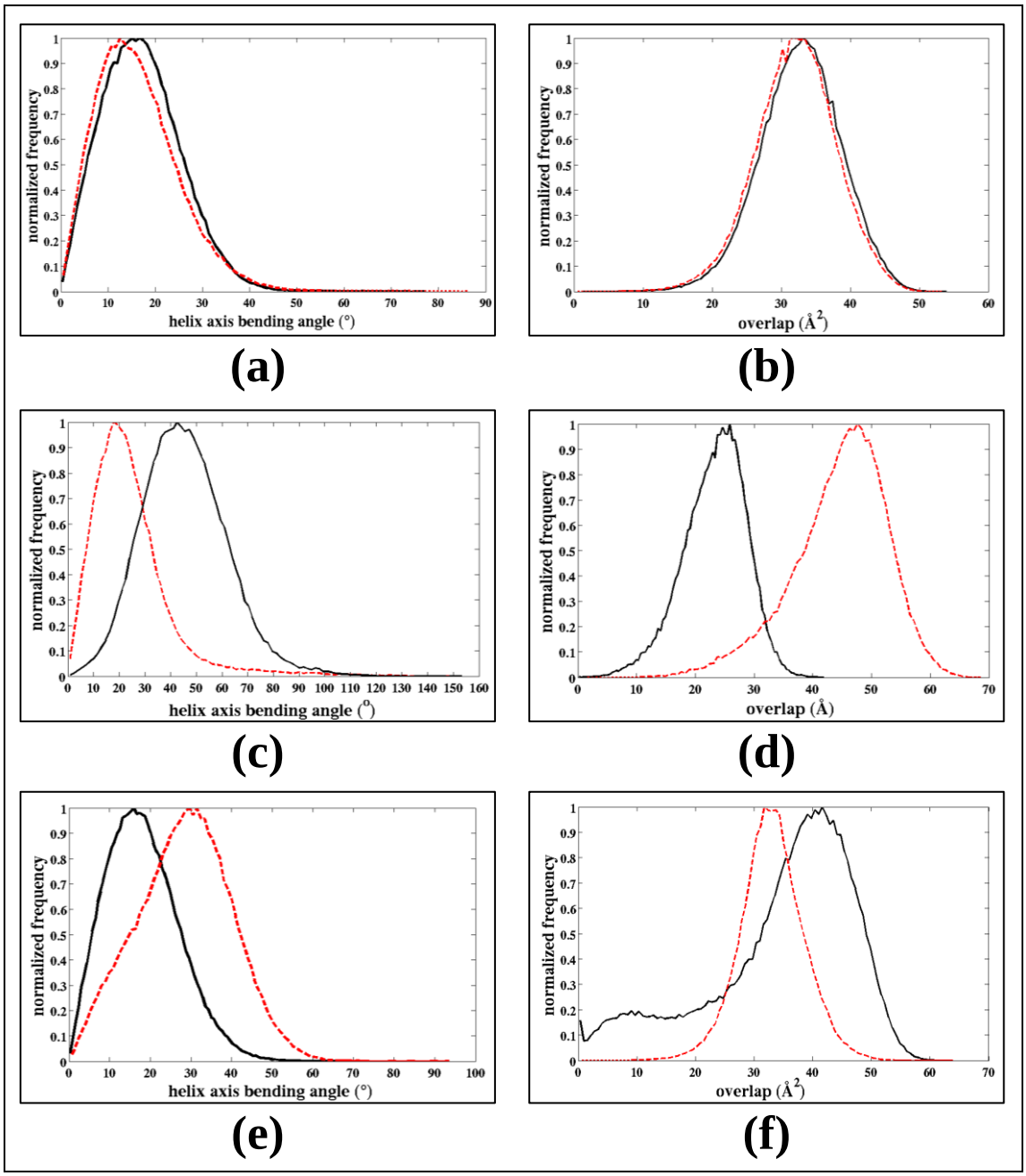


Figure 3.14: Histogram for helix axis bending angle of (a) systems 1a and 1b (c) systems 2a and 2b, and (e) systems 3a and 3b. Histogram for overlap of base pairs at the junction of (b) systems 1a and 1b, (d) systems 2a and 2b, and (f) systems 3a and 3b. Black line represent systems 1a, 2a and 3a; and Red lines represent systems 1b, 2b and 3b.

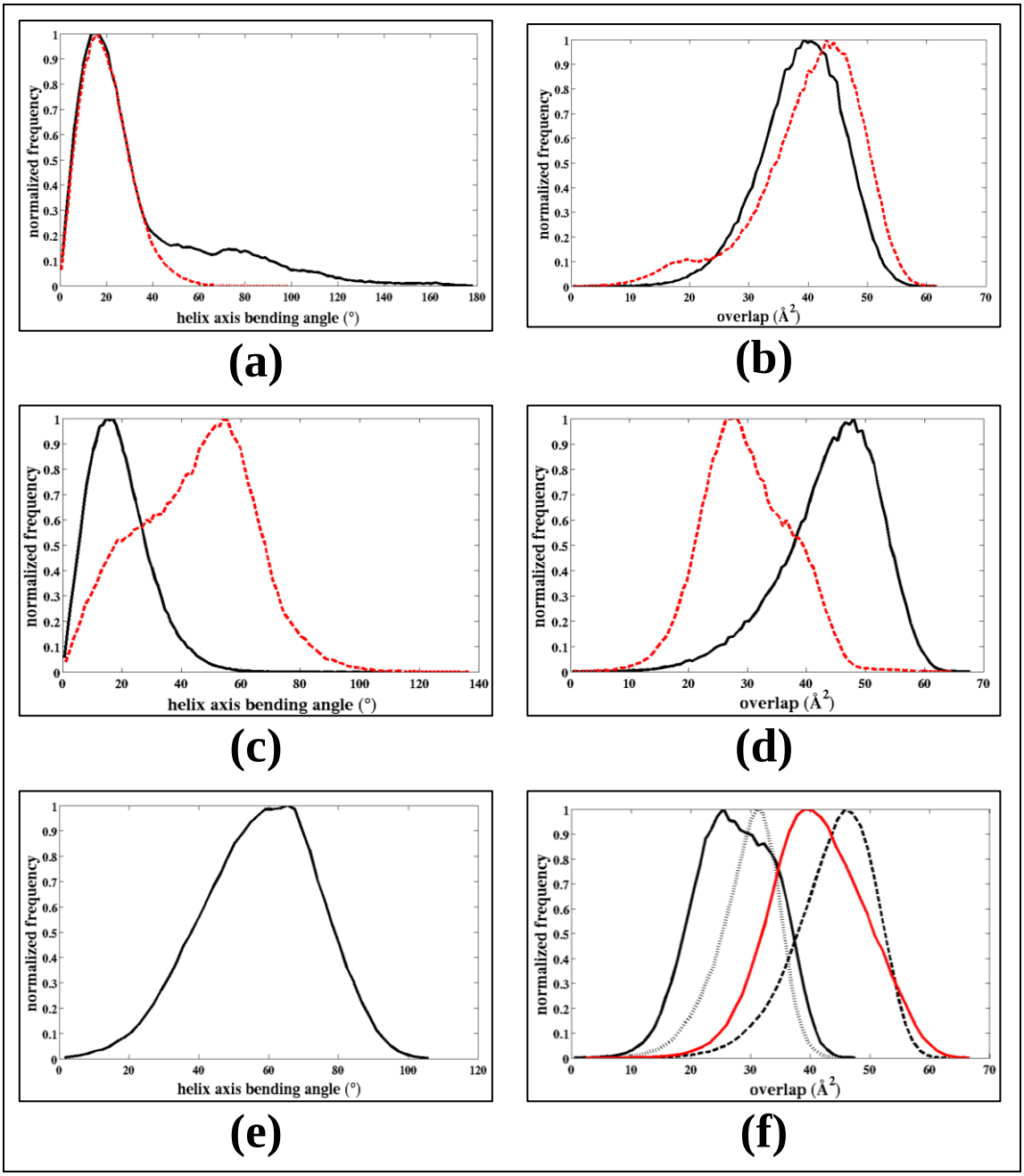


Figure 3.15: Frequency Histograms of : (a) helix axis bending angle and (b) overlap at the junction for System 4a (Black) and System 4b (Red dotted); (c) helix axis bending angle and (d) overlap at the junction for System 5a (Black) and System 5b (Red dotted); and (e) helix axis bending angle for System 7 and (f) overlap for 140U141G/162A^114G base pair step at the junction (black); the base pair steps above and below the junction, *i.e.* 139C140U/114G115G (black - - -) and 141G142U/162A161G (black dotted); and 129A130C/124G125G closing base pair step below the tri-loop region.

Similarly, for System 3a, where the bulge residue is involved in base-triple and dinucleotide platform formation, low helix-axis bending angle and high overlap of the base pair step at junction was observed **(Figure 3.14e and 3.14f)**. The overlap values, however, show nearly bimodal distribution, a change to near zero overlap is

noted at 350 to 450ns, when slide also becomes too large **(Figure 3.16)**. The average twist value at this step, however, is found to be 5° with the other parameters being in usual range of double helical RNA which may be due to base triple formation. The expected value of twist and other parameters for stacking between non WC base pairs, however, is poorly understood and hence it is impossible to attribute the reason behind such low twist. The base pair parameters of the non-canonical base pairs (A:A H:H T, U:A W:H T and G:U s:h C) in systems 3a and 3b are all indicative of strong and planar nature **(Table 3.6)**. However, when the bugle residue was absent in case of System 3b, the same overlap decreased due to improper stacking of base pair at the junction thereby increasing the helix-axis bending angle of the RNA introducing a kink in the helix. Thus, System 3b deviated strongly from standard RNA stretch in absence of the well H-bonded bulge residue. The removal of bulge residue (3078G) of System 3a resulting in System 3b affected the latter's base pair and base pair step parameters significantly which is reflected from the high standard deviations of buckle and open parameters corresponding to base pair 3079U:3103A of System 3b. It is to be mentioned here that although the RMSD plot for System 4b indicated 120ns as the equilibration time, this base pair at the junction required more time to stabilized **(Figure 3.17)**, thus there remains a scope for further extending this simulation. We however did not extend this as it represents only a counter model to System 4a.

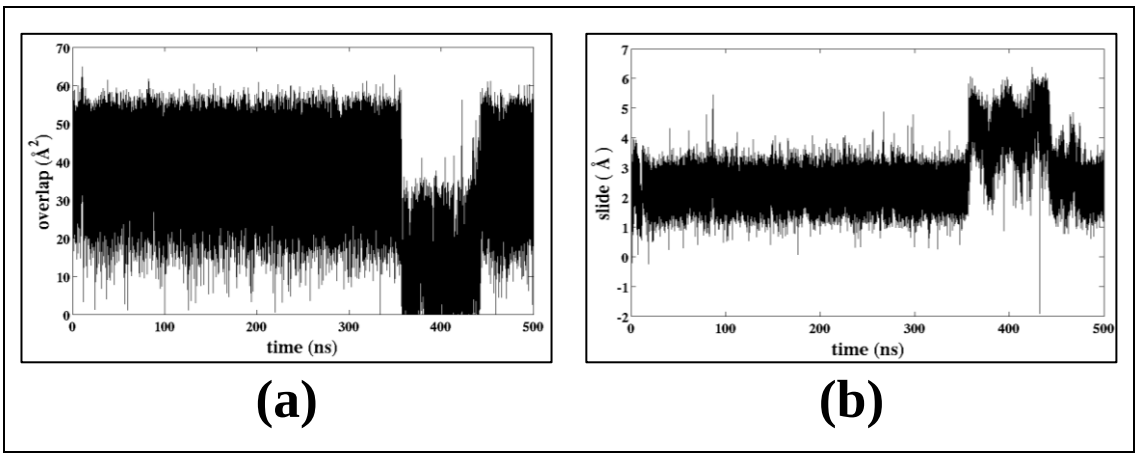


Figure 3.16: time evolution plot for (a) overlap and (b) slide plots for base pair step 3077G^3079U/3103A3104A at the junction of System 3a.

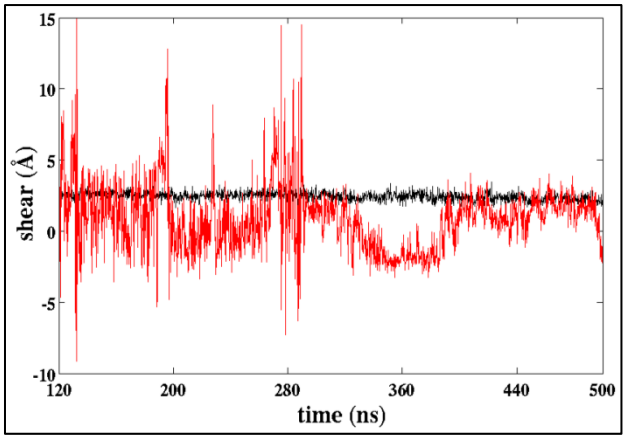


Figure 3.17: Shear vs. time plot for 3079U:3103A base pair at the junction of systems 3a (black) and 3b (red).

3.3.1.2.3.2 Pseudo-continuous Helical Systems

The base pair and base pair step parameters from the simulations of pseudo-continuous helical fragments are given in **Tables 3.5 and 3.6**. The average helix-axis bending angle for both the systems 4a and 4b are found to be around 15°. Moreover, high overlaps of 38.9Å^2 and 40.23Å^2 for base pair steps at the junctions of systems 4a and 4b respectively **(Figure 3.15)**. The average base pair step parameters such as large positive twist, positive roll, and small negative slide for the 502U503C/588G^611A step **(Table 3.5)** and base pair parameters such as small buckle, negative propeller twist and stretch around 3Å for the base pairs at the junction accompanied by small standard deviations for both systems 4a and 4b **(Table 3.6)**, corroborates with those reported earlier from the crystal structures. These findings establish a similarity in the base pair geometries attained by standard A-form double helix and the pseudo-continuous stack motif, in turn exhibiting similarities between the overall duplex structures. Although, the helix-axis bending angle for System 5b is calculated to be very high around 60°, the same for System 5a is ~15°. Again there exists lower overlap of 29.9Å^2 for the base pair step at the junction of System 5b as compared to higher value of 44.1Å^2 in System 5a. There exists an interface of non-canonical (A:G W:W C) and canonical (A:U W:W C) pairing at the junction of systems 5a and 5b. The other base-pair step parameters are also found to be different for systems 5a and 5b, as evident from roll value of 6.25° and negative slide of -2.63Å for 2653A2654A/2712G^2625U step of System 5a, "^" indicates presence of a pseudohelix region or discontinuity in the stand, as compared to roll and

slide values of 12.44° and 3.76Å for the 2653A2654A/2712G2625U step of systems 5b. The average twist value corresponding to this interface is found to be small (8.40(6.7)) for System 5a whereas in System 5b the value reaches near regular of 34.17° **(Table 3.4)**, indicating that the extra-flanking residues of System 5a make the entire RNA-stretch behave structurally more like a pseudo-continuous helix (different from a continuous A-RNA stretch). Moreover, unusual stretch of non-canonical 2653A:2652U base pair at the junction and large standard deviations for the propeller twist, buckle and open angle seen for junction base pairs of system 5b, are indicative of the unstable nature of the System 5b as compared to that of System 5a **(Figure 3.17)**.

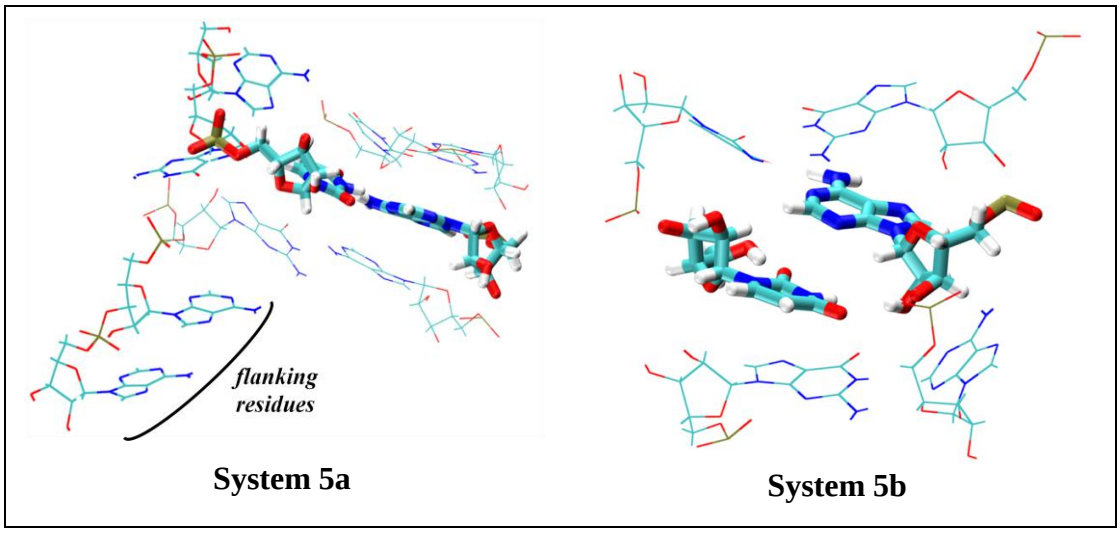


Figure 3.17: The difference in base pair geometry in presence (System 5a) and absence (System 5b) of extra flanking residues at the junction of the pseudohelix.

The pseudo-continuous helix, System 6, contains a three-residue loop motif connecting the terminal base pair of the duplex. The dinucleotide step 140U141G/162A^114G at the junction has an overlap of 27.49Å^2 (6.6) and the helical stretch has an average helix-axis bending angle of 60° **(Figure 3.15e and 3.15f)**. This bending angle is much higher than the other aforementioned pseudohelices. The unusually high rise of 6.02Å (0.6) and twist of 68.13° (16.9) for the junction base pair step is in line with low overlap for the corresponding base pair step. Apart from this, unusually high negative roll and tilt values are observed at the same junction. Moreover, the base pair 141G:162A at junction adopted unusually high buckle of 24.73° (9.28), high positive propeller twist of 23.36° (15.03), high stagger of -6.5Å (0.48) and irregular stretch of -1.46Å (1.14). The base pair stacks present above and

below the junction region has overlap of ~45Å^2 and ~30Å^2 respectively. The stacking interaction for the base pairs present near the loop region was seen to be higher with overlap of 42Å^2, the reason for which could be the presence of the capping by loop connecting the base pairs.

3.3.2 MD Simulation by charmm36 force field

There is almost no theoretical data available regarding structural analysis of RNA bulges. Thus, the bulge containing RNA motifs (systems 1a, 2a and 3a) have also been simulated for 100ns each with charmm36 force field and NAMD software with the sole purpose of comparing the results with those obtained from AMBER simulations and ensure that the reported AMBER results are not simulation artifact.

3.3.2.1 Dynamics and Fluctuation

Similar to AMBER simulations, initial idea about the dynamics and overall behaviour of the trajectories for the charmm simulations were also obtained from the RMSD **(Figure 3.18.1)** and RMSF **(Figure 3.18.2)** plots.

The systems 1a (PDB ID: 3U5B (3)) and 3a (PDB ID: 1VQO (9)) were both well stabilized within 20ns of the respective 100ns MD-run **(Figure 3.18.1a and 3.18.1c)**. The residues at the terminal showed high RMSF due to terminal fraying effect. In System 1a the bulge residue 1657U showed ~3Å fluctuation **(Figure 3.18.2a)** due to its bulged-out conformation and in System 3a the 3078G residue showed too showed slightly higher fluctuation **(Figure 3.18.2b)**. These are in line with the observations of AMBER simulations. Furthermore, In case of System 2a (PDB ID: 3U5D (3)), the increase in RMSD from 2Å to 6Å **(System 3.18.1b)** indicated a major readjustment of conformation, especially after about 70ns of MD-run. As stated earlier this system is only twelve base pairs long and the bulged residue is stacked inside the helix. This system indicated similar major readjustments and breaking of the 727G:742G H:S T base pair during simulations with AMBER-99sb force field. Thus, it is clear that the simulations with both the force fields indicate similar structural features.

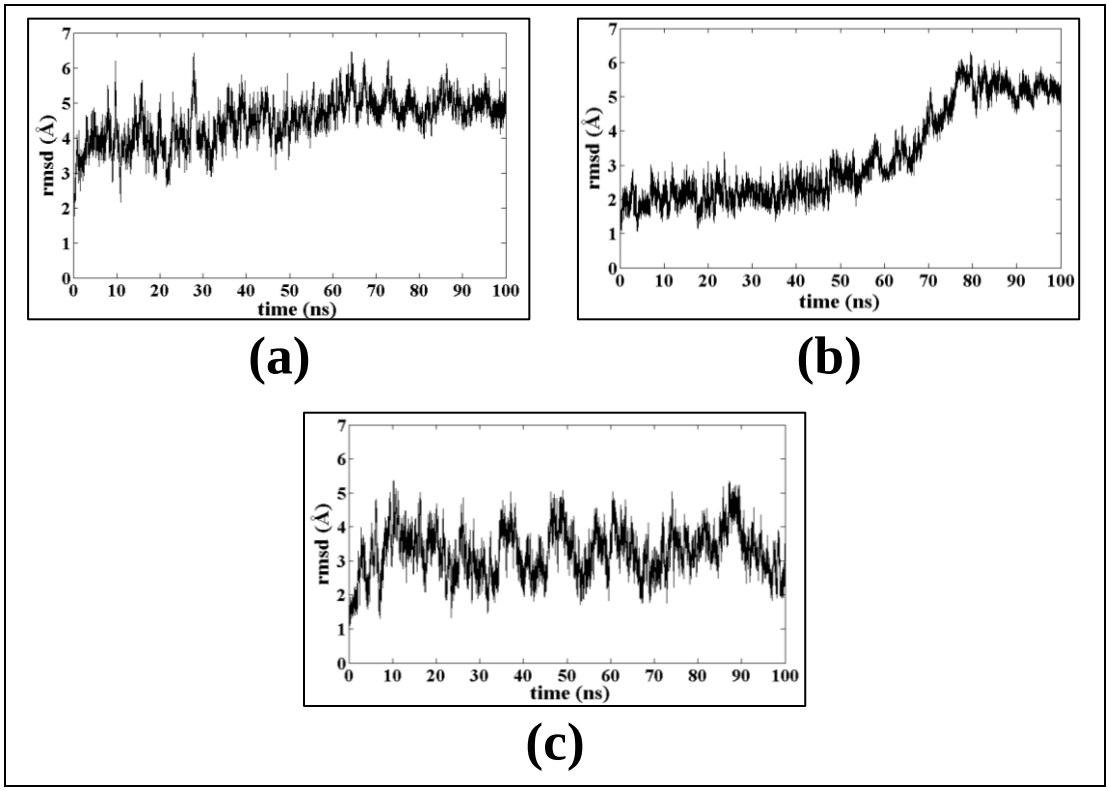


Figure 3.18.1: RMSD plots for (a) System 1a, (b) System 2a and (c) System 3a.

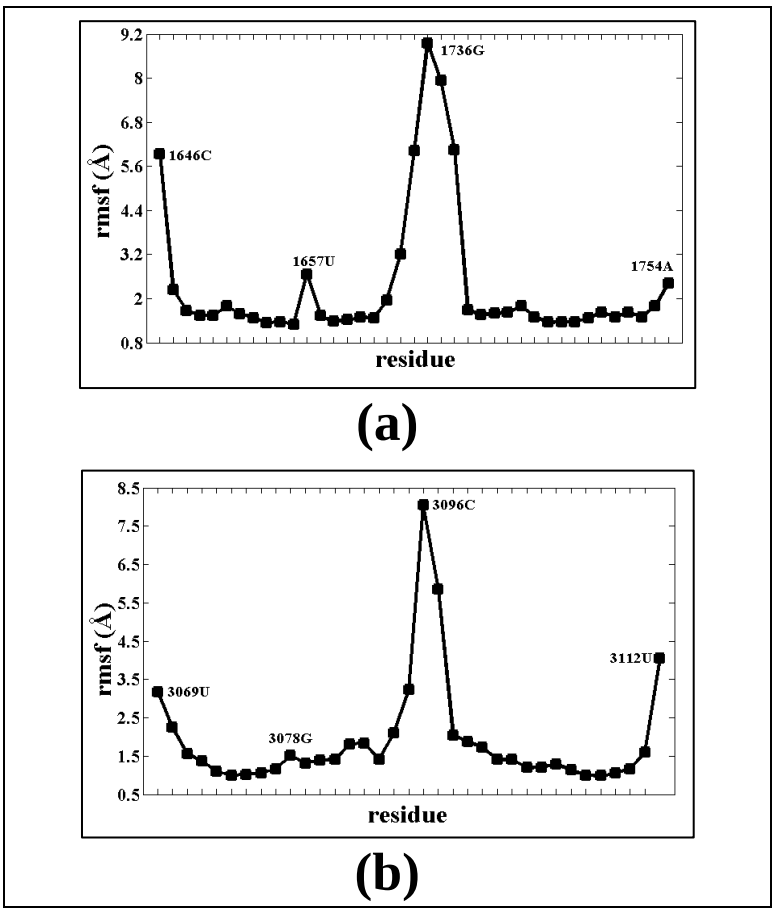


Figure 3.18.2: RMSF for (a) System 1a and (b) System 3a

3.3.2.2 Helix Axis Bending and Stacking Overlap

The helix axis bending angle and stacking overlap at the junction was calculated for the three simulations performed with charmm36 force field. A comparative frequency histogram plot for the bending angle and overlap, considering the entire MD-run for both AMBER and charmm simulations have been plotted **(Figure 3.19)**.

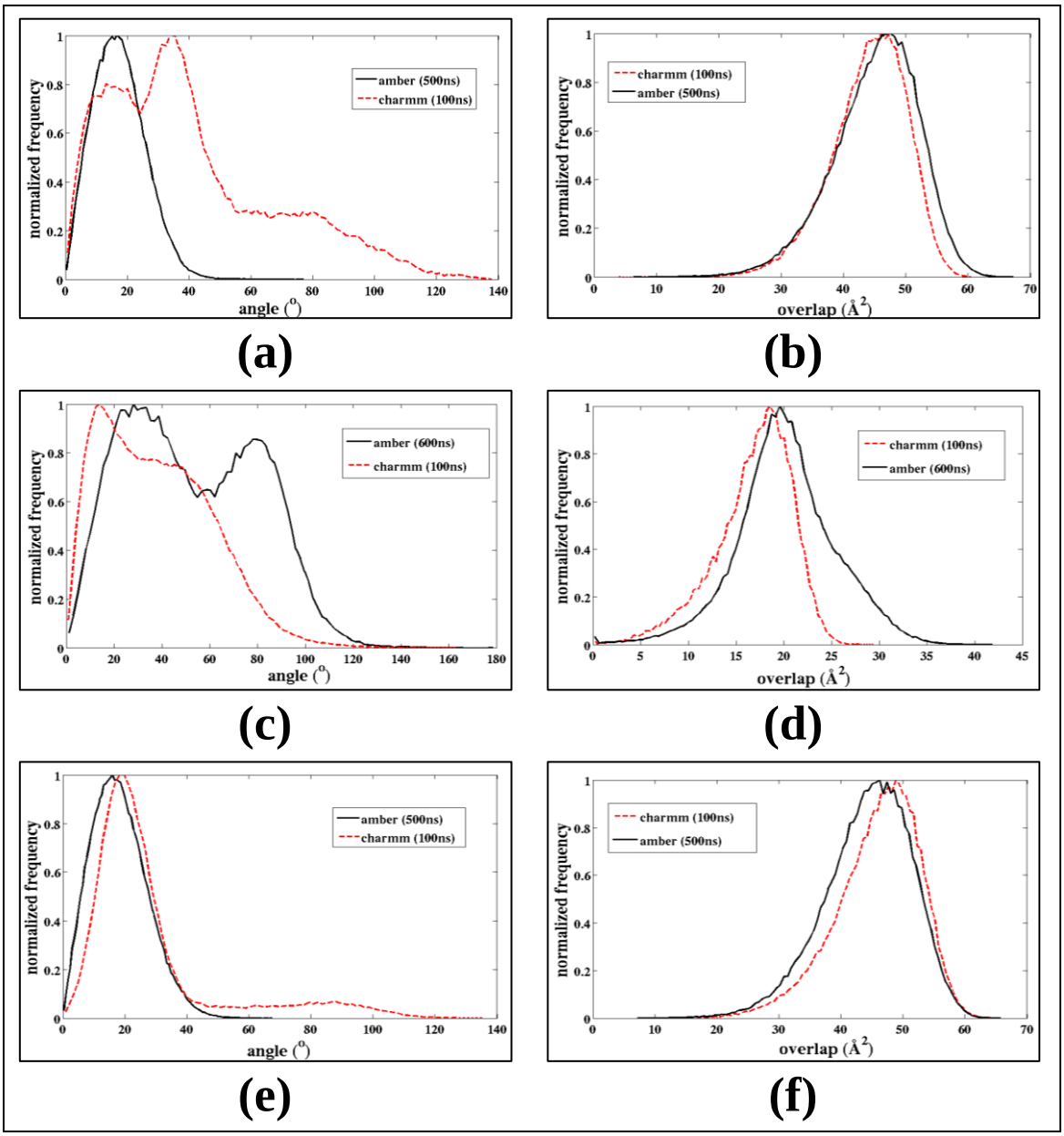


Figure 3.19: frequency histogram plot for helix axis bending angles of (a) System 1a, (b) System 2a and (e) System 3a; and frequency histogram plot for overlap of base pairs at the junction region of (b) System 1a, (d) System 2a and (f) System 3a.

The stacking overlap at the junction of System 1a is similar (~50Å^2) for both the charmm and AMBER simulations **(Figure 3.19a)**. However, the helix axis bending angle histogram is slightly different for the two cases. The AMBER simulation shows a single peak at 20° whereas; the charmm simulation shows an additional peak at a higher bending of 40° and a hump at 80° **(Figure 3.19b)**. The reason for such broad distribution and high bending angle with charmm36 force field is the extensive terminal fraying of the helix indicated by high fluctuations in open and shear values of 732C:737G extreme terminal base pair **(Figure 3.20)**. In case of System 2a both the overlap at the junction shows almost similar distribution for AMBER and charmm simulations. The frequency histogram of helix axis bending angle is slightly different for the two set of simulations which may be due to difference in MD-run time, *i.e.* the second peak in for charm simulation is not prominent due to lesser number of snapshots corresponding to only 100ns MD as opposed to 600ns MD of AMBER simulation **(Figure 3.19d)**. The System 3a exhibits similar distribution for both the helix axis bending angle and stacking overlap obtained from charm and AMBER simulations **(Figure 3.19e and 3.19f)**. These exhibits the nature of the bulge residues and its impact on the overall helix irrespective of the force field employed in simulation.

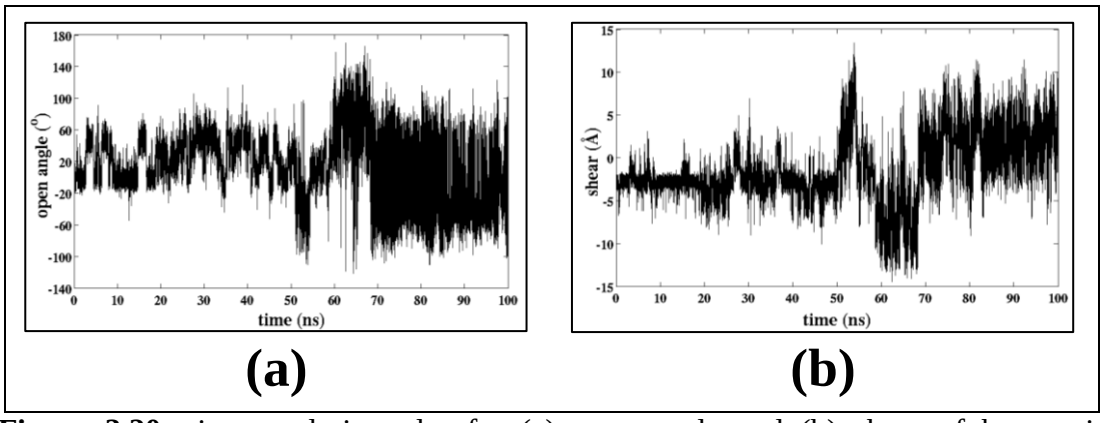


Figure 3.20: time evolution plot for (a) open angle and (b) shear of base pair 732C:737G of System 2a.

3.4 Discussion

The MD simulations of bulge containing RNA systems (systems 1a, 2a and 3a) indicated the effect of the bulge residue on the geometry of the base pairs in specific and on the overall structure of the duplex in general. It may be concluded that the bulged-out unpaired residues neither alter the base pairing at the junction region nor affect the overall structure of the duplex. However, the bulged-in residues induce alterations at the junction and affect the overall RNA-duplex to a high extent. Moreover, it was observed that the helix containing a bulged-out unpaired residue (System 1a) has features closer to the standard continuous A-form double helix RNA. Thus, there still exists scope of further probing specifically into the effect of orientation of the unpaired bulge residue on the overall structure of RNA duplex. The system containing base-triple inside the double-helix (System 3a) behaves well during the simulation run and has structural variability similar to a standard A-RNA, which is probably due to appropriate stacking of the triplet in between two base pairs at the junction. . The effects of bulge residues were found to be quiet similar for simulations with the AMBER-99sb and charmm36 force fields. The pseudo-continuous helical systems behave similar to the standard continuous RNA when there are extra-flanking residues at the junction. The type of base pairing (canonical or non-canonical) at the junction of pseudo-continuous helices can determine the structural variability. This study can be further extended by studying a set of coaxial stacks with all possible base pairing patterns at the junction. It is seen from the simulation of System 6 that loop acting as capping of the terminals may add to stability of the entire helix, in terms of less fluctuation. On the whole the present study has reported an initial overview of the various parameters related to base pairs at the junction of bulges and pseudo-continuous helices and has been compared to a standard RNA wherever possible.

Chapter IV

Simulation Studies of the wild-type and mutant human lamin A Ig fold

4.1 Introduction

The nucleus of eukaryotic cells contains chromatin in a tightly packed and highly organized fashion. The nucleus is composed of two main compartments: the nucleoplasm and the nuclear envelope (NE). The major structural elements of the NE are: the inner nuclear membrane (INM), the outer nuclear membrane (ONM), the nuclear pore complexes (NPCs), and the nuclear lamina (Dechat *et al.*, 2008). The INM encloses the nucleoplasm, and is covered by the nuclear lamina. The main constituents of the lamina are the type V intermediate filament (IF) proteins, the nuclear lamins (Aebi *et al.*, 1986; Goldman *et al.*, 1986; McKeon *et al.*, 1986). Lamins divided into A and B types based on sequence homologies and confer proper shape and mechanical rigidity to the nucleus.

The lamins are composed of a long central α-helical rod domain, flanked by globular short N-terminal (head) and longer C-terminal (tail) domains **(Figure 4.1)** (Prokocimer *et al.*, 2009). The C-terminal tail domain of lamins contains a structural motif similar to a type s immunoglobulin fold (Ig-fold) (Dhe-Paganon *et al.*, 2002; Krimm *et al.*, 2002). The small N-terminal segment consists of about 30 residues, the following α-helical rod domain is of about 40kDa and the C-terminal tail is of about 20 kDa size. A 16 amino acid segment at the rod domain N-terminal end and a 30 amino acid segment at the rod domain C-terminal end are highly conserved (Fisher *et al.*, 1986).

The topology of C-terminal domain of lamin A/C protein consists of nine β strands, forming two β sheets each of four and five strands, packed into a classical β sandwich **(Figure 4.2)**. The β strands 1, 4, 5, 8, and 9 form the first β sheet and the strands 2, 3, 6, and 7 form the second β sheet, β7 being parallel to β6. Strands 2, 3, 4, 5, 6, 8, and 9 compose the classical Ig fold of type s and following the Ig fold nomenclature, corresponds to strands A, B, C, C', E, F, and G, respectively. The β1 and β7 are additional strands typical of the lamin fold (Krimm et al., 2002).

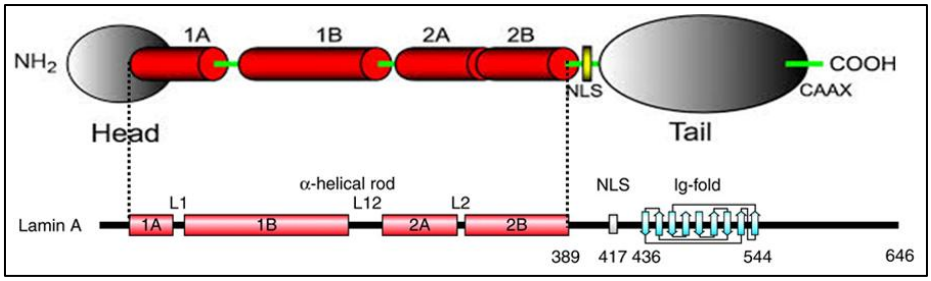


Figure 4.1: Lamins composed of α-helical rod domain, flanked by globular short N-terminal and longer C-terminal domains.

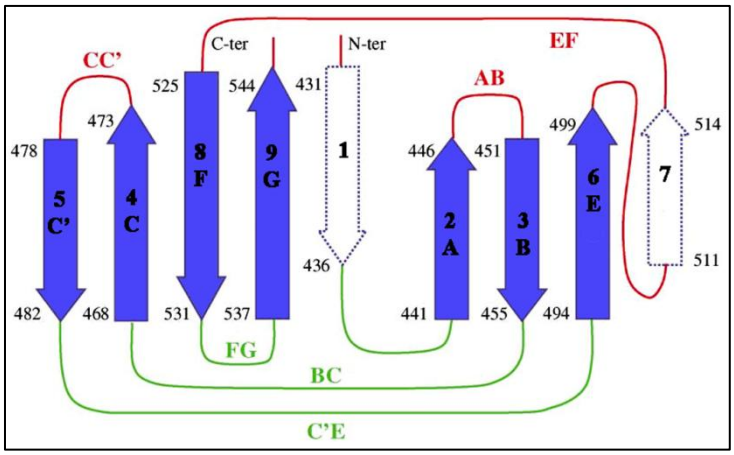


Figure 4.2: Topology of C-terminal domain of lamin A/C protein.

The β sandwich of the lamin A/C domain is characterized by a large tilt of 45° between the axes of two β sheets. These two regular β sheets show two classic β bulges: (i) the β bulge formed by Arg 453 of β3 and residues Glu 443 and Glu 444 of β2. These three residues are located within the anti-parallel beta-sheet formed by strands β2 and β3. These three residues are protruded outwards from the core of the protein into the solvent but due to formation of this β bulge, the three amide porotons are protected against solvent exchange and the beta sheet structure is maintained; (ii) the β bulge formed by Ile 531 in β8 and the residues Val 538 and Ala 539 in β9 has similar features where the amide protons of Ile and Val are protected.

Lamin proteins are known to be related diversely to basic nuclear processes like replication, DNA damage repair and transcription. More than 400 mutations in human lamin A protein alone have been known to produce at least eleven different disease conditions jointly termed as laminopathies. These mutations in lamin A are scattered throughout its helical rod domain as well as the C-terminal domain **(Figure 4.3a)**. The commonality of phenotypes in all these diseases is characterized by misshapen nuclei of the affected tissues.

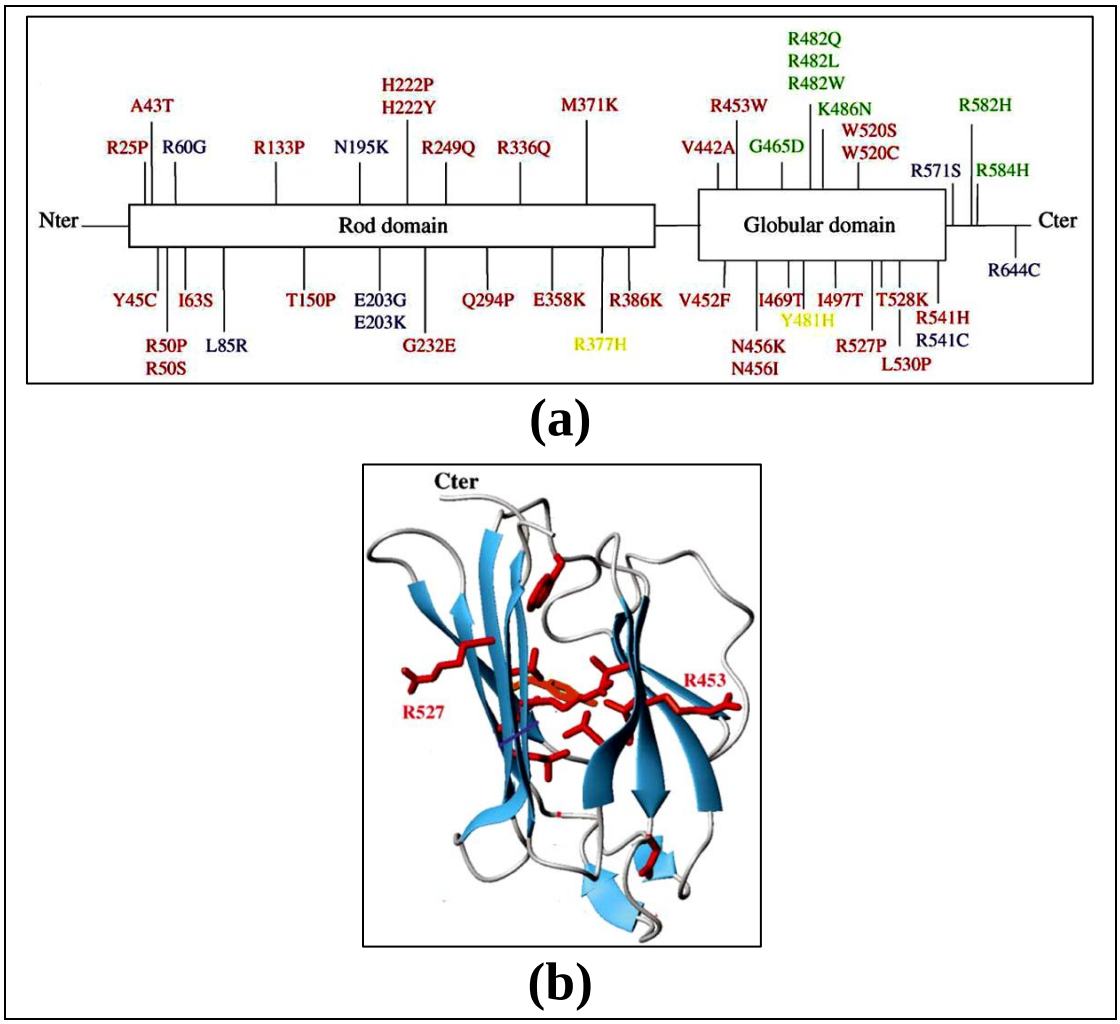


Figure 4.3: (a) All mutations in lamin A. (b) Mutations R453W and R527P are exposed to the solvent.

The C-terminal domain of lamin A/C is mutated in four heritable diseases: Emery-Dreifuss muscular dystrophy (EDMD), limb girdle muscular dystrophy type 1B, and dilated cardiomyopathy, are all characterized by cardiac abnormalities, and, in the case of the two muscular dystrophies, progressive skeletal muscle wasting, Dunnigan-type familial partial lipodystrophy is characterized by regional fat loss, insulin resistance, and diabetes mellitus with no striated muscle abnormalities. The thirteen missense mutations in the C-terminal domain that can cause EDMD are: V442A, V452F, R453W, N456K and N456I, I469T, W520S AND W520C, R27P, T528K, L530P AND R541H **(Figure 4.3a)** (Bonne *et al.*, 2000; Brown *et al.*, 2001; Raffaele *et al.*, 2000). All the mutations except R453W and R527P are concerned with buried residues localized in the core of the 3D structure **(Figure 4.3b)**.

In this work we have particularly focused on the R453W mutation **(Figure 4.4)** causing autosomal dominant EDMD laminopathy. The rationale behind choosing this mutation is the highly conserved nature of the site within Ig fold across the invertebrate and vertebrate species (Erber *et al.*, 1999). Also, this mutation disrupts two salt bridges in the lamin A Ig fold (Krimm *et al.*, 2002). Thus, since R453W has a strong destabilizing effect on the structure of the protein, we have investigated by force-induced simulation studies how mechanical perturbation modify the unfolding mechanism or stretching behaviour of Ig-fold domain and compared the mechano-elastic properties of the mutant R453W with that of the wild-type. Force-induced pulling SMD simulations with R453W Ig showed interesting differences in unfolding behaviour compared to the wild type and were in agreement with data from AFM experiments.

4.2 Methodology

The coordinates of the wild-type protein were directly obtained from crystal structure reported in RCSB-PDB with PDB ID: 1IFR (Dhe-Paganon *et al.*, 2002) and the coordinates of the mutant protein were generated using Modeller 9.11 (Eswar *et al.*, 2006; Martí-Renom *et al.*, 2000). The structures of wild-type and mutant protein difference only at position 453, the wild-type has arginine (Arg) (**Figure 4.4a)** whereas the mutant has tryptophan (Trp) **(Figure 4.4b)** residue at that position.

4.2.1 Molecular Dynamics Simulations

Equilibrium Molecular Dynamics (MD) simulations of both the wild type and mutant proteins were carried out using NAMD software (Phillips *et al.*, 2005) with charmm22 force field (Mackerell *et al.*, 2004). In each simulation, the protein molecules were explicitly solvated with TIP3P water molecules in a periodic box whose boundaries extended at least 10Å from any solute atom. Additional charge neutralizing Cl^- counter-ions were added in such a manner that neutralizes the charge of the protein molecule. The initial round of equilibration with explicit solvent and ions was performed by CHARMM (Brooks *et al.*, 1983) and involved 100 steps of steepest descent, followed by 1000 steps of conjugate gradient and 20000 steps of adopted basis Newton-Raphson energy minimization. The particle mesh Ewald

(PME) summation method (Darden *et al.*, 1993), with width of the Gaussian distribution as 0.35, was used to treat the long-range electrostatic interactions. A force-switch method was applied for non-bonded interactions (van der Waals) with a 12Å cutoff. After energy minimization the systems were gently heated from 0K to 300K in 30ps with a time step of 1.0fs followed by the final production run for 100ns by Constant Pressure Temperature (CPT) dynamics algorithm (Feller *et al.*, 1995) at 1atm pressure, with 1fs timestep. The trajectories were made up by frames collected every 1.0ps and then analyzed using CHARMM and VMD (Humphrey *et al.*, 1996).

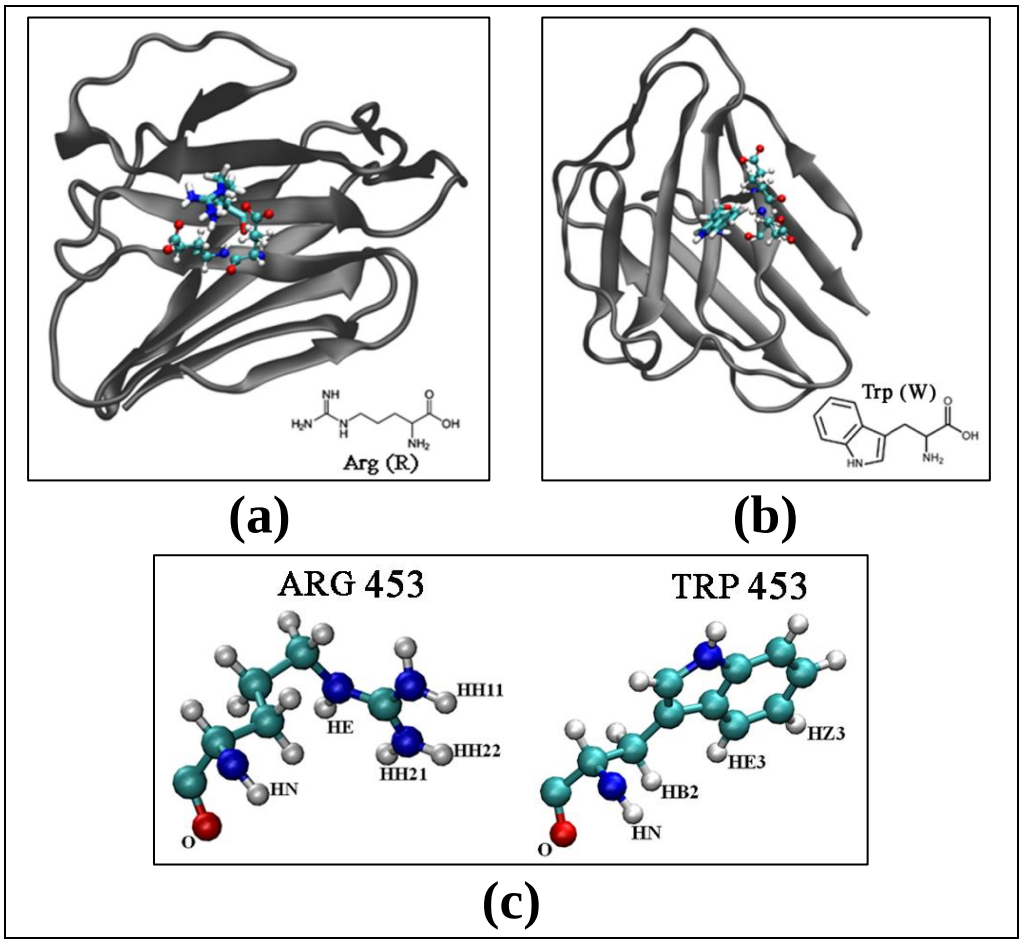


Figure 4.4: 3D representation of the (a) wild-type protein containing Argenine and (b) mutant protein containing Tryptophan residue, at 453 position, and (c) structure of Argenine and Trptophan.

4.2.2 Steered Molecular Dynamics Simulations

Along with equilibrium MD simulations, force-induced unfolding simulations were carried out separately for both the proteins by Steered Molecular Dynamics (SMD) simulations (Balsera *et al.*, 1997; Evans and Ritchie, 1997; Izrailev *et al.*,

1997). The coordinates of the explicitly solvated protein molecules at 300K, were considered for SMD using NAMD software with charmm22 force field and Generalized Born Implicit Solvent (GBIS) model for water (Feig and Brooks, 2004; Guvench *et al.*, 2002; Qiu *et al.*, 1997). The GBIS model was chosen since it was difficult to have an adequate explicit solvation for the protein chain whose dimensions vary from nearly a sphere of 40Å diameter to a fully stretched molecule of 350Å length.

The simulations were performed with a timestep of 1fs, a solvent dielectric constant of 78.5, and a cut-off of Coulomb forces with a switching function starting at a distance of 15 Å and reaching zero at 16Å. SMD simulations of constant velocity stretching (SMD-CV protocol) were carried out by fixing the C-alpha atom of N-terminal residue (resid: 432) and applying external forces to the C-alpha atom of C-terminal residue (resid: 544) which was the SMD atom, along the direction of the vector from residue 432 to residue 544. This force was applied in the direction parallel to the hydrogen bonds (H-bonds) of first β-sheet (β-sheet1). It is to be mentioned here that, in such simulations the SMD atom is attached to a dummy atom via a virtual spring. This dummy atom is then moved at a constant velocity and the force experienced by the pulled terminal atom is:

$$F = k (v t – x),$$

where x is the displacement of the pulled atom from its original position, v is the pulling velocity, and *k* is the spring constant.

Three sets of simulations were performed each for the wild-type and the mutant proteins considering spring constant of 0.6Kcal/mol/$Å^2$ and three different pulling velocities **(Table 4.1)**. As expected the time required for complete stretching depends on the pulling velocity. Considering the slow rate of experimental pulling, which needs to be compared with our simulation results, data from set-3 has been used for further analysis. Nevertheless, the three sets of data show similar trend. The trajectories were made up by frames collected every 1.0ps and analyzed using CHARMM and VMD. The extension d(t), defined as the increase of the end-to-end distance from that of the native fold, was monitored along with the force F(t) and secondary structure analysis of the simulation snapshots were done following DSSP algorithm (Kabsch and Sander, 1983) as implemented in CHARMM.

Table 4.1: SMD simulation set-up.

	velocity of pulling in Å/timestep (timestep=1fs)	required duration of simulation for complete stretching (ns)
set-1	0.00025	10
set-2	0.000025	50
set-3	0.0000025	140

4.3 Results

4.3.1 Molecular Dynamics Simulation Analysis

The wild type protein has Argenine residue at position 453 and the EDMD causing mutant has Tryptophan. The mutant adopts the same structure as the wild-type probable due to high beta-sheet propensity of tryptophan (Chou and Fasman, 1974; Schulz and Schirmer, 1979). The secondary structure of both the wild-type and the mutant has nine β-strands along with some coil/unstructured regions but no alpha-helix **(Figure 4.2)**. The β1 (432-436), β9 (537-544), β8 (525-531), β4 (468-473) and β5 (478-482) strands form the β-sheet1 and β-sheet2 is formed by strands β2 (441-446), β3, β6 (494-499) and β7 (511-514). The β3 strand consists of residues 451-455 which includes the mutation position.

4.3.1.1 Dynamics and Fluctuation

The root-mean-square deviations (RMSD) of the trajectories with respect to the energy-minimized structures of the corresponding wild-type and mutant proteins were calculated to understand the behaviour of the simulations and compare the structures of the two proteins. The structural integrity of the two systems persisted during the complete course of 100ns of MD simulation as the all atom RMSD values rarely reached 2.2Å for either of the proteins **(Figure 4.5a)**.

The time evolution plot for radius of gyration **(Figure 4.5b)** indicates similar features for both the proteins after about 40ns of MD-run. The solvent accessible surface area (SASA) of the mutant protein was initially smaller as compared to that of the wild-type; however in the last 30ns of the MD-run the two proteins retain the same value of SASA **(Figure 4.5c)**. A closer look at the SASA values of the R453

(Arg 453) and W453 (Trp 453) residues of the wild-type and mutant proteins respectively, indicate that during the entire course of MD-run, R453 has much lower solvent accessibility (~50Å^2) than W453 (~100Å^2) **(Figure 4.5d)**, *i.e.* R453 is buried deep into the core of the protein.

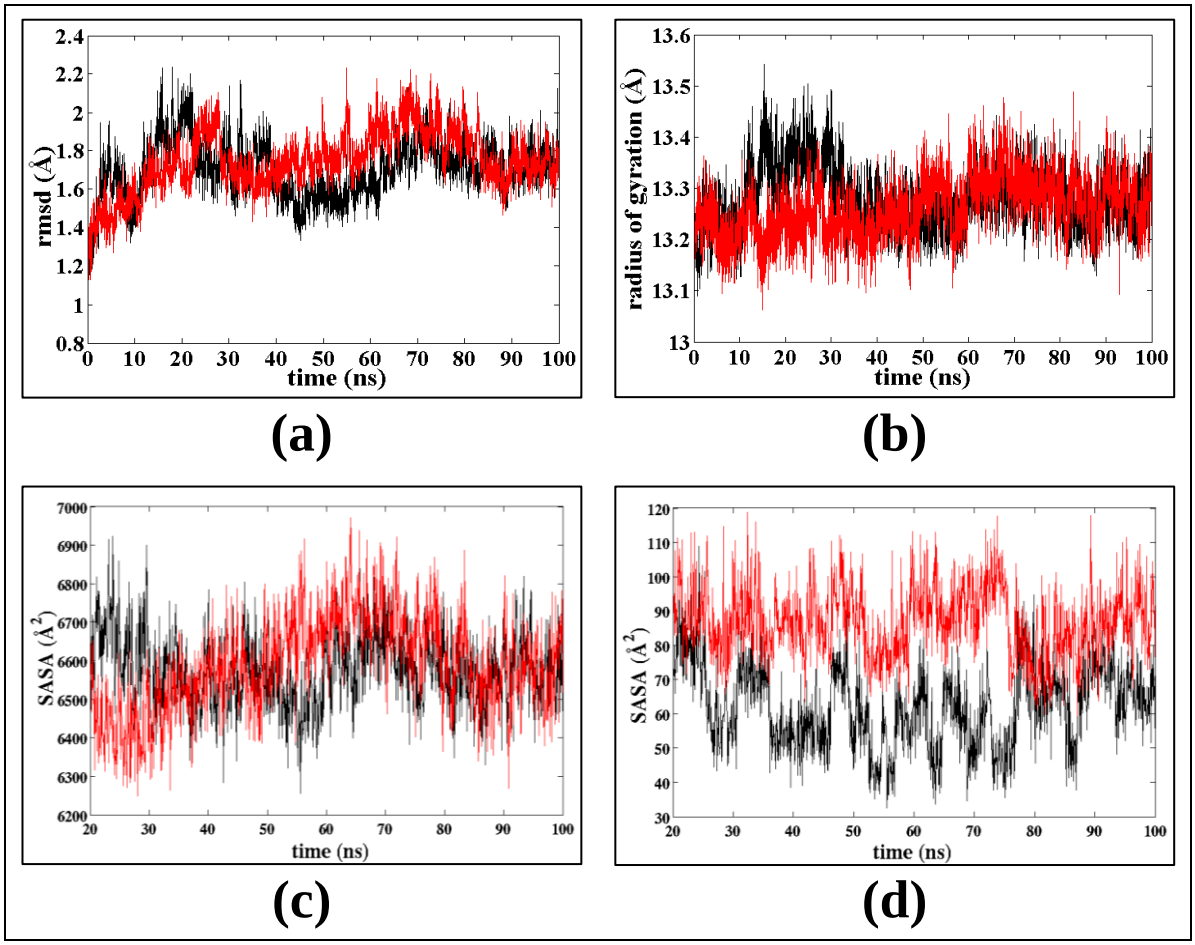


Figure 4.5: (a) all atom RMSD plot, (b) time evolution plot for radius of gyration, (c) overall solvent accessible surface area (SASA), for the wild-type (black line) and the mutant (red line) proteins; and (d) SASA for Arg 453 (black line) and Trp 453 (red line) residues.

The root-mean-square fluctuations (RMSF) of the two proteins were calculated considering: (i) all atom, (ii) protein backbone atoms and (iii) side-chain atoms of the protein and plotted with respect to residue ID **(Figure 4.6)**. Almost all the residues, except a few showed similar fluctuations for the wild-type and the mutant proteins. The residues 475-479 and 501-505, which showed higher RMSF for the mutant protein **(Figure 4.6a and 4.6b)**, are involved in helix formation making them inherently more flexible. In particular, the residues Glu 443 and Glu 444 which were involved in salt-bridge formation with Arg 453 of the wild-type, showed a

difference in RMSF **(Figure 4.6a and 4.6b)** when compared to that of the mutant wherein the salt-bridge is broken.

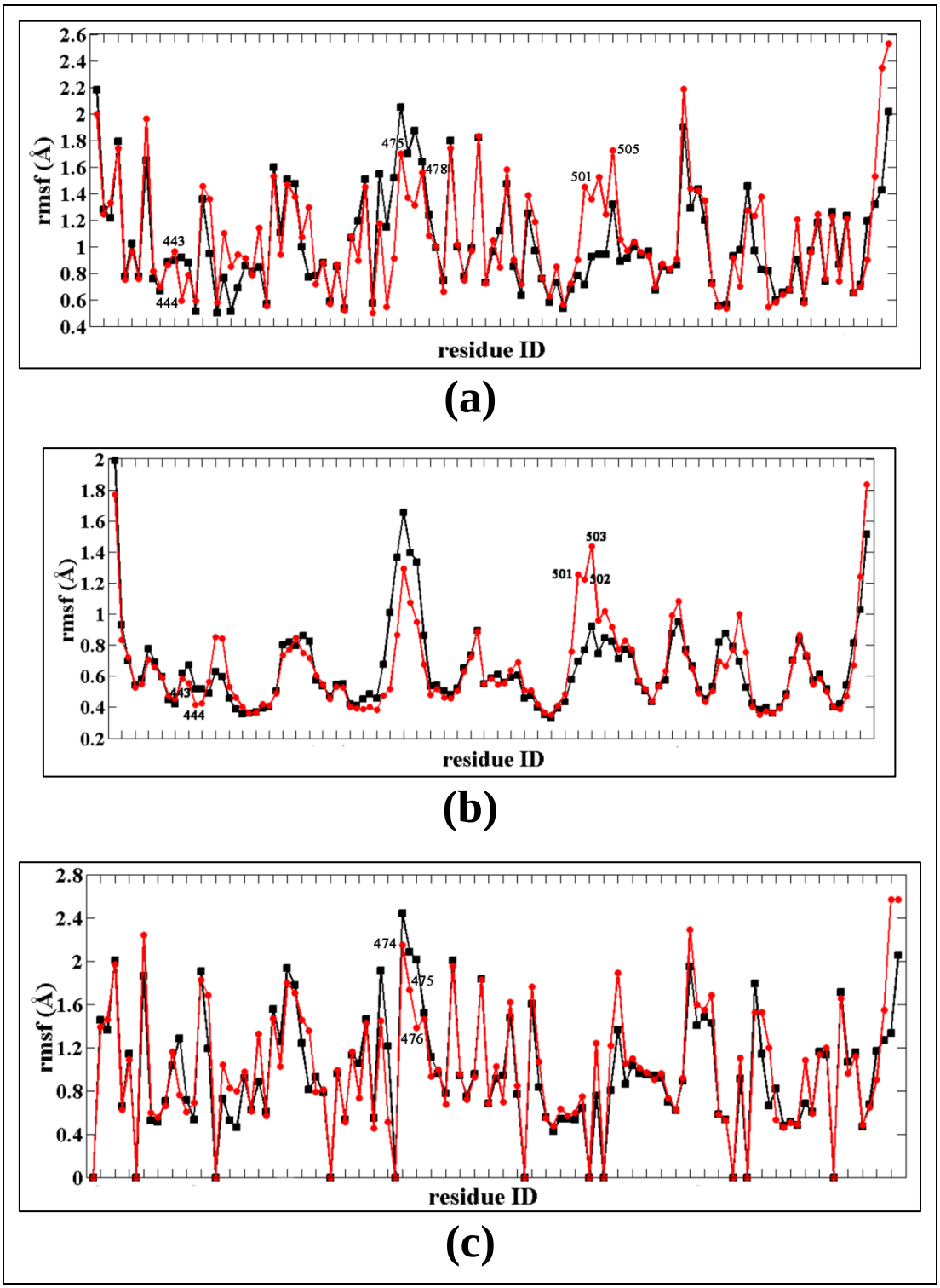


Figure 4.6: RMSF of the wild-type (black) and mutant (red) proteins considering (a) all atoms, (b) backbone atoms, (c) side-chain atoms of the respective proteins.

4.3.1.2 Secondary Structure Analysis

The 100ns long MD simulations carried out for both the wild-type protein (1IFR) and its mutant (R453W) using explicit water model and periodic boundary condition assisted in understanding the C-terminal domain of both the proteins. Analyses of the contact maps of the proteins indicate similar tertiary folding of the two types **(Figure 4.7.1)**. Furthermore, the variations in secondary structure **(Figure 4.7.2)** indicate that almost all the β-strands of both the wild-type and the mutant are maintained throughout their respective 100ns MD-run except β2 of the mutant protein, which is disrupted after about 10ns. It is to be noted here that this β2 contains the residue Glu 443 which was involved in salt bridge formation with Arg 453 of wild-type, but such interaction is absent in the mutant.

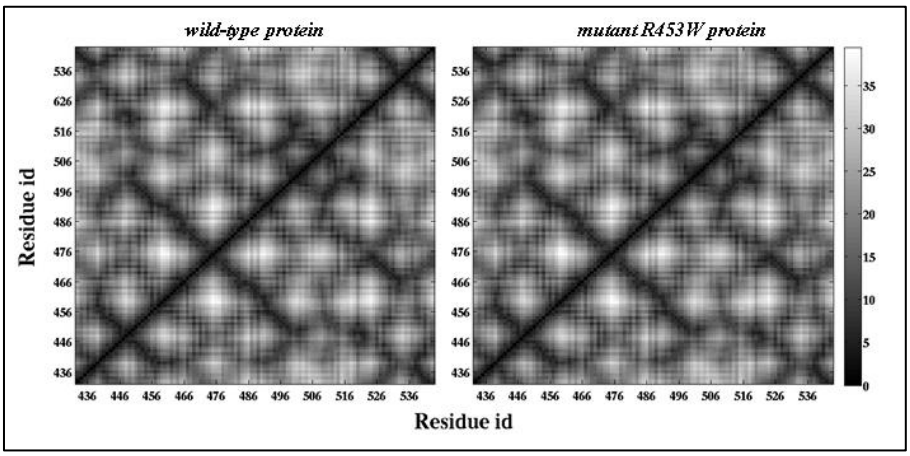


Figure 4.7: Contact map plot (L to R) for the wild-type and mutant proteins.

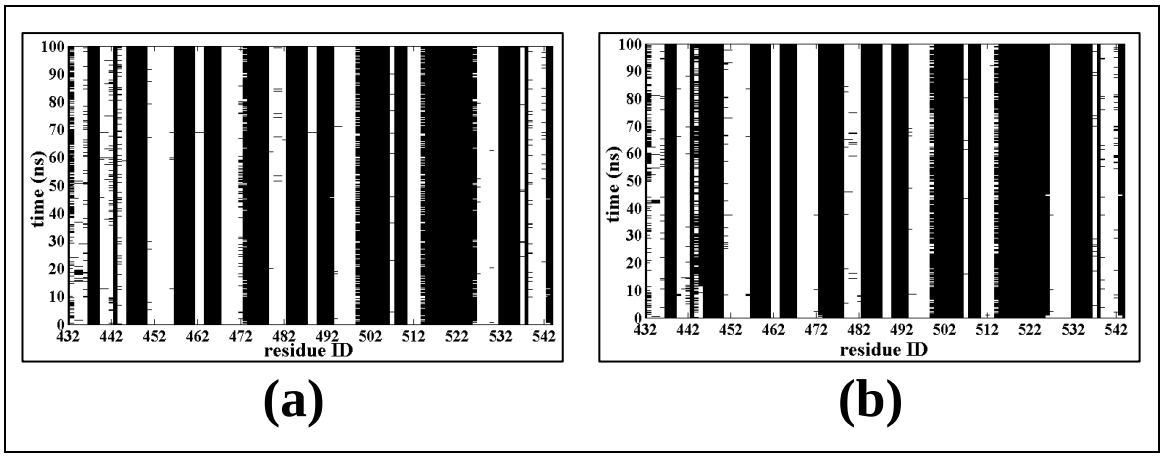


Figure 4.8: Variation in secondary structure during 100ns MD-run for (a) wild-type protein and (b) mutant protein.

4.3.1.3 Backbone Structure Analysis

The main difference in structure of the wild-type and mutant protein is at the region involving the residues Arg 453/Trp 453, Glu 443 and Glu 444. Thus, we have tried to look into these residues from the perspective of [ϕ,ψ] torsion angles. The ϕ and ψ values of the residues Glu 443 and 444 do not differ much for the wild-type and mutant as seen from the MD-simulation **(Table 4.2)**.

Table 4.2: mean and standard deviation (bracket) of ϕ and ψ angles of Glu 443 and 444 residues of the wild-type and mutant proteins.

	ϕ(°)		ψ (°)	
	Glu 443	Glu 444	Glu 443	Glu 444
wild-type	-80.18 (14.4)	-151.62 (26.0)	-46.13 (10.5)	144.76 (10.2)
mutant	-97.22 (19.8)	-146.56 (46.7)	-34.08 (15.2)	147.73 (8.7)

The [ϕ,ψ] plots for Arg 453 and Trp 453 **(Figure 4.10.1a and 4.10.1b)** are almost similar and the angles lie well within the allowed region of the Ramachandran plot. The [ϕ,ψ] values of Glu 443 residue of the wild-type and mutant proteins **(Figure 4.10.1c and 4.10.d)** fall mostly in the all allowed region however, this residue attains certain conformations during the MD-run which are forbidden. The Glu 444 residue also has [ϕ,ψ] values in the allowed region for both the wild-type and the mutant **(Figure 4.10.1.e and 4.10.1f)**.

The χ angle describes the rotation around the bonds between the side chain atoms. The first χ angle (χ_1) is defined as the rotation between the C_α and C_β atoms, higher χ angles are defined by the rotation between the following atoms if the side chain. It is seen from the backbone-independent rotamer library, as given by Dunbrack and Karplus (Dunbrack Jr and Karplus, 1993) that highest % χ_1 for Arg, Trp and Glu all lie in the range χ_1= -60±60 but with different percentages of 58.3%, 51.8% and 55.2% respectively. Although, the time evolution plot for χ_1 of Arg 453 of the wild-type protein maintains 180° as the dihedral throughout the MD-run which has 30.9% of χ_1 as per the same rotamer library but the Trp 453 residue of the mutant maintains -60° dihedral **(Figure 4.10.2a and 4.10.2b)**. The time evolution plot for χ_1 indicated similar trend for Glu 443 residue of the wild-type and mutant **(Figure 4.10.2c and 4.10.2d)**; however, the trend was different for Glu 444 **(Figure 4.10.2e and 4.10.2f)**. The Glu 444 of the wild-type attained the major -60° dihedral along

with 180° angle. In case of the mutant this residue was confined to 180° of the χ_1 dihedral.

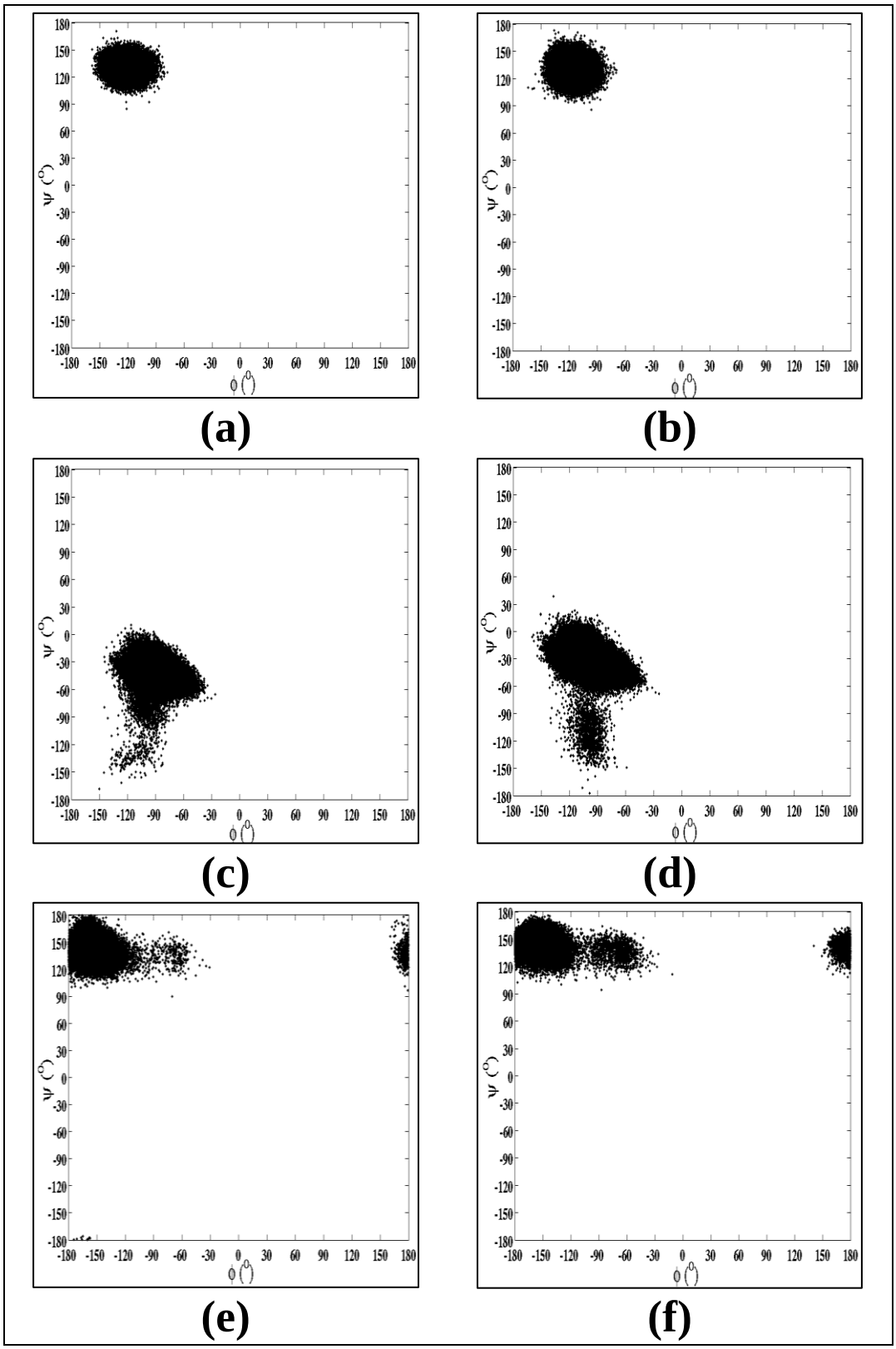


Figure 4.9: [ϕ,ψ] plot for (a) Arg 453, (b) Trp 453, (c) Glu 443 of wild-type, (d) Glu 443 of mutant, (e) Glu 444 of wild type and (f) Glu 444 of mutant.

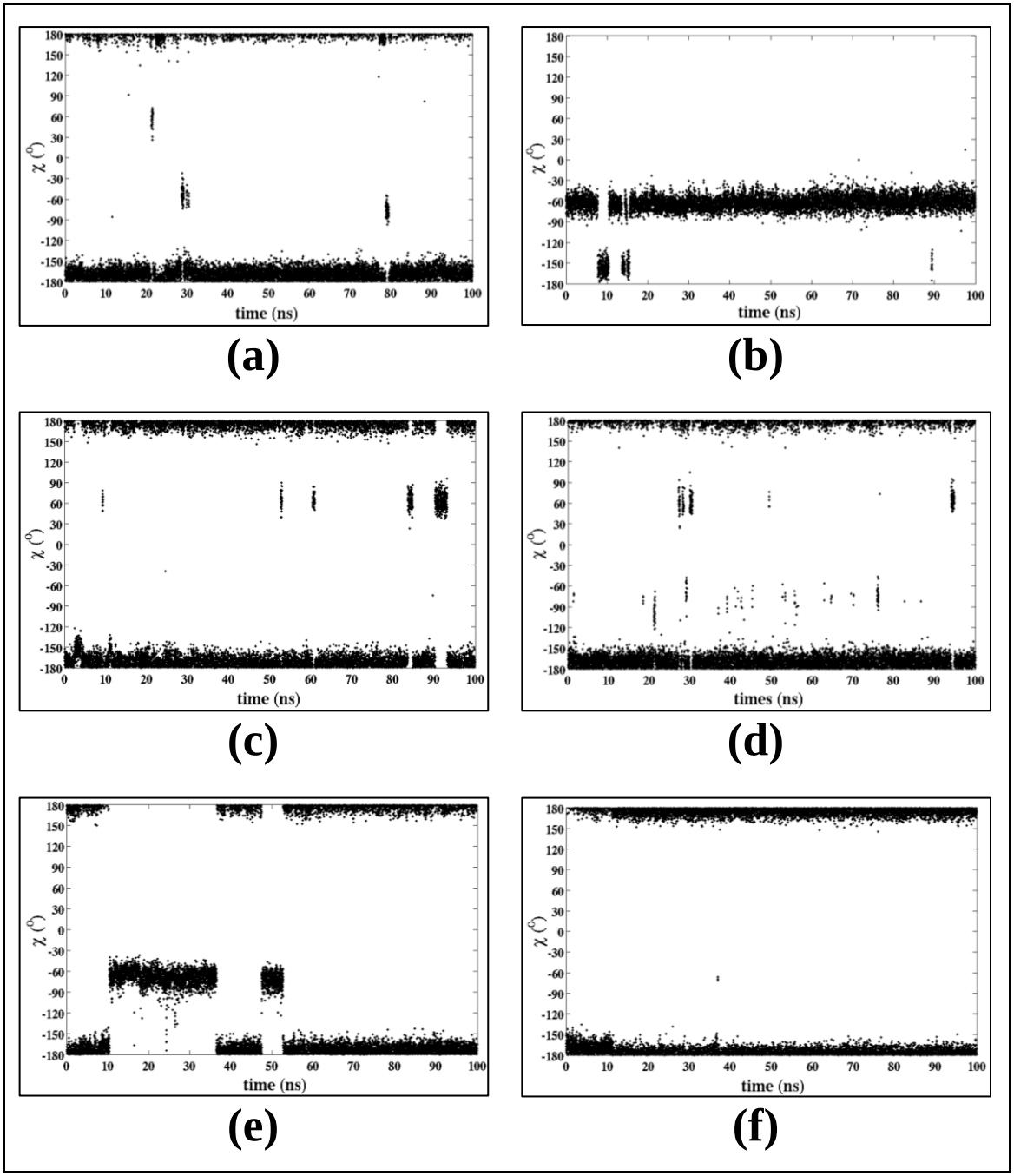


Figure 4.10: time evolution plot for χ torsion angle of residues: (a) Arg 453, (b) Trp 453, (c) Glu 443 of wild-type, (d) Glu 443 of mutant, (e) Glu 444 of wild type and (f) Glu 444 of mutant.

4.3.2 Steered Molecular Dynamics Simulation Analysis

The CV-SMD pulling was performed on both the wild-type and mutant proteins with three different velocities (i) 0.00025Å/fs, (ii) 0.000025Å/fs and (iii) 0.000025Å/fs, independently until the respective polypeptide chains were completely straightened out **(Table 4.1)**.

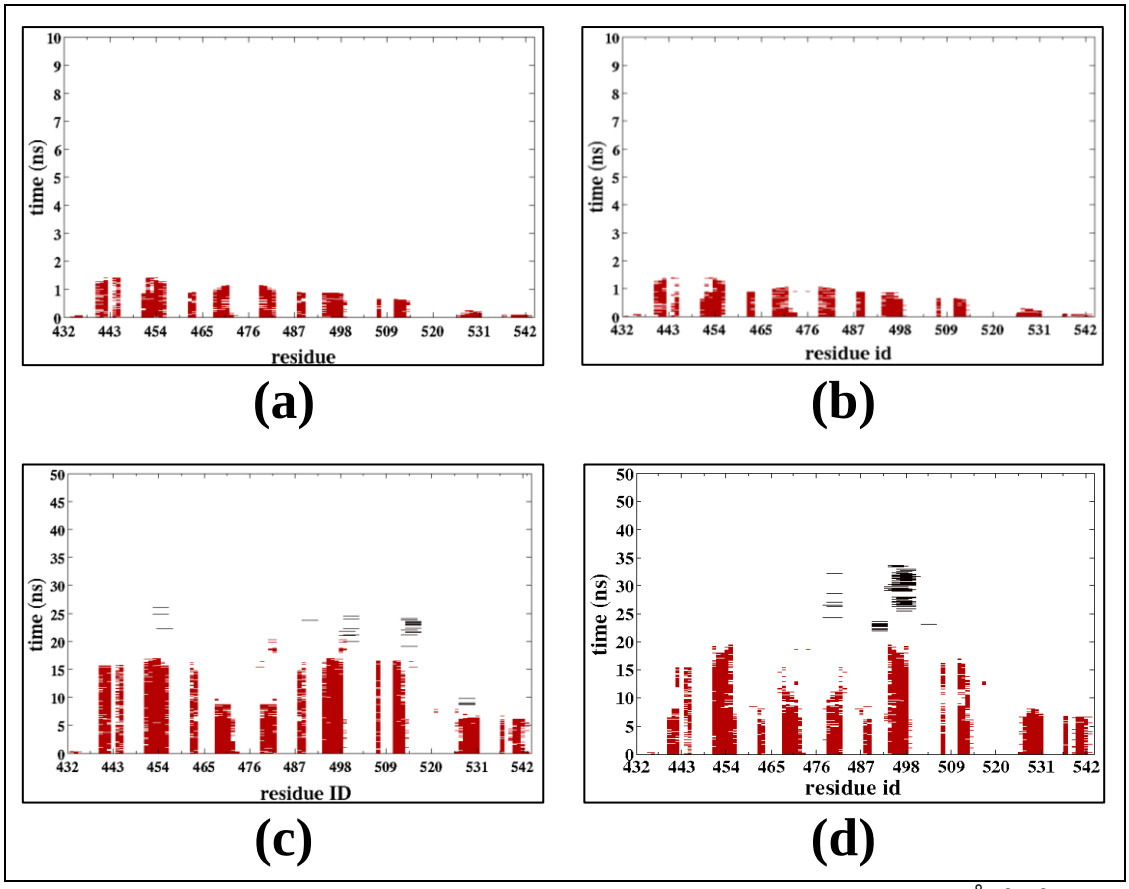


Figure 4.11: secondary structure plots when pulling velocity is 0.00025Å/fs for (a) wild-type and (b) mutant protein; and when pulling velocity is 0.000025Å/fs for (c) wild-type (d) mutant protein. Red colour represents β-sheet, black colour represents α-helix/coil and white region represents the unstructured protein.

The pulling velocity of 0.00025Å/fs produced time dependent gradual elongations of both the wild-type and the mutant proteins until the polypeptide chains were completely straightened out and attained maximum extension of ~350Å at around 10ns. The secondary structure plots **(Figure 4.11a and 4.11b)** clearly indicate that in case of both the wild-type and mutant proteins, all the beta-sheets are complete broken into unstructured form within the first 2ns of pulling MD. Thus, the velocity applied is too fast to get hold of any crucial intermediate state indicative of transition from β-sheet to α-helix or coil formation. Moreover, no structural difference between the wild-type and the mutant is visible. A slightly lower pulling velocity of 0.000025Å/fs produced similar time dependent gradual elongations of both the wild-type and the mutant proteins but required a longer time of ~50ns to completely straighten out and attain the maximum extension. The secondary structure plots obtained from these simulations **(Figure 4.11c and 4.11d)** indicate that the proteins (wild-type or mutant) attain complete unstructured form within 40ns of the SMD-run.

The secondary structure plots obtained from these SMD-runs indicated some intermediate states representing α-helix. Also, the plots are slightly different for the two proteins. The β2 strand is intact for ~15ns in case of the wild-type, however for the mutant it starts breaking at an early stage. The β4, β5, β6 and β7 strands also behave differently during SMD of the two different proteins. However, a far lower velocity was next applied to capture the unfolding process with far more details.

The pulling velocity of 0.0000025Å/fs during SMD of both the wild-type and the mutant proteins produced time dependent gradual elongations until the polypeptide chains were completely straightened out and attained maximum extension of ~350Å at around 140ns **(Figure 4.12)**. As tryptophan has higher β sheet forming propensity (Chou and Fasman, 1974; Schulz and Schirmer, 1979), the experimental CD data of R453W mutant indicated a slight percentage increase in β sheet structure (Bera *et al.*, 2014). Both the forms however, retained their respective shape and compactness up to 20ns as observed from their radius of gyration vs. extension plots **(Figure 4.13)**.

A similar kind of unfolding pattern (with two transition peaks) for both wild type and R453W mutant protein was observed from the force extension analyses **(Figures 4.12a and 4.12b)**. For wild type, the first peak appeared at 50Å extension corresponding to a force of 250pN and unfolding force for the latter transition peak, was calculated to be around 150pN at 150Å extension. But for the mutant, first transition peak appeared as a doublet between 50-75Å extensions with unfolding force around 200pN. Interestingly, the mutant exhibited the same peak corresponding to150Å extension at a lower unfolding force of ~75pN. Hence, the mutant R453W is mechanically weaker compared to the wild type as is corroborated from single molecule force spectroscopic data and also from earlier structural studies of the mutant (Krimm *et al.*, 2002). The secondary structure variations of both the wild type and the mutant proteins for the entire simulation trajectories were analysed. Since, the first and the last β-strands, located at the periphery were H-bonded to each other, and the force was applied parallel to this H-bond vector, β1 at the N-terminal and β9 at the C-terminal strands were uncoiled at the earliest compared to the other β-strands for both the simulations. This also induced early rupturing of the nearby β8 strand. The simulation trajectories showed significantly different features at short extension (within 180Å) in the both cases, whereas at longer extensions those domains were almost identical. The variation of different secondary structure elements of all the

residues of both the wild-type and mutant protein, with simulation time scale are shown in **Figures 4.12c and 4.12d**. For both the wild type and the mutant, an initial force of ~100pN was required to completely rupture the β1-strand at an extension less than 10Å. A high force of ~250pN and ~200pN for the wild type and the mutant proteins respectively corresponded to the disruption of β8-β9 H-bonding and ultimately rupturing of these β-strands. All these events were completed within the first 40ns of simulation for both the cases. As indicated earlier, the Arg 453 residue of β3 forms two salt-bridges with Glu 443 and Glu 444 of β2, which were lost due to R453W mutation. As a result, the β2 strand of the mutant protein was rupturing simultaneously along with β8 and β9 at around the same time indicated by a broader peak at ~50Å extension in the force vs. extension plot **(Figures 4.12a and 4.12b)**. On the other hand, stability of the wild type β2 strand, in comparison to its mutant was higher, as it was completely stretched only after 60ns and corresponded to a force as high as ~150pN. Rupturing of β2 was followed by β3 which was again more stable for the wild type as compared to the mutant. The β6 and β7 strands were ruptured next and the process occurred faster in case of mutant protein compared to the wild type under similar conditions of force. The β4 and β5 strands of the first beta sheet layer of the wild-type domain remained stable till ~85ns (up to 200Å extension) whereas, for the mutant these strands unfolded completely within ~75ns (up to 150Å extension) at comparable forces. Similarly, β2 and β7 ruptured quickly in case of mutant protein compared to the wild type under similar conditions. This is further corroborated from the radius of gyration vs. time plot **(Figure 4.13)** where the mutant showed an altogether different compactness compared to the wild type. We observed an interesting conformational transition from β-strand to α-helix prior to complete stretching in both the cases. The secondary structure alteration from β-strand to alpha helix was observed more in case of the mutant-protein. However, the alpha-helices were not stable in either case though their lengths were quite significant. In conclusion the SMD studies are in good agreement with our single molecule force spectroscopy results.

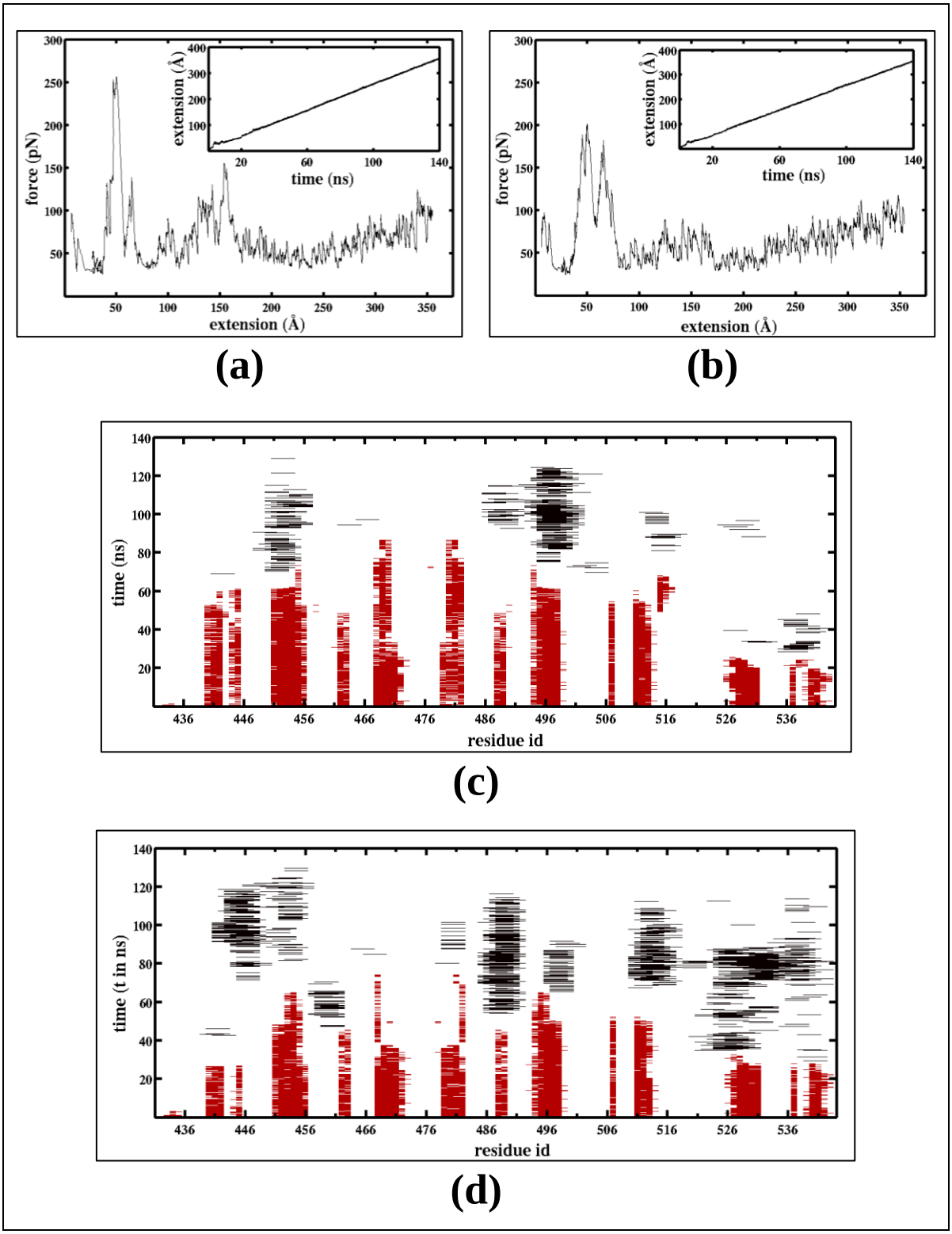


Figure 4.12: Force vs. extension curve for (a) wild type lamin Ig fold (PDB ID 1IFR) and (b) R453W mutant. The corresponding extensions vs. time plots have been shown in the inset. Variations in secondary structures are shown in (c) for wild type and (d) for R453W mutant. Red colour represents β-sheet, black colour represents α-helix/coil and white region represents the unstructured form.

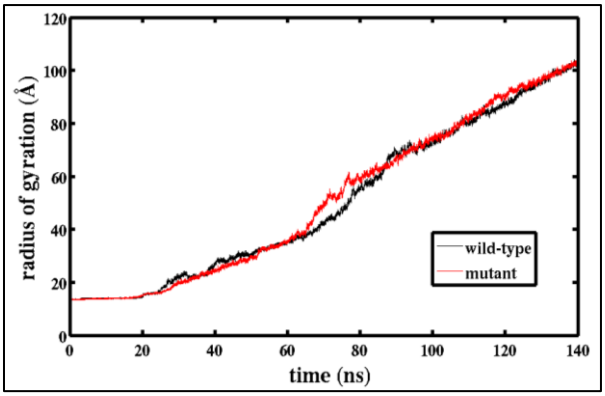


Figure 4.13: Within the first 20ns of SMD the wild-type (black) and mutant (red) proteins retained their compactness and after that the radius of gyrations kept increasing with time indicating destruction in the globular shape.

4.4 Discussion

The wild-type protein involved two salt bridges: (Arg 453-Glu 443) and (Arg 453-Glu 444), whereas in the mutant due to replacement of argentine by tryptophan these are broken. However, our MD simulations of 100ns indicate structural similarity between the wild-type and the mutant. It might be expected that longer simulations will indicate a structural difference involving the β3 strand of β-sheet2. Our experimental collaborators checked the thermal stability of (I27-wt Ig/Ig R453W)$_4$ by differential scanning calorimetry and although the secondary structure did not change significantly, the monomer of Ig and Ig R453W had significantly different denaturation temperatures of 62ºC and 43ºC respectively. As the R453W mutation resulted in the loss of two salt bridges at Glu 443 and Glu 444, the expected change in the force mediated unfolding was observed by SMD simulations. The unfolding forces obtained from theoretical calculations were in complete agreement to the data obtained by our experimental collaborators from Single Molecule Force Spectroscopy. We also observed an interesting conformational transition from β-strand to α-helix prior to complete stretching of the proteins in both the cases. The secondary structure alteration from β-strand to α-helix was observed more in case of the mutant-protein. However, the α-helices were not stable in either case though their lengths were quite significant. In conclusion, we observed that there was little mechanical resistance from the first and last two β-strands and the mutant yielded a less stable intermediate in the process of unfolding. The SMD studies were in good agreement with single molecule force spectroscopy results.

Chapter V

Quantum Chemical Studies on Graphene/GO sheets

5.1 Introduction

Graphene is a 2D planar sheet of sp^2-hybridized carbon atoms covalently bonded and arranged in a honeycomb lattice, having excellent physicochemical properties (Balandin *et al.*, 2008; Bolotin *et al.*, 2008; Hernández Rosas *et al.*, 2011; Service, 2009). Due to these properties a graphene sheet has various applications in the fields of electronics; energy storage and conversion, engineering nano-composites, *etc.* A single layer graphene is found to be more reactive than the multi layered stack, with the edges of the graphene showing more reactivity than its surface. These nanosize 2D graphene sheets have huge bio-applications such as biosensors, drug delivery, cellular imaging, *etc.* due to their edges.

Bulk scale production of graphene for commercial applications has always been a challenge for the researchers, a perfect solution to which is not known. However, a consecutive oxidation and reduction method appears to be a good solution (Bagri *et al.*, 2010; Eda *et al.*, 2008; Gilje *et al.*, 2007; Larciprete *et al.*, 2011; Stankovich *et al.*, 2006). Upon oxidation, graphite readily exfoliates as single sheets in water, forming graphene oxide (GO). Chemical or thermal reduction of these GO to obtain graphene is an economical and large-scale production method (Bao *et al.*, 2012; Eda *et al.*, 2008; McAllister *et al.*, 2007; Stankovich *et al.*, 2007; Xu *et al.*, 2008; Zhu *et al.*, 2010). The mechanism for an oxidative breakup of graphene sheet has already been proposed (Ajayan and Yakobson, 2006; Li *et al.*, 2006), where, the epoxy groups formed during oxidation were shown to have aligned in a line inducing a rupture in the underlying C–C bonds (Li *et al.*, 2009). During the oxidation procedure, many functional groups such as –COOH, –O–, –OH, –C=O, *etc.* get attached onto the graphene sheet, leading to breaking of its intrinsic C–C π bonds which in turn causes the GO sheet to automatically tear into smaller pieces of random shape and sizes after reduction (Liu *et al.*, 1998; McAllister *et al.*, 2007). The exact pattern of functional group positions on the graphene skeleton is still unidentified but some NMR studies on GO show that carbonyl (–C=O and –COOH) are preferably present at the edges (Cai *et al.*, 2008; Lerf *et al.*, 1998), whereas, hydroxyl (–OH) and epoxy (–O–) groups are abundant on the faces of the GO sheet.

The GO contains oxygenated functional groups on its basal plane and edges making it functionalized through covalent interactions, which in turn makes the GO suitable as a building block for synthesizing functional nanomaterials (Cai *et al.*,

2008; Dreyer *et al.*, 2010; Loh *et al.*, 2010). The presence of many structural defects and functional groups on reduced GO makes it appropriate for various electrochemical usage (Shao *et al.*, 2010). Moreover, the single-layer graphene sheets obtained by tearing of GO mostly is terminated at the edges by functional groups like –OH, –NH2, –COOH, *etc.* to maintain valency/neutrality.

It is known that an arbitrarily cut graphene sheet has two edges; arm-chair (*cis*) and zig-zag (*trans*) edge (Banerjee *et al.*, 2005; Kobayashi *et al.*, 2005) with distinctly different electronic properties (Banerjee *et al.*, 2006). The *trans*-edge is more hydrophillic as compared to the overall *cis*-edge, which has been established both experimentally and theoretically (Banerjee and Bhattacharyya, 2008; Panigrahi *et al.*, 2011; Sarkar *et al.*, 2010). It was theoretically shown that chain of epoxy groups are responsible for oxidative unzipping of GO sheets (Li *et al.*, 2006; Li *et al.*, 2009; Sun and Fabris, 2012; Xu and Xue, 2010; Yang *et al.*, 2007) but none of the studies indicated any preference in generation of *cis*-edge or *trans*-edge after reduction

A nano-scale graphene sheet has hydrophobic faces and polar edges (Banerjee *et al.*, 2005; Panigrahi *et al.*, 2011; Yang *et al.*, 2007), with *trans*-edge being more polar. Thus, functionality of a graphene sheet depends predominantly on the type of edge, *i.e.* a graphene sheet with higher amount of exposed *trans*-edges has higher hydrophilicity and in turn might have higher chances of functionalization. Since, molecular recognitions in biological systems are based on hydrophobic and hydrophilic interactions; the edge-type of the graphene sheet becomes important. In this study we have therefore, carried out quantum chemical calculation involving graphene/GO of different sizes and types to address in particular the effect of type of edge on the process of GO formation and thereafter its tearing. We have shown that after complete tearing of the GO, smaller size graphene sheets graphene sheets are formed predominantly with *trans*-edge. Generation of large number of *trans*-edges upon reduction causes an increase in magnetism and these experimental magnetization data have supported out theoretical predictions.

Furthermore, graphene has proved to be an attractive material for desalination filters because of its high tensile strength and low thickness. Recent investigations reveal that single-layer nano-graphene can effectively filter NaCl salt from water. It has been suggested that presence of commonly occurring –OH functional groups at the edges of graphene/functionalized graphene sheets can almost double the water

flux due to their hydrophilicity. The water permeability of these graphene-based materials is much higher than conventional reverse osmosis membranes (Cohen-Tanugi & Grossman, 2012).

5.2 Tearing of Graphene Oxide Sheets

5.2.1 Methodology

Graphene sheets of different dimensions were obtained by varying the number of basic unit of benzene rings –(i) 3-ring (3 X 3), (ii) 4-ring (4 X 4), (iii) 5-ring (5 X 5), (iv) 6-ring (6 X 6) and (v) 8-ring (8 X 8). The atomic coordinates of all the C-atoms arranged in hexagonal lattice with H-terminated *trans*- and *cis*- edges were generated by the molecular modeling software MOLDEN (Schaftenaar and Noordik, 2000). Standard bond lengths (b(C–C) = 1.421Å and b(C–H) = 1.009Å) and all angles equal to 120° were considered for the modeling. There are studies reporting that in absence of proper termination at the edges (Koskinen *et al.*, 2008), the graphene sheet with dangling electrons would behave like a radical, increasing the complexity of the study. Thus, the edges were terminated with H-atoms forming –C–H bonds so as to neutralize the valencies of all the C-atoms. We successfully optimized all the graphene sheets except the (8 X 8)-sheets, which were optimized by standard dispersion corrected density functional theory (DFT-D) based method ωB97XD/6-31G(2d,2p) (Chai and Head-Gordon, 2008) using Gaussian09 (g09) (Frisch *et al.*, 2009). The (8 X 8) GO system containing five epoxy groups, was optimized with B3LYP/6-31G(2d,2p) (Becke, 1993; Rassolov *et al.*, 2001; Stephens and Greene, 1994) as it was computationally difficult for such a large system to be optimized by the expensive ωB97XD/6-31G(2d,2p) method. Considering overall properties of the (3 X 3), (4 X 4) and (5 X 5) graphene sheets, and our available computational power, the (5 X 5) graphene sheets were considered as standard miniature model for carrying out further studies. However, in some cases (4 X 4), (6 X 6) and (8 X 8) systems were used as standard models for complete understanding of the processes under study.

Several GO sheets of (5 X 5) dimensions were then modeled with different number of epoxy (–O–) groups placed at desired specific positions on the earlier optimized (5 X 5) graphene sheet with b(C–O) = 1.40Å and epoxy angles a(C–O–C) = 110°. A few GO sheets of dimension (4 X 4), (6 X 6) and (8 X 8) were also

developed (APPENDIX III), whenever required following the same protocol for both modeling and optimization.

Completion of geometry optimization was confirmed by calculating the vibrational frequencies of the respective optimized structures. Diagrammatic representations of all the optimized geometries are provided at the end of this section in **Table 5.5**. Since, electronic distribution within the highly conjugated graphene sheet is sensitive to its environment; Natural Bond Orbital (NBO) (Glendening *et al.*, 2012; Reed *et al.*, 1988) analysis was carried out to understand the charge distribution on the graphene/GO sheets and thereby the effect of epoxidation. The –C–C– bonds in graphene/GO sheets have extended conjugation and partial double bond character which might have an influence on the tearing pattern. Thus, bond orders of the systems were also calculated following the Wiberg method (Wiberg, 1968) using g09.

5.2.2 Results

Graphene is generally an infinite 2D sheet but since, we were specifically interested in understanding the structural changes it undergoes during the tearing of GO, we adopted dispersion corrected-DFT formalism suitable for small molecules without periodicity; using localized atomic orbital as basis set instead of the plane wave pseudopotential method. A graphene sheet has two different types of edges **(Figure 5.1a)** involving four different types of –C–C– linkages; indicated in this work as H–C–(C–C)–H, H–C–(C–C)–C–H, H–(C–C)–H and *corner*-H–(C–C)–H **(Figure 5.1b)**. The *trans*-edge involves only H–C–(C–C)–H bonding whereas, the *cis*-edge contains both H–(C–C)–H and H–C–(C–C)–C–H types of bonds. The corners of the graphene sheet also have similar H–(C–C)–H linkage but its properties are slightly different, hence named *corner*-H–(C–C)–H.

We started by optimizing the (3 X 3), (4 X 4), (5 X 5) and (6 X 6) graphene sheets and observed that all the sheets retained their geometric planarity and maintained consistency of –C–C– bond length and bond order. The NBO charges of the C-atoms were similar in all the graphene sheets independent of its size **(Table 5.1)**. The previous work (Panigrahi *et al.*, 2011) from our group, suggested that H-bond interaction between a water molecule and a graphene sheet is stronger at the *trans*-edge. This is in accordance with our present NBO charge analysis that the *trans*-edge has comparatively higher net positive charge which attracts the partially

negatively charged oxygen atom of water molecule. Hence, we decided to consider the (5 X 5)-sheet as a standard for proceeding with our work.

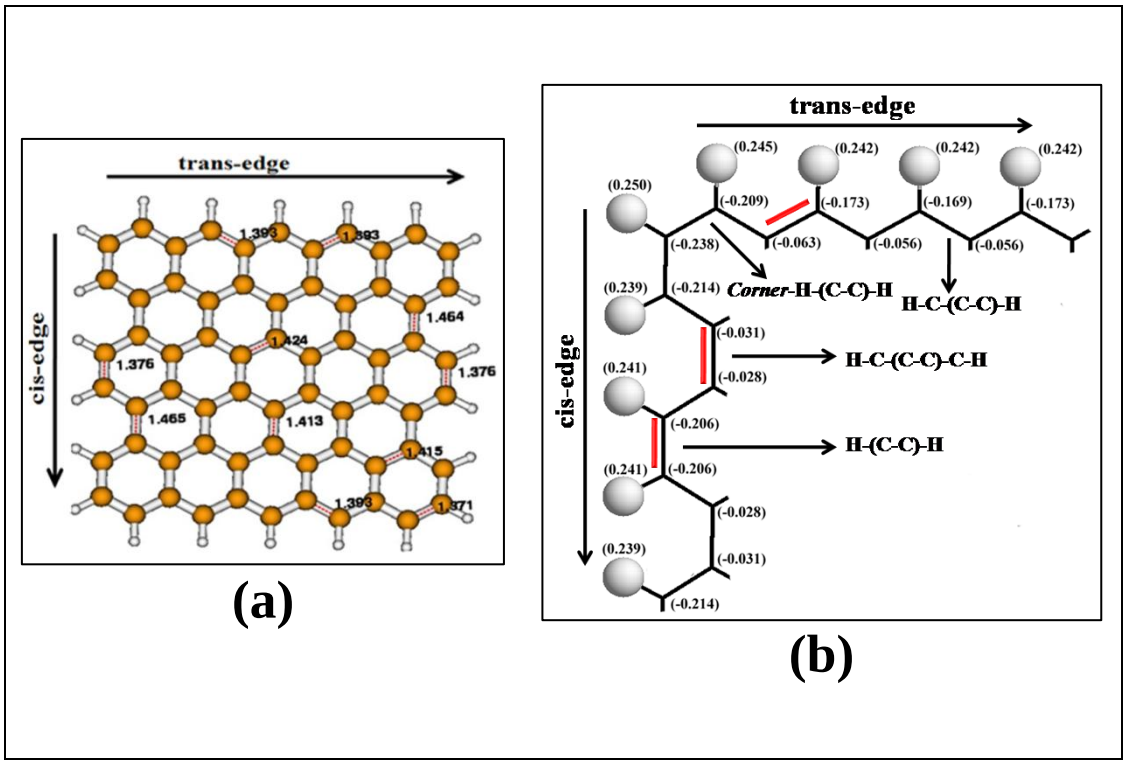


Figure 5.1: (a) A graphene sheet contains *cis*- and *trans*- edges marked by different C–C bond lengths (b) The *trans*-edge is formed by H–C–(C–C)–H linkage and *cis*-edge is formed by both H–C–(C–C)–C–H and H–(C–C)–H linkages. Also, there exists H–(C–C)–H bonds at the corners, atoms of which have a slightly different property, thus labeled as corner-H–(C–C)–H linkage. The NBO charges on terminal C- and H-atoms atoms clearly indicate the polarity difference in the two edges.

Along with difference in bond length, bond orders and NBO charges, the C–C bonds at the edges can also be distinguished on the basis of Infra-red (IR) absorption frequencies **(Figure 5.2)**. The normal mode analysis for the (5 X 5)-graphene sheet showed comparatively higher IR intensity for the frequencies (~3220-3240 cm^{-1}) that correspond to the *cis*-edge C–H bond stretching as compared to the *trans*-edge C–H bonds **(Figure 5.2)**. Thus, for a graphene sheet having two distinct edges *viz.*, *cis*-edge and *trans*-edge with different electrostatic and structural properties, epoxidation might lead to placement of epoxy groups at either of the two edges or both the edges simultaneously.

Table 5.1: The average bond lengths and bond orders of –(C–C)– bonds in different type of linkages along with the NBO charges on these C- and H-atoms are given. The standard deviations are given in bracket.

dim of graphene sheet	edge-type[1]	–C–C– bond length (Å)	–C–C– bond order	NBO charge on terminal H-atom	NBO charge on C-atoms
(3 X 3)	**H–(C–C)–H**	1.381	1.476	0.246 (0.00)	-0.215 (0.00)
	H–C–(C–C)–H	1.393 (0.02)	1.373 (0.11)	0.242 (0.00)	-0.057 (0.00)[a] -0.185 (0.00)[b]
	H–C–(C–C)–C–H	1.474	1.08	0.239 (0.00)	-0.028 (0.00)
	***corner*-H–(C–C)–H**	1.358 (0.01)	1.626 (0.10)	0.247 (0.00)	-0.218 (0.01)
(4 X 4)	**H–(C–C)–H**	1.387 (0.00)	1.418 (0.00)	0.243 (0.00)	-0.209 (0.00)
	H–C–(C–C)–H	1.394 (0.02)	1.365 (0.11)	0.242 (0.00)	-0.057 (0.00)[a] -0.182 (0.00)[b]
	H–C–(C–C)–C–H	1.469 (0.00)	1.088 (0.00)	0.240 (0.00)	-0.028 (0.00)
	***corner*-H–(C–C)–H**	1.355 (0.01)	1.643 (0.09)	0.247 (0.00)	-0.224 (0.02)
(5 X 5)	**H–(C–C)–H**	1.376 (0.00)	1.395 (0.06)	0.241 (0.00)	-0.206 (0.00)
	H–C–(C–C)–H	1.397 (0.00)	1.094 (0.00)	0.242 (0.00)	-0.064 (0.00)[a] -0.172 (0.00)[b]
	H–C–(C–C)–C–H	1.464 (0.00)	1.330 (0.01)	0.240 (0.00)	-0.029 (0.01)
	***corner*-H-(C-C)-H**	1.371 (0.00)	1.523 (0.00)	0.247 (0.00)	-0.223 (0.01)
(6 X 6)	**H-(C-C)-H**	1.380 (0.00)	1.452 (0.00)	0.242 (0.00)	-0.205 (0.00)
	H–C–(C–C)–H	1.396 (0.01)	1.345 (0.07)	0.242 (0.00)	-0.061 (0.00)[a] -0.179 (0.00)[b]
	H–C–(C–C)–C–H	1.462 (0.00)	1.104 (0.01)	0.241 (0.00)	-0.029 (0.00)
	***corner*-H–(C–C)–H**	1.357 (0.01)	1.628 (0.09)	0.247 (0.00)	-0.223 (0.02)

[1]*Four different types of linkages are possible along the two edges of a graphene sheet. The H-C-(C-C)-H linkage has two different types of C-atoms, viz., one C-atom is attached to two other C-atoms and a C-atom is attached to a H-atom having different charges.*
[a]*charge on the C-atom which is attached to two other C-atoms.*
[b]*charge on the C-atom that is attached to a H-atom.*

During epoxidation the epoxy groups are certainly placed on the two faces of a graphene sheet along with the edges. However, during oxidation of graphitic material the edges are far more accessible than the faces. Thus, in this study we have mimicked epoxidation by initially considering only the edges and then moved onto the faces.

The C-C bonds of the *trans*-edge retained the length of 1.40Å even after optimization of the (5 X 5) nascent graphene sheet but some of the *cis*-edge C–C bond lengths were increased to 1.464Å and some decreased to 1.376Å. This was in accordance to the presence of two types of C–C bonds: H–(C–C)–H and H–C–(C–C)–C–H along the *cis*-edge. This also indicated an increase in number of possible epoxidation sites giving rise to three possibilities as positions for epoxidation along

the two edges, *viz.*, one along the *trans*-edge and two others along the *cis*-edge **(Figure 5.3)**.

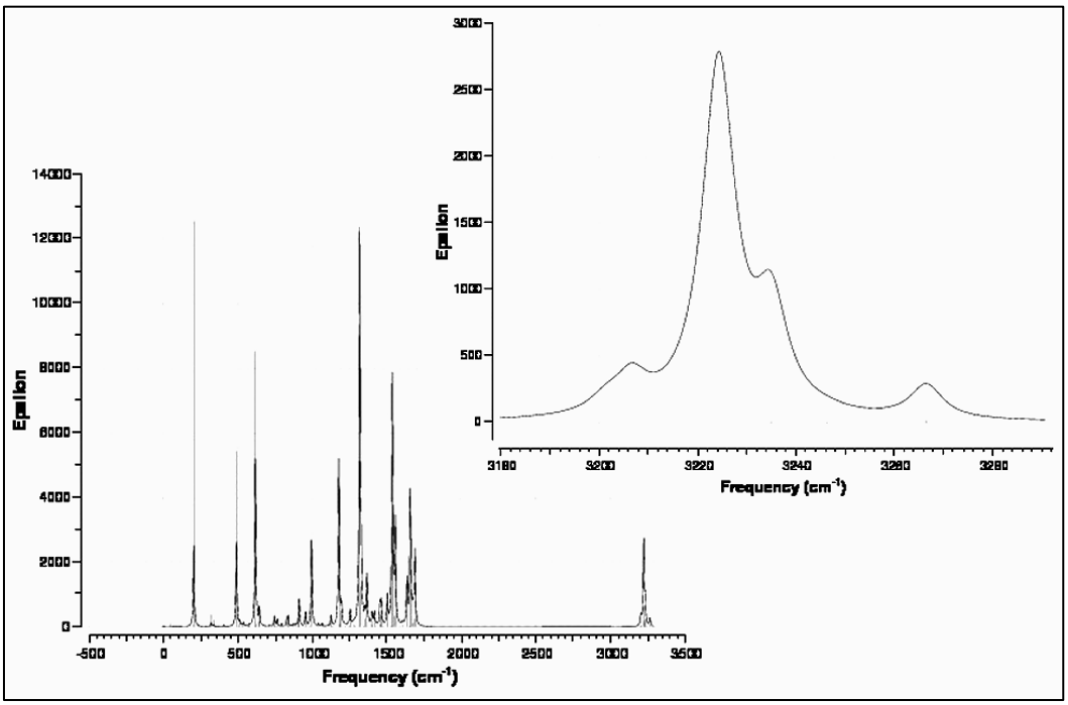


Figure 5.2: The IR spectra corresponding to a (5 X 5) graphene sheet. Frequencies less than 2000cm^{-1} correspond to the –C–C– bond movements in the graphene plane whereas, frequencies above 3000 cm^{-1} correspond to the vibrations of terminal C–H bonds. In the inset plot, the highest intensity peak at 3224.23cm^{-1} and other peaks above this frequency correspond to the *cis*-edge C–H bond vibrations whereas the *trans*-edge C–H bond vibrations are indicated by lower intensity peak at 3206.72cm^{-1}.

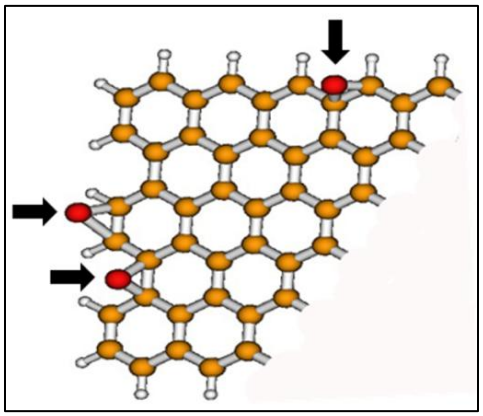


Figure 5.3: The three possible positions for epoxidation.

Thus, considering the probability that tearing of GO may be influenced by the position of the epoxy group on the graphene sheet, we firstly generated three GO models. Each of these models had one epoxy group placed along the edges of the optimized (5 X 5) nascent graphene sheet. At the *cis*-edge epoxy groups were placed

at two distinct positions such that: (i) the epoxy group is linked to the C*-atoms of the H–C–(C*–C*)–C–H linkage as shown in *model-1*; and (ii) the epoxy group is linked to the two C*-atoms of H–(C*–C*)–H linkage as in *model-2*. The optimization of both these GO models indicated that *model-1* was energetically favourable over *model-2* by 10.60kcal/mol, *i.e.* epoxidation is preferred at the H–C–(C*–C*)–C–H linkage along the *cis*-edge. Moreover, the optimization of *model-3* wherein an epoxy group was placed along the *trans*-edge involving C*-atoms of the H–C–(C*–C*)–H linkage indicated that *model-1* was favoured by 7.99kcal/mol over this system (*model-2)* too. The optimized structure of *model-1,* showed complete breakage of the C–C bond involved at the epoxidation site inducing a huge bending in the GO sheet, thus indicating an initiation of the tearing process. There was however a small increment in the C–C bond distance at epoxidation site at the *trans*-edge of the optimized *model-3* with slight bend in the GO sheet. Both these scenarios elucidate that during epoxidation of graphitic material, initially the exposed graphene layers get oxidized first introducing a bend/kink in that sheet that readily exposes the lower sheets to oxidation ultimately leading to tearing and exfoliation into individual graphene/GO sheets.

We next tried to figure out why the optimized *model-1* is energetically most favoured. The NBO charges for the three atoms (C*, O and C*) forming the epoxy linkage (–C*–O–C*–) indicated that in *model-1* the partially negatively charged oxygen was attached to two partially positively charged carbon atoms, however in *model-2* and *model-3* the concerned carbon atoms have very low positive charge. Also, the molar heat capacity at constant volume for optimized *model-1* was found to be the lowest confirming that it is energetically easier to for epoxidation and in turn tearing to happen as in *model-1* **(Table 5.2)**. Moreover, the NBO charges on the C-atoms of the (5 X 5) nascent graphene sheet, clearly showed that the C*-atoms of the H–C–(C*–C*)–C–H linkage have lowest negative charge making them most suitable for initiation of epoxidation **(Figure 5.1b)** and; the bond orders indicated that the C–C bonds formed by the same C*-atoms of H–C–(C*–C*)–C–H linkage is weak with low partial double bond character, making these bonds a preferable tearing site **(Table 5.1)**. Thus, the GO sheet with epoxy group at the H–C–(C*–C*)–C–H position of the *cis*-edge seemed to be the best model by far for addressing the effect of type of edge involved in the process of GO formation and ultimately leading to tearing of graphene sheets.

Apart from these situations there exist two more possibilities for positioning of an epoxy group; one along the *cis*-edge (*model-4*) and another along the *trans*-edge (*model-5*). In both these situations an oxygen atom replaces two hydrogen atoms. We developed *model-4* by connecting the two C*-atoms of the H–C*–C–C–C*–H linkage and removing the terminal H-atoms for maintaining valency. It was expected that optimization might give stability to the five member ring, *i.e.* decreasing the associated –(C–C)– bond lengths causing breaks in some of the nearby –(C–C)– bonds. But the optimized geometry of *model-4* showed only slight increase in –(C–C)– bond length of the bonds adjacent to five-member ring. The *model-5* was not studied as it seemed to be possible only theoretically since a four-member ring formation on a graphene plane would be energetically too unstable.

Table 5.2: Comparison of three models: *model-1*, *model-2* and *model-3* on the basis of molar heat capacities at constant volume for the optimized structures along with NBO charges on the atoms involved in epoxidation.

model	Charge of the atoms forming epoxy linkage			CV (molar heat capacity at const. vol.) in Cal/mol-K
	C*	O	C*	
model-1	0.32	-0.521	0.315	169.747
model-2	0.045	-0.517	0.045	169.971
model-3	0.071	-0.518	0.179	170.059

C-atoms are the C-atoms connected via epoxy linkage -C*-O-C*-*

Usually under experimental conditions the epoxidation of a graphene sheet is not restricted to a single epoxy group but leads to placement of a number of such groups on the sheet (He *et al.*, 1998). Thus, obtaining smaller graphene sheets from tearing of large size GO sheets require a number of epoxy groups to be present. The NBO charges and bond order calculations performed on optimized *model-1* indicated that the C–C bonds next in parallel to the first epoxidation site have lowest bond order along with corresponding carbon atoms having lowest negative charge **(Figure 5.4)** making them susceptible to further epoxidation. In order to confirm this, we developed two more models by placing another epoxy group on the optimized *model-1*, but at different positions (*model-6a* and *model-6b*). The optimized *model-6a* containing two epoxy groups at parallel positions proved to be energetically

favourable by 44.52kcal/mol compared to *model-6b* containing the two groups in nearly adjacent positions.

Furthermore, with the aim to clarify the influence of number of epoxy groups on tearing of GO, we optimized two more model systems where four epoxy groups were placed, starting from the *cis*-edge (*model-7a*) and *trans*-edge *(model-7b)* respectively. All the epoxy groups were placed on the same side of the sheet for both the models and only the H–C–(C*–C*)–C–H linkage of the *cis*-edge was considered while building *model-7a* as tearing along it was arguably more favoured. Optimized structures of both *model-7a* and *model-7b* showed breakage in the C–C bonds at the epoxidation sites along with a bending in the sheet, although a detail clearly suggested that tearing of GO sheet was favoured by 50.39kcal/mol, when the oxidation occurs involving the *cis*-edge. Another significant observation was that with increase in number of epoxy groups the possibility of tearing GO sheet also increases. The C–C bond of *trans*-edge with only one epoxy group was not broken in *model-3,* whereas the similar C–C bonds were elongated when four epoxy groups were lined up from the *trans*-edge.

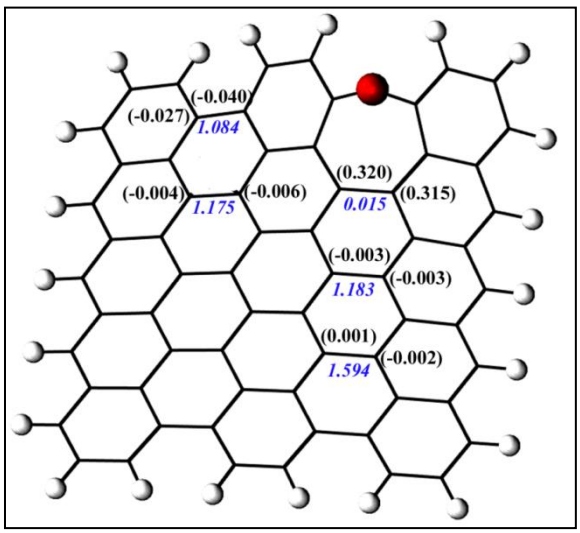


Figure 5.4: After breaking of a –C–C– bond due to epoxidation, the properties like charge on C-atoms and bond order of –C–C– adjacent to the epoxidation site changes. The values given in bracket are NBO charges of the corresponding C-atoms and the values in italics and blue are bond order of the –C–C– bonds.

Until now we optimized systems involving epoxy groups at the edges but epoxidation occurring only on the plane and away from the edges of the graphene sheet may be important. Hence, we developed *model-8a* and *model-8b* by removing

epoxy groups from the *cis*-edge of *model-7a* and trans-edge of *model-7b* respectively. Only three epoxy groups were taken this time to reduce the computational expense as the effect of number of epoxy groups had already been addressed. The analysis of optimized structures corresponding to these two models showed that *model-8a* was energetically favoured over *model-8b* by 15.96kcal/mol, generalizing the earlier result that tearing is easier with epoxidation involving the influence of *cis*-edge types.

We also tried to probe into directionality of the tearing pattern by developing two model systems: *model-9a* and *model-9b* with (4 X 4) graphene sheet where each contains two epoxy groups, to capture the directional influence with more clarity and lower computational cost. The optimized *model-9a* was energetically favoured over optimized *model-9b* by 30.54kcal/mol establishing that tearing is favoured in a particular direction **(Figure 5.5)** starting from H–C–(C–C)–C–H edge towards the H–(C–C)–H edge. The average –C–C– distance and –C–O–C– angles at the epoxidation sites for *model-7a*, *model-7b*, *model-9a* and *model-9b* **(Table 5.3)** emphasized that tearing is favoured only by breaking-up of particular type of bonds. This allowed us to conclude that tearing is most favoured by breaking the C*–C* bonds of H–C–(C*–C*)–C–H and moving towards H–C*–C*–H. We also modeled and optimized two (8 X 8) GO sheets containing five epoxy groups on the plane of each sheet but none at the edges and with only difference in the position of the epoxy groups, *model-10a* was found energetically favourable to *model-10b* by 33.04kcal/mol, thus generalizing the above stated hypothesis for GO sheets of all sizes.

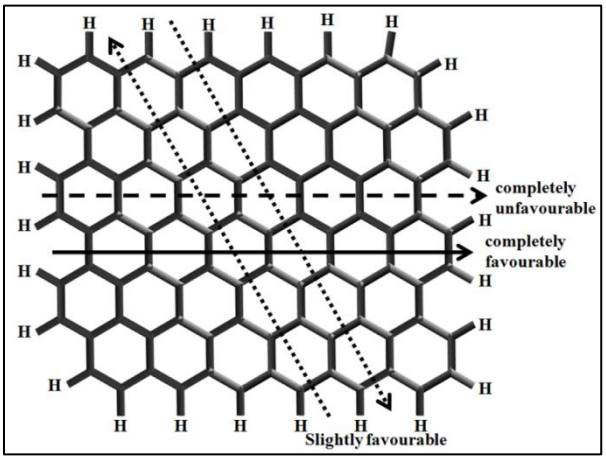


Figure 5.5: Schematic representation of favourable directions for tearing of GO sheet. Epoxidation initiating at H–C–(C*–C*)–C–H linkage favours tearing over epoxidation involving C*-atoms of H–(C*–C*)–H linkage.

Table 5.3: Comparison of four models: *model-7a*, *model-7b*, *model-9a* and *model-9b* on the basis of –C–C– bond distance and –C–O–C– angle at the epoxidation site.

model	C-C distance at epoxy site (Å) [distance between C*-atoms in C*–O–C* link]	C–O–C angle at epoxy site (°) [C*–O–C* angle]
model-7a	2.317 (0.05)	115.775 (3.35)
model-7b	1.913 (0.37)	87.148 (24.94)
model-9a	2.287(0.06)	113.450 (3.84)
model-9b	2.230 (0.08)	109.710 (4.97)

The tearing of GO sheets produce smaller graphene sheets having functionalization at the edges (Schniepp *et al.*, 2006). These smaller graphene sheets may have either only *cis*- or *trans*-, or both the edges. The following part of our study involved prediction of the type of edge formation, after the process of tearing is completed, *i.e.* to theoretically predict which edge formation is more plausible after complete tearing **(Figure 5.6)**. Keeping in mind that it was impossible to place four epoxy groups on a (5 X 5) graphene sheet along the most preferred position (*cis*-edge H–C–(C*–C*)–C–H influence) without involving the edges, we developed *model-11a1* which had somewhat H–C–(C*–C*)–C–H influence. The bond length, bond-order, IR frequency of C-H stretching and NBO charges on C-atoms of the *corner*-H–(C–C)–H bonds were intermediate between the *cis*-edge and the *trans*-edge. Hence, we developed *model-11a2* containing four epoxy groups on a (5 X 5) graphene sheet but with H–(C*–C*)–C–H linkage influence of the *trans*-edge. Thus, we modeled three GO sheets each with four epoxy groups placed on the same side of the sheet in such a way that tearing of one leads to formation of *cis*-edged sheets *model-11b* and the other two gives a *trans*-edged sheet. The optimizations suggested that tearing of both *model-11a1* and *model-11a2* were preferred over *model-11b* by 35.33kcal/mol and 29.86kcal/mol of energy respectively, indicating that formation of *cis*-edge upon tearing is not favourable. This was in accordance with the observation that the optimized geometries of only *model-11a1* and *model-11a2* showed breaks in the C–C bonds eventually suggesting formation of *trans*-edged sheet after tearing of GO. The average C–C distance and C–O–C angle at the epoxidation sites in optimized geometry of *model-11b* were seen to be 1.866Å (0.39) and 86.007° (25.13) respectively, the same in case of optimized *model-11a1* geometry being 2.089Å (0.35) and 100.034° (22.06) eventually making up for the fact that ultimate tearing of GO will lead to formation of smaller graphene sheets with *trans*-edge. This is in

accordance with the experimental report that thermal reduction favours unzipping process that leads to increase in exposed zig-zag edges (Bagani *et al.*, 2014).

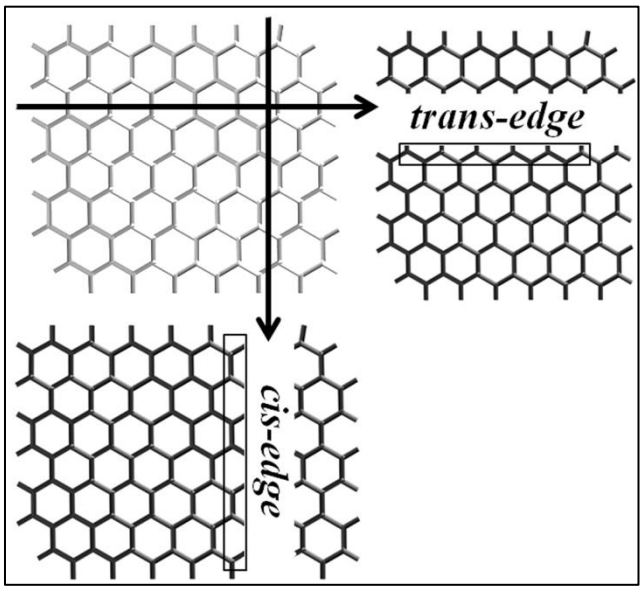


Figure 5.6: Schematic diagram indicating that tearing of a large-size graphene sheet would lead to formation of smaller sheets having either *trans*-edge or *cis*-edge.

Furthermore, experimentally the positioning of epoxy groups can be random; hence the four epoxy groups might not line-up on the same side of the GO sheet. We tackled this aspect by optimizing two more model systems: *model-12a* and *model-12b*, where we placed the alternate epoxy groups on the opposite faces of the graphene sheets. Since, *model-11a1* was preferred over *model-11a2*; the epoxy group placement pattern in *model-12a* was similar to that of *model-11a1*. The analysis of the optimized structures showed that *model-12a* is unflavoured over *model-12b* by 21.3172kcal/mol but unlike in *model-11a1* or *model-11a2* the bonds were intact at the epoxidation sites of both *model-12a* and *model-12b* with slight change in the concerned C-C bond lengths. A comparison of energies of all the five models, *model-11a1, model-11a2, model-11b, model-12a* and *model 12b*, clearly suggested that *model-11a1* was energetically the most favourable situation for tearing of GO.

With the aim to cover up all possible aspects we finally developed two more models; *model-13a* and *model-13b* which had (6 X 6) dimension of the GO sheet and ten epoxy groups randomly placed. In *model-13a* epoxy groups were arranged in a way that upon tearing it may lead to either *cis*- or *trans*-edge smaller sheets or both simultaneously. After its optimization the C-C bonds at the epoxidation sites were broken in such a way that if extrapolated it will lead to formation of only *trans*-edge

smaller sheets **(Figure 5.7a)**. In *model-13b* the epoxy groups were far more randomly placed yet after optimization only those C-C bonds were broken **(Figure 5.7b)** which in future will lead to tearing in a way that theoretically confirms the higher probability of forming smaller sheets with enhanced *trans*-edges.

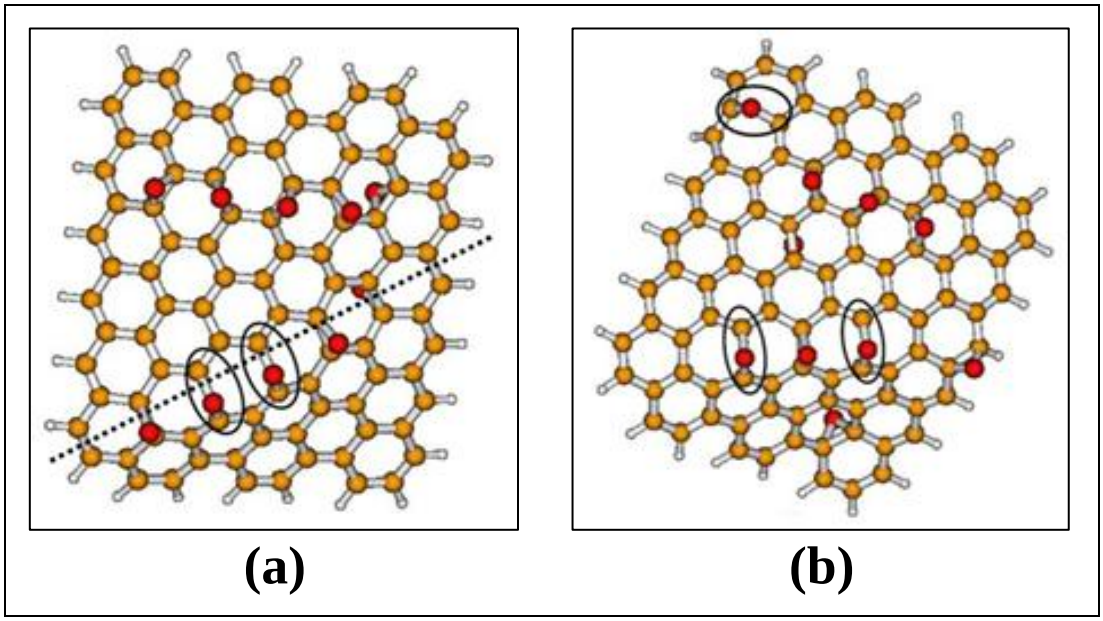


Figure 5.7: Optimized structure of (a) *model-13a* and (b) *model-13b*. The highlighted broken -C-C- bonds at some epoxidation site indicate that tearing will always lead to *trans*-edge sheets.

5.2.3 Discussion

The calculated relative energies of all the model systems are tabulated in **Table 5.4**. The most feasible representative situation (or models) can be understood from this table. The main outcome of this investigation was that we could narrow down the type of edge possible to obtain, after tearing of a graphene sheet. Along with this a preference in the tearing pattern was also noted. The tearing of a graphene sheet prefers to happen along the *cis*-edge with C*–C* bonds of H–C–(C*–C*)–C–H linkage breaking at the lowest cost of energy. This was indicated by the energies obtained from quantum chemical calculations and supported by the NBO charges. A close analysis all possible tearing patterns **(Figure 5.8a)**, suggested that progressive tearing of a large graphene sheet into smaller and smaller ones would lead to smaller graphene sheets with higher percentage of *trans*-edge, *i.e.* –(C–C–C)– bonding predominant at the edges with some functionalization at the terminal carbon atoms. A

closer look at the expected tearing patterns also indicates that there exists a possibility of obtaining triangular graphene sheets containing n^2-number (*i.e.* 9, 16, 25, 36, *etc.*) of carbon atoms. It would be interesting in future to experimentally detect the existence of such a pattern in number of C-atoms forming the nano-graphene sheets. However, as a conclusive statement, it is possible to obtain a functionalized graphene sheet containing only exposed hydrophilic *trans*-edge (zig-zag edge) **(Figure 5.8b)** making the sheet highly exploitable in biological systems.

Table 5.4: A Comparison of the optimized models on the basis of energy. Similar models having same number of atoms and similar pattern of epoxidation have been grouped in one set. Energy of one model in each system is set to zero and the energies of the rest are given in relative to that.

set	Dimension of GO sheet	Total number of			models	Δ(kcal/mol)
		C-atoms	H-atoms	O-atoms		
1	(5 X 5)	64	22	1	*model-1*	0.00
					model-2	10.60
					model-3	7.99
2	(5 X 5)	64	22	2	*model-6a*	0.00
					model-6b	44.52
3	(5 X 5)	64	22	4	*model-7a*	0.00
					model-7b	50.39
4	(5 X 5)	64	22	3	*model-8a*	0.00
					model-8b	30.54
5	(4 X 4)	48	18	2	*model-9a*	0.00
					model-9b	15.96
6	(8 X 8)	160	34	5	*model-10a*	0.00
					model-10b	33.04
7	(5 X 5)	64	22	4	*model-11a1*	-35.33
					model-11a2	-29.86
					model-11b	0.00
8	(5 X 5)	64	22	4	*model-12a*	21.32
					model-12b	0.00

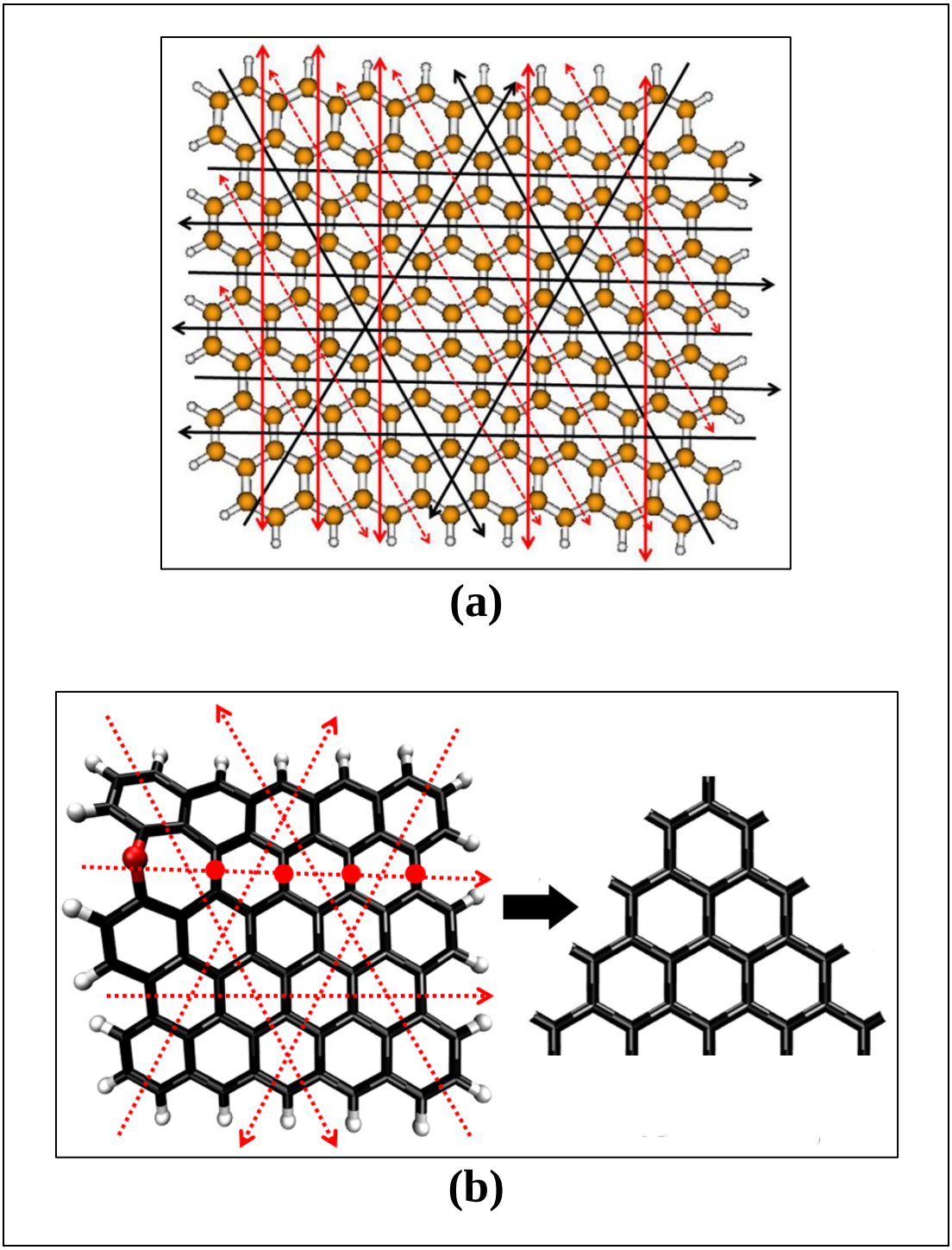
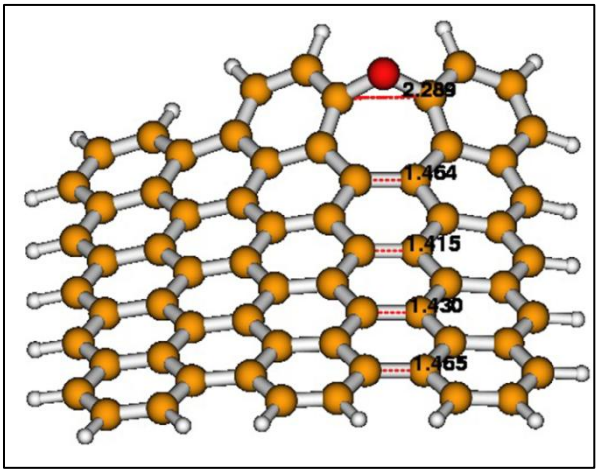
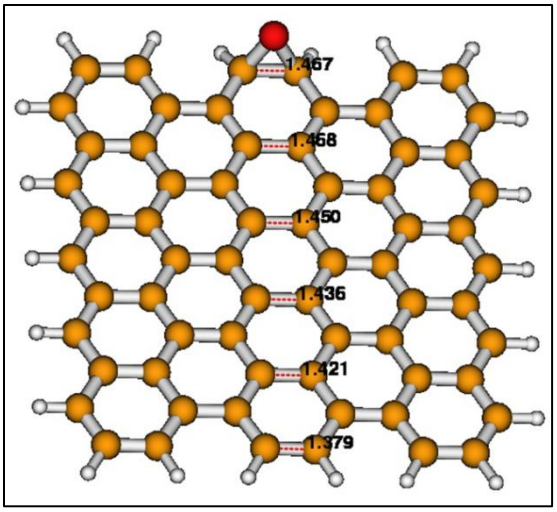


Figure 5.8: (a) All possible tearing patters are indicated. Red (bold) lines indicate the tearing directions/pattern that is not possible. Red (dotted) lines indicate possible tearing pattern but at the cost of high energy whereas, black lines represent the most favourable tearing pattern. (b) Considering only the favourable tearing directions the expected smaller size graphene sheet obtained after complete tearing of a GO-sheet is shown.

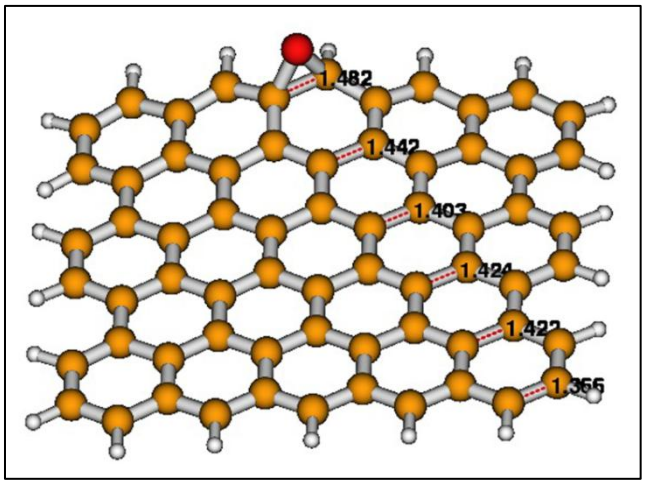
Table 5.5: Optimized geometries of all the model systems



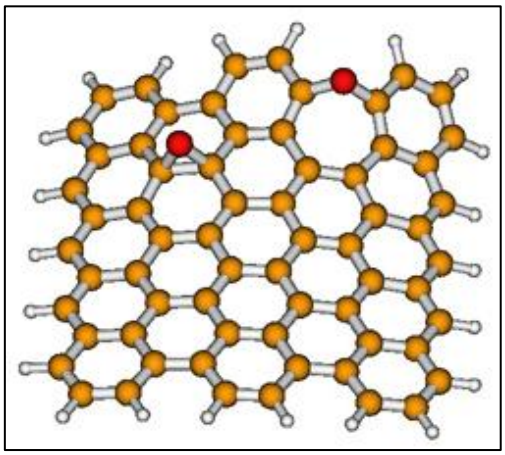
model-1



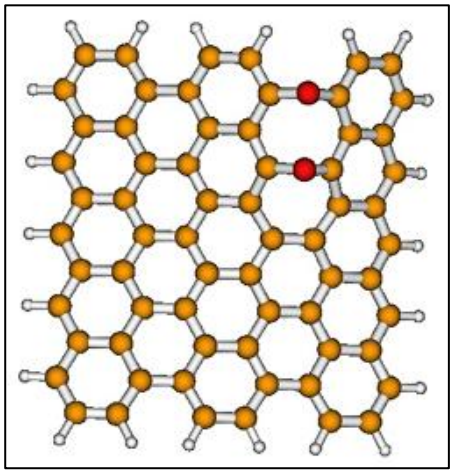
model-2



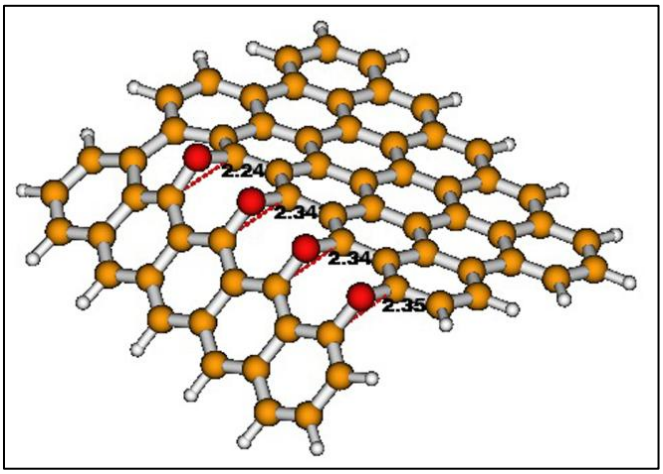
model-3



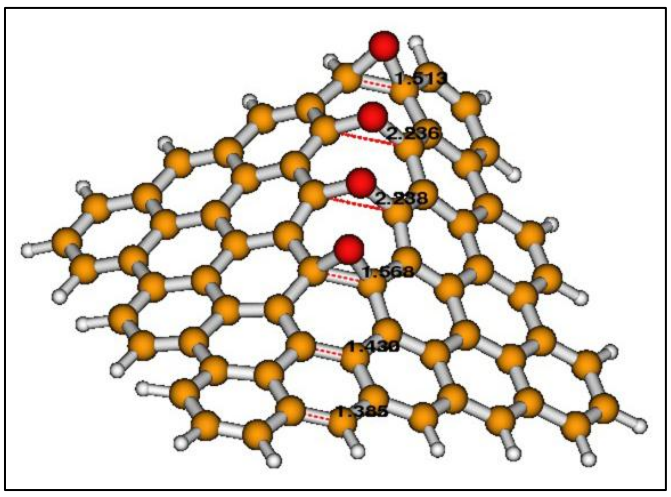
model-6a



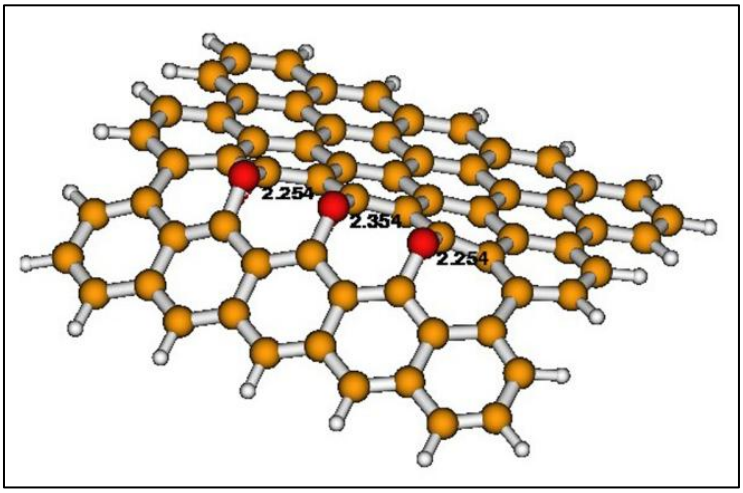
model-6b



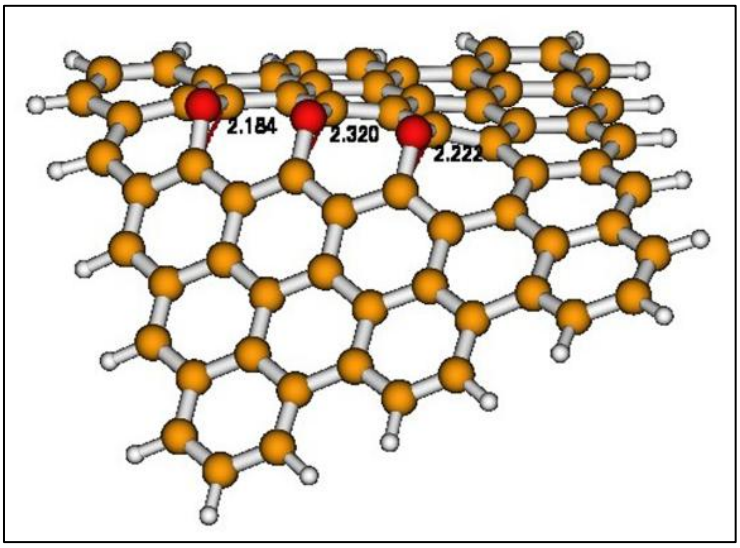
model-7a



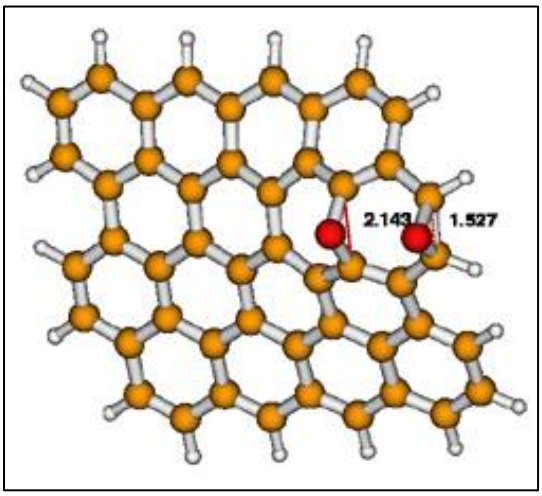
model-7b



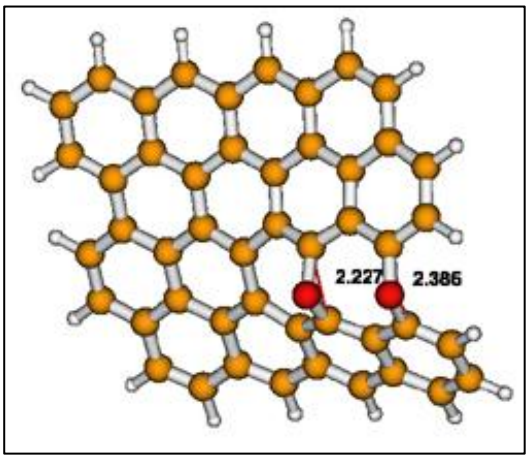
model-8a



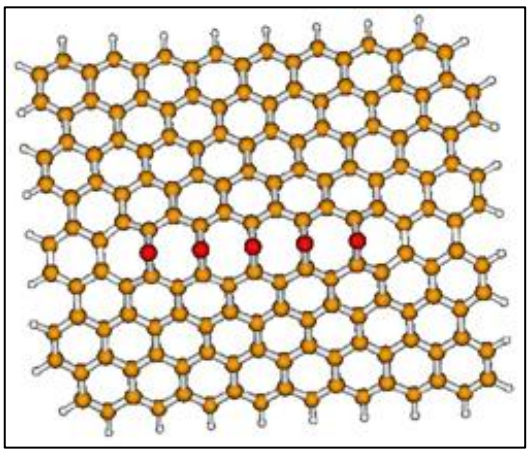
model-8b



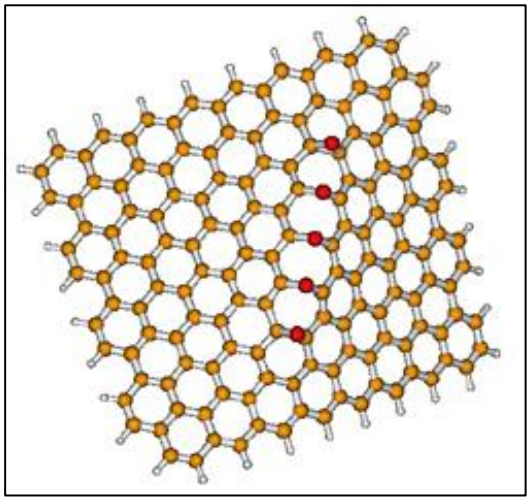
model-9a



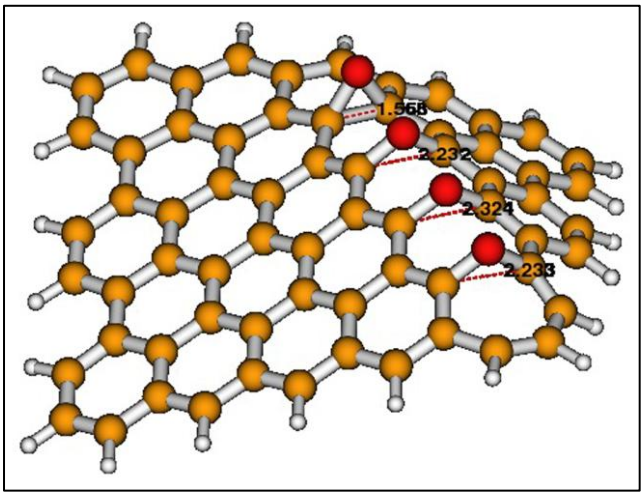
model-9b



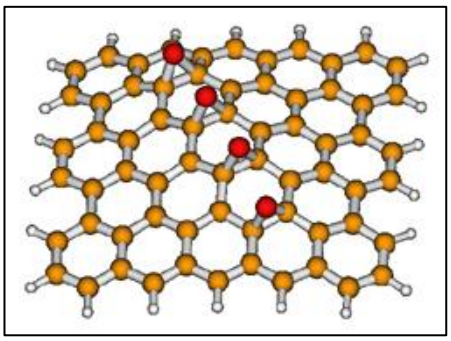
model-10a



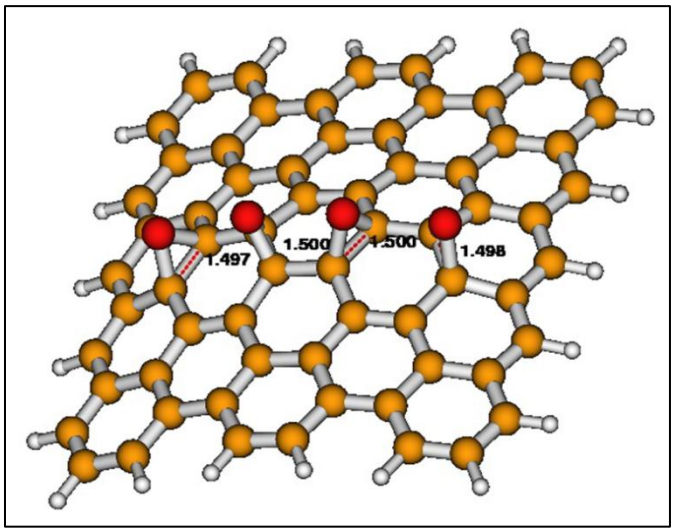
model-10b



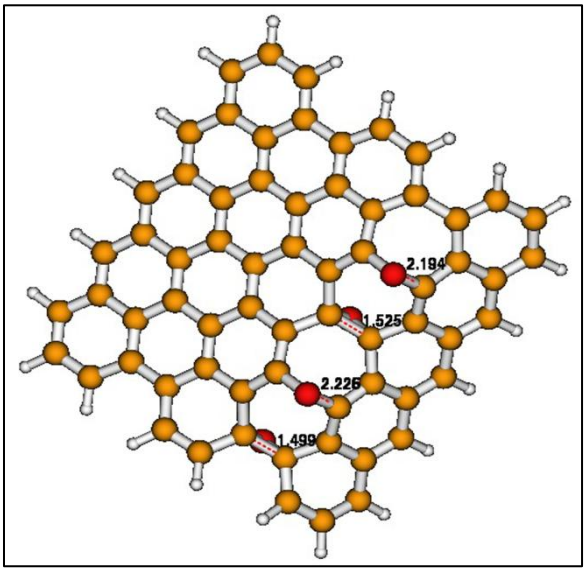
model-11a1



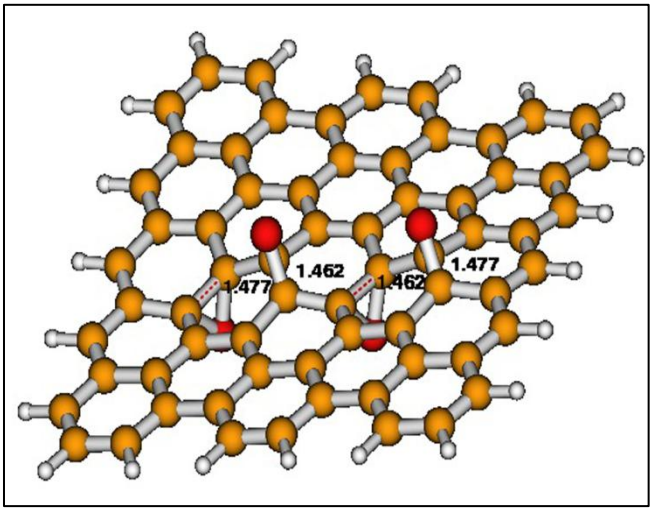
model-11a2



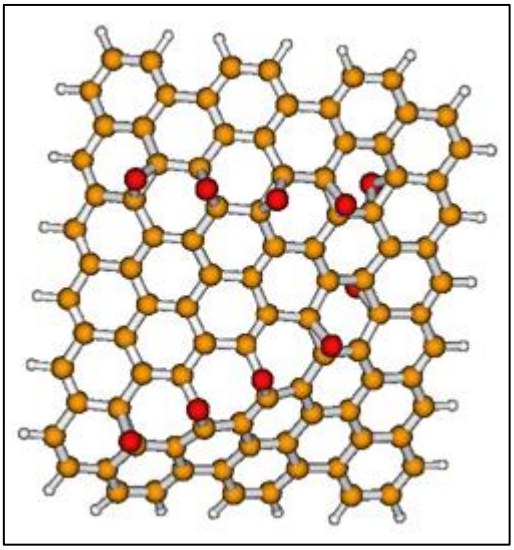
model-11b



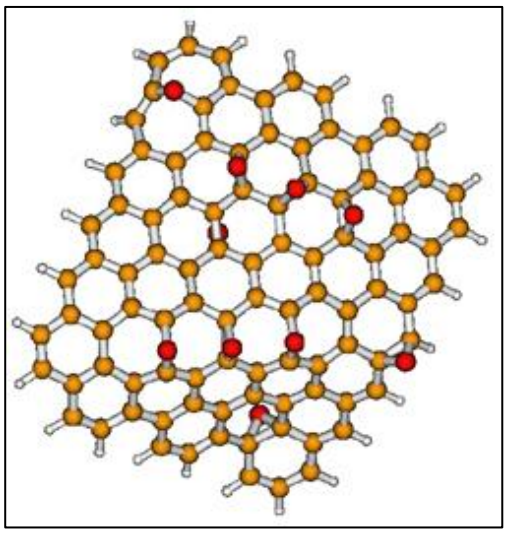
model-12a



model-12b



model-13a

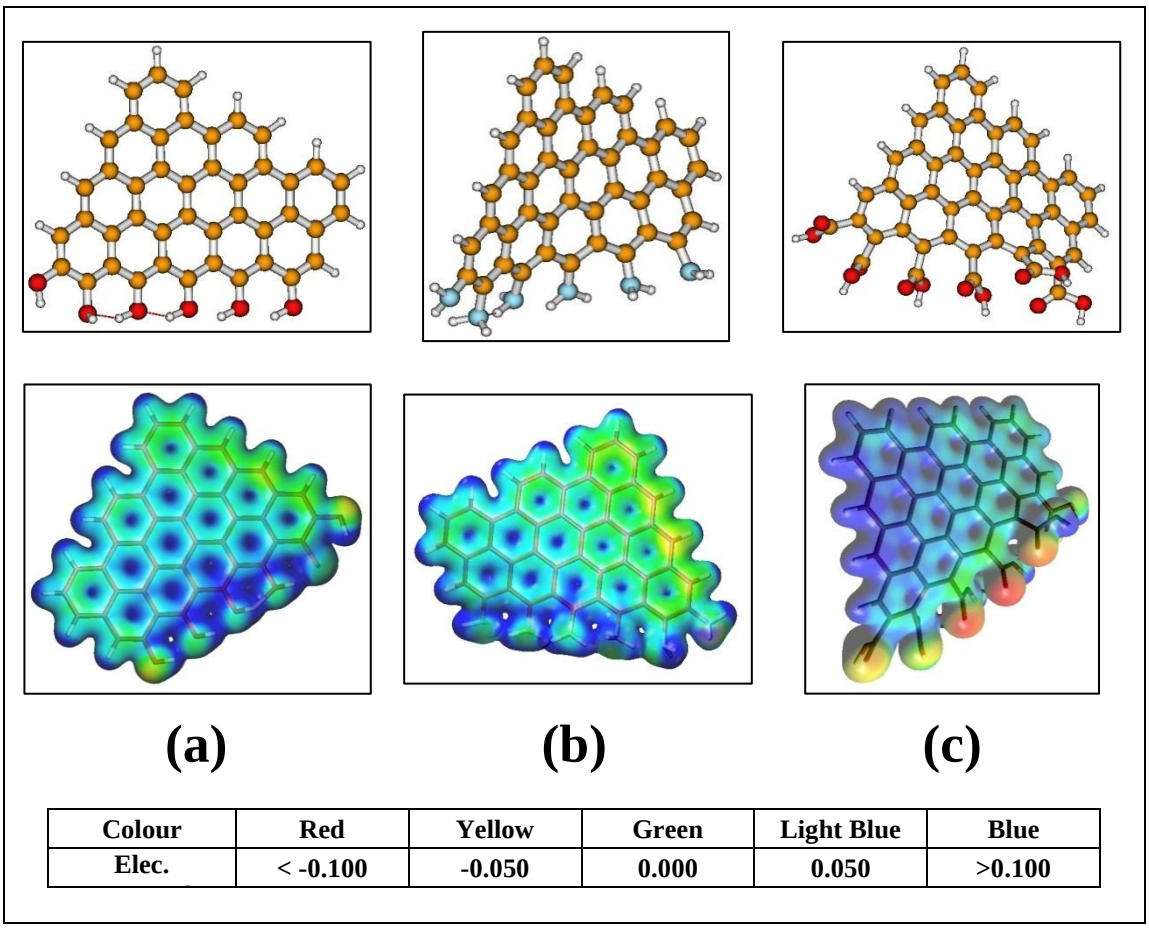


model-13b

5.3 Functionalized Graphene Sheets

As discussed earlier, chemically modified graphenes are obtained by addition of different functional groups to GO platelets by various chemical reactions that provide for covalent or non-covalent attachment to the resulting graphene sheet. Also, during the process of chemical tearing of GO sheets, the edges of the resulting smaller sheets are terminated by functional groups such as –OH, $–NH_2$, –COOH, *etc.* Addition of such functionality to groups already present on GO, makes graphene/GO more versatile precursor for a wide range of applications (Dreyer *et al.*, 2010).

We modeled and optimized three graphene sheets each with functionalization at one of the edges by –OH, –NH2, and –COOH groups respectively. The electrostatic potential surface was then generated which clearly indicated the difference in electrostatic potential **(Figure 5.9)** at the two edges which differ in functionality and the plane of the sheet.



Colour	Red	Yellow	Green	Light Blue	Blue
Elec.	< -0.100	-0.050	0.000	0.050	>0.100

Figure 5.9: Optimized functionalized graphene sheet and corresponding electrostatic potential distribution for edge-termination by (a) –OH, (b) –NH2, and (c) –COOH.

5.4 Graphene Sheet as Filter

5.4.1 Methodology

Our earlier studies revealed that a graphene sheet of (5 X 5) dimension can serve as a standard miniature model representing a nano-graphene sheet (Ray *et al.*, 2015). Thus, in this study we have considered such a (5 X 5) graphene sheet. The atomic coordinates of all the carbon atoms arranged in hexagonal lattice with *trans*- and *cis*-edges were generated by the molecular modeling software MOLDEN (Schaftenaar and Noordik, 2000) using standard bond lengths (b(C C) = 1.421Å and b(C-H) = 1.009Å) and all angles equal to 120°. The edges were terminated with hydrogen atoms forming C-H bonds neutralizing the valencies of all the carbon atoms. This also eliminated the possibility that, the graphene sheet would behave like radical due to presence of dangling electrons, increasing the complexity of the study. We have successfully optimized the three model systems (i) graphene sheet with H_2O molecule, (ii) graphene sheet with Na+ ion and (iii) graphene sheet with Cl- ion, by standard dispersion corrected density functional theory (DFT-D) based method ωB97XD/6-31g(2d,2p) using g09 (Frisch *et al.*, 2009). The H_2O-molecule, Na^+ ion and Cl^- ion were initially place near the (i) *trans*-edge and (ii) *cis*-edge of the graphene sheet.

5.4.2 Results and Discussion

The earlier studies from our group which reported that a water molecule moves towards the *trans*-edge from of the graphene sheet were based on HF and MP2 methods (Panigrahi *et al.*, 2011). Our present study with dispersion corrected density functional theory (DFT-D) based method involving H_2O at the *cis*- or *trans*-edge indicates a similar situation with the optimized graphene(*trans*-edge)-H_2O system being favoured by -5.072kcal/mol over the graphene(*cis*-edge)-H_2O system.

The optimized geometries of the four graphene-ion systems are given in **Figure 5.10**. A comparative energies of graphene(*trans*/*cis*-edge)-Na^+ and graphene(*trans*/*cis*-edge)-Cl^- **(Table 5.6)** indicated that the ion position, independent of its type, is favourable at the *trans*-edge by -15.602 kcal/mol and -19.005 kcal/mol respectively for Cl^- and Na^+ ions.

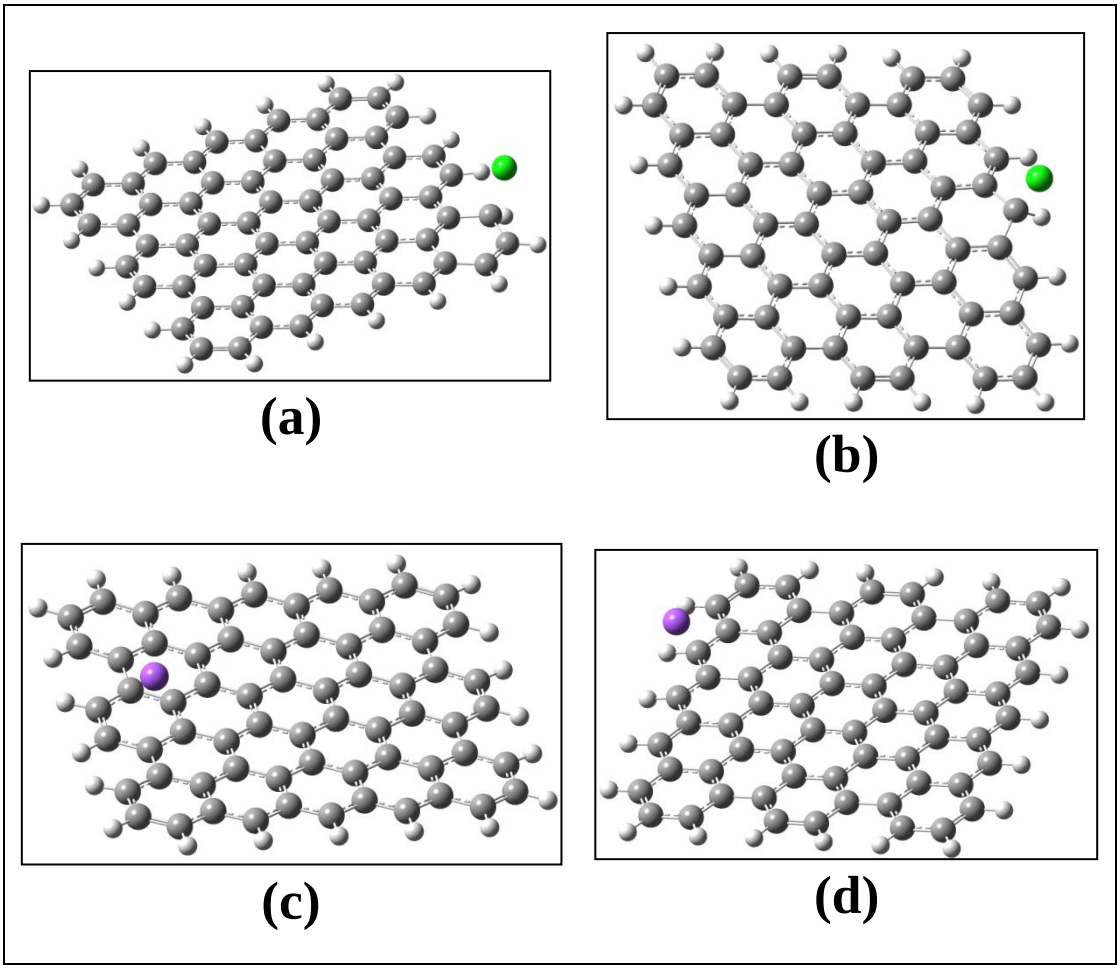


Figure 5.10: Cl^- (green colour) is present at the (a) *cis*-edge and (b) *trans*-edge; and Na^+ (violet colour) is present at the (c) *cis*-edge and (d) *trans*-edge.

Table 5.6: The energies of the systems optimized.

Molecule/Ion	E (in Hartree) of the system when		ΔE (*trans*-edge—*cis*-edge)
	interaction is at *cis*-edge	interaction is at *trans*-edge	
H2O	-2604.240	-2604.248	-5.072
Na^+	-2689.966	-2689.996	-19.005
Cl^-	-2988.136	-2988.161	-15.602

The ions (Na^+ and Cl^-) however, always remained close to their initial positions at the *trans*-/*cis*-edge after optimization. But, depending on the initial position of the H_2O molecule, in some cases it was found that in the optimized system the molecule moved away from the edge and onto the plane of the graphene sheet which may be indicative of water passing through a graphene sheet **(Figure 5.11)**. It is to be mentioned here that this could be due to consideration of dispersion interaction, which dominates when more atoms are in contact, *i.e.* three atoms of a

water molecule coming in close contact with ring carbon atoms of graphene. However, it can clearly be said that ions prefer getting trapped to the *trans*-edge of a graphene sheet.

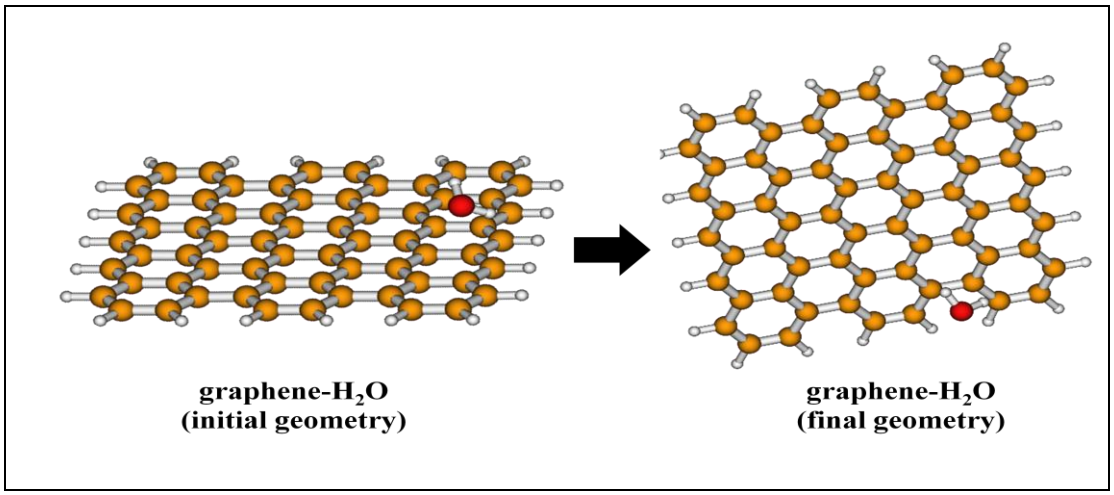


Figure 5.11: H_2O molecule moving from *trans*-edge towards the plane and *cis*-edge of the graphene sheet.

Chapter VI

Study of molecules BDBPZ and NAD^+/NADH by Quantum Chemical Approach

Chapter VI
Section I

Study of H-bonding probe BDBPZ by Quantum Chemical Approach

6.1.1 Introduction

Hydrogen bonds (H-bonds) play a significant role in structure and function of biological systems (Desiraju and Steiner, 2001; Jeffrey, 1997; Schuster *et al.*, 1976). The spectroscopic character of an organic molecule can change significantly when it is involved in H-bond formation. When H-bonding occurs in the electronically excited state, the changes in spectroscopic properties of a fluorophore become even more prominent (Han and Zhao, 2011). Generally fluorophore-solvent interactions are classified into two main categories: (i) non-specific interactions due to dielectric of the medium and (ii) specific interactions such as H-bonds between hydrogen donor and acceptor molecules. The dynamic behaviour of intermolecular H-bonds in the electronically excited state helps in understanding of microscopic structure and function in many molecular systems. Both these interactions play important role during the salvation process. The dynamic behaviour of intermolecular H-bonds in the electronically excited state helps in understanding the microscopic structure and function of many molecular systems which makes the investigation of H-bond dynamics of photoexcited chromophores very important in understanding the effect of environment on their photophysical and photochemical behaviour (Bhattacharyya, 2008; Glasbeek and Zhang, 2004; Pal and Zewail, 2004).

It has already been reported that the fluorophore dibenzo[a,c]phenazine (DBPZ) acts as a polarity insensitive H-bond acceptor probe. Its benzoyl derivative: 11-benzoyl-dibenzo[a,c]phenazine (BDBPZ) **(Figure 6.1.1)** has similar or even better H-bond accepting character as compared to its parent. In addition BDBPZ can also sense the polarity of the solvent medium. On interaction with H-bond donating solvents the fluorescence intensity of BDBPZ increases abruptly and the extent of red shift of λ_{max} depends on H-bond donating ability of the solvent associated. Further the organic amine such as N,N-dimethyl aniline (DMA) interacts very weakly with DBPZ in the first excited singlet state (Dey *et al.*, 2007b). On the contrary DMA can efficiently interact with photoexcited free BDBPZ as well as with the excited state H-bonded species. The mode of interaction is the photo-induced electron transfer from the amine to the fluorophore. Both the improved excited state H-bonding efficiency and better sensibility of the environment are attributed to the structure of the molecule.

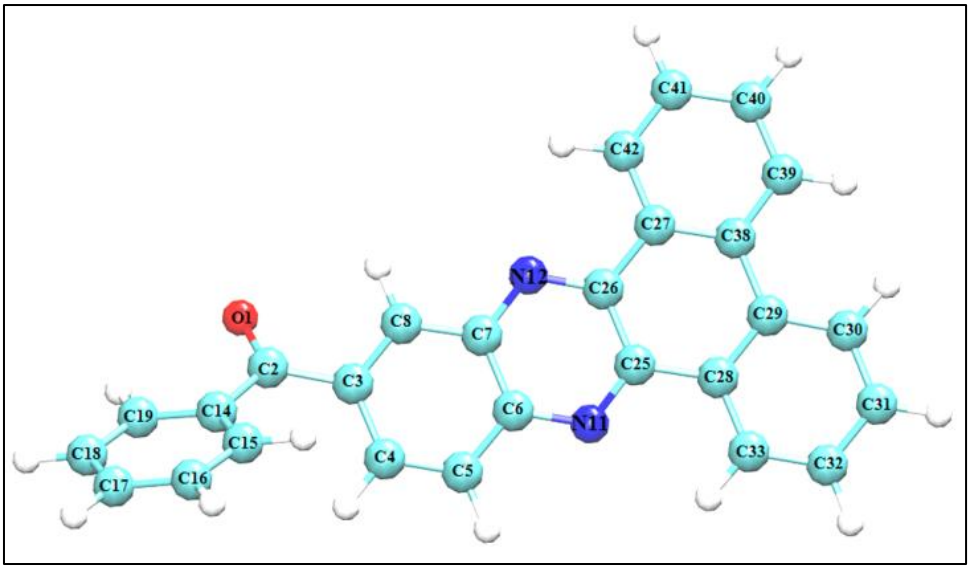


Figure 6.1.1: 11-benzoyl-dibenzo[a,c]phenazine (BDBPZ)

6.1.2 Methodology

The models of free BDBPZ and its various H-bonded complexes were built using MOLDEN software (Schaftenaar and Noordik, 2000) with standard bond lengths (b(C-C) = 1.421Å, b(C-N) = 1.33 Å, b(C-H) = 1.09Å and b(C=O) = 1.22Å) and bond angles. The ground state (GS) optimizations of the modeled structures were carried out with Density Functional Theory (DFT) based method B3LYP/cc-PVDZ (Becke, 1993; Dunning Jr, 1989; Lee et al., 1988) using Gaussian09 (g09) (Frisch *et al.*, 2009). The excited state (ES) geometry optimizations were carried out by Time Dependant-Self Consistent Field (TD-SCF) DFT method using the same functional and basis set. The geometry optimizations were confirmed by calculating the vibrational frequencies of the respective optimized structures. The solvent effect was studied considering the Conductor-like Polarizable Continuum Model (CPCM) (Cossi *et al.*, 2003) for the solvents ethanol (ε = 24.8520) and acetonitrile (ε = 35.6880). Electrostatic Potential (ESP)-fit dipole moment and Natural Bond Orbital (NBO) charges (Reed *et al.*, 1988) were also obtained using the same g09.

6.1.3 Results

There are three probable sites for H-bonding in BDBPZ, viz., N11, N12 and O1 **(Figure 6.1.1)**. However, the NBO charges obtained after GS geometry

optimization of BDBPZ in vacuum indicated that N11 (-0.464) and O1 (-0.550) are preferred for H-bonding over N12 (-0.454). Moreover, N11 is at anti-position to O6 which reduces the steric clash during H-bond formation. Thus, four H-bonded complexes of BDBPZ with EtOH and H_2O were modeled considering the two discussed possible H-bonding sites **(Figure 6.1.2)**.

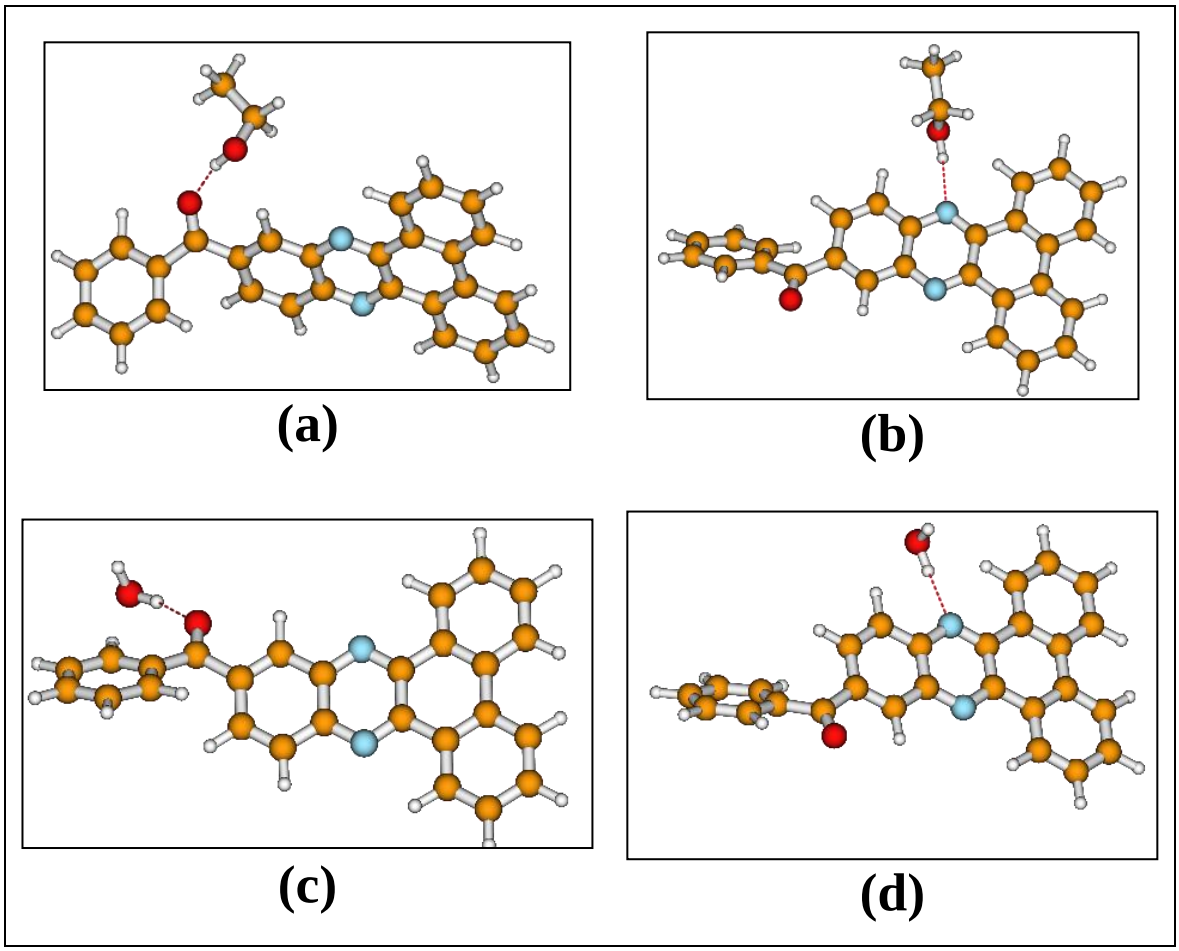


Figure 6.1.2: GS optimized structures of the four models in vacuum; (a) BDBPZ(O1)-EtOH, (b) BDBPZ(N11)-EtOH, (c) BDBPZ(O1)-H_2O and (d) BDBPZ(N11)-H_2O.

Along with GS and ES geometry optimizations in vacuum, free BDBPZ was also optimized in EtOH and MeCN medium. The dipole moment of the molecule increases with increasing solvent polarity both in the GS and ES **(Table 6.1.1)**. The energy difference between the GS and the ES optimized structures of free BDBPZ indicated a fluorescence maximum at 470.53nm **(Table 6.1.2)** whereas, when the optimizations were done in acetonitrile medium (MeCN), the peak shifted to 414.46nm.

Table 6.1.1: Dipole moment (in D) of free BDBPZ in different dielectric medium.

state	vacuum	EtOH	MeCN
GS	2.80	4.04	4.06
ES	11.54	16.22	16.31

Table 6.1.2: Fluorescence peak of the free-BDBPZ and the different H-bonded systems optimized in MeCN medium.

System [medium]	Fluorescence peak (nm)
BDBPZ [vacuum]	470.53
BDBPZ [MeCN]	414.46
BDBPZ(O1)-EtOH [MeCN]	421.61
BDBPZ(N11)-EtOH [MeCN]	419.51
BDBPZ(O1)-H2O [MeCN]	419.61
BDBPZ(N11)-H2O [MeCN]	418.73

The four H-bonded BDBPZ complexes **(Figure 6.1.2)** optimized in vacuum with BSSE correction (counterpoise correction) have been further analyzed in details. The dipole moments of all the four complexes are much greater in the ES compared to the GS **(Table 6.1.3)** suggesting an increase in charge separation on photoexcitation and in turn reflect stronger H-bonding in the excited state. Also, shorter H-bond lengths in the ESs are observed whether the bonding is through nitrogen (N11) or oxygen (O1). As a consequence the –O-H bond length of the solvent molecules is increased in the excited state and corresponding vibrational frequencies of –O-H bond is decreased.

Now the question is amongst the two potential sites of H-bonding which one makes H-bond more efficiently with the solvent? Apparently it seems that the oxygen is much more accessible to the external solvent system. Further the higher electronegativity of oxygen (3.44) compared to nitrogen (3.04) suggests stronger H-bonding with the oxygen atom. The H-bond energy (E_{HB}) in the GS also shows that the strength of H-bonding is greater with both the solvents (H_2O and EtOH) when it is through the oxygen atom. Similarly, the H-bond energy in the ES (E_{HB}^{*}) is lower when nitrogen atom is involved in H-bonding indicating better H-bonding situation. The NBO charges on oxygen (O1) and nitrogen (N11) atoms of the optimized BDBPZ system in vacuum are respectively, -0.550 and -0.464 in the GS and -0.579

and -0.531 in ES. Thus, it is evident that excited state H-bonding seems more efficient involving the oxygen atom. This is in line with the fact that the benzoyl group of BDBPZ is flanked outside the core aromatic skeleton and is very much accessible to the solvents and thus it can sense the polarity of the medium.

Table 6.1.3: The H-bonding parameters obtained from GS and ES optimized structures (in vacuum) of the four BDBPZ H-bonded systems.

parameter	state	BDBPZ(O1)-EtOH	BDBPZ(N11)-EtOH	BDBPZ(O1)-H_2O	BDBPZ(N11)-H_2O
HB length (Å)	GS	1.8965	1.9244	1.9387	1.9680
	ES	1.8168	1.8370	1.8951	1.8829
Δ(–O–H bond length) (Å)	GS	0.00652	0.01224	0.00706	0.01136
	ES	0.01064	0.01943	0.00944	0.01796
Δ(ν (–O–H)) (cm^{-1})	GS	-110.3	-358.1	-43.7	-54.2
	ES	-197.2	-410.8	-48.0	-59.6
E_{HB} (kcal/mol)	GS	-3.698	-2.867	-3.801	-2.717
	ES	4.245	6.680	-1.696	0.038
μ (D)	GS	3.66	1.65	4.50	1.42
	ES	12.37	11.21	14.91	11.16

We have also performed GS and ES optimizations for DMA-BDBPZ complex in vacuum and the fluorescence intensity is found to be quenched for DMA-BDBPZ as compared to free-BDBPZ **(Table 6.1.4)** as indicated from the reduced oscillator strength (0.0053 from 0.0225 of free BDBPZ). Analysis of total Mulliken and NBO charges, on DMA and BDBPZ molecules in the GS of DMA-BDBPZ complex indicates near zero values of the individual molecules, both in aqueous medium and vacuum. In ES, however, we noticed nearly one full electron transfer from DMA to BDBPZ, as the total atomic charges on DMA and BDBPZ become around 0.9 and -0.9, respectively **(Figure 6.1.3)**. We have also observed from the individual charges on the atoms of BDBPZ that, in the ES electron density over O1 and N12 atoms of BDBPZ increases by ~0.12**.** The geometry of DMA-BDBPZ in the GS has pyramidalization of N-atom of DMA, arising due to presence of lone-pair of electrons on the atom. However, in the ES, a slight rotation in -N-C- bond of DMA is seen, which is possible only when the bond has single bond character after an electron is donated from lone-pair of nitrogen atom of DMA to BDBPZ.

Table 6.1.4: Fluorescence peak (oscillator strength) of the free-BDBPZ and DMA-BDBPZ complex after optimization in vaccum

system	Fluorescence peak in nm (oscillator strength)
free-BDBPZ	470.53 (0.0225)
DMA-BDBPZ complex	518.34 (0.0013)

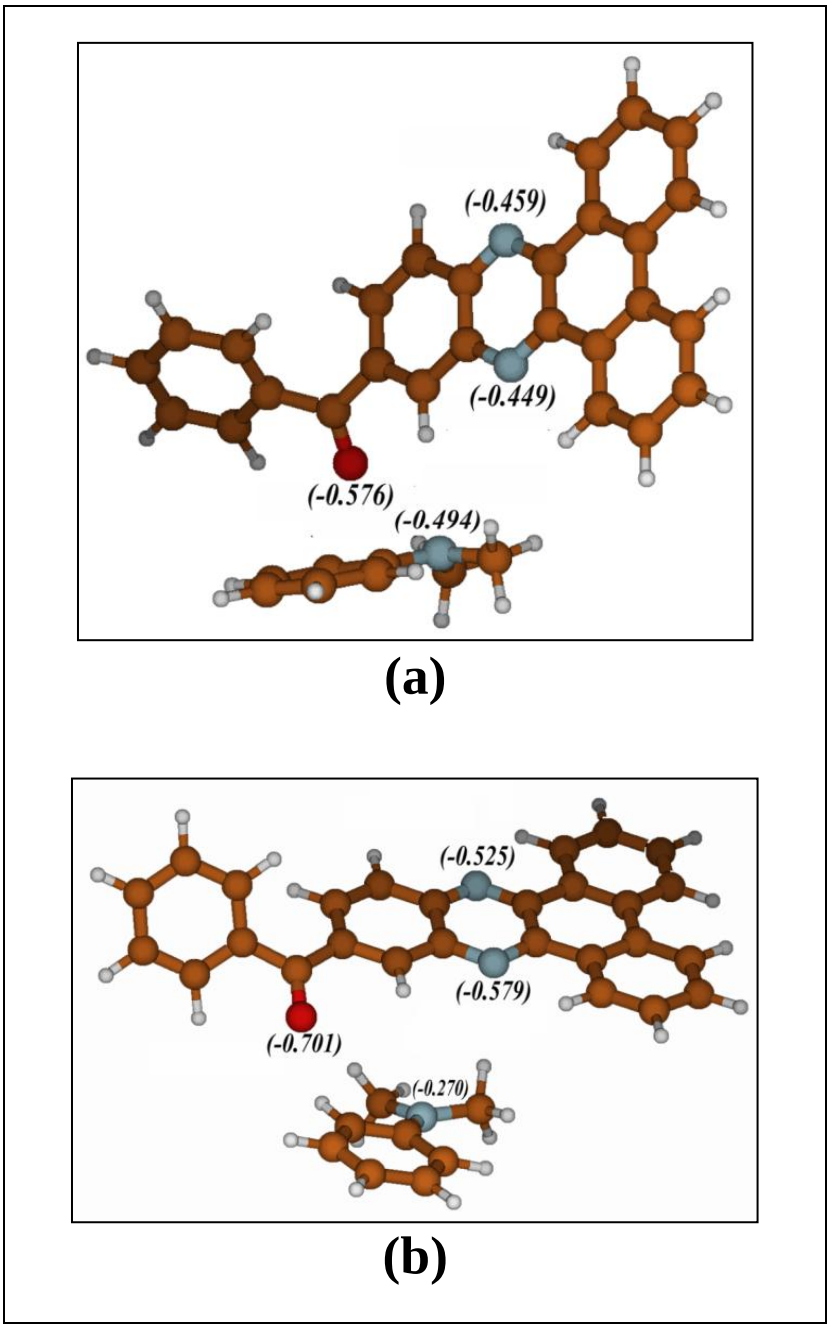


Figure 6.1.3: Optimized structure (in vacuum) of DMA-BDBPZ complex in (a) GS showing pyramidalization at N-atom of DMA and (b) ES showing rotation involving N-C bond in DMA. The NBO charges of O1, N11 and N12 atoms of BDBPZ and the nitrogen atom of DMA are given in parenthesis.

6.1.4 Discussion

It is observed from experiments (Dey *et al.*, 2015) that the derivative BDBPZ is much more interactive than DBPZ due to the benzoyl group that is flanked outside the core skeletal aromatic rings, and is very much accessible to the solvents thereby, helps to sense the environment properly and thus shows better ES H-bonding capacity than DBPZ. It is clearly shown from the present quantum chemical calculations and the optimized geometry of BDBPZ **(Figure 6.1.4)** that the H-bonding takes place through this benzoyl oxygen (O1) atom both in GS and ES. A new exciplex has been identified in the first excited singlet state between DMA and BDBPZ in cyclohexane medium by steady state fluorescence quenching experiments. It is expected that the fluorescence quenching occurs on addition of DMA to BDBPZ due to photoinduced electron transfer from DMA to photoexcited BDBPZ. It is revealed from our quantum chemical calculations that the carbonyl group (benzoyl oxygen, O1) of BDBPZ accepts the electron from DMA. Thus, it can be stated that here the mode of interaction is the photo-induced electron transfer from the amine to the fluorophore. Both the improved excited state H-bonding efficiency and better sensibility of the environment are attributed to the structure of the molecule.

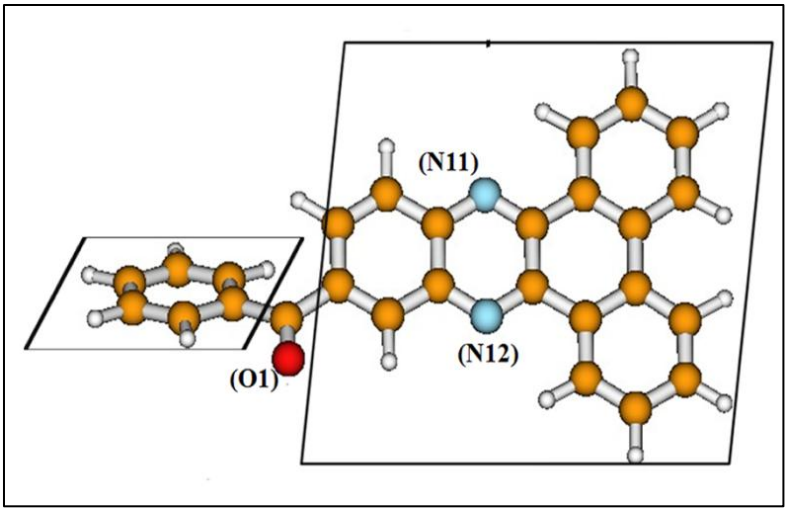


Figure 6.1.4: GS optimized geometry of BDBPZ indicating out-of plane benzoyl group making the O1 atom and its electron density easily accessible to solvents for H-bonding.

Chapter VI
Section II

Structural Study of Coenzyme NAD^+/NADH by Quantum Chemical Approach

6.2.1 Introduction

As we already know the nucleus of a cell contains proteins along with DNA and RNA. Unlike structural proteins, the biologically active proteins also known as enzymes catalyze biochemical reactions in cells (Cooper, 2000). Enzymes are usually linear chains of amino acids, *i.e.* proteins, that generally fold to produce a 3D globular structure and function alone or in large complexes. A non-protein chemical compound is often required for the protein's biological activity. It is called cofactor and can be considered as "helper molecules" that assist in biochemical transformations. There are two types of cofactors: (i) coenzymes and (ii) prosthetic groups. A coenzyme is a small organic non-protein molecule loosely bound to the protein.

One such coenzyme found in all living cells is nicotinamide adenine dinucleotide (NAD). It consists of two nucleotides: adenine and nicotinamide along with their five-member ribose sugar and connected through their phosphate groups, hence dinucleotide. It exists in two forms inside cells: oxidized NAD^+ and reduced NADH **(Figure 6.2.1a)**. In metabolism, nicotinamide adenine dinucleotide is involved in redox reactions such that NAD^+ accepts electrons from other molecules and becomes reduced to NADH, which then acts as reducing agent donating electrons. The enzyme dehydrogenase, belonging to the group of oxidoreductase which oxidizes a substrate by reduction reaction also consists of $NADP^+$. Nicotinamide adenine dinucleotide phosphate ($NADP^+$) is also a cofactor and requires NADPH as a reducing agent. $NADP^+$ differs from NAD^+ in the presence of an additional phosphate group on the 2' position of the ribose ring that carries the adenine moiety **(Figure 6.2.1b)**.

When bound to a protein, NAD^+ and NADH are usually held within a structural motif known as the Rossmann fold (Lesk, 1995). The structure is composed of up to seven mostly parallel β-strands **(Figure 6.2.2)**. The first two strands are connected by a single α-helix. This initial βαβ fold is the most conserved segment of Rossmann folds and remains in contact with the adenosine diphosphate (ADP) portion of NAD **(Figure 6.2.3a)** (Hanukoglu, 2015). In the NAD binding site, the tight turn between the first β-strand and the α-helix is in contact with the negatively charged oxygens of the two phosphate groups **(Figure 6.2.3b)** (Hanukoglu, 2015).

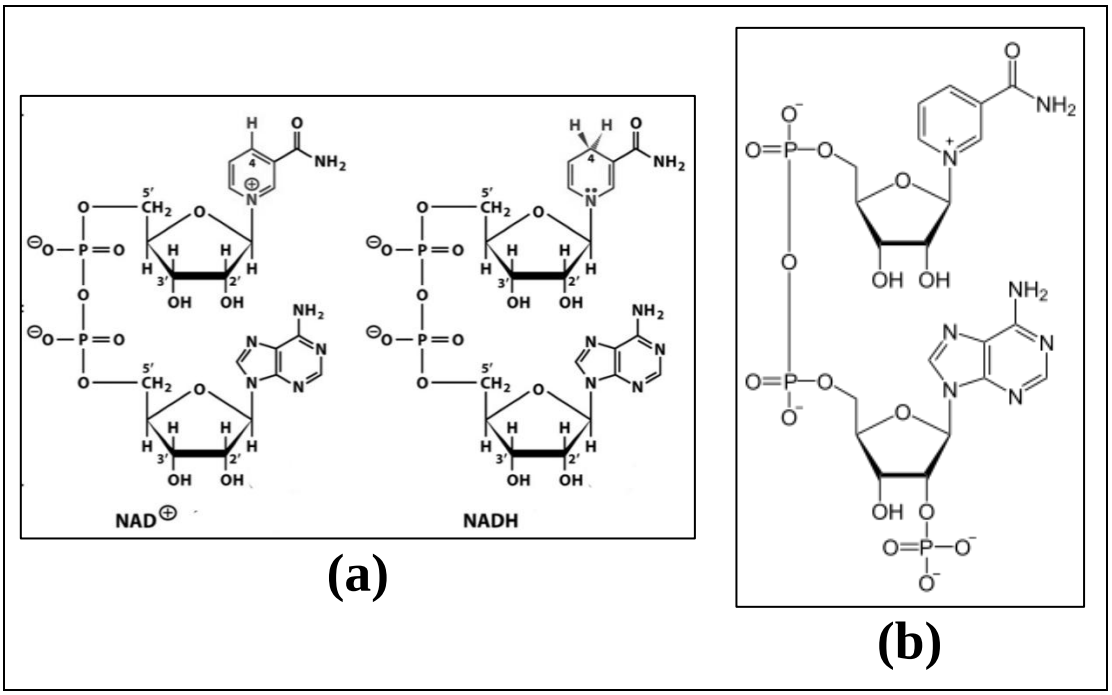


Figure 6.2.1: (a) The two forms of nicotinamide adenine dinucleotide in cell (Left to Right: NAD+ and NADH). (b) Nicotinamide adenine dinucleotide phosphate ($NADP^+$).

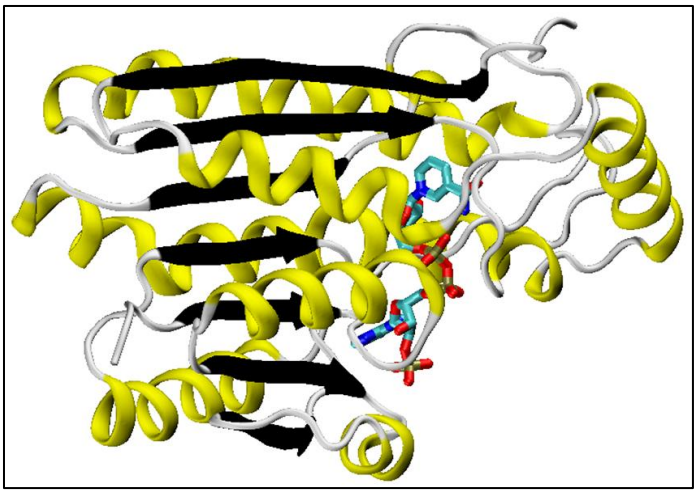


Figure 6.2.2: complete Rosmann fold with NADP (box).

The absorption spectrum of a NADH solution shows two maxima at the UV end of the visible spectrum, one at 250nm and the other at about 340nm. NAD^+, on the other hand has an absorption maximum at 250nm and almost does not absorb light above 300nm (Renault *et al.*, 1982). Upon excitation with UV-light NADH, unlike NAD^+, fluoresces in the blue (broad-band emission centered around 460nm) (Ince *et al.*, 1992). With the aim of understanding the NAD-protein interaction and the UV-

VIS spectra associated with these, we started by performing structural analysis and obtaining theoretical electronic spectra associated with NAD.

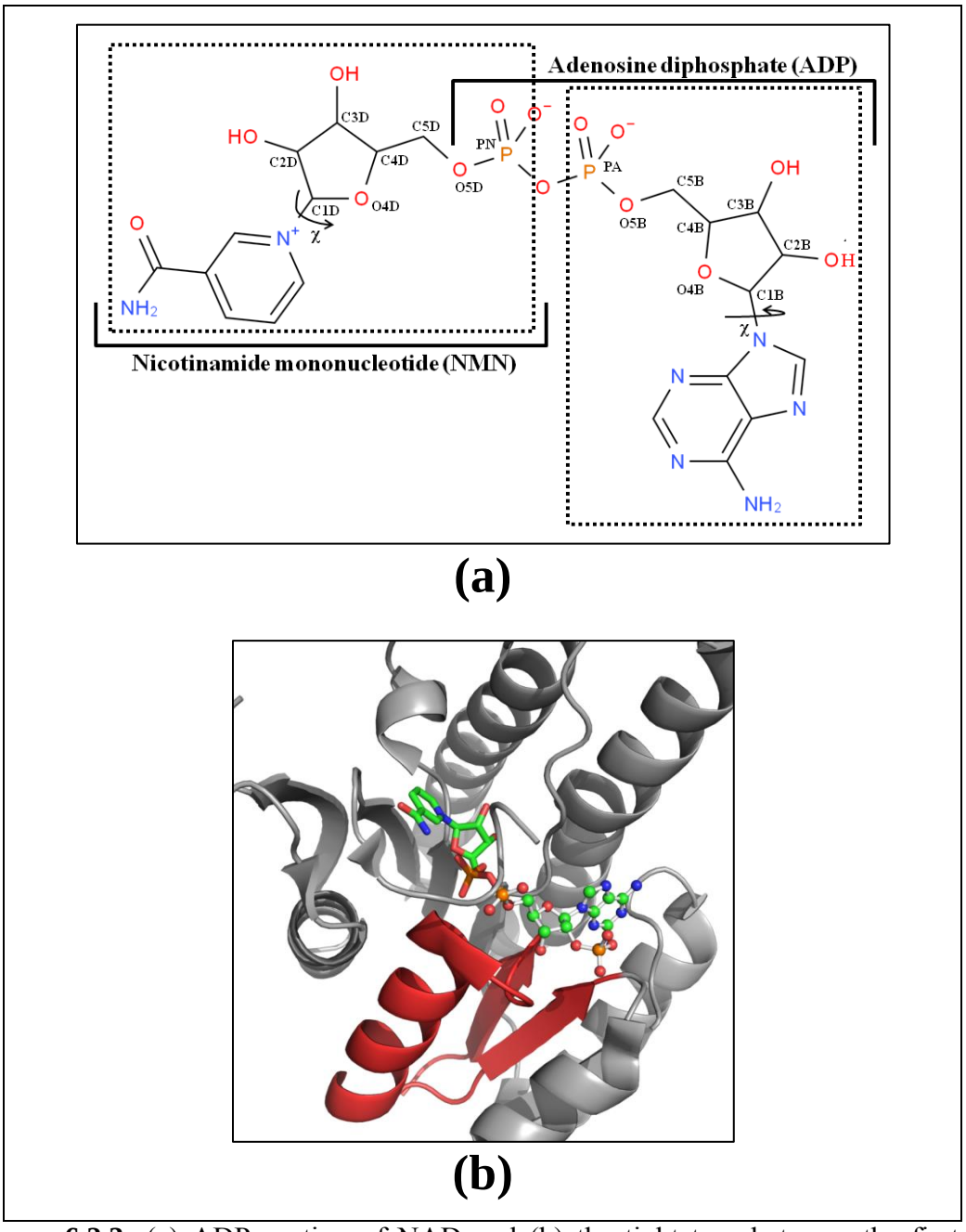


Figure 6.2.3: (a) ADP portion of NAD and (b) the tight turn between the first β-strand and the α-helix (of βαβ motif) is in contact with the negatively charged oxygens of the two phosphate groups.

In the present study we have particularly tried to understand the structure of NAD+/NADH system. The motif was taken from crystal structure with PDB ID: 3P19 **(Figure 6.2.4a)** which contains NADP bound to Rossmann fold **(Figure 6.2.4b)**. However, for the simplicity of the initial studies we have modeled this coenzyme as

NAD by removing the additional phosphate group on the 2' position of the ribose ring that carries the adenine moiety and adding an –OH group instead.

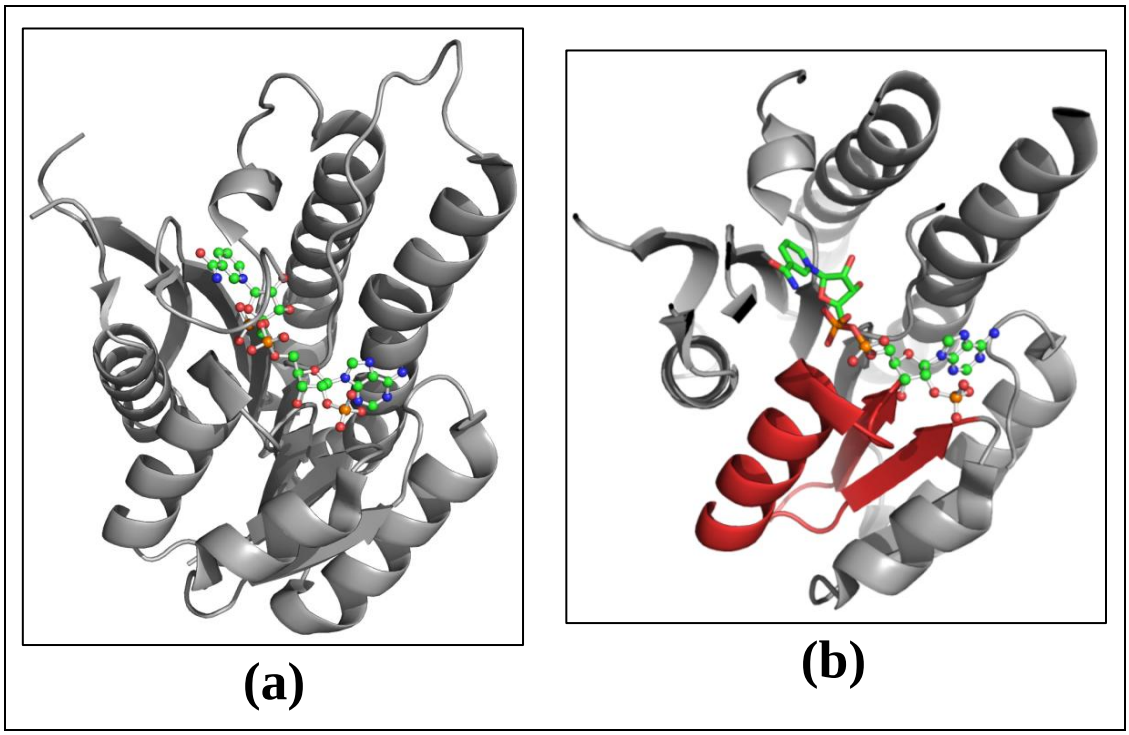


Figure 6.2.4: Coordinates are taken from PDB ID: 3P19.

6.2.2 Methodology

An initial search in RCSB-PDB database with the keyword NADH and refinement criteria: (i) Organism- Homo sapiens, (ii) Experimental Method- X-ray crystallography and (iii) X-ray Resolution as 1.0-2.0Å, gave a set of 17 PDB-files. These were then analysed for NADH moiety and the amino acid residues present in its surrounding with a cut off of 4.0Å using RASMOL (Sayle and Milner-White, 1995).

The initial coordinates of complete NAD was were taken from crystal structure (PDB ID: 3P19) and slightly modified to generate NAD^+ and NADH molecules, using the molecular modeling software MOLDEN (Schaftenaar and Noordik, 2000) with standard bond lengths (b(C-C) = 1.421Å, b(C-N) = 1.33 Å, b(C-H) = 1.09Å, b(N-H)= and b(C=O) = 1.22Å) and bond angles and appropriate phosphate groups (PO_4^{3-}). The ground and excited state optimizations of the modeled structures were initially carried out in vacuum with Density Functional Theory (DFT) and Time Dependant-Self Consistent Field (TD-SCF) DFT based method B3LYP/6-

31G(2d,2p) (Becke, 1993; Lee *et al.*, 1988) respectively, using Gaussian09 (g09) (Frisch *et al.*, 2009). The models were also optimized in aqueous solvent water (ε = 78.39) considering the Conductor-like Polarizable Continuum Model (CPCM) (Cossi *et al.*, 2003) for the solvent. Normal-mode frequency calculations were then carried out to confirm the energy minimized geometries. Theoretical emission spectra were also obtained from TD-SCF optimizations to compare with available experimental spectra for the NAD^+/NADH systems. Electrostatic Potential (ESP)-fit dipole moment and Natural Bond Orbital (NBO) charges (Glendening *et al.*, 2012) were also obtained to further characterize the systems.

6.2.3 Result and Discussion

The initial search in RCSB-PDB resulted in 17 PDBs (PDB IDs: 4X4L, 4RLS, 4JNK, 3N80, 3N81, 3N83, 2FZW, 1UMK, 1PL8, 1O02, 1I3L, 1I3M, 1I3N, 1HZJ, 1F0Y, 1EK6, 1GRB). It was observed that the PDBs: 3N80.pdb, 3N81.pdb, 3N83.pdb and 1UMK.pdb did not contain any NAD motif. The remaining 13 PDBs were then further analysed. A cut off of 4.0Å around NAD residue of only chain A of the crystal structure was applied and the amino acid residues surrounding the NAD were tabulated in **Table 6.2.1.**

A single NAD can be divided into three regions **(Figure 6.2.5)**: adenine-region, phosphate-region and the nicotinamide-region. A cut off of 3.7A was applied around each region and the surrounding amino acid residues were tabulated **(Table 6.2.2)**.

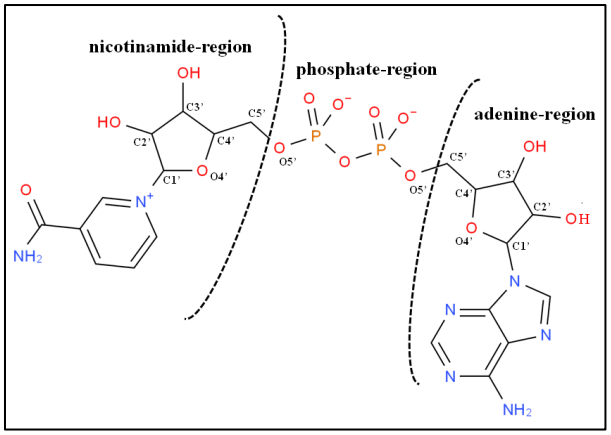


Figure 6.2.5: A single NAD molecule divided into three sections.

Table 6.2.1: Amino acid residues surrounding NAD/NAD in different proteins.

PDB ID	amino acid residues																				
1EK6	ALA	*GLY*	ILE	VAL	PHE	ASN	ASP	LYS	SER	TYR	**PRO**	HIS		MET							
1GRB	ALA	*GLY*	ILE	VAL	PHE			LYS		TYR					***ARG***		**GLU**	**LEU**			
1HZJ	ALA	*GLY*	ILE	VAL	PHE	ASN	ASP	LYS	SER	TYR	**PRO**	HIS		MET							
1I3L	ALA	*GLY*	ILE	VAL	PHE	ASN	ASP	LYS	SER	TYR	**PRO**	HIS		MET							
1I3M	ALA	*GLY*	ILE	VAL	PHE	ASN	ASP	LYS	SER	TYR	**PRO**	HIS		MET							
1I3N	ALA	*GLY*	ILE	VAL	PHE	ASN	ASP	LYS	SER	TYR	**PRO**	HIS		MET							
1O02	ALA	*GLY*	ILE		PHE	ASN		LYS	SER		**PRO**					***CYS***	**GLU**		GLN	TRP	
1PL8	ALA	*GLY*	ILE	VAL	PHE		ASP				**PRO**		THR		***ARG***	***CYS***		**LEU**			
2FZW	ALA	*GLY*	ILE	VAL	PHE	ASN	ASP	LYS		TYR			THR		***ARG***	***CYS***					
4JNK	ALA	*GLY*	ILE	VAL		ASN	ASP		SER			HIS	THR		***ARG***			**LEU**			
4RLS	ALA	*GLY*	ILE	VAL	PHE	ASN	ASP		SER	TYR		HIS	THR		***ARG***						
4X4L	ALA	*GLY*	ILE	VAL	PHE	ASN		LYS	SER	TYR	**PRO**		THR			***CYS***	**GLU**		GLN	TRP	
1F0Y	ALA	*GLY*	ILE	VAL	PHE	ASN	ASP	LYS	SER			HIS	THR	MET			**GLU**	**LEU**	GLN		

Table 6.2.2: Amino acid residues specifically surrounding the three different regions of NAD in different proteins.

	PDB ID	amino acid residues																		
phosphate-region	4RLS	GLY					VAL	ARG		ALA										
	4JNK	GLY					VAL	ARG		ALA										
	2FZW	GLY					VAL	ARG				HIS								
	1O02								SER		TRP		PHE	ILE						
	4X4L	GLY					VAL		SER		TRP		PHE		GLU					
	1PL8	GLY	ILE					ARG									PRO			
	1I3L	GLY	ILE	TYR	LYS	ASN														
	1I3M	GLY	ILE	TYR	LYS	ASN														
	1I3N	GLY	ILE	TYR	LYS	ASN														
	1HZJ	GLY	ILE	TYR	LYS	ASN														
	1EK6	GLY	ILE	TYR	LYS	ASN														
	1GRB	GLY	ILE	TYR	LYS				SER			HIS						LEU	MET	
	1F0Y	GLY																		
adenine-region	4RLS	GLY	ILE	ASP	ALA				VAL											
	4JNK	GLY	ILE	ASP	ALA				VAL											
	2FZW	GLY	ILE	ASP																
	1O02	GLY	ILE							PHE	LYS									
	4X4L	GLY	ILE		ALA				VAL	PHE	LYS									
	1PL8			ASP								ARG	GLU	THR	LEU					
	1I3L	GLY	ILE	ASP	ALA	ASN	HIS	MET	VAL	PHE										
	1I3M	GLY	ILE	ASP	ALA	ASN	HIS	MET												
	1I3N	GLY	ILE	ASP	ALA	ASN	HIS	MET	VAL											

	1HZJ	GLY	ILE	ASP	ALA	ASN	HIS	MET		PHE									
	1EK6	GLY	ILE	ASP	ALA	ASN	HIS	MET											
	1GRB	GLY	ILE		ALA							ARG				GLN			
	1F0Y	GLY	ILE	ASP															
nicotinamide-region	4RLS		SER				VAL	THR	ASN	ILE							HIS		
	4JNK		SER				VAL	THR	ASN	ILE			LEU		ARG	ALA	HIS		
	2FZW	PHE		PRO			VAL	THR		ILE	GLY	CYS				ALA			
	1O02	PHE	SER	PRO					ASN		GLY	CYS	LEU	GLU				GLN	
	4X4L	PHE	SER		LYS			THR	ASN		GLY	CYS		GLU				GLN	
	1PL8	PHE					VAL				GLY	CYS	LEU		ARG				
	1I3L	PHE	SER	PRO	LYS	TYR													
	1I3M	PHE	SER	PRO	LYS	TYR													
	1I3N	PHE	SER	PRO	LYS	TYR													
	1HZJ	PHE	SER	PRO	LYS	TYR													
	1EK6	PHE	SER	PRO	LYS	TYR													
	1GRB					TYR	VAL			ILE			LEU	GLU	ARG				
	1F0Y	PHE	SER		LYS		VAL	THR	ASN					GLU		ALA	HIS		MET

The Rossmann fold of chain A and the bound NDP in crystal structure with PDB ID: 3P19 has been mainly considered in this study. The residues present around 4.0Å of NDP in this protein are: GLY, ALA, SER, ILE, ARG, VAL, ASP, ASN, MET, TYR, LYS, PRO, THR, GLU and LEU **(Figure 6.2.6)**. In this case, the residues surrounding the adenine-region are GLY, SER, ARG, VAL and ASP; the residues around nicotinamide-region are ILE, ASN, GLY, SER, TYR, LYS, PRO, ALA, VAL, THR and LEU. The phosphate-region is surrounded by SER, GLY, ILE, THR, GLU and LEU. Interestingly, none of these residues except TRP, are capable of electron transfer. Hence, isolated NAD has been studied in vaccum and in presence of implicit solvent model. We have taken the coordinates of this NADP from chain A of 3P1P.pdb and performed quantum chemical analysis.

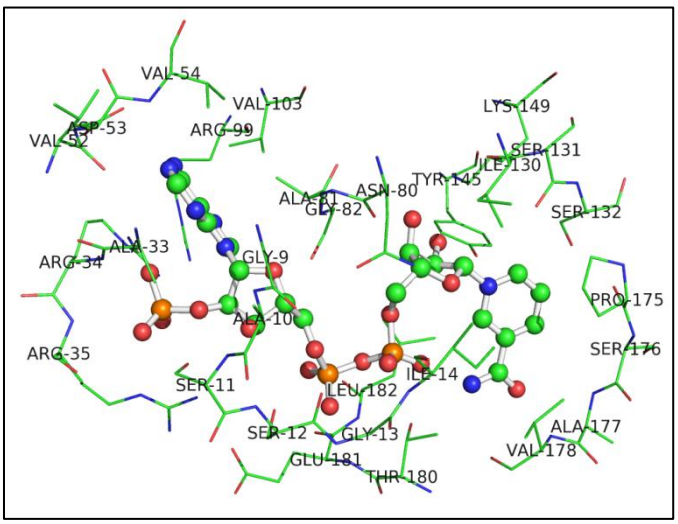


Figure 6.2.6: The amino acid residues surround NAD as in PDB ID: 3P19.

The backbone torsion angles and sugar pucker of NAD is similar to that of nucleic acids, hence we could define these **(Table 6.2.3)** and obtain an idea about the puckering and backbone conformation **(Table 6.2.4)** for the crystal structure.

Table 6.2.3: Definition of sugar pucker and backbone torsion angles

sugar pucker and	adenine-part	NAD^+/NADH-part
ν_0	C4B-O4B-C1B-C2B	C4D-O4D-C1D-C2D
ν_1	O4B-C1B-C2B-C3B	O4D-C1D-C2D-C3D
ν_2	C1B-C2B-C3B-C4B	C1D-C2D-C3D-C4D
ν_3	C2B-C3B-C4B-O4B	C2D-C3D-C4D-O4D
ν_4	C3B-C4B-O4B-C1B	C3D-C4D-O4D-C1D
γ	O5B-C5B-C4B-C3B	O5D-C5D-C4D-C3D
β	PA-O5B-C5B-C4B	PN-O5D-C5D-C4D

Table 6.2.4: Details of sugar pucker and backbone torsion.

system	optimized in	ν_0	ν_1	ν_2	ν_3	ν_4	χ	γ	β	P	*puckering*
					crystal (3P19)						
adenine-part		-30.548	40.727	-34.883	18.272	7.290	-123.558	-67.092	167.149	150.681	
NAD-part		-46.021	47.566	-31.714	4.984	25.788	-131.568	51.836	127.114	130.473	
					NAD+ (GS)						
adenine-part	vacuum	4.22	-24.463	33.86	-32.373	18.12	-148.397	52.267	-174.3	11.821	c3'-endo
NAD^+-part		-11.725	31.429	-38.495	32.588	-13.414	-148.25	-176.322	-103.364	-178.623	c3'-exo/c2'-endo
					NADH (GS)						
adenine-part	vacuum	2.947	-18.415	25.509	-24.507	13.981	-140.087	51.21	-137.27	12.306	c3'-endo
NADH-part		-18.269	35.727	-38.717	29.24	-7.2	56.48	50.049	-107.715	171.619	c2'-endo
					model systems						
GS NAD^+-part	vacuum	-61.135	49.422	11.053	-20.614	49.862	176.881			79.358	
GS NADH-part	vacuum	-27.974	40.997	-37.430	23.010	2.745	-166.631			157.081	
GS adenine-part	vacuum	17.588	-45.883	53.818	-45.786	17.473	-153.292			-0.073	
ES NAD^+-part	vacuum	-27.974	40.997	-37.43	23.01	2.745	-166.631			157.081	c2'-endo
ES NAD^+-part	water	-25.006	36.92	-33.927	21.025	2.183	-167.985			157.578	
ES NADH-part	vacuum	-31.69	40.905	-33.94	17.027	9.23	-130.377			148.187	c2'-endo
ES NADH-part	water	-27.046	39.046	-35.396	21.199	3.686	-174.813			155.966	
ES adenine-part	vacuum	3.963	-24.643	34.336	-33.169	18.776	-152.89			12.454	c3'-endo
ES adenine-part	water	3.874	-23.325	32.423	-31.176	17.584	-150.162			12.192	

It is known that reduced NAD (*i.e.* NADH) is fluorescent while its oxidized counterpart (NAD^+) is not. Thus, we started by successful GS optimizations of NAD^+ and NADH systems in vacuum condition **(Figure 6.2.7)**, however when the optimizations were tried for ES, the convergence was difficult to attain. Thus, to recognize spectral properties of NAD^+/NADH, we developed three model systems **(Figure 6.2.8)**: (i) the adenine-part **(Figure 6.2.8a)**, (ii) the NAD^+-part **(Figure 6.2.8b)**, and (iii) the NADH-part **(Figure 6.2.8c)**. The adenine-part is common for both NAD^+ and NADH.

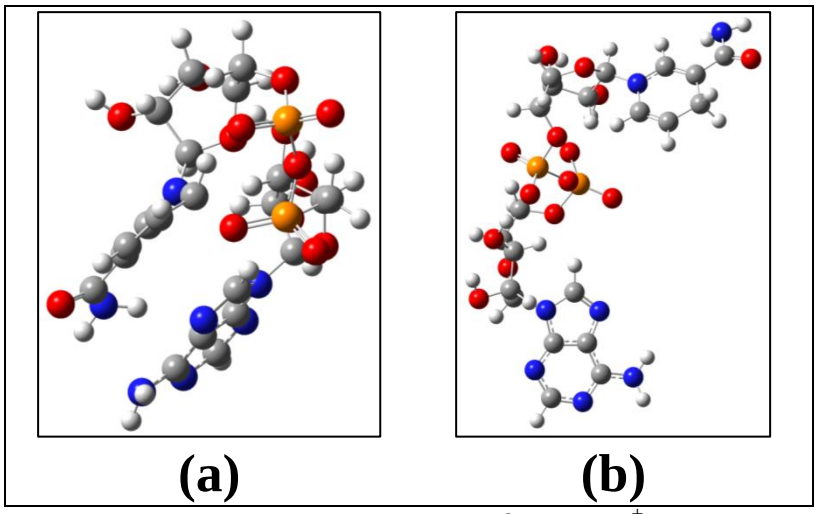


Figure 6.2.7: GS optimized geometries of (a) NAD^+ and (b) NADH.

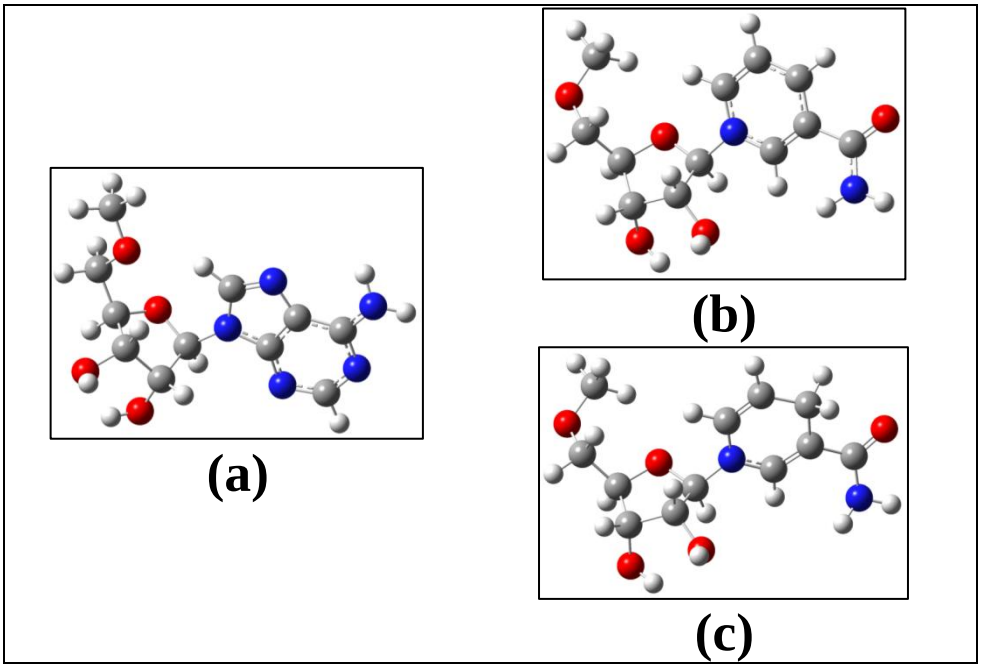


Figure 6.2.8: GS optimized geometries (in vacuum) of (a) adenine-part, (b) NAD^+-part and (c) NADH-part.

As is known from Jablonski diagram, the energy difference between GS and ES electronic configurations corresponding to GS geometry of a molecule is proportional to its absorption wavelength. The absorption wavelength for adenine base (or adenine-part) is found to be in the IR regions, as is also known from experimental findings (Table 5). The absorption wavelengths of the NAD^+-part and NADH-part appears in the blue-green region. We have also calculated such absorption wavelengths of common amino acid residues which may absorb photon, such as side chains of TYR or TRP residues **(Table 6.2.5)** and a tripeptide **(Figure 6.2.9)** by optimizations in GS in vacuum by DFT based B3LYP/6-31G(2d,2p) method, followed by single point calculations in ES.

Table 6.2.5: Details of absorption of different systems.

amino acid	wavelength (nm)	oscillator strength
Ade	240.98	0.1714
NAD^+-part	398.47	0.0017
NADH-part	322.27	0.1175
tripeptide	220.73	0.0035
TRP	264.44	0.0569
TYR	234.46	0.0871
HIS	209.69	0.0066
PHE	235.76	0.0501

[$]all optimization in GS and further ES calculations have been performed in vacuum

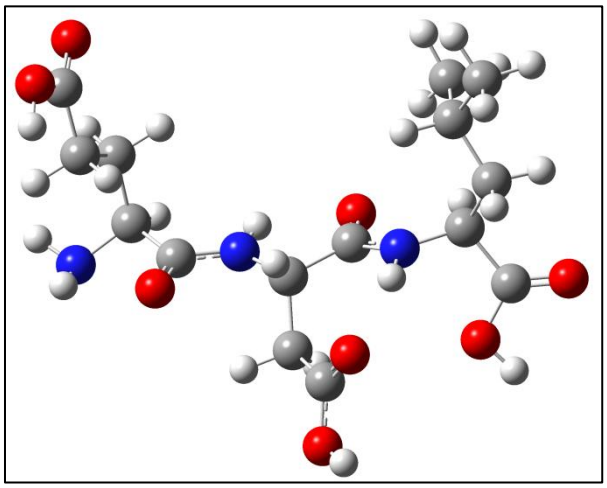


Figure 6.2.9: A tripeptide considered in this study.

We have also optimized the three model systems shown in **Figure 8** in ES. The energy difference between ES electronic configuration (singlet state with unpaired electron in LUMO) and ES electronic configuration (with paired electrons in

HOMO and no electron in LUMO) corresponds to fluorescence emission. The fluorescence emission frequencies of the model systems, along with their oscillator strengths, indicative of intensity of emission, are shown in **Table 6.2.6**. The emission wavelength of adenine-part both in vacuum and aqueous medium is ~270nm. It appears that even adenine base can fluoresce, in principle, with very large intensity, although it is quite difficult to observe this frequency by experimental techniques. It is thus, indicated from ES optimizations (in vacuum) that the fluorescence obtained for NAD^+/NADH is due to the NAD^+/NADH-part. The NADH-part fluoresces at 590nm with significant oscillator strength while the oscillator strength for fluorescence at 482nm of NAD^+-part is very small. It is seen that the dipole moment is very much increased in the ES for NAD^+-part whereas, there is not much change in dipole moment between GS and ES of the NADH-part.

The emission frequencies of both NADH-part and NAD^+-part are blue-shifted in aqueous environment **(Table 6.2.6)** but the fluorescence intensity of NADH-part is seen to increase while that of NAD^+-part decreases. Thus, even though one may be able to detect fluorescence signal of NAD^+ in vacuum or in hydrophobic medium, its detection at water environment may not be possible. Furthermore, the amino acid residues, namely PHE, TYR or TRP, may absorb this fluorescence emission from NAD^+ as the fluorescence wavelength of NAD^+ and the absorption wavelengths of the above mentioned amino acid residues are very close. Hence, one might need to design a protein with a hydrophobic cavity devoid of the aromatic side chains to observe the weak fluorescence from NAD^+.

Table 6.2.6: Details of quantum chemical studies.

model systems	optimized in	dipole (ES)	dipole (ES)	Wavelength (nm)	Oscillator Strength
adenine-part	vacuum	5.3866	3.2009	267.29	0.2719
NAD^+-part		5.2522	13.697	482.02	0.0057
				959.63nm	*0.0044*
NADH-part		5.5494	3.3546	589.85	0.0448
adenine-part	water	6.9681	4.096	276.42	0.4621
NAD^+-part		6.8075	14.2218	372.09	0.0034
NADH-part		5.5494	5.8825	502.82	0.1177

We have also optimized the NAD^+-part and NADH-part in ES in vacuum by ωB97XD/6-31G(2d,2p) method and observed that the wavelength corresponding to the NAD^+-part also falls in the UV-VIS region but its oscillator strength is very low **(Table 6.2.7)**. Thus, till now it is clear that NADH can fluoresce but NAD^+ cannot, however one may look into conditions wherein due to environment the NAD^+ may contribute to fluorescent emissions.

Table 6.2.7: Quantum chemical studies performed with ωB97XD/6-31G(2d,2p) method.

	dipole moment in ES	wavelength (nm)	Oscillator Strength
NADH-part	3.383	462.72	0.0746
		269.76	0.0137
NAD^+-part	3.941	667.83	0.0015
		330.59	0.0016

The quantum chemical calculations performed on NAD/NADH systems and reported in this section are in its initial stage. The region around NAD/NADH inside a protein has to be considered in details while performing further GS and ES calculations. This would lead to better probing into the UV-VIS spectra of these systems.

General Discussion

CONCLUSION

The complete work presented in this thesis is based on structural characterization of quadruple-helix DNA, RNA bulges and pseudo-continuous helices and human lamin Ig-fold domain by Molecular Dynamics and Steered Molecular Dynamics Simulations. Besides this, Quantum Chemical Analysis has been performed for studying G-quartets/G-quartet stacks, nanosize graphene/GO systems, BDBPZ and NAD^+/NADH molecules.

Conclusion (Chapter II)

- The quantum chemical studies indicated that the presence of ions (Na+/K+) near the G-tetrad core gives it a planar geometry in turn adding to its stability.
- The planarity of the G-tetrad is dependent on the type and position of the ion. A single K^+ ion inside G-tetrad core or two Na^+ ions on either side of it gives the G-tetrad a planar conformation.
- A stack of two G-quartets containing a central metal ion (Na+/K+) at the core of the G-tetrad stack, which is lined by eight O6 atoms, attains a square antiprism geometry.
- A few structural factors concerning the telomeric quadruplex-DNA from the perspective of base pair and base pair step parameters were obtained after performing MD simulations with charmm27 force field.
- The G-quadruplexes can attain four different topologies, *viz.*, parallel, anti-parallel, mixed-(3+1)-form1 and mixed-(3+1)-form2. The orientation parameters of the G:G H:W C base pairs constituting the G-tetrads were found to be dependent on the topology of the G-quadruplex. The twist values for all the base pair steps of type W:H::W:H were found to be close to 30° whereas, for the steps of type W:H::H:W or H:W::W:H the twist values were found to be around 20°.
- In some of the G-quadruplexes, Ade:Ade or Ade:Thy base pairs are formed by the loop region residues which acts as a capping over the outer G-quartets. The anti-parallel and mixed-(3+1)-form1 and mixed-(3+1)-form2 topologies contain such capping over both the extreme G-quartets. This gives

compactness to the quadruplexes making them suitable as the major conformation.

- The MD simulation of hTERT promoter sequence was found to be inherently unstable with anomalous values of the base pair and base pair step parameters. Thus, it was concluded that all G-rich sequences may not adopt a topology similar to the ones adopted by G-rich sequences of the telomeric region.
- Combining the MD and SMD simulation studies it could be concluded that the anti-parallel topology is the most suited as a telomeric topology, although the mixed-(3+1)-form1 should not be completely neglected while considering the major conformation.

Conclusion (Chapter III)

- The MD simulations of 1-residue bulge containing RNA systems with both AMBER-99sb and charmm36 force fields indicated the effect of the bulge residue on the geometry of base pairs in specific and on the overall structure of the duplex in general.
- The bulged-out unpaired residue does not alter the base pairing at the junction or overall structure of the duplex to an extent as high as that induced by presence of a bulged-in or intercalated residue. The RNA stretch containing bulged-out residue maintains features similar to a standard continuous RNA stretch.
- The RNA stretch containing base-triple stacked inside the double-helix behaves well during the MD-run and has structural variability similar to a continuous RNA double helix. The reason is probably appropriate stacking of the triplet in between two base pairs.
- The pseudo-continuous helical systems behaves well during the MD-run when there are extra-flanking residues are present at the junction. Also, the type of base pairing (canonical or non-canonical) at the junction of pseudohelices can determine the structural variability.
- The simulation of RNA pseudohelix containing a 3-residue loops acting as capping at the terminal is seen to add stability to the system, indicated by less fluctuation during its MD-run.

- An initial overview of the various parameters related to base pairs at the junction of bulges and pseudo-continuous helices and has been reported and compared to a standard RNA wherever possible.

Conclusion (Chapter IV)

- The wild-type protein involved two salt bridges: (Arg 453-Glu 443) and (Arg 453-Glu 444), which are broken in the mutant R453W. However, the MD simulations of 100ns indicated structural similarity between the wild-type and the mutant proteins.
- As the R453W mutation resulted in the loss of two salt bridges at 443 Glu and 444 Glu, the expected difference in the force mediated unfolding was observed during SMD simulations. The mutant was found to be mechanically weaker compared to the wild type when the unfolding forces corresponding to different extensions were seen.
- The secondary structure variations of both the wild type and the mutant proteins for the entire simulation trajectories showed significantly different secondary structural features at short extension (within 180Å) in both the cases, whereas at longer extensions those domains were almost identical.
- An interesting conformational transition from β-strand to α-helix prior to complete stretching of the proteins was observed in both the cases. However, this secondary structure alteration from β-strand to α-helix was observed more in case of the mutant-protein. The α-helices were not stable in either case though their lengths were quite significant.
- There was little mechanical resistance from the first and last two β-strands and the mutant yielded a less stable intermediate in the process of unfolding.
- The SMD studies were in good agreement with single molecule force spectroscopy results.

Conclusion (Chapter V)

- It is already know that a graphene sheet has two edges, viz., *cis*- and *trans*-edge. However, our studies involving NBO charge analysis and bond order analysis indicate that there are four different types of –C–C– bonds: (i) H–C–

(C–C)–H, (ii) H–C–(C–C)–C–H, (iii) H–(C–C)–H and (iv) *corner*-H–(C–C)–H, which are the probable sites for initial epoxidation prior to tearing.

- Quantum chemical studies indicated that progressive tearing of a large graphene sheet into smaller and smaller ones would lead to smaller graphene sheets with a gradual increase in the percentage of *trans*-edge, *i.e*. –(C–C–C)– bonding predominant at the edges with some functionalization at the terminal C-atoms.
- A preference in the tearing pattern was also observed. The tearing of a graphene sheet prefers to happen along the *cis*-edge with –C*–C*– bonds of H–C–(C*–C*)–C–H type breaking at the lowest cost of energy.
- The Na^+/Cl^- ions prefer getting trapped to the *trans*-edge of a graphene sheet whereas; a water molecule is seen to move towards the plane of a graphene sheet during optimization. Thus, a graphene sheet acts a desalination filter.

Conclusion (Chapter VI)

- Unlike the parent molecule DBPZ, its benzoyl derivate BDBPZ can sense the polarity of the medium and act as a better Excited State (ES) hydrogen bonding (H-bonding) probe.
- It is clearly shown from the quantum chemical calculations and the optimized geometry of BDBPZ that the H-bonding takes place through the benzoyl oxygen (O1) atom both in GS and ES. This benzoyl group is flanked outside the core skeletal aromatic rings, and is very much accessible to the solvents thereby. This helps to sense the environment properly and thus shows better H-bonding capacity than DBPZ in ES.
- Steady state fluorescence quenching on addition of DMA due to photoinduced electron transfer from DMA to photoexcited BDBPZ was seen by our experimental collaborators and the product radical ions was identified using laser flash photolysis. The electron transfer from DMA to BDBPZ in ES was also confirmed by out theoretical studies.
- An initial idea about the amino acid residues surrounding NAD/NADH inside a protein was obtained. The amino acid residues specifically present around and can interact with the adenine, nicotinamide and phosphate regions of NAD/NADH was also noted.

- The structure of NAD/NADH systems with respect to backbone conformation was studied for both the crystal structure and optimized geometries.
- The theoretical calculations are clearly in line with experimental evidence that NADH can fluoresce but NAD^+ cannot. The initial quantum chemical studies indicate that the reason for this might be that some of the amino acid residues surrounding the NAD molecule absorb the emission from NAD^+. However, the reason for this still need to be understood in details from further quantum studies.

SCOPE OF FUTURE WORK

Understanding structure and dynamics of nucleic acids (DNA/RNA), protein and other biologically important macromolecules (graphene/GO, BDBPZ, NAD^+/NADH) involves multiple aspects, all of which could not be adequately covered in the present Ph.D. work.

The present thesis involved in-depth structural study of four types of G-quadruplexes formed by human telomeric G-rich sequences. But there exist non-human and non-telomeric G-rich sequences that are to be completely characterized from the perspective of structural variability (base pair and base pair parameters) and backbone conformations. It would also be interesting to study unfolding and folding of G-quadruplex structures with different topology in the light of helicoidal parameters and ion coordination. It would be interesting to study the unfolding process in explicit solvent (water) and mimic the experimental condition.

RNA double helical stretches containing only 1-residue bulge has been studied and reported. This study can be extended to 2-residue or 3-residue bulge containing systems. Since, it is expected that the type of base pairing (canonical or non-canonical) at the junction of pseudohelices can determine the structural variability at the junction, it would be interesting to understand what type of stacking arrangement between non-canonical base pairs is most favourable. This study can be further extended by studying a set of coaxial stacks with all possible base pairing patterns at the junction.

The 100ns long MD-run of the two proteins; wild-type and mutant did not indicate any significant structural dissimilarity. It would thus be suitable to extend

these simulations and notice the interesting structural difference involving the β3-strand, containing the mutation, of β-sheet2. Such SMD simulations can also be carried out with other mutants of the Ig-domain, especially, the one involving R527P as it similar to R453W in the sense that this residue is also facing outwards into the solvent.

The present work deals extensively with tearing pattern and preference of graphene/GO sheet. However, the situation after complete tearing still has to be understood, *i.e.* the termination or functionalization at the edges should be studied in future to understand how the valency is maintained in smaller newly produced graphene sheets. The study of functionalized graphene sheets has been initiated here but a complete conclusive work is in future plans. Studying desalination properties of graphene sheet and to see if GO can also act as one such filter would be interesting.

H-bonding probe DBPZ and then it derivative BDBPZ has been well studied. Similarly, other derivatives can be synthesized experimentally and their H-bonding capabilities can be studied. The work related to NAD^+/NADH enzyme has been initiated in this thesis, however, it requires further studies involving quantum chemical approach and MD simulations of entire protein-NAD^+/NADH complex.

References

(1983). IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Abbreviations and symbols for the description of conformations of polynucleotide chains. Recommendations 1982. *Eur J Biochem* 131:9

A

Abu Almakarem, A. S., Petrov, A. I., Stombaugh, J., Zirbel, C. L., & Leontis, N. B. (2012). Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic Acids Res*, *40*(4), 1407–1423.

Adrian, M., Heddi, B., & Phan, A. T. (2012). NMR spectroscopy of G-quadruplexes. *Methods*, *57*(1), 11–24.

Aebi, U., Cohn, J., Buhle, L., & Gerace, L. (1986). The nuclear lamina is a meshwork of intermediate-type filaments. *Nature*, *323*(6088), 560–564.

Agarwal, A. (2015). Structural Analysis of RNA Bulges and Pseudohelices: A Molecular Dynamics Simulation Approach. PhD thesis, West Bengal University of Technology, Kolkata.

Aich, P., Labiuk, S. L., Tari, L. W., Delbaere, L. J., Roesler, W. J., Falk, K. J., … Lee, J. S. (1999). M-DNA: A complex between divalent metal ions and DNA which behaves as a molecular wire. *J Mol Bio*, *294*(2), 477–85.

Ajayan, P. M., & Yakobson, B. I. (2006). Material science: oxygen breaks into carbon world. *Nature*, 441(7095), 818-819.

Allemand, J. F., Bensimon, D., Lavery, R., & Croquette, V. (1998). Stretched and overwound DNA forms a Pauling-like structure with exposed bases. *Proc Natl Acad of Sci U S A*, *95*(24), 14152–14157.

Altona, C., & Sundaralingam, M. (1972). Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J Am Chem Soc*, *94*(23), 8205–8212.

Arnott, S. (1999). Polynucleotide secondary structures: an historical perspective. *The Oxford Handbook of Nucleic Acid Structure*. S. Neidle (Ed.), Oxford University Press: New York.

Arnott, S., Hukins, D. W. L., & Dover, S. D. (1972). Optimised parameters for RNA double-helices. *Biochem Biophys Res Commun*, *48*(6), 1392–1399.

Auffinger, P., & Westhof, E. (1998). Hydration of RNA base pairs. *J Biomol Struct Dyn*, *16*(3), 693–707.

Avery, O. T., Macleod, C. M., & McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types : induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med*, *79*(2), 137–158.

B

Babcock, M. S., & Olson, W. K. (1994). The effect of mathematics and coordinate system on comparability and "dependencies" of nucleic acid structure parameters. *J Mol Biol*, *237*(1), 98–124.

Babcock, M. S., Pednault, E. P., & Olson, W. K. (1994). Nucleic acid structure analysis. Mathematics for local Cartesian and helical structure parameters that are truly comparable between structures. *J Mol Biol*, *237*(1), 125–156.

Baeyens, K. J., De Bondt, H. L., & Holbrook, S. R. (1995). Structure of an RNA double helix including uracil-uracil base pairs in an internal loop. *Nature Struct Biol*, *2*(1), 56–62.

Baeyens, K. J., De Bondt, H. L., Pardi, A., & Holbrook, S. R. (1996). A curved RNA helix incorporating an internal loop with G.A and A.A non-Watson-Crick base pairing. *Proc Natl Acad Sci U S A*, *93*(23), 12851–12855.

Bagani, K., Ray, M. K., Satpati, B., Ray, N. R., Sardar, M., & Banerjee, S. (2014). Contrasting Magnetic Properties of Thermally and Chemically Reduced Graphene Oxide. *J Phys Chem C*, *118*(24), 13254–13259.

Bagri, A., Grantab, R., Medhekar, N. V., & Shenoy, V. B. (2010). Stability and formation mechanisms of carbonyl- and hydroxyl-decorated holes in graphene oxide. *J Phys Chem C*, *114*(28), 12053–12061.

Balandin, A. A., Ghosh, S., Bao, W., Calizo, I., Teweldebrhan, D., Miao, F., & Lau, C. N. (2008). Superior thermal conductivity of single-layer graphene. *Nano Lett*,

8(3), 902–907.

Balsera, M., Stepaniants, S., Izrailev, S., Oono, Y., & Schulten, K. (1997). Reconstructing potential energy functions from simulated force-induced unbinding processes. *Biophys J*, *73*(3), 1281–1287.

Ban, N., Nissen, P., Hansen, J., Moore, P. B., & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science*, *289*(5481), 905–920.

Bandyopadhyay, D. (2005). Different non-local interactions in stabilizing nucleic acid structures. PhD Thesis, University of Calcutta, Kolkata.

Bandyopadhyay, D., & Bhattacharyya, D. (2000). Effect of neighboring bases on base-pair stacking orientation: a molecular dynamics study. *J Biomol Struct Dyn*, *18*(1), 29–43.

Banerjee, S., & Bhattacharyya, D. (2008). Electronic properties of nano-graphene sheets calculated using quantum chemical DFT. *Comput Mater Sci*, *44*(1), 41–45.

Banerjee, S., Sardar, M., Gayathri, N., Tyagi, A. K., & Raj, B. (2005). Conductivity landscape of highly oriented pyrolytic graphite surfaces containing ribbons and edges. *Phys Rev B*, *72*(7), 075418.

Banerjee, S., Sardar, M., Gayathri, N., Tyagi, A. K., & Raj, B. (2006). Enhanced conductivity in graphene layers and at their edges. *Appl Phys Lett*, *88*(6), 062111.

Bansal, M., Bhattacharyya, D., & Ravi, B. (1995). NUPARM and NUCGEN: software for analysis and generation of sequence dependent nucleic acid structures. *Comput Appl Biosci*, *11*(3), 281–287.

Bao, H., Pan, Y., Ping, Y., Sahoo, N. G., Wu, T., Li, L., … Gan, L. H. (2011). Chitosan-functionalized graphene oxide as a nanocarrier for drug and gene delivery. *Small*, *7*(11), 1569–1578.

Bao, H., Pan, Y., & Li, L. (2012). Recent advances in graphene-based nanomaterials for biomedical applications. *Nano LIFE*, *02*(01), 1230001.

Barciszewski, J., & Clark, B. F. C. (1999). *RNA Biochemistry and Biotechnology*.

Springer Netherlands.

Bartel, D. (2004). MicroRNAsGenomics, Biogenesis, Mechanism, and Function. *Cell*, *116*(2), 281–297.

Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., … Williamson, J. R. (1996). Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science*, *273*(5281), 1547–1551.

Bazzicalupi, C., Ferraroni, M., Bilia, A. R., Scheggi, F., & Gratteri, P. (2013). The crystal structure of human telomeric DNA complexed with berberine: An interesting case of stacked ligand to G-tetrad ratio higher than 1:1. *Nucleic Acids Res*, *41*(1), 632–638.

Beal, P. A., & Dervan, P. B. (1991). Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. *Science*, *251*(4999), 1360–1363.

Becke, A. D. (1993). Density-Functional Thermochemistry .3. the Role of Exact Exchange. *J Chem Phys*, *98*(7), 5648–5652.

Becker, M. M., & Wang, Z. (1989). B----A transitions within a 5 S ribosomal RNA gene are highly sequence-specific. *J Biol Chem*, *264*(7), 4163–4167.

Bennigton-Castro, J. (2014). A Step Closer to Graphene Desalination Filters. *Materials 360 Online*. http://www.materials360online.com/newsDetails/44792

Bera, M., Kotamarthi, H. C., Dutta, S., Ray, A., Ghosh, S., Bhattacharyya, D., … Sengupta, K. (2014). Characterization of unfolding mechanism of human lamin A Ig fold by single-molecule force spectroscopy-implications in EDMD. *Biochemistry*, *53*(46), 7247–7258.

Bergues-Pupo, A. E., Arias-Gonzalez, J. R., Morón, M. C., Fiasconaro, A., & Falo, F. (2015). Role of the central cations in the mechanical unfolding of DNA and RNA G-quadruplexes. *Nucleic Acids Res*, (26), 7638-7647.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., … Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res*, *28*(1), 235–242.

Bertrand, H. O., Ha-Duong, T., Fermandjian, S., & Hartmann, B. (1998). Flexibility of the B-DNA backbone: Effects of local and neighbouring sequences on

pyrimidine-purine steps. *Nucleic Acids Res*, *26*(5), 1261–1267.

Beveridge, D. L., Barreiro, G., Byun, K. S., Case, D. A, Cheatham, T. E., Dixit, S. B., … Young, M. A. (2004). Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of DNA Oligonucleotides. I. Research Design and Results on d (C(p)G). *Biophys Journal*, *87*(6), 3799–3813.

Bhattacharya, S., Mittal, S., Panigrahi, S., Sharma, P., S P, P., Paul, R., … Mitra, A. (2015). RNABP COGEST: a resource for investigating functional RNAs. *Database (Oxford)*, doi: 10.1093/database/bav011.

Bhattacharyya, D., & Bansal, M. (1990). Local variability and base sequence effects in DNA crystal structures. *J Biomol Struct Dyn*, *8*(3), 539–572.

Bhattacharyya, K. (2008). Nature of biological water: a femtosecond study. *Chem Commun*, (25), 2848–2857.

Bian, Y., Tan, C., Wang, J., Sheng, Y., Zhang, J., & Wang, W. (2014). Atomistic picture for the folding pathway of a hybrid-1 type human telomeric DNA G-quadruplex. *PLoS Comput Biol*, *10*(4), e1003562.

Björk, G. R. (1995). Biosynthesis and Function of Modified Nucleosides. *tRNA: Structure, Biosynthesis, and Function*, 165–205.

Blackburn, E. H. (1990). Minireview Telomeres : Structure and Synthesis. *J Biol Chem*, *265*(11), 5919–5921.

Bloomfield, V A; Crothers, D M; Tinoco, I, J. (2000). *Nucleic Acids: Structures, Properties, and Functions*. University Science Books.

Bolotin, K. I., Sikes, K. J., Jiang, Z., Klima, M., Fudenberg, G., Hone, J., … Stormer, H. L. (2008). Ultrahigh electron mobility in suspended graphene. *Solid State Commun*, *146*(9-10), 351–355.

Bonne, G., Mercuri, E., Muchir, A., Urtizberea, A., Bécane, H. M., Recan, D., … Muntoni, F. (2000). Clinical and molecular genetic spectrum of autosomal dominant Emery-Dreifuss muscular dystrophy due to mutations of the lamin A/C gene. *Ann Neurol*, *48*(2), 170–80.

Boyle, J., Robillard, G. T., & Kim, S.-H. (1980). Sequential folding of transfer RNA. *J Mol Biol*, *139*(4), 601–625.

Branden, C. and Tooze, J. (1992). *Introduction to protein tructure*. *Biochemical Education* (Vol. 20). Garland Publishing, New York.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem*, *4*(2), 187–217.

Brown, C. A., Lanning, R. W., McKinney, K. Q., Salvino, A. R., Cherniske, E., Crowe, C. A., … Spence, J. E. (2001). Novel and recurrent mutations in lamin A/C in patients with Emery-Dreifuss muscular dystrophy. *Am J Med Genet*, *102*(4), 359–367.

Bunch, J. S., Verbridge, S. S., Alden, J. S., Van Der Zande, A. M., Parpia, J. M., Craighead, H. G., & McEuen, P. L. (2008). Impermeable atomic membranes from graphene sheets. *Nano Lett*, *8*(8), 2458–2462.

Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., & Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res*, *34*(19), 5402–5415.

Bustamante, C., Chemla, Y. R., Forde, N. R., & Izhaky, D. (2004). Mechanical processes in biochemistry. *Annu Rev Biochem*, *73*, 705–748.

Butcher, S. E., Dieckmann, T., & Feigon, J. (1997). Solution structure of the conserved 16 S-like ribosomal RNA UGAA tetraloop. *J Mol Biol*, *268*(2), 348–358.

Butcher, S. E., & Pyle, A. M. (2011). The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc Chem Res*, *44*(12), 1302–1311.

C

Cai, W., Piner, R. D., Stadermann, F. J., Park, S., Shaibat, M. A., Ishii, Y., … Ruoff, R. S. (2008). Synthesis and Solid-State NMR Structural Characterization of 13C-Labeled Graphite Oxide. *Science*, *321*(5897), 1815–1817.

Calladine, C. R. (1982). Mechanics of sequence-dependent stacking of bases in B-DNA. *J Mol Biol*, *161*(2), 343–352.

Calladine, C. R., & Drew, H. R. (1984). A base-centred explanation of the B-to-A

transition in DNA. *J Mol Biol*, *178*(3), 773–782.

Calladine, C. R., Drew, H. R., Luisi, B. F., Travers, A. A. (2004). *Understanding DNA* (3rd Edition). Elsevier.

Car, R., & Parrinello, M. (1985). Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.*, *55*(22), 2471–2474.

Carell, T. (2004). Course-Chemical Biology of Nucleic Acids, 1–40.

Castro Neto, A., Guinea, F., Peres, N., Novoselov, K., & Geim, A. (2009). The electronic properties of graphene. *Rev Mod Phys*, *81*(1), 109–162.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., … Doudna, J. A. (1996a). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, *273*(5282), 1678–1685.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., … Doudna, J. A. (1996b). RNA Tertiary Structure Mediation by Adenosine Platforms. *Science*, *273*(5282), 1696–1699.

Cavaillé, J., & Bachellerie, J. P. (1998). SnoRNA-guided ribose methylation of rRNA: Structural features of the guide RNA duplex influencing the extent of the reaction. *Nucleic Acids Res*, *26*(7), 1576–1587.

Cech, T. R. (1990). Self-splicing of group I introns. *Annu Rev Biochem*, *59*, 543–68.

Chai, D. (2008). RNA structure and modeling: progress and techniques. *Prog Nucleic Acid Res Mol Biol*, *82*, 71–100.

Chai, J. Da, & Head-Gordon, M. (2008). Systematic optimization of long-range corrected hybrid density functionals. *J Chem Phys*, *128*(8), 084106.

Chang, H., Tang, L., Wang, Y., Jiang, J., & Li, J. (2010). Graphene fluorescence resonance energy transfer aptasensor for the thrombin detection. *Anal Chem*, *82*(6), 2341–2346.

Chang, K. Y., & Tinoco, I. (1994). Characterization of a "kissing" hairpin complex derived from the human immunodeficiency virus genome. *Proc Natl Acad Sci U S A*, *91*(18), 8705–8709.

Chargaff, E; Zamenhof, S; Green, C. (1950). Human desoxypentose nucleic acid:

Composition of human desoxypentose nucleic acid. *Nature*, *165*(4202), 756–757.

Chawla, M., Sharma, P., Halder, S., Bhattacharyya, D., & Mitra, A. (2011). Protonation of base Pairs in RNA: Context analysis and quantum chemical investigations of their geometries and stabilities. *J Phys Chem B*, *115*(6), 1469–1484.

Chen, B., Liu, M., Zhang, L., Huang, J., Yao, J., & Zhang, Z. (2011). Polyethylenimine-functionalized graphene oxide as an efficient gene delivery vector. *J Mater Chem*, *21*(21), 7736-7741.

Chen, J.-L., & Greider, C. W. (2005). Functional analysis of the pseudoknot structure in human telomerase RNA. *Proc Natl Acad Sci U S A*, *102*(23), 8080–8085; discussion 8077–8079.

Cheong, C., Varani, G., & Tinoco, I. (1990). Solution structure of an unusually stable RNA hairpin, 5GGAC (UUCG) GUCC. *Nature*, *346*(6285), 680–682.

Chou, P. Y., & Fasman, G. D. (1974). Prediction of protein conformation. *Biochemistry*, *13*(2), 222–245.

Chowdhury, S., & Bansal, M. (2000). Effect of coordinated ions on structure and flexiblity of parallel G-quandruplexes: a molecular dynamics study. *J Biomol Struct Dyn*, *18*(1), 11–28.

Clark, G. R., Squire, C. J., Baker, L. J., Martin, R. F., & White, J. (2000). Intermolecular interactions and water structure in a condensed phase B-DNA crystal. *Nucleic Acids Res*, *28*(5), 1259–1265.

Cohen-Tanugi, D., & Grossman, J. C. (2012). Water desalination across nanoporous graphene. *Nano Lett*, *12*(7), 3602–3608.

Constantinou, A., Davies, A. A., & West, S. C. (2001). Branch migration and Holliday junction resolution catalyzed by activities from mammalian cells. *Cell*, *104*(2), 259–268.

Convery, M. A., Rowsell, S., Stonehouse, N. J., Ellington, A. D., Hirao, I., Murray, J. B., … Stockley, P. G. (1998). Crystal structure of an RNA aptamer-protein complex at 2.8 A resolution. *Nature Struct Biol*, *5*(2), 133–139.

Cooper, D. R., D'Anjou, B., Ghattamaneni, N., Harack, B., Hilke, M., Horth, A., …

Yu, V. (2012). Experimental Review of Graphene. *ISRN Condensed Matter Physics*, 2012 (2012), Article ID 501686, 56 pages.

Cooper, G. M. (2000). *The Cell* (2nd edition). Sinauer Associates.

Correll, C. C., & Swinger, K. (2003). Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 A resolution. *RNA (New York, N.Y.), 9*(3), 355–363.

Cossi, M., Rega, N., Scalmani, G., & Barone, V. (2003). Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J Comput Chem*, *24*(6), 669–681.

Cramer, C. J. (2013). *Essentials of Computational Chemistry: Theories and Models* (2nd edition). John Wiley & Sons.

Creighton, T. E. (1992). *Proteins : structures and molecular properties* (2nd edition). W.H. Freeman.

Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*(5258), 561–563.

Dahm, R. (2008). Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum Genet*, *122*(6), 565–581.

D

Dao, N. T., Haselsberger, R., Michel-Beyerle, M. E., & Phan, A. T. (2011). Following G-quadruplex formation by its intrinsic fluorescence. *FEBS Lett*, *585*(24), 3969–3977.

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys*, *98*(12), 10089.

Das, J., Mukherjee, S., Mitra, A., & Bhattacharyya, D. (2006). Non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. *J Biomol Struct Dyn*, *24*(2), 149–61.

Dassonneville, L. (1997). Binding of Hoechst 33258 to the TAR RNA of HIV-1. Recognition of a pyrimidine bulge-dependent structure. *Nucleic Acids Res*, *25*(22), 4487–4492.

De Messieres, M., Chang, J. C., Brawn-Cinani, B., & La Porta, A. (2012). Single-molecule study of g-quadruplex disruption using dynamic force spectroscopy. *Phys Rev Lett*, *109*(5), 1–5.

De Santis, P., Palleschi, a, Savino, M., & Scipioni, a. (1990). Validity of the nearest-neighbor approximation in the evaluation of the electrophoretic manifestations of DNA curvature. *Biochemistry*, *29*(39), 9269–9273.

Dechat, T., Pfleghaar, K., Sengupta, K., Shimi, T., Shumaker, D. K., Solimando, L., & Goldman, R. D. (2008). Nuclear lamins: Major factors in the structural organization and function of the nucleus and chromatin. *Genes Dev 22*(7), 832–853.

Deng, H., & Braunlin, W. H. (1996). Kinetics of sodium ion binding to DNA quadruplexes. *J. Mol. Biol.*, *255*, 476–483.

Desalination.com (2013). Graphyne may outperform graphene for water desalination. http://www.graphene-info.com/graphyne-may-outperform-graphene-water-desalination

Desiraju, G. R., & Steiner, T. (2001). *The Weak Hydrogen Bond: In Structural Chemistry and Biology*. Oxford University Press.

Dey, D., Bose, A., Bhattacharyya, D., Basu, S., Maity, S. S., & Ghosh, S. (2007a). Dibenzo[a,c]phenazine: A polarity-insensitive hydrogen-bonding probe. *J Phys Chem A*, *111*(42), 10500–10506.

Dey, D., Bose, A., Chakraborty, M., & Basu, S. (2007b). Magnetic field effect on photoinduced electron transfer between dibenzo[a,c]phenazine and different amines in acetonitrile-water mixture. *J Phys Chem A*, *111*(5), 878–884.

Dey, D., Sarangi, M. K., Ray, A., Bhattacharyya, D., & Maity, D. K. (2015). Excited state hydrogen bonding fluorescent probe: Role of structure and environment. *J Lumin*, doi: 10.1016/j.jlumin.2015.12.011.

Dhakal, S., Cui, Y., Koirala, D., Ghimire, C., Kushwaha, S., Yu, Z., … Mao, H. (2013). Structural and mechanical properties of individual human telomeric G-quadruplexes in molecularly crowded solutions. *Nucleic Acids Res*, *41*(6), 3915–3923.

Dhe-Paganon, S., Werner, E. D., Chi, Y.-I. I., & Shoelson, S. E. (2002). Structure of the globular tail of nuclear lamin. *J Biol Chem*, *277*(20), 17381–17384.

Dickerson, R. E. (1989). Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res*, *17*(5), 1797–1803.

Dickerson, R. E. (1998). DNA bending: The prevalence of kinkiness and the virtues of normality. *Nucleic Acids Research*, *26*(8), 1906–1926.

Diener, J. L., & Moore, P. B. (1998). Solution Structure of a Substrate for the Archaeal Pre-tRNA Splicing Endonucleases: The Bulge-Helix-Bulge Motif. *Mol Cell*, *1*(6), 883–894.

Ditzler, M. A., Otyepka, M., Sponer, J., & Walter, N. G. (2010). Molecular dynamics and quantum mechanics of RNA: conformational and chemical change we can believe in. *Acc Chem Res*, *43*(1), 40–7.

Dixit, S. B., Beveridge, D. L., Case, D. A., Cheatham, T. E., Giudice, E., Lankas, F., … Varnai, P. (2005). Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys J*, *89*(6), 3721–3740.

Djuranovic, D., & Hartmann, B. (2004). DNA fine structure and dynamics in crystals and in solution: the impact of BI/BII backbone conformations. *Biopolymers*, *73*(3), 356–68.

Djuranovic, D., Oguey, C., & Hartmann, B. (2004). The role of DNA structure and dynamics in the recognition of bovine papillomavirus E2 protein target sequences. *J Mol Biol*, *339*(4), 785–96.

Donohue, J., & Trueblood, K. N. (1960). Base pairing in DNA. *J Mol Biol*, *2*, 363–371.

Drew, H. R., & Dickerson, R. E. (1981). Structure of a B-DNA dodecamer. III. Geometry of hydration. *J Mol Biol*, *151*(3), 535–556.

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1981). Structure of a B-DNA dodecamer: conformation and dynamics. *Proc Natl Acad Sci U S A*, *78*(4), 2179–2183.

Dreyer, D. R., Park, S., Bielawski, C. W., & Ruoff, R. S. (2010). The chemistry of graphene oxide. *Chem Soc Rev*, *39*(1), 228–240.

Duarte, C. M., & Pyle, A. M. (1998). Stepping through an RNA structure: A novel approach to conformational analysis. *J Mol Biol*, *284*(5), 1465–1478.

Dunbrack Jr, R. L., & Karplus, M. (1993). Backbone-dependent rotamer library for proteins application to side-chain prediction. *J. Mol. Biol*, *230*, 543-574.

Dunning Jr, T. H. (1989). Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, *90*, 1007.

E

Echols, H. (2001). *Operators and Promoters: The Story of Molecular Biology and Its Creators*. The versatility of RNA.

Eda, G., Fanchini, G., & Chhowalla, M. (2008). Large-area ultrathin films of reduced graphene oxide as a transparent and flexible electronic material. *Nat Nanotechnol*, 3, 270-274.

El Hassan, M. A., & Calladine, C. R. (1997). Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. *Phil Trans R Soc A*, *355*(1722), 43-100.

Elias, D. C., Nair, R. R., Mohiuddin, T. M. G., Morozov, S. V, Blake, P., Halsall, M. P., … Novoselov, K. S. (2009). Control of graphene's properties by reversible hydrogenation: evidence for graphane. *Science*, *323*(5914), 610–613.

Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C., & Dumas, P. (2001). Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Biol*, *8*(12), 1064–1068.

Ennifar, E., Yusupov, M., Walter, P., Marquet, R., Ehresmann, B., Ehresmann, C., & Dumas, P. (1999). The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure*, *7*(11), 1439–1449.

Erber, A., Riemer, D., Hofemeister, H., Bovenschulte, M., Stick, R., Panopoulou, G.,

… Weber, K. (1999). Characterization of the Hydra lamin and its gene: A molecular phylogeny of metazoan lamins. *J Mol Evol*, *49*(2), 260–271.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M.-Y., … Sali, A. (2006). Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et al.]*, *Chapter 5*, Unit 5.6.

Evans, E., & Ritchie, K. (1997). Dynamic strength of molecular adhesion bonds. *Biophys J*, *72*(4), 1541–1555.

Evans, T., Schon, E., Gora-Maslak, G., Patterson, J., & Efstratiadis, A. (1984). S1-hypersensitive sites in eukaryotic promoter regions. *Nucleic Acids Res*, *12*(21), 8043–8058.

F

Fadrná, E., Špačková, N., Sarzyñska, J., Koča, J., Orozco, M., Cheatham, T. E., … Šponer, J. (2009). Single stranded loops of quadruplex DNA as key benchmark for testing nucleic acids force fields. *J Chem Theory Comput*, *5*(9), 2514–2530.

Feig, M., & Brooks, C. L. (2004). Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr Opin Struct Biol*, *14*(2), 217–24.

Feller, S. E., Zhang, Y., Pastor, R. W., & Brooks, B. R. (1995). Constant pressure molecular dynamics simulation: the Langevin piston method. *J Chem Phys*, *103*(11), 4613–4621.

Feng, L., Zhang, S., & Liu, Z. (2011). Graphene based gene transfection. *Nanoscale*, *3*(3), 1252–1257.

Ferré-D'Amaré, A. R., Zhou, K., & Doudna, J. A. (1998). Crystal structure of a hepatitis delta virus ribozyme. *Nature*, *395*(6702), 567–574.

Fisher, D. Z., Chaudhary, N., & Blobel, G. (1986). cDNA sequencing of nuclear lamins A and C reveals primary and secondary structural homology to intermediate filament proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(17), 6450–6454.

Foloppe, N., & MacKerell, A. D. J. (2000). All-Atom Empirical Force Field for Nucleic Acids : I . Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J Comput Chem*, *21*(2), 86–104.

Frank-Kamenetskii, M. D., & Mirkin, S. M. (1995). Triplex DNA structures. *Annu Rev Biochem*, *64*, 65–95.

Franklin, R. E., & Gosling, R. G. (1953). Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature*, *172*(4369), 156–157.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., … Fox, D. J. (2009). Gaussian 09, Revision A.02. *Gaussian Inc Wallingford CT*.

Frith, M. C., Pheasant, M., & Mattick, J. S. (2005). The amazing complexity of the human transcriptome. *Eur J Hum Genet*, *13*(8), 894-897.

Fry, M., & Loeb, L. a. (1994). The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc Natl Acad Sci U S A*, *91*(11), 4950–4954.

Fu, T. J., Tse-Dinh, Y. C., & Seeman, N. C. (1994). Holliday junction crossover topology. *J Mol Biol*, *236*(1), 91–105.

Fujii, S., Kono, H., Takenaka, S., Go, N., & Sarai, A. (2007). Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucl Acids Res*, *35*(18), 6063–6074.

G

Gagniuc, P. a, & Ionescu-Tirgoviste, C. (2012). Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters. *BMC Genomics*, *13*(1), 512.

Gao, M., Wilmanns, M., & Schulten, K. (2002). Steered molecular dynamics studies of titin I1 domain unfolding. *Biophys J*, *83*(6), 3435–3445.

Gehring, K., Leroy, J. L., & Guéron, M. (1993). A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature*, *363*(6429), 561–565.

Geim, A. K., & Novoselov, K. S. (2007). The rise of graphene. *Nat Mater*, *6*(3), 183–191.

Gellert, M., Lipsett, M. N., & Davies, D. R. (1962). Helix formation by guanylic acid. *Proc Natl Acad Sci U S A*, *48*(12), 2013–2018.

Gendron, P., Lemieux, S., & Major, F. (2001). Quantitative analysis of nucleic acid three-dimensional structures. *J Mol Biol*, *308*(5), 919–936.

George A. Jeffrey. (1997). *An Introduction to Hydrogen Bonding*. Oxford University Press.

Gilje, S., Han, S., Wang, M., Wang, K. L., & Kaner, R. B. (2007). A chemical route to graphene for device applications. *Nano Lett*, *7*(11), 3394–3398.

Gkionis, K., Kruse, H., Platts, J. A., Mládek, A., Koča, J., & Šponer, J. (2014). Ion binding to quadruplex DNA stems. comparison of MM and QM descriptions reveals sizable polarization effects not included in contemporary simulations. *J Chem Theory Comput*, *10*(3), 1326–1340.

Glasbeek, M., & Zhang, H. (2004). Femtosecond studies of solvation and intramolecular configurational dynamics of fluorophores in liquid solution. *Chem Rev*, *104*(4), 1929–54.

Glendening, E. D., Landis, C. R., & Weinhold, F. (2012). Natural bond orbital methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science*.

Gogonea, V., Westerhoff, L. M., & Merz, K. M. (2000). Quantum mechanical/quantum mechanical methods. I. A divide and conquer strategy for solving the Schrödinger equation for large molecular systems using a composite density functional–semiempirical Hamiltonian. *J Chem Phys*, *113*(14), 5604.

Goldman, A. E., Maul, G., Steinert, P. M., Yang, H. Y., & Goldman, R. D. (1986). Keratin-like proteins that coisolate with intermediate filaments of BHK-21 cells are nuclear lamins. *Proc Natl Acad Sci U S A*, *83*(11), 3839–3843.

Goodsell, D. S., Grzeskowiak, K., & Dickerson, R. E. (1995). Crystal structure of C-T-C-T-C-G-A-G-A-G. Implications for the structure of the Holliday junction. *Biochemistry*, *34*(3), 1022–1029.

Gorin, A. A., Zhurkin, V. B., & Olson, W. K. (1995). B-DNA twisting correlates with base-pair morphology. *J Mol Biol*, *247*(1), 34–48.

Grate, D., & Wilson, C. (1997). Role REVersal: understanding how RRE RNA binds

its peptide ligand. *Structure*, *5*(1), 7–11.

Gräter, F., Shen, J., Jiang, H., Gautel, M., & Grubmüller, H. (2005). Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. *Biophys J*, *88*(2), 790–804.

Grochowski, P., Lesyng, B., Bala, P., & McCammon, J. A. (1996). Density functional based parametrization of a valence bond method and its applications in quantum-classical molecular dynamics simulations of enzymatic reactions. *Int J Quant Chem*, *60*(6), 1143–1164.

Grosjean, H., Sprinzl, M., & Steinberg, S. (1995). Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies. *Biochimie*, *77*(1-2), 139–141.

Gupta, G., Bansal, M., & Sasisekharan, V. (1980). Conformational flexibility of DNA: polymorphism and handedness. *Proc Natl Acad Sci U S A*, *77*(11), 6486–6490.

Guschlbauer, W., Chantot, J. F., & Thiele, D. (1990). Four-stranded nucleic acid structures 25 years later: from guanosine gels to telomer DNA. *J Biomol Struct Dyn*, *8*(3), 491–511.

Guvench, O., Weiser, J., Shenkin, P., Kolossváry, I., & Still, W. (2002). Application of the frozen atom approximation to the GB/SA continuum model for solvation free energy. *J Comput Chem*, *23*(2), 214–221.

H

Hagerman, K. R., & Hagerman, P. J. (1996). Helix rigidity of DNA: the meroduplex as an experimental paradigm. *J Mol Biol*, *260*(2), 207–223.

Haider, S. M., Neidle, S., & Parkinson, G. N. (2011). A structural analysis of G-quadruplex/ligand interactions. *Biochimie*, *93*(8), 1239–1251.

Haider, S., Parkinson, G. N., & Neidle, S. (2002). Crystal structure of the potassium form of an Oxytricha nova G-quadruplex. *J Mol Biol*, *320*(2), 189–200.

Haider, S., Parkinson, G. N., & Neidle, S. (2008). Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophys J*,

95(1), 296–311.

Halder, A., Halder, S., Bhattacharyya, D., & Mitra, A. (2014). Feasibility of occurrence of different types of protonated base pairs in RNA: a quantum chemical study. *Phys Chem Chem Phys*, *16*(34), 18383–96.

Halder, S., & Bhattacharyya, D. (2010). Structural stability of tandemly occurring noncanonical basepairs within double helical fragments: molecular dynamics studies of functional RNA. *J Phys Chem B*, *114*(44), 14028–14040.

Halder, S., & Bhattacharyya, D. (2012). Structural variations of single and tandem mismatches in RNA duplexes: a joint MD simulation and crystal structure database analysis. *J Phys Chem B*, *116*(39), 11845–11856.

Halder, S., & Bhattacharyya, D. (2013). RNA structure and dynamics: a base pairing perspective. *Prog Biophys Mol Biol*, *113*(2), 264–283.

Han, K.-L., & Zhao, G.-J. (2011). *Hydrogen Bonding and Transfer in the Excited State (Vol. I & II)*. John Wiley & Sons.

Hanukoglu, I. (2015). Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochem Mol Biol Educ*, *43*(3), 206–209.

Hartmann, B., Piazzola, D., & Lavery, R. (1993). BI-BII transitions in B-DNA. *Nucleic Acids Res*, *21*(3), 561–568.

Hazel, P., Huppert, J., Balasubramanian, S., & Neidle, S. (2004). Loop-length-dependent folding of G-quadruplexes. *J Am Chem Soc*, *126*(50), 16405–16415.

Hazel, P., Parkinson, G. N., & Neidle, S. (2006). Predictive modelling of topology and loop variations in dimeric DNA quadruplex structures. *Nucleic Acids Res*, *34*(7), 2117–2127.

He, H., Riedl, T., Lerf, A., & Klinowski, J. (1996). Solid-State NMR Studies of the Structure of Graphite Oxide. *J. Phys. Chem.*, *100*(51), 19954–19958.

He, H. Y., Klinowski, J., Forster, M., & Lerf, A. (1998). A new structural model for graphite oxide. *Chem Phys Lett*, *287*(1-2), 53–56.

He, L., & Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet*, *5*(7), 522–531.

He, Y., Neumann, R. D., & Panyutin, I. G. (2004). Intramolecular quadruplex conformation of human telomeric DNA assessed with 125I-radioprobing. *Nucleic Acids Res*, *32*(18), 5359–5367.

Heddi, B., & Phan, A. T. (2011). Structure of human telomeric DNA in crowded solution. *J Am Chem Soc*, *133*(25), 9824–9833.

Heinemann, U., Alings, C., & Bansal, M. (1992). Double helix conformation, groove dimensions and ligand binding potential of a G/C stretch in B-DNA. *EMBO J*, *11*(5), 1931–1939.

Hendrix, D. K., Brenner, S. E., & Holbrook, S. R. (2005). RNA structural motifs: building blocks of a modular biomolecule. *Q Rev Biophys*, *38*(3), 221–243.

Hermann, T., & Patel, D. J. (1999). Stitching together RNA tertiary architectures. *J Mol Biol*, *294*(4), 829–849.

Hermann, T., & Patel, D. J. (2000). RNA bulges as architectural and recognition motifs. *Structure*, *8*(3), R47–R54.

Hermann, T., & Westhof, E. (1999). Non-Watson-Crick base pairs in RNA-protein recognition. *Chem Biol*, *6*(12), R335–R343.

Hernández Rosas, J. J., Ramírez Gutiérrez, R. E., Escobedo-Morales, A., & Chigo Anota, E. (2011). First principles calculations of the electronic and chemical properties of graphene, graphane, and graphene oxide. *J Mol Model*, *17*(5), 1133–1139.

Hershey, A. D., & Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol*, *36*(1), 39–56.

Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput*, *4*(3), 435–447.

Hesselmann, A., Jansen, G., & Schütz, M. (2006). Interaction energy contributions of H-bonded and stacked structures of the AT and GC DNA base pairs from the combined density functional theory and intermolecular perturbation theory approach. *J Am Chem Soc*, *128*(36), 11730–11731.

Holbrook, S. R., Sussman, J. L., Warrant, R. W., & Kim, S. H. (1978). Crystal

structure of yeast phenylalanine transfer RNA. II. Structural features and functional implications. *J Mol Biol*, *123*(4), 631–660.

Holliday, R. (1964). The Induction of Mitotic Recombination by Mitomycin C in Ustilago and Saccharomyces. *Genetics*, *50*(3), 323–35.

Hoogsteen, K. (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica*, *16*(9), 907–916.

Hopper, A. K., & Phizicky, E. M. (2003). tRNA transfers to the limelight. *Genes Dev*, *17*(2), 162–180.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, *65*(3), 712–725.

Hud, N. V., Smith, F. W., Anet, F. a L., & Feigon, J. (1996). The selectivity for K+ versus Na+ in DNA quadruplexes is dominated by relative free energies of hydration: A thermodynamic analysis by 1H NMR. *Biochemistry*, *35*(48), 15383–15390.

Hummers, W. M. J., & Offeman, R. E. (1958). Preparation of Graphitic Oxide. *J. Am. Chem. Soc*, *80*(6), 1339-1339.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph*, *14*(1), 33–38.

Hunter, C. A., & Lu, X. J. (1997a). Construction of double-helical DNA structures based on dinucleotide building blocks. *J Biomol Struct Dyn*, *14*(6), 747–756.

Hunter, C. A., & Lu, X. J. (1997b). DNA base-stacking interactions: a comparison of theoretical calculations with oligonucleotide X-ray crystal structures. *J Mol Biol*, *265*(5), 603–619.

Huppert, J. L. (2007). Four-stranded DNA: cancer, gene regulation and drug development. *Phil Trans A Math Phys Eng Sci*, *365*(1861), 2969–84.

Huppert, J. L., & Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic Acids Res*, *33*(9), 2908–2916.

I

Ida, R., & Wu, G. (2008). Direct NMR detection of alkali metal ions bound to G-quadruplex DNA. *J Am Chem Soc*, *130*(11), 3590–3602.

Ince, C.; Coremans, J. M. C. C.; Bruining, H. A. (1992). *In vivo NADH fluorescence,. In Adv. Exp. Med.:Oxygen Transport to Tissue XIV*. (W. Erdmann & D. F. Bruley, Eds.) (Vol. 317). Plenum Press, New York.

Islam, B., Sgobba, M., Laughton, C., Orozco, M., Sponer, J., Neidle, S., & Haider, S. (2013). Conformational dynamics of the human propeller telomeric DNA quadruplex on a microsecond time scale. *Nucl Acids Res*, *41*, 2723–2735.

Ivanovskii, A. L. (2012). Graphene-based and graphene-like materials. *Russ Chem Rev*, *81*(7), 571.

Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y., & Schulten, K. (1997). Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys J*, *72*(4), 1568–1581.

J

Jagannathan, B., & Marqusee, S. (2013). Protein folding and unfolding under force. *Biopolymers*, *99*(11), 860–869.

Jain, A., Wang, G., & Vasquez, K. M. (2008). DNA triple helices: Biological consequences and therapeutic potential. *Biochimie*, *90*(8), 1117-1130.

Jang, H., Kim, Y. K., Kwon, H. M., Yeo, W. S., Kim, D. E., & Min, D. H. (2010). A graphene-based platform for the assay of duplex-DNA unwinding by helicase. *Angew Chem Int Ed*, *49*(33), 5703–5707.

Jensen, F. (2006). *Introduction to Computational Chemistry* (2nd Edition). John Wiley & Sons.

Jiang, L., & Patel, D. J. (1998). Solution structure of the tobramycin-RNA aptamer complex. *Nature Struct Biol*, *5*(9), 769–774.

Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H., & Pardi, A. (1996). A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol*, *264*(5), 968–

980.

Jucker, F. M., & Pardi, A. (1995). GNRA tetraloops make a U-turn. *RNA*, *1*(2), 219–222.

Jung, J. H., Cheon, D. S., Liu, F., Lee, K. B., & Seo, T. S. (2010). A graphene oxide based immuno-biosensor for pathogen detection. *Angew Chem Int Ed*, *49*(33), 5708–5711.

K

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*(12), 2577–2637.

Kailasam, S., Bhattacharyya, D., & Bansal, M. (2014). Sequence dependent variations in RNA duplex are related to non-canonical hydrogen bond interactions in dinucleotide steps. *BMC Res Notes*, *7*(1), 83.

Kapranov, P., Willingham, A. T., & Gingeras, T. R. (2007). Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet*, *8*(6), 413–423.

Karplus, M. (2003). Molecular dynamics of biological macromolecules: A brief history and perspective. *Biopolymers*, *68*(3), 350-358.

Khare, R., Mielke, S. L., Paci, J. T., Zhang, S., Ballarini, R., Schatz, G. C., & Belytschko, T. (2007). Coupled quantum mechanical/molecular mechanical modeling of the fracture of defective carbon nanotubes and graphene sheets. *Phys Rev B*, *75*(7), 075412.

Kilpatrick, M. W., Torri, A., Kang, D. S., Engler, J. A., & Wells, R. D. (1986). Unusual DNA structures in the adenovirus genome. *J Biol Chem*, *261*(24), 11350–11354.

Kim, J., Walter, A. E., & Turner, D. H. (1996). Thermodynamics of coaxially stacked helixes with GA and CC mismatches. *Biochemistry*, *35*(43), 13753–61.

Kim, S. H., Sussman, J. L., Suddath, F. L., Quigley, G. J., McPherson, A., Wang, A. H., … RICH, A. (1974). The general structure of transfer RNA molecules. *Proc*

Natl Acad Sci U S A, 71(12), 4970–4974.

Klosterman, P. S., Hendrix, D. K., Tamura, M., Holbrook, S. R., & Brenner, S. E. (2004). Three-dimensional motifs from the SCOR, structural classification of RNA database: Extruded strands, base triples, tetraloops and U-turns. *Nucleic Acids Res, 32*(8), 2342–2352.

Klosterman, P. S., Tamura, M., Holbrook, S. R., & Brenner, S. E. (2002). SCOR: a Structural Classification of RNA database. *Nucleic Acids Res, 30*(1), 392–394.

Kobayashi, Y., Fukui, K. I., Enoki, T., & Kusakabe, K. (2006). Edge state on hydrogen-terminated graphite edges investigated by scanning tunneling microscopy. *Phys Rev B, 73*(12), 125415.

Kobayashi, Y., Fukui, K. I., Enoki, T., Kusakabe, K., & Kaburagi, Y. (2005). Observation of zigzag and armchair edges of graphite using scanning tunneling microscopy and spectroscopy. *Phys Rev B, 71*(19), 193406.

Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science (New York, N.Y.), 184*(139), 868–871.

Koskinen, P., Malola, S., & Häkkinen, H. (2008). Self-passivating edge reconstructions of graphene. *Phys Rev Lett, 101*(11).

Krimm, I., Östlund, C., Gilquin, B., Couprie, J., Hossenlopp, P., Mornon, J. P., … Zinn-Justin, S. (2002). The Ig-like structure of the C-terminal domain of lamin A/C, mutated in muscular dystrophies, cardiomyopathy, and partial lipodystrophy. *Structure, 10*(6), 811–823.

L

Lane, A. N., Chaires, J. B., Gray, R. D., & Trent, J. O. (2008). Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res, 36*(17), 5482–5515.

Langridge, R., & Rich, A. (1963). Molecular structure of helical polycytidylic acid. *Nature, 198*, 725–728.

Larciprete, R., Fabris, S., Sun, T., Lacovig, P., Baraldi, A., & Lizzit, S. (2011). Dual path mechanism in the thermal reduction of graphene oxide. *J Am Chem Society, 133*(43), 17315–17321.

Laughlan, G., Murchie, A. I., Norman, D. G., Moore, M. H., Moody, P. C., Lilley, D. M., & Luisi, B. (1994). The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science*, *265*(5171), 520–524.

Lavery, R. (2005). *Multiple aspects of DNA and RNA from Biophysics to Bioinformatics*. *Elsevier Book Series on ScienceDirect*.

Lavery, R., & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dyn*, *6*(1), 63–91.

Leach, A. R. (2001). *Molecular Modelling: Principles and Applications* (2nd Edition), *second edition*. Pearson Education EMA.

Lech, C. J., Heddi, B., & Phan, A. T. (2013). Guanine base stacking in G-quadruplex nucleic acids. *Nucleic Acids Res*, *41*(3), 2034–2046.

Lee, A. J., & Crothers, D. M. (1998). The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence. *Structure*, *6*(8), 993–1005.

Lee, C., Wei, X., Kysar, J. W., & Hone, J. (2008). Measurement of the elastic properties and intrinsic strength of monolayer graphene. *Science*, *321*(5887), 385–388.

Lee, C., Yang, W., & Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B*, *37*(2), 785–789.

Lee, J. C., Cannone, J. J., & Gutell, R. R. (2003). The lonepair triloop: a new motif in RNA structure. *J Mol Biol*, *325*(1), 65–83.

Lee, J. C., & Gutell, R. R. (2004). Diversity of base-pair conformations and their occurrence in rRNA structure and RNA structural motifs. *J Mol Biol*, *344*(5), 1225–1249.

Lee, J. C., & Gutell, R. R. (2014). Helix capping in RNA structure. *PloS One*, *9*(4), e93664.

Lee, J. Y., Okumus, B., Kim, D. S., & Ha, T. (2005). Extreme conformational diversity in human telomeric DNA. *Proc Natl Acad Sci U S A*, *102*, 18938–18943.

Lemieux, S., & Major, F. (2002). RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res*, *30*(19), 4250–4263.

Leontis, N. B., Lescoute, A., & Westhof, E. (2006). The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol*, *16*(3), 279-287.

Leontis, N. B., Stombaugh, J., & Westhof, E. (2002). The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res*, *30*(16), 3497–3531.

Leontis, N. B., & Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *RNA*, *7*(4), 499–512.

Leontis, N. B., & Westhof, E. (2002). The annotation of RNA motifs. *Comp Funct Genomics*, *3*(6), 518–524.

Lerf, A., He, H., Forster, M., & Klinowski, J. (1998). Structure of Graphite Oxide Revisited. *J Phys Chem B*, *102*(23), 4477–4482.

Leroy, J. L., Guéron, M., Mergny, J. L., & Hélène, C. (1994). Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif. *Nucleic Acids Res*, *22*(9), 1600–1606.

Lescoute, A., & Westhof, E. (2005). Riboswitch structures: Purine ligands replace tertiary contacts. *Chem Biol*, *12*(1), 10-13.

Lesk, A. M. (1995). NAD-binding domains of dehydrogenases. *Curr Opin Struct Biol*, *5*(6), 775–783.

Levitt, M., & Chothia, C. (1976). Structural patterns in globular proteins. *Nature*, *261*(5561), 552–558.

Li, H., Cao, E. H., & Gisler, T. (2009). Force-induced unfolding of human telomeric G-quadruplex: A steered molecular dynamics simulation study. *Biochem Biophys Res Commun*, *379*(1), 70–75.

Li, J. L., Kudin, K. N., McAllister, M. J., Prud'homme, R. K., Aksay, I. A., & Car, R. (2006). Oxygen-driven unzipping of graphitic materials. *Phys Rev Lett*, *96*(17), 176101.

Li, Z., Zhang, W., Luo, Y., Yang, J., & Hou, J. G. (2009). How graphene is cut upon oxidation? *J Am Chem Soc*, *131*(18), 6320–6321.

Lietzke, S. E., Barnes, C. L., Berglund, J. A., & Kundrot, C. E. (1996). The structure of an RNA dodecamer shows how tandem U-U base pairs increase the range of stable RNA structures and the diversity of recognition sites. *Structure*, *4*(8), 917–930.

Lilley, D. M. (1998). Folding of branched RNA species. *Biopolymers*, *48*(2-3), 101–112.

Lilley, D. M. (2000). Structures of helical junctions in nucleic acids. *Q Rev Biophys*, *33*(2), 109–159.

Lim, K. W., Lacroix, L., Yue, D. J. E., Lim, J. K. C., Lim, J. M. W., & Phan, A. T. (2010). Coexistence of two distinct G-quadruplex conformations in the hTERT promoter. *J Am Chem Soc*, *132*(35), 12331–12342.

Lim, K. W., Ng, V. C. M., Martín-Pintado, N., Heddi, B., & Phan, A. T. (2013). Structure of the human telomere in Na+ solution: An antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Res*, *41*(22), 10556–10562.

Lipanov, A., Kopka, M. L., Kaczor-Grzeskowiak, M., Quintana, J., & Dickerson, R. E. (1993). Structure of the B-DNA decamer C-C-A-A-C-I-T-T-G-G in two different space groups: conformational flexibility of B-DNA. *Biochemistry*, *32*(5), 1373–1389.

Liu, J., Rinzler, A. G., Dai, H., Hafner, J. H., Bradley, R. K., Boul, P. J., … Smalley, R. E. (1998). Fullerene Pipes. *Science*, *280*(5367), 1253 –1256.

Liu, Z., Robinson, J. T., Sun, X., & Dai, H. (2008). PEGylated nanographene oxide for delivery of water-insoluble cancer drugs. *J Am Chem Soc*, *130*(33), 10876–10877.

Loh, K. P., Bao, Q., Ang, P. K., & Yang, J. (2010). The chemistry of graphene. *J Mater Chem*, *20*(12), 2277-2289.

Long, X., Parks, J. W., Bagshaw, C. R., & Stone, M. D. (2013). Mechanical unfolding of human telomere G-quadruplex DNA probed by integrated fluorescence and

magnetic tweezers spectroscopy. *Nucleic Acids Res*, *41*(4), 2746–2755.

Lu, C. H., Yang, H. H., Zhu, C. L., Chen, X., & Chen, G. N. (2009). A graphene platform for sensing biomolecules. *Angew Chem Int Ed*, *48*(26), 4785–4787.

Lu, H., & Schulten, K. (1999). Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins*, *35*(4), 453–463.

Lu, X., Huang, H., Nemchuk, N., & Ruoff, R. S. (1999). Patterning of highly oriented pyrolytic graphite by oxygen plasma etching. *Appl Phys Lett*, *75*(2), 193.

Lu, X. J., El Hassan, M. A., & Hunter, C. A. (1997). Structure and conformation of helical nucleic acids: analysis program (SCHNAaP). *J Mol Biol*, *273*(3), 668–680.

Lu, X. J., & Olson, W. K. (2003). 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res*, *31*(17), 5108–5121.

Lu, X.-J., Olson, W. K., & Bussemaker, H. J. (2010). The RNA backbone plays a crucial role in mediating the intrinsic stability of the GpU dinucleotide platform and the GpUpA/GpA miniduplex. *Nucleic Acids Res*, *38*(14), 4868–4876.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, *389*(6648), 251–260.

Luu, K. N., Phan, A. T., Kuryavyi, V., Lacroix, L., & Patel, D. J. (2006). Structure of the human telomere in K+ solution: An intramolecular (3 + 1) G-quadruplex scaffold. *J Am Chem Soc*, *128*(30), 9963–9970.

M

Macgregor, R. B., & Poon, G. M. K. (2003). The DNA double helix fifty years on. *Comp Biol Chem*, *27*(4-5), 461-467.

Mackerell, A. D., Feig, M., & Brooks, C. L. (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation. *J Comput Chem*, *25*(11), 1400–1415.

Maden, B. E. H. (1990). The Numerous Modified Nucleotides in Eukaryotic Ribosomal RNA. *Prog Nucleic Acid Res Mol Biol*, *39*(C), 241–303.

Major, F., & Griffey, R. (2001). Computational methods for RNA structure determination. *Curr Opin Struct Biol*, *11*(3), 282-286.

Malkov, V. A., Voloshin, O. N., Veselkov, A. G., Rostapshov, V. M., Jansen, I., Soyfer, V. N., & Frank-Kamenetskii, M. D. (1993). Protonated pyrimidine-purine-purine triplex. *Nucleic Acids Res*, *21*(1), 105–111.

Manning, G. S. (1978). The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys*, *11*(2), 179–246.

Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., & De Vries, A. H. (2007). The MARTINI force field: Coarse grained model for biomolecular simulations. *J Phys Chem B*, *111*(27), 7812–7824.

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*, *29*, 291–325.

Maruyama, Y., Matsushita, T., Ueoka, R., & Hirata, F. (2011). Solvent and salt effects on structural stability of human telomere. *J Phys Chem B*, *115*(10), 2408–2416.

Mattick, J. S., & Makunin, I. V. (2006). Non-coding RNA. *Hum Mol Gen*, *15*(Spec No 1), R17–29.

McAllister, M. J., Li, J. L., Adamson, D. H., Schniepp, H. C., Abdala, A. A., Liu, J., … Aksay, I. A. (2007). Single sheet functionalized graphene by oxidation and thermal expansion of graphite. *Chem Mater*, *19*(18), 4396–4404.

McKeon, F. D., Kirschner, M. W., & Caput, D. (1986). Homologies in both primary and secondary structure between nuclear envelope and intermediate filament proteins. *Nature*, *319*(6053), 463–468.

Mei, H. Y., Cui, M., Heldsinger, A., Lemrow, S. M., Loo, J. A., Sannes-Lowery, K. A., … Czarnik, A. W. (1998). Inhibitors of protein-RNA complexation that target the RNA: specific recognition of human immunodeficiency virus type 1

TAR RNA by small organic molecules. *Biochemistry*, *37*(40), 14204–14212.

Meselson, M; Stahl, F W; Vinograd, J. (1957). Equilibrium sedimentation of macromolecules in density gradients. *Proc Natl Acad Sci U S A*, *43*(7), 581–587.

Meyer, M., Steinke, T., Brandl, M., & Suhnel, J. (2001). Density functional study of guanine and uracil quartets and of guanine quartet/metal ion complexes. *J Comput Chem*, *22*(1), 109–124.

Michel, F., & Westhof, E. (1990). Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol*, *216*(3), 585–610.

Mohanty, D., & Bansal, M. (1991). DNA polymorphism and local variation in base-pair orientation: a theoretical rationale. *J Biomol Struct Dyn*, *9*(1), 127-142.

Møller, C., & Plesset, M. S. (1934). Note on an Approximation Treatment for Many-Electron Systems. *Phys Rev*, *46*(7), 618–622.

Mondal, M., Mukherjee, S., Halder, S., & Bhattacharyya, D. (2015). Stacking geometry for non-canonical G:U wobble base pair containing dinucleotide sequences in RNA: dispersion-corrected DFT-D study. *Biopolymers*, *103*(6), 328–38.

Moore, P. B. (1999). Structural motifs in RNA. *Annu Rev Biochem*, *68*, 287–300.

Mukherjee, S., Bansal, M., & Bhattacharyya, D. (2006). Conformational specificity of non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. *J Comput Aided Mol Des*, *20*(10-11), 629–645.

Mukherjee, S., & Bhattacharyya, D. (2013). Influence of divalent magnesium ion on DNA: molecular dynamics simulation studies. *J Biomol Struct Dyn*, *31*(8), 896–912.

Mukherjee, S., Kailasam, S., Bansal, M., & Bhattacharyya, D. (2014). Energy hyperspace for stacking interaction in AU/AU dinucleotide step: Dispersion-corrected density functional theory study. *Biopolymers*, *101*(1), 107–120.

Mukherjee, S., Kailasam, S., Bansal, M., & Bhattacharyya, D. (2015). Stacking interactions in RNA and DNA: Roll-slide energy hyperspace for ten unique dinucleotide steps. *Biopolymers*, *103*(3), 134–147.

Murphy, F. L., Wang, Y. H., Griffith, J. D., & Cech, T. R. (1994). Coaxially stacked RNA helices in the catalytic center of the Tetrahymena ribozyme. *Science*, *265*(5179), 1709–1712.

N

Nagaswamy, U. (2002). NCIR: a database of non-canonical interactions in known RNA structures. *Nucleic Acids Res*, *30*(1), 395–397.

Nagaswamy, U., Gao, X., Martinis, S. A., & Fox, G. E. (2001). NMR structure of a ribosomal RNA hairpin containing a conserved CUCAA pentaloop. *Nucleic Acids Res*, *29*(24), 5129–5139.

Nair, R. R., Blake, P., Grigorenko, A. N., Novoselov, K. S., Booth, T. J., Stauber, T., … Geim, A. K. (2008). Fine structure constant defines visual transparency of graphene. *Science*, *320*(5881), 1308.

Nair, R. R., Ren, W., Jalil, R., Riaz, I., Kravets, V. G., Britnell, L., … Geim, A. K. (2010). Fluorographene: A two-dimensional counterpart of Teflon. *Small*, *6*(24), 2877–2884.

Naryshkin, N. A., Gait, M. J., & Ivanovskaya, M. G. (1998). RNA recognition and regulation of HIV-1 gene expression by viral factor Tat. *Biochemistry* (Mosc), *63*(5), 489–503.

Neidle, S. (2002). *Nucleic Acid Structure and Recognition.*Oxford University Press.

Neidle, S. (2007). *Principles of Nucleic Acid Structure* (1st Edition). Elsevier

Neidle, S., & Parkinson, G. N. (2003). The structure of telomeric DNA. *Curr Opin Struct Biol*, *13*(3), 275–283.

Nelson, D. L., & Cox, M. M. (2004). *Lehninger Principles of Biochemistry* (4th Edition). W. H. Freeman.

Nguyen, T., Brunson, D., Crespi, C. L., Penman, B. W., Wishnok, J. S., & Tannenbaum, S. R. (1992). DNA damage and mutation in human cells exposed to nitric oxide in vitro. *Proc Natl Acad Sci U S A*, *89*(7), 3030–3034.

Nissen, P. (2000). The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science*, *289*(5481), 920–930.

Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., & Steitz, T. A. (2001). RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci U S A*, *98*(9), 4899–4903.

Noller, H. F. (2005). RNA structure: reading the ribosome. *Science (New York, N.Y.)*, *309*(5740), 1508–1514.

Novoselov, K. S., Fal′ko, V. I., Colombo, L., Gellert, P. R., Schwab, M. G., & Kim, K. (2012). A roadmap for graphene. *Nature*.

Novoselov, K. S., Geim, A. K., Morozov, S. V, Jiang, D., Zhang, Y., Dubonos, S. V, … Firsov, A. A. (2004). Electric field effect in atomically thin carbon films. *Science*, *306*(5696), 666–669.

O

Oberhauser, A. F., & Carrión-Vázquez, M. (2008). Mechanical biochemistry of proteins one molecule at a time. *J Biol Chem*, *283*(11), 6617–6621.

Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., … Berman, H. M. (2001). A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol*, *313*(1), 229–237.

Olson, W. K., Esguerra, M., Xin, Y., & Lu, X. J. (2009). New information content in RNA base pairing deduced from quantitative analysis of high-resolution structures. *Methods*, *47*(3), 177-186.

Ou, T. M., Lu, Y. J., Tan, J. H., Huang, Z. S., Wong, K. Y., & Gu, L. Q. (2008). G-Quadruplexes: Targets in Anticancer Drug Design. *ChemMedChem*, 3(5), 690-713.

P

Packer, M. J., Dauncey, M. P., & Hunter, C. A. (2000). Sequence-dependent DNA structure: dinucleotide conformational maps. *J Mol Biol*, *295*(1), 71–83.

Packer, M. J., & Hunter, C. A. (1998). Sequence-dependent DNA structure: the role of the sugar-phosphate backbone. *J Mol Biol*, *280*(3), 407–420.

Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., & Lipps, H. J. (2005).

Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat Struct Mol Biol*, *12*(10), 847–854.

Pal, S. K., & Zewail, A. H. (2004). Dynamics of water in biological recognition. *Chem Rev*, *104*(4), 2099–123.

Palacký, J., Vorlíčková, M., Kejnovská, I., & Mojzeš, P. (2013). Polymorphism of human telomeric quadruplex structure controlled by DNA concentration: A Raman study. *Nucleic Acids Res*, *41*, 1005–1016.

Pan, Y., Bao, H., Sahoo, N. G., Wu, T., & Li, L. (2011). Water-soluble poly(N-isopropylacrylamide)-graphene sheets synthesized via click chemistry for drug delivery. *Adv Funct Mater*, *21*(14), 2754–2763.

Panigrahi, S., Bhattacharya, A., Bandyopadhyay, D., Grabowski, S. J., Bhattacharyya, D., & Banerjee, S. (2011). Wetting property of the edges of monoatomic step on graphite: Frictional-force microscopy and ab initio quantum chemical studies. *J Phys Chem C*, *115*(30), 14819–14826.

Parkinson, G. N., Lee, M. P. H., & Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, *417*(6891), 876–880.

Pauling, L., & Corey, R. B. (1953). A Proposed Structure For The Nucleic Acids. *Proc Natl Acad Sci U S A*, *39*(2), 84–97.

Penedo, J. C., Wilson, T. J., Jayasena, S. D., Khvorova, A., & Lilley, D. M. J. (2004). Folding of the natural hammerhead ribozyme is enhanced by interaction of auxiliary elements. *RNA (New York, N.Y.)*, *10*(5), 880–888.

Perdew, J. P., Burke, K., & Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Phys. Rev. Lett.*, *77*(18), 3865–3868.

Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A., & Orozco, M. (2007). Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J*, *92*(11), 3817–29.

Petraccone, L., Spink, C., Trent, J. O., Garbett, N. C., Mekmaysy, C. S., Giancola, C., & Chaires, J. B. (2011). Structure and stability of higher-order human telomeric quadruplexes. *J Am Chem Soc*, *133*(51), 20951–20961.

Petrov, A. I., Zirbel, C. L., & Leontis, N. B. (2013). Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. *RNA*, *19*(10), 1327–1340.

Phan, A. T., Kuryavyi, V., Luu, K. N., & Patel, D. J. (2007). Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K+ solution. *Nucleic Acids Res*, *35*(19), 6517–6525.

Phan, A. T., Kuryavyi, V., & Patel, D. J. (2006). DNA architecture: from G to Z. *Curr Opin Struct Biol*, *16*(3), 288–298.

Phan, A. T., & Leroy, J.-L. (2000). Intramolecular i-Motif Structures of Telomeric DNA. *J Biomol Struct Dyn*, *17*(Suppl 1), 245-251 .

Phan, A. T., Luu, K. N., & Patel, D. J. (2006). Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K+ solution. *Nucleic Acids Res*, *34*(19), 5715–5719.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., … Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J Comput Chem*, *26*(16), 1781–1802.

Phys.org. (2012). Graphene may enable the ultimate water desalination device. http://www.graphene-info.com/graphene-may-enable-super-water-desalination-devices

Pilch, D. S., Plum, G. E., & Breslauer, K. J. (1995). The thermodynamics of DNA structures that contain lesions or guanine tetrads. *Curr Opin Struct Biol*, *5*(3), 334–42.

Pingali, P. K., Halder, S., Mukherjee, D., Basu, S., Banerjee, R., Choudhury, D., & Bhattacharyya, D. (2014). Analysis of stacking overlap in nucleic acid structures: algorithm and application. *J Comput Aided Mol Des*, *28*(8), 851–67.

Pley, H. W., Flaherty, K. M., & McKay, D. B. (1994). Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature*, *372*(6501), 111–113.

Podbevšek, P., Hud, N. V., & Plavec, J. (2007). NMR evaluation of ammonium ion movement within a unimolecular G-quadruplex in solution. *Nucleic Acids Res*, *35*(8), 2554–2563.

Podbevšek, P., Šket, P., & Plavec, J. (2008). Stacking and not solely topology of T3 loops controls rigidity and ammonium ion movement within d(G4T3G4) 2 G-quadruplex. *J Am Chem Soc*, *130*(43), 14287–14293.

Portmann, S., Grimm, S., Workman, C., Usman, N., & Egli, M. (1996). Crystal structures of an A-form duplex with single-adenosine bulges and a conformational basis for site-specific RNA self-cleavage. *Chem Biol*, *3*(3), 173–184.

Proctor, D. J., Ma, H., Kierzek, E., Kierzek, R., Gruebele, M., & Bevilacqua, P. C. (2004). Folding thermodynamics and kinetics of YNMG RNA hairpins: specific incorporation of 8-bromoguanosine leads to stabilization by enhancement of the folding rate. *Biochemistry*, *43*(44), 14004–14.

Prokocimer, M., Davidovich, M., Nissim-Rafinia, M., Wiesel-Motiuk, N., Bar, D. Z., Barkan, R., … Gruenbaum, Y. (2009). Nuclear lamins: Key regulators of nuclear structure and activities. *J Cell Mol Med*, *13*(6), 1059–1085.

Protozanova, E., & Macgregor, R. B. (1998). Circular dichroism of DNA frayed wires. *Biophys J*, *75*(2), 982–989.

Q

Qin, Y., & Hurley, L. H. (2008). Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*, *90*(8), 1149–1171.

Qiu, D., Shenkin, P. S., Hollinger, F. P., & Still, W. C. (1997). The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J Phys Chem A*, *101*(16), 3005–3014.

Quigley, G. J., & Rich, A. (1976). Structural domains of transfer RNA molecules. *Science*, *194*(4267), 796–806.

Quintana, J. R., Grzeskowiak, K., Yanagi, K., & Dickerson, R. E. (1992). Structure of a B-DNA decamer with a central T-A step: C-G-A-T-T-A-A-T-C-G. *J Mol Biol*, *225*(2), 379–395.

R

Raffaele Di Barletta, M., Ricci, E., Galluzzi, G., Tonali, P., Mora, M., Morandi, L., … Toniolo, D. (2000). Different mutations in the LMNA gene cause autosomal dominant and autosomal recessive Emery-Dreifuss muscular dystrophy. *Am J Hum Genet*, *66*(4), 1407–1412.

Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J Mol Biol*, *7*, 95–99.

Rassolov, V. A., Ratner, M. A., POPLE, J. A., Redfern, P. C., & Curtiss, L. A. (2001). 6-31G*basis set for third-row atoms. *J Comput Chem*, *22*(9), 976–984.

Ray, A., Bagani, K., Banerjee, S., & Bhattacharyya, D. (2015). Oxidative Tearing of Graphene Sheets: Insights into the Probable Situations by Computational and Experimental Studies. *J Phys Chem C*, *119*(2), 951–959.

Ray, S. S., Halder, S., Kaypee, S., & Bhattacharyya, D. (2012). HD-RNAS: An Automated Hierarchical Database of RNA Structures. *Front Genet*, *3*, 59.

Reddy, S. Y., Obika, S., & Bruice, T. C. (2003). Conformations and dynamics of Ets-1 ETS domain-DNA complexes. *Proc Natl Acad Sci U S A*, *100*(26), 15475–15480.

Reed, A. E., Curtiss, L. A., & Weinhold, F. (1988). Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem Rev*, *88*(6), 899–926.

Renault, G., Raynal, E., Sinet, M., Berthier, J. P., Godard, B., & Cornillault, J. (1982). A laser fluorimeter for direct cardiac metabolism investigation. *Opt Laser Technol*, *14*(3), 143–148.

Renčiuk, D., Kejnovská, I., Školáková, P., Bednářová, K., Motlová, J., & Vorlíčková, M. (2009). Arrangements of human telomere DNA quadruplex in physiologically relevant K+ solutions. *Nucleic Acids Res*, *37*(19), 6625–6634.

Reuters. (2013). Lockheed Martin developed a new graphene based water desalination technology, hopes to commercialize it by 2014-2015. http://www.graphene-info.com/lockheed-martin-developed-new-graphene-based-water-desalination-technology-hopes-commercialize-it-20

Richardson, J. S. (1977). beta-Sheet topology and the relatedness of proteins. *Nature*, *268*, 495–500.

Richardson, J. S. (1981). The Anatomy & Taxonomy of Protein Structure. *Advances In Protein Chemistry*, *34*, 167–339.

Richardson, J. S., Getzoff, E. D., & Richardson, D. C. (1978). The beta bulge: a common small unit of nonrepetitive protein structure. *Proc Natl Acad Sci U S A*, *75*(6), 2574–8.

Rowsell, S., Stonehouse, N. J., Convery, M. A., Adams, C. J., Ellington, A. D., Hirao, I., … Phillips, S. E. (1998). Crystal structures of a series of RNA aptamers complexed to the same protein target. *Nat Struct Biol*, *5*(11), 970–975.

Roy, A., Panigrahi, S., Bhattacharyya, M., & Bhattacharyya, D. (2008). Structure, stability, and dynamics of canonical and noncanonical base pairs: quantum chemical studies. *J Phys Chem B*, *112*(12), 3786–3796.

RSC. (2015). Graphene oxide "teabags" to remove mercury from water. http://www.graphene-info.com/graphene-oxide-teabags-remove-mercury-water

Ryoo, S. R., Kim, Y. K., Kim, M. H., & Min, D. H. (2010). Behaviors of NIH-3T3 fibroblasts on graphene/carbon nanotubes: Proliferation, focal adhesion, and gene transfection studies. *ACS Nano*, *4*(11), 6587–6598.

S

Saenger, W. (1984). *Principles of Nucleic Acid Structure*. New York, NY: Springer-Verlag.

Sagui, C., & Darden, T. A. (1999). Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu Rev Biophys Biomol Struct*, *28*, 155–179.

Sahoo, N. G., Bao, H., Pan, Y., Pal, M., Kakran, M., Cheng, H. K. F., … Tan, L. P. (2011). Functionalized carbon nanomaterials as nanocarriers for loading and delivery of a poorly water-soluble anticancer drug: a comparative study. *Chem Commun*, *47*(18), 5235–5237.

Samanta, D., Mukhopadhyay, D., Chowdhury, S., Ghosh, J., Pal, S., Basu, A., … DasGupta, C. (2008). Protein folding by domain V of Escherichia coli 23S

rRNA: specificity of RNA-protein interactions. *J Bacteriol*, *190*(9), 3344–3352.

Sarkar, S., Patra, S., Gayathri, N., & Banerjee, S. (2010). Effect of self-affine fractal characteristics of surfaces on wetting. *Appl Phys Lett*, *96*(6), 063112.

Sayle, R., & Milner-White, E. J. (1995). RasMol: Biomolecular graphics for all. *Trends Biochem Sci*, *20*(9), 374.

Schaftenaar, G., & Noordik, J. H. (2000). Molden: A pre- and post-processing program for molecular and electronic structures. *J Comput Aided Mol Des*, *14*(2), 123–134.

Schlick, T. (2010). *Molecular Modeling and Simulation: An Interdisciplinary Guide* (2nd Edition). New York, NY: Springer-Verlag.

Schneider, B., Neidle, S., & Berman, H. M. (1997). Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers*, *42*(1), 113–124.

Schniepp, H. C., Li, J.-L., McAllister, M. J., Sai, H., Herrera-Alonso, M., Adamson, D. H., … Aksay, I. A. (2006). Functionalized single graphene sheets derived from splitting graphite oxide. *J Phys Chem B*, *110*(17), 8535–8539.

Schultze, P., Smith, F. W., & Feigon, J. (1994). Refined solution structure of the dimeric quadruplex formed from the Oxytricha telomeric oligonucleotide d(GGGGTTTTGGGG). *Structure*, *2*(3), 221–233.

Schulz, G. E., & Schirmer, R. H. (1979). *Principles of protein structure*. New York, NY: Springer-Verlag

Schulz, G. E., & Schirmer, R. H. (2013). *Principles of protein structure*. Springer Science & Business Media.

Schuster, P., Zundel, G., & Sandorfy, C. (1976). *The Hydrogen bond: recent developments in theory and experiments* (Vol 1 & 2). North-Holland Pub. Co.

Sen, D., & Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, *334*(6180), 364–366.

Service, R. F. (2009). Materials science. Carbon sheets an atom thick give rise to

graphene dreams. *Science*, *324*(5929), 875-877.

Shao, Y., Wang, J., Wu, H., Liu, J., Aksay, I. A., & Lin, Y. (2010). Graphene based electrochemical sensors and biosensors: A review. *Electroanalysis*, *22*(10), 1027-1036.

Sharma, P., Sponer, J. E., Sponer, J., Sharma, S., Bhattacharyya, D., & Mitra, A. (2010). On the role of the cis Hoogsteen:sugar-edge family of base pairs in platforms and triplets-quantum chemical insights into RNA structural biology. *J Phys Chem B*, *114*(9), 3307–3320.

Simonsson, T. (2001). G-quadruplex DNA structures - Variations on a theme. *Biol Chem*, *382*(4), 621–628.

Simonsson, T., & Sjoback, R. (1999). DNA Tetraplex Formation Studied with Fluorescence Resonance Energy Transfer. *J Biol Chem*, *274*(24), 17379–17383.

Singh, V., Azarkh, M., Exner, T. E., Hartig, J. S., & Drescher, M. (2009). Human telomeric quadruplex conformations studied by pulsed EPR. *Angew Chem Int Ed*, *48*(51), 9728–9730.

Sket, P., & Plavec, J. (2007). Not all G-quadruplexes exhibit ion-channel-like properties: NMR study of ammonium ion (non)movement within the d(G(3)T(4)G(4))(2) quadruplex. *J Am Chem Soc*, *129*(28), 8794–800.

Smargiasso, N., Rosu, F., Hsia, W., Colson, P., Baker, E. S., Bowers, M. T., … Gabelica, V. (2008). G-quadruplex DNA assemblies: Loop length, cation identity, and multimer formation. *J Am Chem Soc*, *130*(31), 10208–10216.

Snoussi, K., & Halle, B. (2008). Internal sodium ions and water molecules in guanine quadruplexes: magnetic relaxation dispersion studies of [d(G3T4G3)]2 and [d(G4T4G4)]2. *Biochemistry*, *47*(46), 12219–12229.

Soĭfer, V. (1996). *Triple-helical nucleic acids*. New York : Springer.

Špačková, N., Berger, I., & Šponer, J. (1999). Nanosecond Molecular Dynamics Simulations of Parallel and Antiparallel Guanine Quadruplex DNA Molecules. *J Am Chem Soc*, *121*(23), 5519–5534.

Šponer, J., Cang, X., & Cheatham, T. E. (2012). Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures. *Methods*,

57(1), 25–39.

Sponer, J. E., Spacková, N., Kulhanek, P., Leszczynski, J., & Sponer, J. (2005). Non-Watson-Crick base pairing in RNA. quantum chemical analysis of the cis Watson-Crick/sugar edge base pair family. *J Phys Chem A*, *109*(10), 2292–2301.

Sponer, J. E., Spackova, N., Leszczynski, J., & Sponer, J. (2005). Principles of RNA base pairing: structures and energies of the trans Watson-Crick/sugar edge base pairs. *J Phys Chem B*, *109*(22), 11399–11410.

Sponer, J., Jurecka, P., & Hobza, P. (2004). Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *J Am Chem Soc*, *126*(32), 10142–10151.

Sprinzl, M., Horn, C., Brown, M., Loudovltch, A., & Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res*, *26*(1), 148–153.

Srinivasan, A. R., & Olson, W. K. (1987). Nucleic acid model building: the multiple backbone solutions associated with a given base morphology. *J Biomol Struct Dyn*, *4*(6), 895–938.

Srinivasan, A. R., Torres, R., Clark, W., & Olson, W. K. (1987). Base sequence effects in double helical DNA. I. Potential energy estimates of local base morphology. *J Biomol Struct Dyn*, *5*(3), 459–496.

Stahl, F. W. (1994). The holliday junction on its thirtieth anniversary. *Genetics*, *138*(2), 241-246.

Stankovich, S., Dikin, D. A., Dommett, G. H. B., Kohlhaas, K. M., Zimney, E. J., Stach, E. A., … Ruoff, R. S. (2006). Graphene-based composite materials. *Nature*, *442*(7100), 282–286.

Stankovich, S., Dikin, D. A., Piner, R. D., Kohlhaas, K. A., Kleinhammes, A., Jia, Y., … Ruoff, R. S. (2007). Synthesis of graphene-based nanosheets via chemical reduction of exfoliated graphite oxide. *Carbon*, *45*(7), 1558–1565.

Stefl, R., Spackova, N., Berger, I., Koca, J., & Sponer, J. (2001). Molecular dynamics of DNA quadruplex molecules containing inosine, 6- thioguanine and 6-thiopurine. *Biophys. J.*, *80*(1), 455–68.

Stephens, J. A., & Greene, C. H. (1994). Vibrational autoionization in H2 above the v+=3 ionization limit. *J Chem Phys*, *100*(10), 7135.

Sternberg, M. J., & Thornton, J. M. (1977). On the conformation of proteins: hydrophobic ordering of strands in beta-pleated sheets. *J Mol Biol*, *115*(1), 1–17.

Subirana, J. A., & Faria, T. (1997). Influence of sequence on the conformation of the B-DNA helix. *Biophys J*, *73*(1), 333–8.

Sun, T., & Fabris, S. (2012). Mechanisms for oxidative unzipping and cutting of graphene. *Nano Lett*, *12*(1), 17–21.

Sun, X., Liu, Z., Welsher, K., Robinson, J. T., Goodwin, A., Zaric, S., & Dai, H. (2008). Nano-Graphene Oxide for Cellular Imaging and Drug Delivery. *Nano Res.*, *1*(3), 203-212.

Sundquist, W. I., & Klug, A. (1989). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, *342*(6251), 825–829.

Suzuki, M., Amano, N., Kakinuma, J., & Tateno, M. (1997). Use of a 3D structure data base for understanding sequence- dependent conformational aspects of DNA. *J. Mol. Biol.*, *274*(3), 421–435.

Sykes, M. T., & Levitt, M. (2005). Describing RNA structure by libraries of clustered nucleotide doublets. *J Mol Biol*, *351*(1), 26–38.

T

Tamura, M., & Holbrook, S. R. (2002). Sequence and structural conservation in RNA ribose zippers. *J Mol Biol*, *320*(3), 455–474.

te Velde, G., Bickelhaupt, F. M., Baerends, E. J., Fonseca Guerra, C., van Gisbergen, S. J. a., Snijders, J. G., … Gisbergen, S. J. a V. a N. (2001). Chemistry with ADF. *J Comput Chem*, *22*(9), 931–967.

The Structures of DNA and RNA. (2002). http://biology.kenyon.edu/courses/biol63/watson_06.pdf

Thoma, F., Koller, T., & Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol*, *83*(2 Pt 1), 403–427.

Thomas, M., White, R. L., & Davis, R. W. (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc Natl Acad Sci U S A*, *73*(7), 2294–2298.

Tian, B., Bevilacqua, P. C., Diegelman-Parente, A., & Mathews, M. B. (2004). The double-stranded-RNA-binding motif: interference and much more. *Nat Rev Mol Cell Biol*, *5*(12), 1013–1023.

Tinoco, I., & Bustamante, C. (1999). How RNA folds. *J Mol Biol*, *293*(2), 271–281.

Todd, A. K., Johnston, M., & Neidle, S. (2005). Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res*, *33*(9), 2901–2907.

Toor, N., Keating, K. S., Taylor, S. D., & Pyle, A. M. (2008). Crystal structure of a self-spliced group II intron. *Science*, *320*(5872), 77–82.

Tran, P. L. T., Mergny, J.-L., & Alberti, P. (2011). Stability of telomeric G-quadruplexes. *Nucleic Acids Res*, *39*(8), 3282–3294.

Tucker, B. J., & Breaker, R. R. (2005). Riboswitches as versatile gene control elements. *Curr Opin Struct Biol*, *15*(3), 342–348.

Tung, C. S., Soumpasis, D. M., & Hummer, G. (1994). An extension of the rigorous base-unit oriented description of nucleic acid structures. *J Biomol Struct Dyn*, *11*(6), 1327–1344.

V

Vallurupalli, P., & Moore, P. B. (2003). The solution structure of the loop E region of the 5S rRNA from spinach chloroplasts. *J Mol Biol*, *325*(5), 843–856.

van Dam, L., Korolev, N., & Nordenskiöld, L. (2002). Polyamine-nucleic acid interactions and the effects on structure in oriented DNA fibers. *Nucleic Acids Res*, *30*(2), 419–428.

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., … Mackerell, A. D. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*, *31*(4), 671–690.

Varani, L., Hasegawa, M., Spillantini, M. G., Smith, M. J., Murrell, J. R., Ghetti, B.,

… Varani, G. (1999). Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17. *Proc Natl Acad of Sci U S A*, *96*(14), 8229–8234.

Vargason, J. M., Eichman, B. F., & Ho, P. S. (2000). The extended and eccentric E-DNA structure induced by cytosine methylation or bromination. *Nat Struct Biol*, *7*(9), 758–761.

Villa, A., Wöhnert, J., & Stock, G. (2009). Molecular dynamics simulation study of the binding of purine bases to the aptamer domain of the guanine sensing riboswitch. *Nucleic Acids Res*, *37*(14), 4774–4786.

W

Walberer, B. J., Cheng, A. C., & Frankel, A. D. (2003). Structural diversity and isomorphism of hydrogen-bonded base interactions in nucleic acids. *J Mol Biol*, *327*(4), 767–780.

Wang, A. H., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., & Rich, A. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, *282*(5740), 680–686.

Wang, K., Ruan, J., Song, H., Zhang, J., Wo, Y., Guo, S., & Cui, D. (2011). Biocompatibility of Graphene Oxide. *Nanoscale Res Lett*, *6*(1), 1–8.

Wang, Y., & Patel, D. J. (1993). Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure*, *1*(4), 263–282.

Warren, R. A. J. (1980). Modified Bases in Bacteriophage DNAs. *Annu Rev Microbiol*, *34*(1), 137–158.

Warshel, A. (2003). Computer simulations of enzyme catalysis: methods, progress, and insights. *Annu Rev Biophys Biomol Struct*, *32*, 425–443.

Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids. *Nature*, *171*(4356), 737–738.

Wellenzohn, B., Flader, W., Winger, R. H., Hallbrucker, A., Mayer, E., & Liedl, K. R. (2001). Complex of B-DNA with polyamides freezes DNA backbone flexibility. *J Am Chem Soc*, *123*(21), 5044–5049.

Wen, Y., Xing, F., He, S., Song, S., Wang, L., Long, Y., … Fan, C. (2010). A graphene-based fluorescent nanoprobe for silver(I) ions detection by using graphene oxide and a silver-specific oligonucleotide. *Chem Commun*, *46*(15), 2596–2598.

Wiberg, K. B. (1968). Application of the pople-santry-segal CNDO method to the cyclopropylcarbinyl and cyclobutyl cation and to bicyclobutane. *Tetrahedron*, *24*(3), 1083–1096.

Widjajakusuma, E. C., Villa, A., & Stock, G. (2012, February). Effect of the ion treatment on an rna hairpin: molecular dynamics study. *Indo J Chem*, *12*(1), 1-11.

Wilkins, M H F; Seeds, W E; Wilson, H. R. (1953). Helical structure of crystalline deoxypentose nucleic acid. *Nature*, *172*(4382), 759–762.

Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W., & Ramakrishnan, V. (1999). A Detailed View of a Ribosomal Active Site. *Cell*, *97*(4), 491–502.

Winger, R. H., Liedl, K. R., Pichler, A., Hallbrucker, A., & Mayer, E. (1999). Helix morphology changes in B-DNA induced by spontaneous B(I)<==>B(II) substrate interconversion. *J Biomol Struct Dyn*, *17*(2), 223–235.

Woese, C. R., & Gutell, R. R. (1989). Evidence for several higher order structural elements in ribosomal RNA. *Proc Natl Acad Sci U S A*, *86*(9), 3119–3122.

X

Xia, B., Tsui, V., Case, D. a., Dyson, H. J., & Wright, P. E. (2002). Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water. *J Biomol NMR*, *22*(4), 317–331.

Xin, Y., & Olson, W. K. (2009). BPS: A database of RNA base-pair structures. *Nucleic Acids Res*, *37*(Database issue), D83-88.

Xu, Y., Bai, H., Lu, G., Li, C., & Shi, G. (2008). Flexible graphene films via the filtration of water-soluble noncovalent functionalized graphene sheets. *J Am Chem Soc*, *130*(18), 5856–5857.

Xu, Z., & Xue, K. (2010). Engineering graphene by oxidation: a first-principles study. *Nanotechnology*, *21*(4), 045704.

Y

Yakovchuk, P., Protozanova, E., & Frank-Kamenetskii, M. D. (2006). Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res*, *34*(2), 564–574.

Yan, J. A., Xian, L., & Chou, M. Y. (2009). Structural and electronic properties of oxidized graphene. *Phys Rev Lett*, *103*(8), 086802.

Yang, H., Fung, S. Y., Pritzker, M., & Chen, P. (2007). Modification of hydrophilic and hydrophobic surfaced using an ionic-complementary peptide. *PLoS ONE*, *2*(12), e1325.

Ye, X., Gorin, A., Frederick, R., Hu, W., Majumdar, A., Xu, W., … Patel, D. J. (1999). RNA architecture dictates the conformations of a bound peptide. *Chem Biol*, *6*(9), 657–669.

Ye, X., Kumar, R. A., & Patel, D. J. (1995). Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem Biol*, *2*(12), 827–840.

Ying, L., Green, J. J., Li, H., Klenerman, D., & Balasubramanian, S. (2003). Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc Natl Acad Sci U S A*, *100*(25), 14629–14634.

Young, B. T., & Silverman, S. K. (2002). The GAAA tetraloop-receptor interaction contributes differentially to folding thermodynamics and kinetics for the P4-P6 RNA domain. *Biochemistry*, *41*(41), 12271–12276.

Yu, W., He, X., Vanommeslaeghe, K., & MacKerell, A. D. (2012). Extension of the CHARMM General Force Field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J Comput Chem*, *33*(31), 2451–2468.

Yurenko, Y. P., Novotný, J., Sklenář, V., & Marek, R. (2014). Exploring non-covalent interactions in guanine- and xanthine-based model DNA quadruplex

structures: a comprehensive quantum chemical approach. *Phys Chem Chem Phys*, *16*(5), 2072–2084.

Z

Zacharias, M. (2000). Simulation of the structure and dynamics of nonhelical RNA motifs. *Curr Opin Struct Biol*, *10*(3), 311–317.

Zhang, L., Zhang, F., Yang, X., Long, G., Wu, Y., Zhang, T., … Chen, Y. (2013). Porous 3D graphene-based bulk materials with exceptional high surface area and excellent conductivity for supercapacitors. *Sci Rep*, *3*, 1408.

Zhang, M., Yin, B.-C., Tan, W., & Ye, B.-C. (2011). A versatile graphene-based fluorescence "on/off" switch for multiplex detection of various targets. *Biosens Bioelectron*, *26*(7), 3260–3265.

Zhang, Y., Ali, S. F., Dervishi, E., Xu, Y., Li, Z., Casciano, D., & Biris, A. S. (2010). Cytotoxicity effects of graphene and single-wall carbon nanotubes in neural phaeochromocytoma-derived pc12 cells. *ACS Nano*, *4*(6), 3181–3186.

Zhao, X., & Yu, Y.-T. (2004). Detection and quantitation of RNA base modifications. *RNA (New York, N.Y.)*, *10*(6), 996–1002.

Zhou, Y. B., Gerchman, S. E., Ramakrishnan, V., Travers, A., & Muyldermans, S. (1998). Position and orientation of the globular domain of linker histone H5 on the nucleosome. *Nature*, *395*(6700), 402–405.

Zhu, H., Xiao, S., & Liang, H. (2013). Structural dynamics of human telomeric G-quadruplex loops studied by molecular dynamics simulations. *PloS One*, *8*(8), e71380.

Zhu, Y., Murali, S., Cai, W., Li, X., Suk, J. W., Potts, J. R., & Ruoff, R. S. (2010). Graphene and graphene oxide: Synthesis, properties, and applications. *Adv Mater*, *22*(35), 3906–3924.

Appendices

APPENDIX I

Molecular Dynamics (MD) Simulation:

In MD simulations, the potential energy between atoms is used to calculate the forces between them. For these calculations the atoms should have a defined initial position in space. This starting structure can be obtained from X-ray or NMR experiments or from molecular modeling. The atoms also need to have starting velocities, which are generated from a random number series. The Newton's equation of motion is then solved to calculate the atomic movements in time to generate successive configurations of the system. A trajectory is finally obtained that species how the positions and velocities of the particles in the system vary with time. The force F_i exerted on the atom i by the remainder of the system is calculated from the negative gradient of the potential energy function with respect to the position indicated by a 3D vector r_i, where,

$$F_i = -\frac{dV}{dr_i}$$

Newton's equations are then used to calculate the acceleration a_i (d^2x_i/dt^2) of atom i, as,

$$a_i = \frac{d^2x_i}{dt^2} = \frac{F_{x_i}}{m_i}$$

where m_i and xi are the respective mass and position of the atom along a single dimension. The position after a short time interval (Δt) are then calculated using standard Taylor series. There are several algorithms for integrating the equations of motion **[1]** where the integration is broken down into small steps, separated in time by usually few fs (Δt).

The Verlet algorithm is widely used **[2]** where the positions and acceleration at time t, and the positions from the previous step, $\vec{r}(t-\delta t)$, are used to calculate the new position $\vec{r}(t+\delta t)$, at time $t+\delta t$.

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

$$\vec{r}(t-\delta t) = \vec{r}(t) - \delta t\vec{V}(t) + \frac{1}{2}\delta t^2\vec{a}(t) + \cdots\cdots$$

Adding these two equations gives,

$$\vec{r}(t+\delta t) = 2\vec{r}(t) - \vec{r}(t-\delta t) + \delta t^2\vec{a}(t)$$

In Verlet algorithm, velocities do not appear explicitly. The equations for velocities can be derived simply by dividing the differences in positions at time $(t+\delta t)$ and $(t-\delta t)$, as,

$$\vec{V}(t)=\frac{[\vec{r}(t+\delta t)-\vec{r}(t-\delta t)]}{2\delta t}$$

Replacing t to $(t+\frac{1}{2}\delta t)$ gives,

$$\vec{V}(t+\frac{1}{2}\delta t)=\frac{[\vec{r}(t+\delta t)-\vec{r}(t)]}{\delta t}$$

However, positions $\vec{r}(t+\delta t)$ are obtained by adding a small term $\delta t^2\vec{a}(t)$ to the difference of two large terms $2\vec{r}(t)$ and $\vec{r}(t-\delta t)$. This leads to a loss of precision. Furthermore, there is no explicit velocity term in the equation. The Leap-Frog method comes with two advantages over the Verlet algorithm. It explicitly includes the velocity in the equation of motion.

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

$$\vec{V}(t+\frac{1}{2}\delta t)=\vec{V}(t-\frac{1}{2}\delta t)+\delta t\vec{a}(t)$$

Another improvement comes with velocity-Verlet algorithm, where,

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

$$\vec{V}(t+\delta t)=\vec{V}(t)+\frac{1}{2}\delta t[\vec{a}(t)+\vec{a}(t+\delta t)]$$

In this method new velocities requires the accelerations at both t and $(t+\delta t)$. So, in the first step the positions at $(t+\delta t)$ are calculated by using velocities and accelerations at time t. The velocities at time $(t+\frac{1}{2}\delta t)$ are then determined using:

$$\vec{V}(t+\frac{1}{2}\delta t)=\vec{V}(t)+\frac{1}{2}\delta t\vec{a}(t)$$

In the final step velocities at time $(t+\delta t)$ are determined, using

$$\vec{V}(t+\delta t)=\vec{V}(t+\frac{1}{2}\delta t)+\frac{1}{2}\delta t\vec{a}(t+\delta t)$$

We have used Velocity-Verlet algorithm for MD simulations in chapters II, III and IV of this thesis.

Steered Molecular Dynamics (SMD) Simulations

Steered Molecular Dynamics (SMD) employs a pulling force to cause a change in structure in a MD simulation. The simulation runs and all atoms adjust to the forced change in structure so that the conformations may be sampled along a particular pathway. In theory the change would happen spontaneously during the simulation, but might require a very long time t get noticed. Thus, the process is accelerated to understand the change in energy of the system.

As the force is executed and motion occurs along a coordinate, the potential energy of the system is calculated. This potential of mean force (PMF) [3] is related to the free energy change for the process.

There are two protocols of SMD:

(i) Constant velocity pulling or SMD-CV method: In this type of simulation the SMD atom is attached to a dummy atom via a virtual spring. This dummy atom is moved at constant velocity and then the force between both is measured using-

$$\underset{F}{\rightarrow} = -\nabla U \qquad \ldots.(1)$$

$$U = 1/2k\left[vt - \left(\underset{r}{\rightarrow} - \underset{r_0}{\rightarrow}\right).\underset{n}{\rightarrow}\right]^2 \qquad \ldots.(2)$$

where,

U = Potential energy.

k = Spring constant

v = Pulling velocity

t = Time

$\underset{r}{\rightarrow}$ = Actual position of the SMD atom

$\underset{r_0}{\rightarrow}$ = Initial position of the SMD atom

$\underset{n}{\rightarrow}$ = Direction of pulling

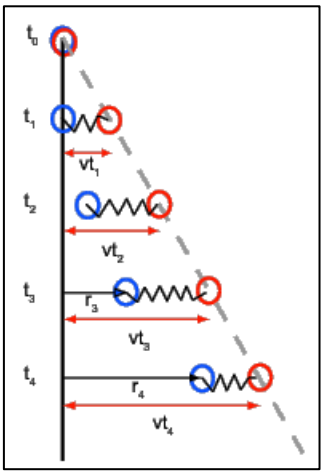


Figure 1: Pulling in a 1D case. As the dummy atom (red) moves at constant velocity the SMD atom (blue) experiences a force that depends linearly on the distance between both atoms.

The SMD-CV protocol has been used in Chapter II and Chapter IV of this thesis to performed pulling simulations (SMD simulations) on G-quadruplex and proteins respectively.

(ii) Constant force pulling or SMD-CF method: In this case SMD simulation is performed by applying a constant force. Here, one of the atoms is kept fixed and the SMD atom (or the pulled one) experiences a constant force in the direction defined by the vector that links both these atoms. In it to be noted that in this case there is no dummy atom or virtual spring.

Force field:

As discussed MD simulations require potential energy of the system of particles involved, described by a force field. The two main components of a force field are: (i) the potential energy function and (ii) the parameters used in the function. The included parameters are mainly derived from experimental work or from high level quantum mechanical calculations on small model compounds. The potential energy is calculated by adding the bonded energy terms (bonds, angles and torsions) to the non-bonded terms (van der Waals, electrostatic),as,

$$V_{total} = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi [1 + \cos(n\chi - \sigma)]$$

$$+ \sum_{\substack{nonbonded \\ pairs,ij}} \left(\varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} \right)$$

where, K_b, K_θ and K_χ are the force constants for bond length, bond angle and dihedral respectively. The values in the current configuration are denoted as b, θ and χ, where subscript zero is used for the equilibrium state. The rotation of a dihedral is periodic with periodicity n and phase σ. The non-bonded energy is the sum of repulsion, attraction and electrostatics between non-bonded atoms. ε_{ij} is related to the well-depth of the Lennard-Jones potential, $R_{min,ij}$ is the distance at which the Lennard-Jones potential has its minimum, q is the partial atomic charge and r_{ij} is the distance between atoms i and j. The Lennard-Jones potential and the coulombic term express the short range non-bonded interactions. The long-range electrostatic interactions are generally evaluated by ignoring these interactions beyond a specific cut-off leading to

approximations in the calculations. With the introduction of the Ewald summation and the particle mesh Ewald (PME) method the long-range electrostatic calculations became more accurate **[4,5]**.

Currently there exist a number of empirical force fields used in simulations of biomolecular systems. Various force fields have been used in this thesis: charmm-27 **[6]** has been used in chapter II, charmm36 **[7]** and AMBER-99sb **[8,9]** have been used in chapter III and Charmm-22 has been used in chapter IV. The functional forms to describe specific interactions are similar between the various force fields but there are differences in the values of the specific parameters and the way they are obtained **[10]**.

Ewald summation method:

The interactions that decay slower than r^{-n}, where n is the dimensionality of the system are known as long-range interactions. These long-range interactions can be a problem as their range is often greater than half of the box length. During the simulations of charged species it is extremely important to properly model these long-range forces. The charge-charge interaction, which decays as (1/r), is extremely difficult to handle properly during the simulations of charged species. Different methods, *e.g*. Ewald summations, the reaction field method, cell multiple method etc. have been developed to handle long-range interaction. The Ewald summation technique of handling long-range electrostatic interactions has been used in chapters II, III and IV of this thesis.

This method was developed by Ewald to study the energetics of ionic crystals. In this method, a particle interacts with all other particles in the simulation box and with all possible images. The position of the image boxes can be related to the central box (assumed a cube of side L) by specifying a vector (±nxL, ±nyL, ±nzL); a, b, c = 0, 1, 2 etc. The potential energy due to the charge-charge interaction in the central box containing 'N' charges can be written as:

$$V = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}},$$

where r_{ij} is the minimum distance between the charges i and j. In general, for a box at a cubic lattice point,

$\vec{n}$ ($= (n_x L, n_y L, n_z L)$ with n_x, n_y, n_z being integers):

$$V = \frac{1}{2}\sum\nolimits'_{|N|=0}\sum_{i=1}\sum_{j=1}\frac{q_i q_j}{4\pi\varepsilon_0 \mid \vec{r}_{ij} + \vec{n} \mid}$$

The 'prime' on the first summation indicates that the charge-charge interaction, i=j, for the primary simulation box is excluded. Now the series is conditionally convergent. The sum of a conditionally convergent series depends on the order in which its terms considered. The sum is converted into two summation series, each of which converges much more rapidly. In this case we can write,

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r}$$

where, f(r) is an appropriate function which deal with the rapid variation of (1/r) at small r and slow decay at long r.

In Ewald sum method each charge is considered to be surrounded by a neutralizing charge distribution of equal magnitude but of opposite sign. The commonly used functional form is:

$$\rho_i(\vec{r}) = \frac{q_i \alpha^3}{\pi^{3/2}} \exp(-\alpha^2 r^2)$$

.

The sum over point charges is now converted to a sum of the interactions between the charges plus the neutralizing distributions. The real space summation is given by:

$$V_{real} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum\nolimits'_{|n|=0}\frac{q_i q_j}{4\pi\varepsilon_0}\frac{erfc(\alpha \mid \vec{r}_{ij} + \vec{n} \mid)}{\mid \vec{r}_{ij} + n \mid}$$

where, erfc is the complementary error function, which can be written as:

$$erfc(x) = \frac{2}{\sqrt{\pi}}\int_{x}^{\infty}\exp(-t^2)dt$$

The new summation involving the error function converges very rapidly and beyond some cut-off distance its value can be considered negligible. The rate of convergence depends upon the width of the Gaussian charge distribution. The contribution comes from the neutralizing charge distribution is:

$$V_{reciprocal} = \frac{1}{2}\sum_{k\neq 0}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\pi L^3}\frac{q_i q_j}{4\pi\varepsilon_0}\frac{4\pi^2}{k^2}\exp(-\frac{k^2}{4\alpha^2})\cos(\vec{k}\cdot\vec{r}_{ij})$$

where, the vectors $\vec{k}$ are reciprocal vectors and given by

$$\vec{k} = \frac{2\pi\vec{n}}{L}$$

This summation is performed in reciprocal space. This reciprocal sum also converges much more rapidly than the original point-charge sum. For the reciprocal space summation the number of terms increases with the width of the Gaussian charge

distribution. So, the value of 'α' need to be chosen in such a way that it can balance real-space and reciprocal space summation. A value for α of (5/L)and 100-200 reciprocal vectors $\vec{k}$ provide an optimal solution. The sum of Gaussian functions in real space also includes the interaction of each Gaussian with itself. A third self-term therefore is subtracted:

$$V_{self} = -\frac{\alpha}{\sqrt{\pi}} \sum_{k=1}^{N} \frac{q_k^2}{4\pi\varepsilon_0}$$

A fourth term depending upon the surrounding medium is also included:

$$V_{Corr} = \frac{2\pi}{3L^3} | \sum_{k=1}^{N} \frac{q_i}{4\pi\varepsilon_0} \vec{r}_k |^2$$

.

The final expression of electrostatic interaction is thus:

$$V_{total} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\begin{array}{l} \sum_{|n|=0}^{'} \frac{q_i q_j}{4\pi\varepsilon_0} \frac{erfc(\alpha | \vec{r}_{ij} + \vec{n} |)}{| \vec{r}_{ij} + \vec{n} |} \\ + \frac{1}{2} \sum_{k \neq 0} \frac{1}{\pi L^3} \frac{q_i q_j}{4\pi\varepsilon_0} \frac{4\pi^2}{k^2} \exp(-\frac{k^2}{4\alpha^2}) \cos(k - r_{ij}) \\ - \alpha / \sqrt{\pi} \sum_{k=1}^{N} \frac{q_k^2}{4\pi\varepsilon_0} + 2\pi / 3L^3 \left| \sum_{k=1}^{N} \frac{q_k}{4\pi\varepsilon_0} \vec{r}_k \right|^2 \end{array} \right)$$

Implicit solvent:

Implicit salvation is a method of representing solvent as a continuous medium instead of individual "explicit" solvent molecules. There are two basic types of implicit solvent methods: models based on accessible surface area (ASA) and more recent continuum electrostatics models. The electrostatic forces of a biological system can be expressed as a system of differential equations which can be solved for the electric field caused by a collection of charges. The Poisson Boltzmann equation:

$$\vec{\nabla} \cdot \left[\epsilon(\vec{r}) \vec{\nabla} \Psi(\vec{r}) \right] = -4\pi\rho^f(\vec{r}) - 4\pi \sum_i c_i^{\infty} q_i \lambda(\vec{r}) \cdot \exp\left[\frac{-q_i \Psi(\vec{r})}{k_B T} \right]$$

is a nonlinear equation which solves for the electrostatic field, $\psi(\underset{r}{\rightarrow})$ based on the position dependent dielectric, $\varepsilon(\underset{r}{\rightarrow})$, the position-dependent accessibility of

position $\vec{r}$ to the ions in solution, $\lambda(\underset{r}{\rightarrow})$, the solute charge distribution, $\rho^f(\underset{r}{\rightarrow})$ and the bulk charge density, c_i^{∞} of ion qi . While this equation does exactly solve for the electrostatic field of a charge distribution in a dielectric, it is very expensive to solve, and therefore not suitable for molecular dynamics.

In a GB simulation, the total electrostatic force on an atom, i , is the net Coulomb force on atom i (from nearby atoms) minus the GB force on atom i(also caused by nearby atoms):

$$\vec{F}_i = \vec{F}_i^{Coulomb} - \vec{F}_i^{GB}.$$

Forces are contributed by other nearby atoms within a cutoff. The GB force on atom i is the derivative of the total GB energy with respect to relative atom distances r_{ij},

$$\begin{aligned}
\vec{F}_i^{GB} &= -\sum_j \left[\frac{dE_T^{GB}}{dr_{ij}}\right]\hat{r}_{ji} \\
&= -\sum_j \left[\sum_k \frac{\partial E_T^{GB}}{\partial \alpha_k}\frac{d\alpha_k}{dr_{ij}} + \frac{\partial E_{ij}^{GB}}{\partial r_{ij}}\right]\hat{r}_{ji} \\
&= -\sum_j \left[\frac{\partial E_T^{GB}}{\partial \alpha_i}\frac{d\alpha_i}{dr_{ij}} + \frac{\partial E_T^{GB}}{\partial \alpha_j}\frac{d\alpha_j}{dr_{ij}} + \frac{\partial E_{ij}^{GB}}{\partial r_{ij}}\right]\hat{r}_{ji} .
\end{aligned}$$

where the partial derivatives are included since the Born radius, α , is a function of all relative atom distances. The total GB energy of the system is

$$E_T^{GB} = \sum_i \sum_{j>i} E_{ij}^{GB} + \sum_i E_{ii}^{GB} ,$$

where, E_{ii}^{GB} is the Born radius dependent self energy of atom i , and the GB energy between atoms i and j is given by

$$E_{ij}^{GB} = -k_e D_{ij}\frac{q_i q_j}{f_{ij}} .$$

The dielectric term **[11]** is

$$D_{ij} = \left(\frac{1}{\epsilon_p} - \frac{\exp\left(-\kappa f_{ij}\right)}{\epsilon_s}\right) ,$$

and the GB function **[12]** is

$$f_{ij} = \sqrt{r_{ij}^2 + \alpha_i\alpha_j \exp\left(\frac{-r_{ij}^2}{4\alpha_i\alpha_j}\right)} .$$

As the Born radii of atoms i and j decrease (increasing screening), the effective distance between the atoms (f_{ij}) increases. The implicit solvent implemented in NAMD is the model of Onufriev, Bashford and Case **[13,14]** which calculates the Born radius as

$$\alpha_k = \left[\frac{1}{\rho_{k0}} - \frac{1}{\rho_k} \tanh\left(\delta\psi_k - \beta\psi_k^2 + \gamma\psi_k^3\right) \right]^{-1}$$

where,

$$\psi_k = \rho_{k0} \sum_l H_{kl} \ .$$

and H_{ij} is the piecewise descreening function **[14,15,16]**.

GBSA is a Generalized Born model augmented with the hydrophobic solvent accessible surface area (SA) term. It is among the most commonly used implicit solvent model combinations. The use of this model in the context of molecular mechanics is known as MM/GBSA.

The Generalized Born Implicit Solvent (GBIS) and MM/GBSA model has been used for steered molecular dynamics simulations of chapter IV of this thesis.

APPENDIX II

Hartee-Fock Approximation:

Under Born-Oppenheimer approximation, the focus is solution of the electronic Schrödinger equation. In such cases the main complication in all electronic structure calculations is the electron-electron potential energy, which depends on the electron-electron separations. The electron-electron interaction couples the electronic degrees of freedom, which makes the problem an impossible task to solve. Now, if we neglect the electron-electron interaction, the many body problems would decouple into one-body problem. Under such approximation, one can assume that an electron moving in a given potential. The electronic Hamiltonian can be written as:

$$H = \sum_k \frac{p_k^2}{2m_k} + \sum_k V(\vec{r}_k) + \sum_{k,k'} \frac{e^2}{|\vec{r}_k - \vec{r}_{k'}|} = \sum_k H_k + \sum_{k,k'} H_{kk'}$$

In the above equation the first two terms are single particle (electron) operator. If we neglect the third term, electron-electron interaction, the wave function Φ of the corresponding Schrödinger equation $\sum_k H_k \Phi = E\Phi$ can be written as:

$$\Phi(\vec{r}_1 ..... \vec{r}_k ..... \vec{r}_N) = \varphi_1(\vec{r}_1) ..... \varphi_k(\vec{r}_k) ..... \varphi_N(\vec{r}_N),$$

with $E = \sum_k E_k$, the Schrödinger equation reduces to sum of one-electron equation, $\sum_k H_k \varphi_k(\vec{r}_k) = \sum_k E_k \varphi_k(\vec{r}_k)$. The wave function Φ depends on all the electron coordinates and parametrically on the nuclear locations. At this stage we have taken strong electron-electron interaction, *i.e.* two body term in the Hamiltonian ($H_{kk'}$), neither the spin of the electron nor the requirement that the electronic wave function must obey the Pauli Exclusion Principle.

In the next step, we insert the term $H_{kk'}$, in the Hamiltonian. So, now the Schrödinger equation is $H\Phi = E\Phi$, with $H = \sum_k H_k + \sum_{k,k'} H_{kk'}$ The expectation value of energy, $E = <\Phi | H | \Phi>$, which has $<\varphi_k | H_k | \varphi_k>$ and $<\varphi_{kk'} | H_{kk'} | \varphi_{kk'}>$ matrix elements. So,

$$E = <\Phi | H | \Phi> = \sum_k <\varphi_k | H_k | \varphi_k> + \frac{e^2}{2} \sum_{k,k'} <\varphi_k \varphi_{k'} | \frac{1}{|\vec{r}_k - \vec{r}_{k'}|} | \varphi_k \varphi_{k'}>,$$

φ_k is assumed to be normalized. This is the expectation value of E for an arbitrary wave function φ_k. Now to find out the best set of functions for the ground state, φ_k which minimizes the energy E, one could apply variational principle, *i.e.*

$$\delta[E - \sum_k (< \varphi_k \mid \varphi_{k'} > -1)] = 0$$

Then,

$$< \delta\varphi_j \mid H_j \mid \varphi_j > + \sum_{k(>j)} < \delta\varphi_j \varphi_k \mid \frac{e^2}{\mid \vec{r}_k - \vec{r}_j \mid} \mid \varphi_j \varphi_k > - E_j < \delta\varphi_j \mid \varphi_j >= 0$$

As the above equation is valid for any variation $< \delta\varphi_j \mid$, therefore φ_j satisfy the following equation:

$$[-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r}) + \sum_{k(>j)} \int \frac{\mid \varphi_k(\vec{r}') \mid^2 e^2}{\mid \vec{r} - \vec{r}' \mid} \delta r'] \varphi_j(\vec{r}) = E_j \varphi_j(\vec{r})$$

The above equation is for single electron Schrödinger equation, which is known as Hartee equation. This describes an electron in an ionic potential $V(\vec{r})$ and interacting with all other electrons via an average distribution of electron density. However, separating one electron out of all and treating all other electrons as a smooth charge density is a crude approximation to make.

We now reintroduce the concept of spin-orbital. A spin-orbital is a product of an orbital wave function and a spin-function. Therefore we extend the expansion of $\Phi(\vec{r}_1 ..... \vec{r}_k ..... \vec{r}_N)$ by applying the Pauli Exclusion Principle and the indistinguishibility of quantum particles. The overall wavefunction is then written as the Slater determinant.

$$\Phi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(q_1) \cdots\cdots \varphi_k(q_1) \cdots\cdots \varphi_n(q_1) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \varphi_1(q_k) \cdots\cdots \varphi_k(q_k) \cdots\cdots \varphi_n(q_k) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \varphi_1(q_n) \cdots\cdots \varphi_k(q_n) \cdots\cdots \varphi_n(q_n) \end{vmatrix}$$

where, q is composite index for electronic coordinates and spin, $N!$ are the possible number of ways to distribute N number of valance electrons at the positions $\vec{r}_1, \cdots \vec{r}_k, \cdots \vec{r}_N$. Considering the indistinguishibility of electrons all those possibilities are equally likely.

Using this new form of wavefunction Φ, again we calculate the expectation value of energy E as,

$E = < \Phi \mid H \mid \Phi >$

$$= \sum_k \int \varphi_k^*(q_1) H_k \varphi_k(q_1) d\tau + \frac{e^2}{2} \sum_{k,k'} \int \frac{|\varphi_k(q_1)|^2 |\varphi_{k'}(q_2)|^2}{|\vec{r}_1 - \vec{r}_2|} d\tau_1 d\tau_2$$

$$- \frac{e^2}{2} \sum_{k,k'} \int \frac{\varphi_k^*(q_1)\varphi_k(q_2)\varphi_{k'}^*(q_2)\varphi_{k'}(q_1)}{|\vec{r}_1 - \vec{r}_2|} d\tau_1 d\tau_2$$

The extra term in the above equation compared to Hartee equation has no classical analogue and known as exchange interaction. The Hartee-Fock for spin-orbital $\varphi_k(q_1)$, where we have arbitrary assigned an electron having composite index q_1 to spin-orbital φ_k , is

$$[\frac{\hbar^2}{2m}\nabla_1^2 + V(\vec{r}_1) + e^2 \sum_{k'} \int \frac{|\varphi_{k'}(q_2)|^2}{|\vec{r}_1 - \vec{r}_2|} d\tau_2] \varphi_k(q_1)$$

$$- e^2 \sum_{k'} \int \frac{\varphi_{k'}^*(q_2)\varphi_k(q_2)}{|\vec{r}_1 - \vec{r}_2|} d\tau_2 \varphi_{k'}(q_1) = \sum_{k'} \lambda_{kk'} \varphi_{k'}(q_1)$$

This equation is known as Hartee-Fock equation.

Many Body Perturbation Theory:

Moller and Plesset proposed perturbation theory to tackle the problem of electron correlation. This method is based on adding successive improvements to both energy and wave functions to the HF description. In this case the true Hamiltonian operator is expressed as sum of the zeroth order Hamiltonian Ho and a perturbation vwhich is expressed as

$$H = Ho + v$$

To obtain an improvement of Hartree-Fock energy, it is necessary to obtain Moller-Plesset perturbation theory **[17]** to at least second order, commonly known as MP2 level of theory. Higher orders of calculations such as MP3, MP4 are also possible, but those are computationally expensive, inappropriate for complex species. They are often restricted to calculate the single point energy calculation on geometries obtained through some lower level of theory.

Density Functional Theory (DFT):

The basic idea behind Density Functional Theory (DFT) is that the energy of an electronic system can be written in terms of the electron probability density (ρ). For a system of n electrons, $\rho(\vec{r})$ denotes the total electron density at a particular point in space r. The electronic energy E is said to be a functional of the electron

density, denoted $E[\rho]$, in the sense that for a given function $\rho(\vec{r})$, there is a single corresponding energy.

The Density Functional Theory proposed by Hohenberg and Kohn **[18]** (1964), and Kohn and Sham (1965) **[19]** reduces this many body problem to that for a single electron. This theorem shows that the ground state charge density ρ(r) determines the potential unequally. So, the total ground state energy $E[\rho(\vec{r})]$ for a given external potential is a unique functional of the ground state density. For a fixed external potential $V_{ext}(\vec{r})$, the energy functional $E[\rho(\vec{r})]$ is minimum for the true ground state density $\rho_0(\vec{r})$:

$$\left.\frac{\delta E[\rho(\vec{r})]}{\delta\rho(\vec{r})}\right|_{\rho(\vec{r})=\rho_0(\vec{r})} = 0$$

The total energy functional in the presence of external potential field $V_{ext}(\vec{r})$ can be written as,

$$E[\rho(\vec{r})] = T[\rho(\vec{r})] + \int V_{ext}(\vec{r})\rho(\vec{r})d^3r + \int\int \frac{\rho(\vec{r})\rho(\vec{r}')}{|\vec{r}-\vec{r}'|} d^3\vec{r}d^3\vec{r}' + \hat{E}_{xc}[\rho(\vec{r})]$$

where, $T[\rho(\vec{r})]$ is the kinetic energy. Kohn and Sham assumed that there exists a non-interacting reference system corresponding to the above mentioned interacting system, for those the ground state density is exactly same. In such system,

$$H_{eff} \mid \varphi_i >= [-\sum_i \nabla_i^2 + V_{eff}[\rho(\vec{r})]] \mid \varphi_i >= \varepsilon_i \mid \varphi_i >$$

Kinetic energy for this system

$$T_{eff}[\rho(\vec{r})] = \sum_i n_i \varepsilon_i - V_{eff}[\rho(\vec{r})] \mid \varphi_i >$$

is the single particle wave function and n_i is the number of occupied quantum state labeled by i. The exchange correlation functional defined by Kohn and Sham is

$$E_{xc}[\rho(\vec{r})] = T[\rho(\vec{r})] - T_{eff}[\rho(\vec{r})] + \hat{E}_{xc}[\rho(\vec{r})].$$

The total energy functional becomes

$$E[\rho(\vec{r})] = T_{eff}[\rho(\vec{r})] + \int V_{ext}(\vec{r})\rho(\vec{r})d^3\vec{r} + \int\int \frac{\rho(\vec{r})\rho(\vec{r}')}{|\vec{r}-\vec{r}'|} d^3\vec{r}d^3\vec{r}' + E_{xc}[\rho(\vec{r})]$$

Taking the variation of $E[\rho(\vec{r})]$ with respect to $\rho(\vec{r})$ is

$$\frac{\delta E[\rho(\vec{r})]}{\delta\rho(\vec{r})} = V_{ext}(\vec{r}) + \int \frac{2\rho(\vec{r}')}{|\vec{r}-\vec{r}'|} d^3\vec{r}' + \frac{\delta T_{eff}[\rho(\vec{r})]}{\delta\rho(\vec{r})} + \frac{\delta E_{xc}[\rho(\vec{r})]}{\delta\rho(\vec{r})} = \mu,$$

where, μ is the Lagrange multiplier associated with the constraint $\int \rho(\vec{r})d\vec{r} = N$, total number of electrons.

$$\therefore \frac{\delta E[\rho(\vec{r})]}{\delta\rho(\vec{r})} = V_{eff}[\rho(\vec{r})] + \frac{\delta T_{eff}[\rho(\vec{r})]}{\delta\rho(\vec{r})} = \mu;\ V_{eff} \text{ is the effective potential;}$$

$$V_{eff}[\rho(\vec{r})] = V_{ext}(\vec{r}) + \int \frac{2\rho(\vec{r}')}{|\vec{r}-\vec{r}'|} d^3\vec{r}' + \frac{\delta E_{xc}[\rho(\vec{r})]}{\delta\rho(\vec{r})}.$$

The single electron density can be obtained by solving Schrödinger like equation corresponding to moving an electron in an effective potential V_{eff}, namely the Kohn-Sham equation,

$$[-\nabla^2 + V_{eff}[\rho(\vec{r})]]|\varphi_i> = \varepsilon_i|\varphi_i>.$$

The ground state energy is given by

$$E_0 = \sum_i n_i\varepsilon_i - \iint \frac{2\rho(\vec{r})}{|\vec{r}-\vec{r}'|} d^3\vec{r}' - \int \frac{\delta E_{xc}[\rho]}{\delta\rho(\vec{r})}\rho(\vec{r}) d^3\vec{r} + E_{xc}[\rho(\vec{r})],$$

where, $\frac{\delta E_{xc}[\rho]}{\delta\rho(\vec{r})}$ is the exchange-correlation potential.

Exchange-Correlation Energy:

In Kohn-Sham equation kinetic energy functional is incorporated directly, but the exchange-correlation functional $E_{xc}[\rho]$ is still unknown. The exchange-correlation energy contains (i) kinetic correlation energy, which is the difference in the kinetic energy functional between the real and non-interacting reference system, (ii) the exchange energy (iii) coulombic correlation energy arising from the inter-electronic repulsion and (iv) a self-interaction correction.

The Local Density Approximation (LDA):

Local density approximation (LDA) is the most successful approximation, first introduced by Slater for simplifying the non-local exchange energy in terms of the local $\rho^{1/3}(\vec{r})$ potential. The LDA essentially amounts to the assumption that the XC-energy depends only on the local electron density $\rho(\vec{r})$ around each volume element $d^3\vec{r}$, *i.e.*

$$E_{xc}^{LDA}[\rho(\vec{r})] \cong \int \rho(\vec{r})\varepsilon_{xc}[\rho(\vec{r})] d^3\vec{r}$$

The functional derivative of E_{xc}^{LDA} gives the exchange correlation potential within LDA,

$$v_{xc}^{LDA} = \frac{\delta E_{xc}^{LDA}[\rho(\vec{r})]}{\delta\rho(\vec{r})} = \varepsilon_{xc}[\rho(\vec{r})] + \rho(\vec{r})\frac{\delta\varepsilon_{xc}[\rho(\vec{r})]}{\delta\rho(\vec{r})}.$$

The Generalized Gradient Approximation (GGA):

In generalized gradient approximation (GGA) the functional depends on the density and its gradient,

$$E_{xc}^{GGA}[\rho(\vec{r})] = \int \rho(\vec{r})\varepsilon_{xc}^{GGA}[\rho(\vec{r}),|\nabla\rho(\vec{r})|]dr$$

The exchange-correlation potential in Cartesian coordinates is given by

$$v_{xc}^{GGA} = [\frac{\delta}{\delta\rho(\vec{r})} - \sum_{i=1}^{3}\nabla_i \frac{\delta}{\delta\nabla_i\rho(\vec{r})}]\rho(\vec{r})E_{xc}^{GGA}[\rho(\vec{r},|\nabla\rho(\vec{r})|)],$$

which depends on the first and second derivatives of the electron density.

Dispersion Corrected Density Functional Theory:

Density functional includes electron correlation in an approximate manner. It is unable to describe the long range electron correlation, which plays significant roles in describing the stacked complexes **[20,21]**. The long-range van der Waals interactions were considered by introducing the dispersion interaction either by proposing a suitable non-local functional or by combing the standard functional with empirical dispersion terms **[22,23]**. The empirical potential has been added in the form of C^6R^{-6}. After adding the dispersion correction the total dispersion corrected energy takes the form of

$$E_{DFT\text{-}D} = E_{KS\text{-}DFT} + E_{dis}$$

where, the 1st term in the right hand side represents the self-consistent Kohn-Sham energy obtained from the chosen DFT functional and the 2nd term represents the empirical dispersion correction, which has the form as

$$E_{dis} = -S_6 \sum_{i=1}^{N_{at}-1}\sum_{j=i+1}^{N_{at}} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij})$$

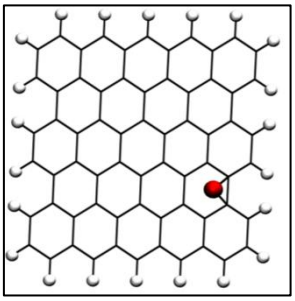
where, N_{at} represents the number of atoms in the system, R is the inter-atomic distances, and C_6 are the dispersion coefficients, the damping function is represented as $f_{dmp,}$ which has the following expression

$$f_{dmp}(R_{ij}) = \frac{1}{1+e^{-d(R_{ij}/R_r - 1)}}$$

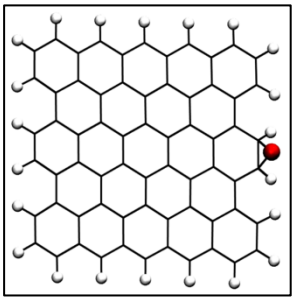
where, R_r is the sum of the atomic van der Waal radii.

The most widely used hybrid functional which include long range dispersion interaction is ωB97XD **[24]** functional, which has been incorporated in a commercial software package Gaussian g09, which has been used by me in my thesis work **[25]**.

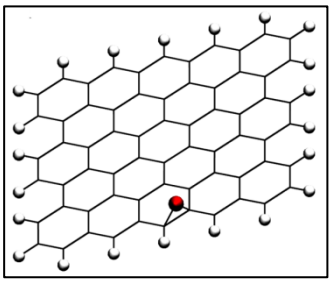
APPENDIX III



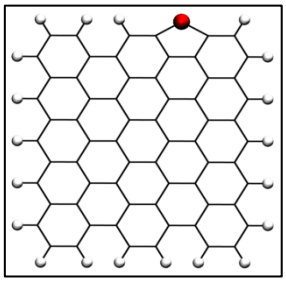
model -1



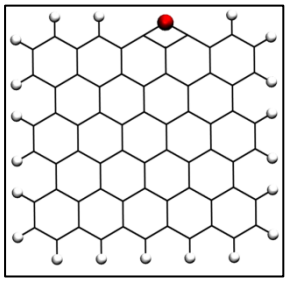
model-2



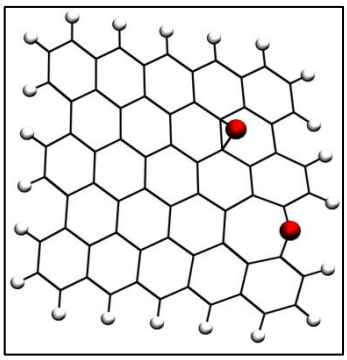
model-3



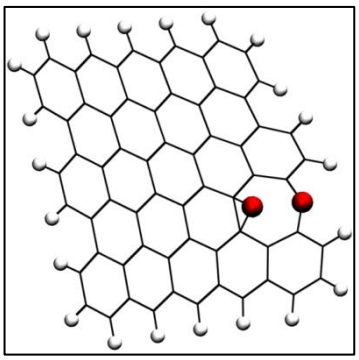
model-4



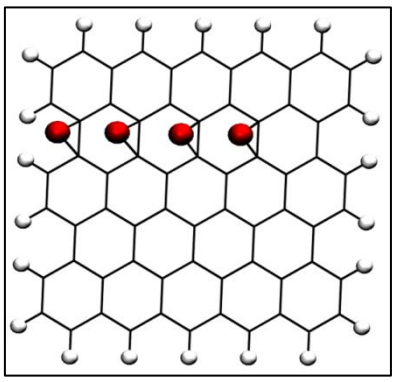
model-5



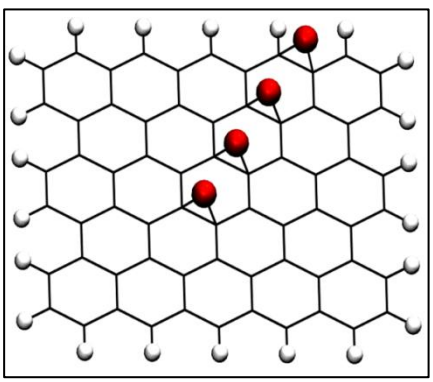
model-6a



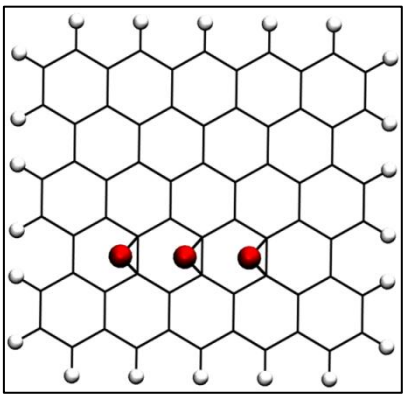
model-6b



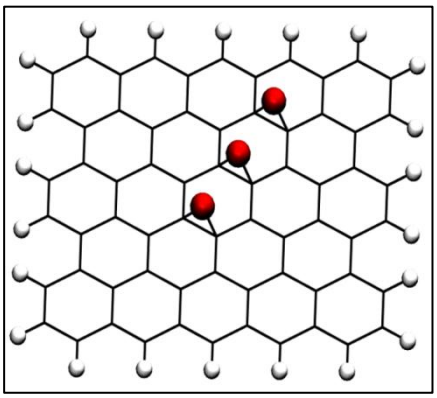
model-7a



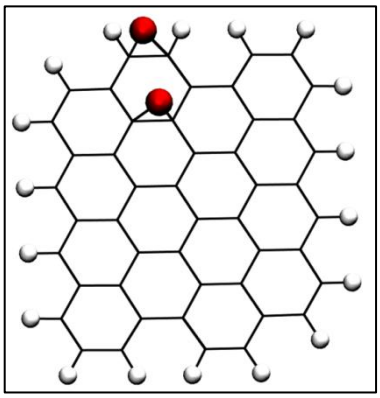
model-7b



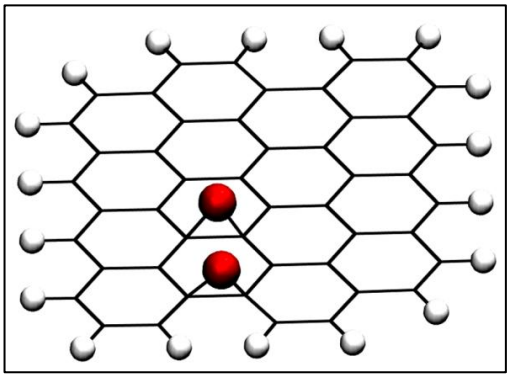
model-8a



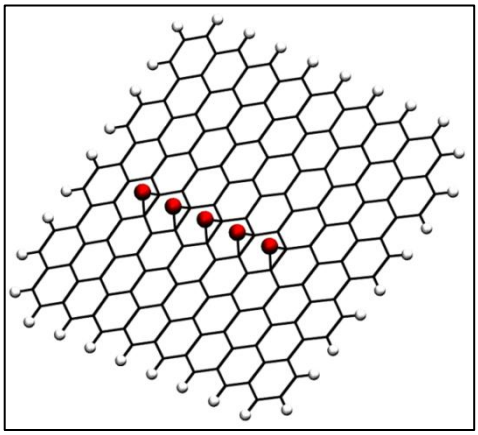
model-8b



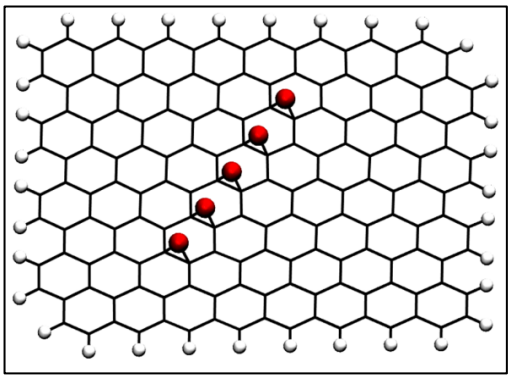
model-9a



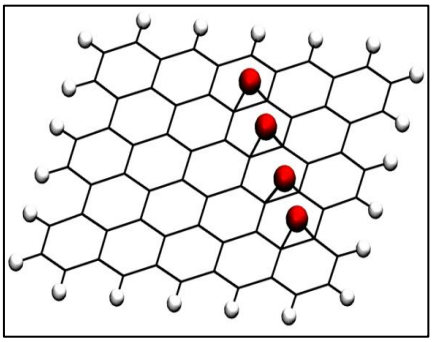
model-9b



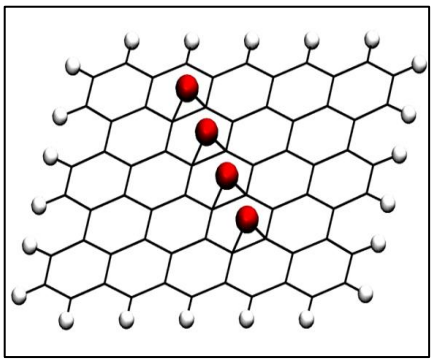
model-10a



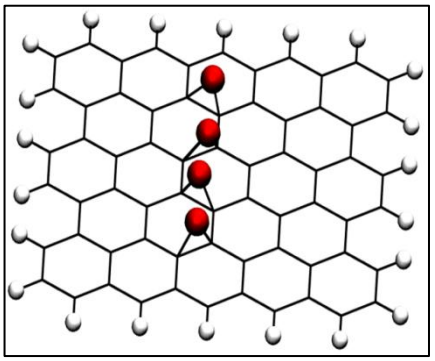
model-10b



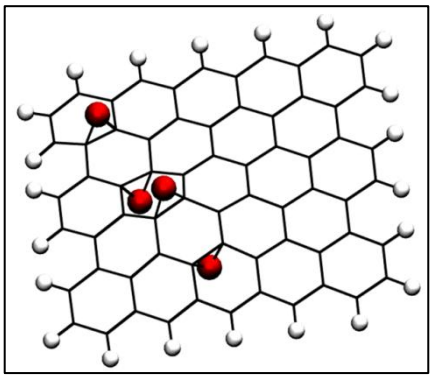
model-11a1



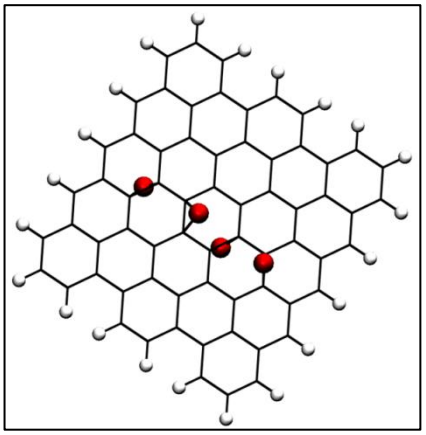
model-11a2



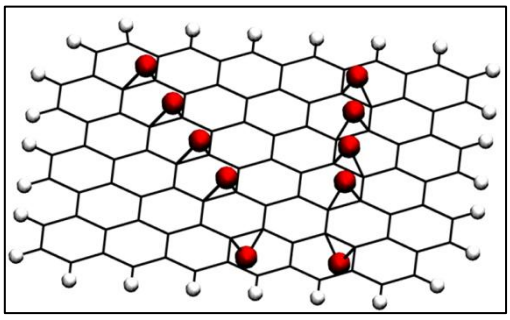
model-11a2



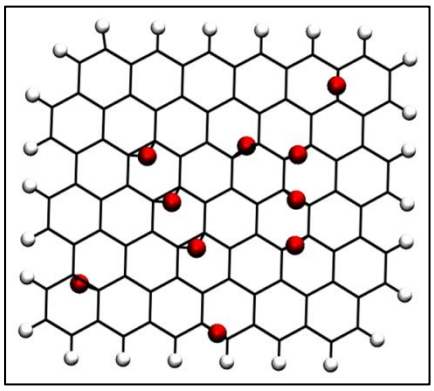
model-12a



model-12b



model-13a



model-13b

APPENDIX IV

Reference for APPENDIX I-II

[1] Leach, A. R. (1996). *Molecular modelling*, Longman Singapore.

[2] Verlet, L. (1967). Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard- Jones molecules. *Phys Rev* 159: 98

[3] Leszczyński, J (2005). *Computational chemistry: reviews of current trends, Volume 9.* pp. 54–56. ISBN 978-981-256-742-0.

[4] Darden, T., York, D. & Pedersen, L. (1993). Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J Chem Phys* 98: 10089

[5] Sagui, C. & Darden, T. A. (1999). Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu Rev Biophys Biomol Struct* 28: 155

[6] Foloppe, N. & MacKerell, A. D. (2000). All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comput Chem* 21: 86

[7] Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., … Mackerell, A. D. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*, *31*(4), 671–690.

[8] Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A. & Orozco, M. (2007b). Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys J* 92: 3817

[9] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, *65*(3), 712–725.

[10] Mackerell, A. D., Jr. (2004). Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem* 25: 1584

[11] J. Srinivasan, M. W. Trevathan, P. Beroza, and D. A. Case., *Theor Chem Acc*, 101:426-434, 1999.

[12] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson., *J. Am. Chem. Soc.*, 112:6127-6129, 1990

[13] A. Onufriev, D. Bashford, and D. A. Case., *J. Phys. Chem.*, 104:3712-3720, 2000

[14] A. Onufriev, D. Bashford, and D. A. Case., *Proteins: Struct., Func., Gen.*, 55:383-394, 2004.

[15] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar., *J. Phys. Chem.*, 100:19824-19839, 1996.

[16] McCammon, J. A., and Harvey, S. C. (1987) Cambridge University Press, Cambridge.

[17] Møller, C. & Plesset, M. S. (1934) Note on an approximation treatment for many-electron systems. *Phys. Rev.,* 46**,** 618

[18] Hohenberg, P. & Kohn, W. (1964) Inhomogeneous electron gas. *Phys. Rev.,* 136**,** B864.

[19] Kohn, W. & Sham, L. J. (1965) Self-consistent equations including exchange and correlation effects. *Phys. Rev,* 140**,** A1133

[20] Kristyán, S. & Pulay, P. (1994). Can (semi) local density functional theory account for the London dispersion forces? *Chem Phys Lett 229*: 175

[21] Cybulski, S., Bledson, T. & Toczyłowski, R. (2002). Comment on "Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory treatment". *J. Chem. Phys. 116*: 11039.

[22] van Mourik, T. & Gdanitz, R. J. (2002). A critical note on density functional theory studies on rare-gas dimers. *J. Chem. Phys. 116*: 9620.

[23] Elstner, M., Hobza, P., Frauenheim, T., Suhai, S. & Kaxiras, E. (2001). Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. *J Chem Phys 114*: 5149

[24] Chai, J. D. & Head-Gordon, M. (2008a). Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys Chem Chem Phys 10*: 6615

[25] Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., Scalmani, G., Barone, V., Mennucci, B. & Petersson, G. (2009). Gaussian G09, Gaussian, Inc. Wallingford, CT.

Publications and Proceedings

List of Publications

1. **Characterization of Unfolding Mechanism of Human Lamin A Ig Fold by Single-Molecule Force Spectroscopy—Implications in EDMD.**

   Bera, M., Kotamarthi, H. M., Dutta, S., **Ray, A.**, Ghosh, S., Bhattacharyya, D., Koti Ainavarapu, S. R., and Sengupta, K.

   ***Biochemistry*, 2014, *53* (46), pp 7247–7258**

2. **Oxidative Tearing of Graphene Sheets: Insights into the Probable Situations by Computational and Experimental Studies.**

   Ray, A., Bagani, K., Banerjee, S., and Bhattacharyya, D.

   ***The Journal of Physical Chemistry C*, 2015, *119* (2), pp 951–959**

3. **A Comparison of Four Different Conformations Adopted by Human Telomeric G-Quadruplex Using Computer Simulations.**

   Ray, A., Panigrahi, S., and Bhattacharyya, D.

   ***Biopolymers*, 2016, *105*(2), pp 83-99**

4. **Excited State Hydrogen Bonding Fluorescent Probe: Role of Structure and Environment.**

   Dey, D., Sarangi, M. K., **Ray, A.**, Bhattacharyya, D., and Maity, D. K. M.

   ***Journal of Luminescence*, [In Press, Accepted Manuscript]**

5. **RNA Bulges and Pseudo-continuous Double Helices: Structural Study by Molecular Dynamics Simulations**

   Ray, A., Agarwal, A., and Bhattacharyya, D.

   [Manuscript in Preparation]

Poster Presentation

1. **Stability Comparison Of Quadruplex Dna: Molecular Dynamics Simulation**

   Ray, A., Bhattacharyya, D.
   Conference on **"Recent Advances in Chemical and Physical Biology"** organized by Saha Institute of Nuclear Physics (SINP), India and The Mechanobiology Institute (MBI), NUS Singapore during March 5-7, 2012.

2. **Structural Features of Telomeric Quadruplex DNA having Different Topologies: Study By Molecular Dynamics Simulation**

   Ray, A., Bhattacharyya, D.
   National Symposium on **"Frontiers of Biophysics, Biotechnology and Bioinformatics" and "37th Annual Meeting of Indian Biophysical Society (IBS)"** jointly organized by Department of Biophysics and Centre for excellence in Basic Sciences, University of Mumbai during January 13-16, 2013.

3. **Tearing of Graphene Sheets by Oxidation: Study by Quantum Chemical Approach**

   Ray, A., Bhattacharyya, D., Banerjee, S.
   Workshop entitled **"Introduction to Gaussian: Theory and Practice"** held at the Hotel Connaught in New Delhi, India from January 6th through 10th 2014. The workshop was co-hosted by Gaussian, Inc. and SCUBE Scientific Software Solutions.

4. **Study of Telomeric Quadruplex DNA having Different Topologies by Molecular Dynamics Simulation**

   Ray, A., Bhattacharyya, D.
   Annual Conference of Indian Biophysical Society entitled "**Molecular Architechure, Dynamics and Assembly in Living Systems"** held during 7-10 February 2014 at SINP, Kolkata.

5. **Molecular Dynamics Simulation of Human Telomeric Quadruplex DNA**

   Ray, A., Bhattacharyya, D.
   The "National Symposium on Biophysics and Golden Jubilee Meeting of Indian Biophysical Society", held during February 14-17, 2015 at Jamia Millia Islamia, New Delhi.

6. **Study of Graphene, Graphene Oxide and Functionalized Graphene Sheets by Quantum Chemical Approach**

   Ray, A., Bhattacharyya, D., Banerjee, S.
   The **14th International Spin Chemistry Meeting (SCM-2015)** held at SINP, Kolkata during 16th-20th March, 2015.

Invited Talk/Oral Presentation

1. Talk titled **"Oxidative tearing of Graphene Sheets: Insights by Computational Approach"** at "Half Day Seminar on aspects of Current Science by Research Scholars" Kalpana Chawla Centre for Space and Nano Sciences, Salt Lake campus on 07th Jan 2015.

Workshop/Seminar Participation

1. Workshop entitled **"Role of Computational Biology in Advancing Modern Medicine"** organized by Centre for Applied Mathematics and Computational Science, Saha Institute of Nuclear Physics, Kolkata during February 2-3, 2012.

2. Seminar on **"Facets of Insilico Chemical Biology for Novel Therapeutics"** held on 18th June 2013 at CSIR-IICB, Kolkata.

3. Seminar on **"Films of Soft Materials"** organized by Surface Physics and Materials Science Division, SINP held on 2nd December 2013.

4. One day workshop entitled **"Bio3D and R scripting for MD trajectory analysis"** at Bose Institute, Kolkata on 16th of December, 2013.

5. Workshop entitled **"Workshop on Electronic Structure, Atomistic and Statistical Modeling in Chemistry, Materials and Life Sciences"**, October 20-22, 2014, IACS, Kolkata with Industry Partner Schrodinger, India.