The Structure and Dynamics of DNA-cation Systems

Bor Yior Yee

MSc by Research

University of York
Physics

November 2015

Abstract

A study of long-term effects of the polycationic spermidine molecules on the conformational stability of the oligonucleotides $d(A)_{20}$ and $d(A\text{-}T)_{20}$ using molecular dynamics simulations has been carried out, with an investigation into the effects of multivalent cations, particularly magnesium, on the structural preferences of the sequences. The preference of B-form DNA over A-form induced by the univalent sodium is shown to be largely disrupted by magnesium, and it is also demonstrated that magnesium ions could potentially be a source of conformational instability.

The introduction of spermidine has been proven to be a source of stabilisation to the DNA structures by restricting their motions and hence greatly reducing the fluctuations in multiple parameters. However it is also shown that the overall stabilising power is not directly proportional to the number of spermidine molecules present around the DNA macromolecule.

The effect of caesium ions on the conformational preference of specific DNA sequences of d(ACCGGCGCCACA) and d(ACCGGCGCCGGT) has been studied using sodium as a reference. It is found that although caesium is monovalent like sodium, it does not induce a strong preference of B-form over A-form, but rather provides extra stability for the initial state. Furthermore, it has also been demonstrated that being palindromic does not seem to provide structural stability for short DNA sequences.

Contents

List of Figures

List of Tables

List of Accompanying Material

Attached to this thesis is a DVD with all the graphics and histograms generated for this thesis. The contents inside the DVD are divided into five folders, namely *misc*, *parameters*, *rms*, *snapshots* and *videos* respectively.

The graphics inside the "misc" folder are mainly ones which were produced using the computer software "CorelDraw" and were used in chapter 1 to aid the explanation of various structural parameters of DNA.

The "parameters" folder includes various structural parameters. The names of the files in this folder are all constructed according to a fixed format. They start with the code of the sequence (explained in chapter 3) and end with either "p-his", "s-his" or "sp". The graphs named "p-his" are histograms for *xdisp*, *incline*, *h-rise* and *h-twist*, whereas the histograms named "s-his" show the distributions of the groove depths and widths. Graphs named "sp" are the time-evolution graphs for the *sugar pucker* parameter. Some of the histograms are used in chapters 3 and 4 for analysis. Though many of them have not been included in the main chapters, they had played an important role in analyses which led to the final conclusion.

All the RMSd graphs presented in chapters 3 and 4 are included inside the folder "rms".

The "snapshots" folder includes the snapshots of 8 different structures at various time steps from the simulations.

Last but not least, the videos of 20 sequences (from which the snapshots were extracted) are included in the folder "videos".

Acknowledgements

I would like to first thank Dr Robert Greenall, not only for his supervision of this work, but also his invaluable knowledge and experience in the field. Keeping me working by maintaining constant weekly meetings between us, he is a superb supervisor who is always liberal and open-minded, listening to my ideas patiently and giving extremely useful feedback. I am also largely beholden to Prof Rex Godby and Dr Matt Probert, who had me explain my ideas of my project from time to time and thus helped me clear my mind a lot.

Dr Paul Sharp has been a great friend since my arrival a year ago, and has taken great care of me in my early life as a researcher - cheers mate! I would also like to thank everybody from N103 and N102, especially Dr Phil Hasnip for his vast knowledge in multiple fields of physics, Jacob for his unceasing yet intelligent sense of humour and his constant aid with my programming skills, and Jonathan (now graduated) for dragging me away from my intense work by inviting me to have a chat now and then. I am sincerely grateful to Ed, Matt, Liam, Mike, Greta and Manuel - my "lunchmates" who have created a friendly environment and made me feel so much at home.

Special thanks shall be given to Mr Jack Shepherd for always being helpful. He gave me every resource available for learning the research tools that I would have to use in this project. He has always been the first person on whom I would rely when my input parameters needed some serious debugging.

I must also say thank you to professors and lecturers at the Hong Kong Baptist University who had taught me and with whom I had worked, especially Prof Michel Van Hove who not only supervised my final year project but guided me to become even more determined as a physicist.

Last but not least, my uttermost thanks go to my parents for their support in everything - I am too much indebted to them. Without them nothing could be made possible.

Author's Declaration

I declare that the work presented in this thesis, except where otherwise stated, is based on my own research and has not been submitted previously for a degree in this or any other university.

Chapter 1

Introduction

Deoxyribonucleic Acid (DNA) and ribonucleic acid (RNA) are collectively known as nucleic acids. They are very important and fundamental building blocks of life forms. Therefore studies of DNA have been a hot topic in the scientific world since its discovery. Moreover DNA has been found to have important interactions with other biological molecules. In this thesis we are going to investigate the interactions of DNA with spermidine, which is a naturally existing biological polyamine.

1.1 The structure of DNA and its biological significance

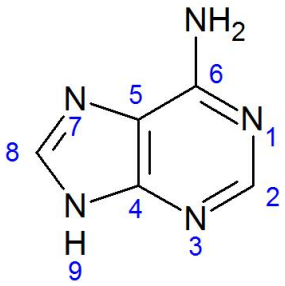
DNA is a long polymeric chain composed of nucleotides. Different organisms have different lengths of DNA. For example, the total length of DNA in a single mammalian cell is typically about 1 metre, whereas that for a fungus is around 1 centimetre and that for a bacterium is of the second order of magnitude in microns [50]. A nucleotide is the basic unit of a nucleic acid. It consists of a base, a five-membered sugar ring (deoxyribose—deoxyfuranose), and a phosphate group. Whilst the sugar ring and the phosphate group are the same for all nucleotides, the bases differ from one nucleotide to another nucleotide. In DNA, the bases are adenine (A), guanine (G), cytosine (C) and thymine (T); in RNA thymine is replaced by uracil (U). They can be further divided, according to their chemical structures, into two groups: purine (consisting of a six-membered aromatic ring) and pyrimidine (consisting of fused five-membered and six-membered rings). Whilst adenine and guanine are purines (Pu), cytosine and thymine are pyrimidines (Py). The nucleotides formed from the four bases are called deoxyadenylate (deoxyadenosine monophosphate,

dAMP), deoxyguanylate (deoxyguanine monophosphate, dGMP), deoxycytydilate (deoxycytosine monophosphate, dCMP) and deoxythymidylate (deoxythymine monophosphate, dTMP) respectively. Figure 1.1 shows the bases (on left hand side) with their corresponding nucleotides on their right.

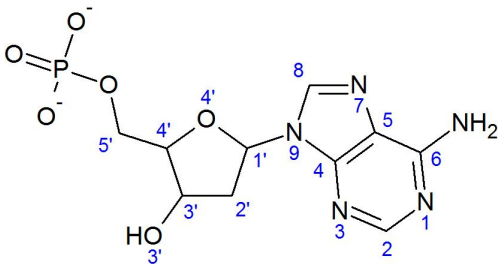
In any arbitrary sequence of DNA, there are free 5′ and 3′ oxygen atoms at the ends of each strand. Nucleotides are joined with one onother with its O5′ atom, via a phosphate residue, with the O3′ atom of its counterpart. Therefore we could assign a directionality for the sequence, and name the polymer such that the direction points in the 5′→3′ sense, as shown on Figure 1.2. As a result, the simple dimer in Figure 1.2 is called "d(AG)·d(CT)", or simply d(AG). The letters inside the parenthesis indicates the sequence in which the nucleotides are aligned. The preceding "d" indicates that this molecule is a DNA, rather than an RNA (whose preceding letter would be naturally an "r").

The sequences of DNA molecules are not trivial, in a sense that there is a criterion for two nucleotides to form a pair (i.e. not every combination of two nucleotides is eligible to form a pair). In 1953, Watson and Crick [63] proposed, that the two bases forming a pair are held together by hydrogen bonds. Moreover they found out that that DNA, as a "stack" of these base pairs, exists as an anti-parallel double helical structure. Furthermore from a chemical model of the molecule that they constructed, they postulated that feasible formations of hydrogen bonds between the bases occur within the pairs A·T and C·G, hence explained the "Chargaff's rules"(or, more specifically, the "Chargaff's first parity rule"), which says that the ratio between purine and pyrimidine bases are always close to unity for DNA [9, 18, 66]. It is also an interesting fact that though the A·T and C·G combinations are the most feasible, there is more than one way that could be adopted in order that the pairs be formed. For example, Hoogsteen first reported a crystal structure in 1963 that the hydrogen bonds could be formed at different sites to those postulated in the Watson-Crick model [27] – The base pairs formed under Hoogsteen's scheme are now known as the "Hoogsteen base pairs". Nevertheless, the Watson-Crick pairs are still the dominant species whilst the Hoogsteen pairs could be observed very rarely in crystals.

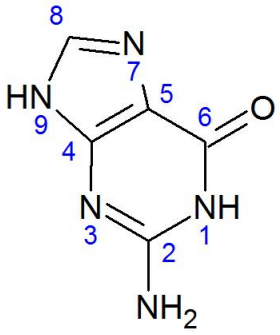
However, the discovery of the double-helical structure of DNA molecules does not mark an end to the story. It is because explanation for the existence of DNA in aqueous solutions (for instance, the nucleoplasm inside nuclei of cells) is not yet provided, in spite of the low polarity which makes it seem more plausible to be stabler in non-polar solvents (such as oil). The answer lies, though, in that at physiological acidity of about pH 7.6 the phosphate groups dissociate and the protons move into the solution leaving net negative charges on the



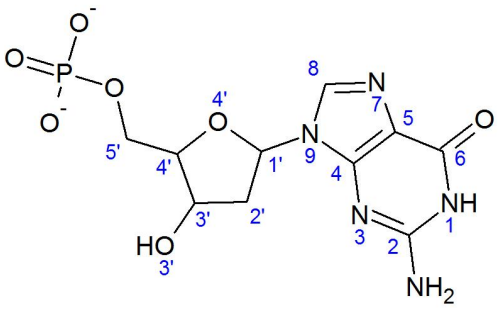
(a) Base: Adenine (Pu)



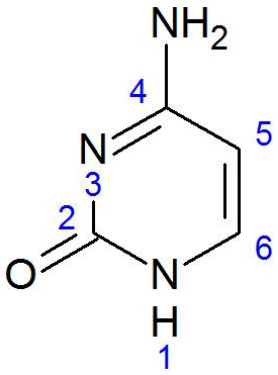
(b) Nucleotide: dAMP



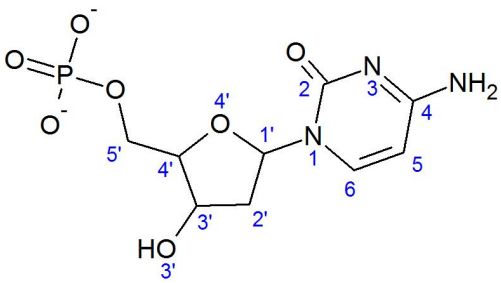
(c) Base: Guanine (Pu)



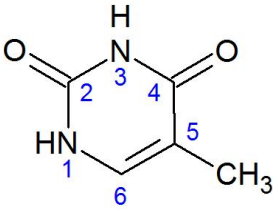
(d) Nucleotide: dGMP



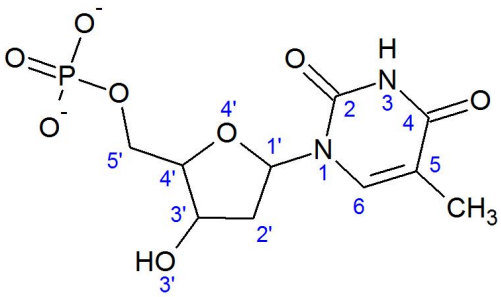
(e) Base: Cytosine (Py)



(f) Nucleotide: dCMP



(g) Base: Thymine (Py)



(h) Nucleotide: dTMP

Figure 1.1: Bases in DNA and their corresponding nucleotides

backbone, hence the name nucleic *acid*.

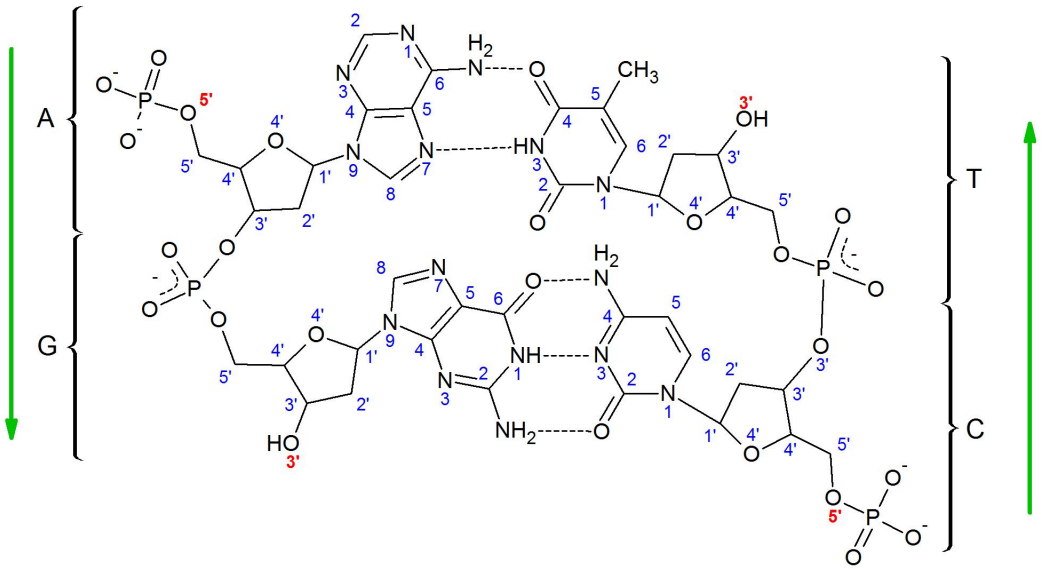


Figure 1.2: A DNA dimer — d(AG)·d(CT). The numbers in bold red fonts are the numbers of atoms which determine the 5′→3′ directionality of the strands.

1.2 DNA polymorphism

In their publication [63], Watson and Crick stated that "Both chains follow right-handed helices". However, over the past half-century since them, scientists discovered from X-ray diffraction that the structure that Watson and Crick modelled is not the only conformation of DNA molecules. In fact, there are at least five conformations, whose transitions depend heavily on the surrounding humidity and the ionic environment.

1.2.1 Canonical forms of DNA

Form	Pitch(Å)	Rise(Å)	Turn(°)	Relative Humidity	Ionic Requirements
A	28.2	2.56	32.7	$< 85\%$	medium, not Li
B	33.8	3.38	36.0	$> 92\%$	high Na, high Li
C	31.0	3.32	38.6	$< 66\%$	low
D	24.3	3.03	45.0	$< 92\%$	heavy ions best
S or Z	43.2	3.60	-30.0	$< 85\%$	high

Table 1.1: DNA conformations with their helical parameters, obtained from X-ray diffraction pattern. Reproduced from [49, 65].

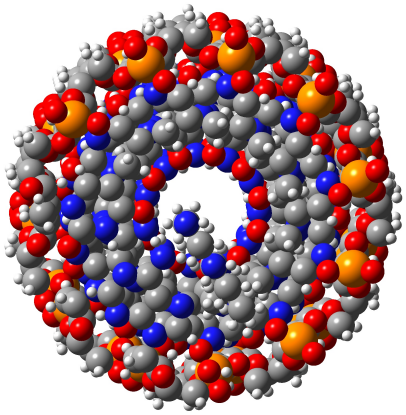
The model that Watson and Crick [63] created was based on the assumption that there is a repetition of the structure per 10 residues, and the "standard configuration" [22] with the vertical distance between two residues (a.k.a. the "rise", see Table 1.1) as 3.4Å, giving the length per repetition (a.k.a the "pitch")

about 34Å. This is now known as the B-form of DNA [20, 33], or simply B-DNA. However, Watson and Crick made very clear that their configuration is an open structure and exists in high water contents — this coincides fairly well with the observation summarised in Table 1.1. They also made a good prediction that if the B-DNA molecule were in slightly lower water contents the bases would tilt, such that the molecule would become more compact and the rise would decrease. This was later found to be true for the case of A-DNA [21], another common form of DNA. As compared to the B-form (33.8Å pitch, 10 base pairs per turn), A-DNA has a pitch of 28.2Å and 11 base pairs per turn. Moreover, the bases in A-DNA are on average 20° tilted away from the helical axis, in turn producing a "hole" around the helical axis. Figure 1.3 shows the difference between the A- and B- forms of the same sequence. It is clear, from the top-down view, that there is a large "hole" near the helical axis for A-form, while the hole is absent in B-form. Moreover, from the side view, the A-form is shown to be more compact ("short and fat") with the nucleoside (the base + the sugar ring, i.e. the "ladder rungs") inside tilted away from the perpendicular of the axis, while the the B-form is less compact and the rungs are less tilted.

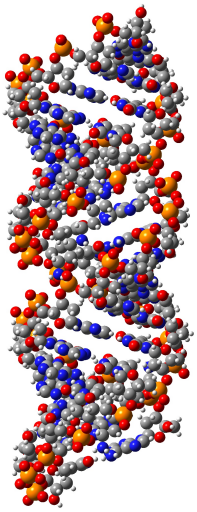
The commonest forms of DNA, i.e. A-DNA and B-DNA, are favoured in environments with relatively high water contents and small surrounding ions such as Li^+ or Na^+. However, in the presence of heavy ions the D-form is much more favoured than the A- and B- forms. From the data in Table 1.1, it can be seen that the D-form is in fact a highly twisted counterpart of the B-form. The turn angle (45°) reveals that there are only 8 base pairs per turn as compared to 10 in the case of B-form. Moreover, D-DNA can be thought to be even more elongated than B-DNA, in view of the rise of the D-form being about 18.3% more than that of B-DNA. It was discovered that the D-form exists in specific sequences such as d(AT)·d(AT) [36].

The S-form (or Z-form) of DNA is very much different to the other forms. While all the other forms take the "right-handed" helical orientation (i.e. a clockwise screwing motion would seem to move the helix away from a top-down observer), the S-form is a "left-handed helix" (i.e. a helix seemingly moving towards a top-down observer in a clockwise screwing motion). Just like the uncommon D-form, the S-form exists in the specific alternating sequence of d(CG)·d(CG) [26].

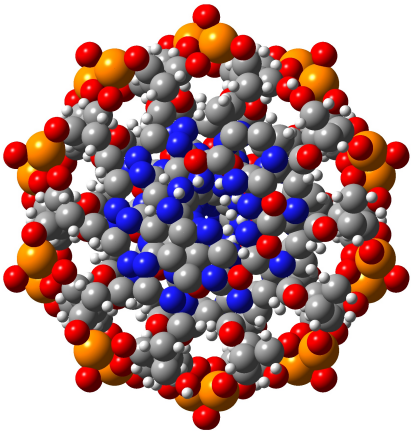
A- and B-forms of DNA, being the most commonly known and mentioned forms, are themselves the most abundant conformations found in organisms. Since DNA molecules tend to take the B-form in relatively humid conditions, it is the dominant form inside cells. However, the A-form is seldom found to be



(a) A-form: top-down view



(b) A-form: side view



(c) B-form: top-down view

(d) B-form: side view

Figure 1.3: Comparison between A-form and B-form DNA. Sequence of DNA icosamer: $d(A_{20})\cdot d(T_{20})$. Colour legends: Blue—Nitrogen, Grey—Carbon, Red—Oxygen, White—Hydrogen, Yellow—Phosphorus

standalone, but rather exists inside DNA/RNA complexes (RNA always takes the A-form). For this reason, it is assumed that the B-form takes the role as a storage of genetic code, while the A-form acts as an intermediary structure during interaction with other molecules.

All the different forms of DNA discussed above are in fact molecular models derived from fibre diffraction which uses very long sections of DNA assuming perfect helical structure, i.e. the same value for one parameter throughout the whole strand (2.56Å rise for A-DNA throughout, for example). However from X-ray crystallography, which uses oligonucleotides (short DNA sections consisting of only a few nucleotides), it was found that in reality there can be short sequence-dependent local variations which means that the DNA can bend and thus the parameters are not constant for each base-pair [14, 15].

1.3 Structural parameters of DNA

From the last section, especially from the figures and the tables, we have a brief glimpse at the overall structure of DNA molecules and the forms that it can take. In this section, various structural parameters of DNA including gross parameters (for instance, rise and twist, etc.) and those which describe the detailed structure of an individual nucleotide (for example, tip and inclination, etc.).

1.3.1 Nomeclature and definition of parameters

A set of parameters was defined in a European Molecular Biology Organisation conference in 1989 [16], which were to be used to describe the position and orientation of the bases and base pairs in a DNA molecule, with respect to each other and the helical axis.

In Figure 1.4, the reference frame for the bases have been clearly defined. This frame will be used frequently in this thesis, to derive further parameters regarding the double-helix and the base pairs. The z-axis, or the $+z$ direction, which stereotypically implies the "rising" direction in normal Cartesian coordinates, points perpendicularly to the the plane of the bases (the bases form a plane because of their aromaticity and that they are held together by hydrogen bonds). The y-axis is defined to be pointing in the direction *away* from the minor groove. The remaining x-axis, is then defined to be perpendicular to both y- and z-axes, such that the three axes form a completely right-handed orthogonal coordinate system.

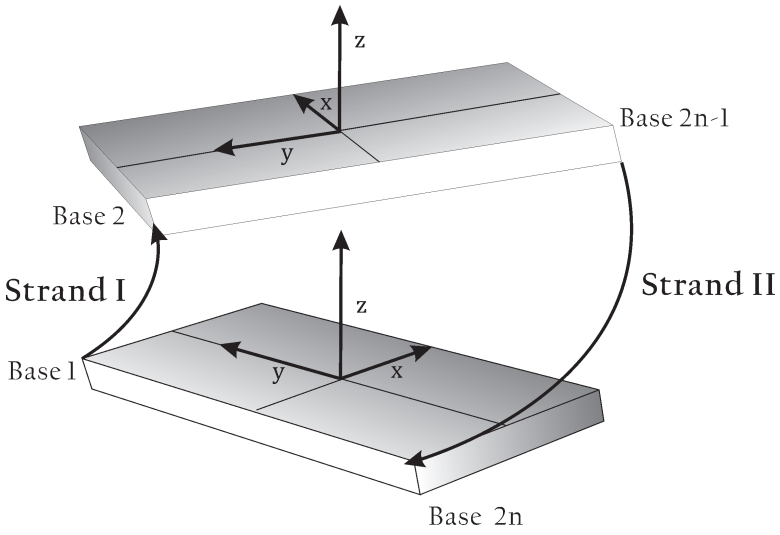


Figure 1.4: Base-pair centred reference frame. Curled arrows indicate the 5′—3′ directionality of backbone strands. Corners of the white edge denote the C1′ atom of the deoxyfuranose ring. Reproduced from [16].

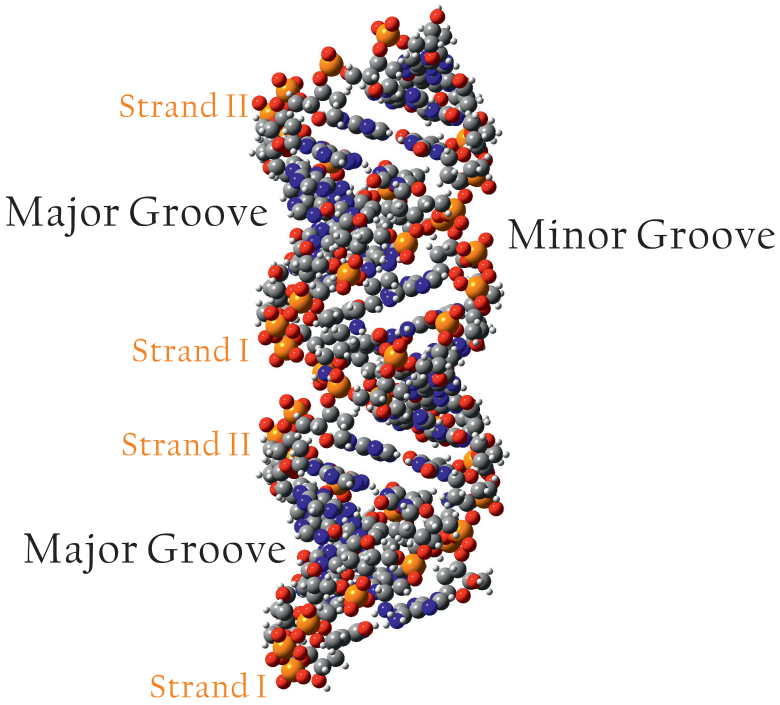


Figure 1.5: Major and minor grooves.

Since the DNA macromolecule has a double-helical structure, if one traverses the edge of the "cylinder", in the $+z$ direction, strands of backbone would be encountered alternately, i.e. Strand I $\rightarrow$ Strand II $\rightarrow$ Strand I $\rightarrow \cdots$. Owing to the tilting effect of the nucleosides, the depth of the groove between the *first* Strand I and the *first* Strand II, is usually different from that of the next groove (i.e. between the *first* Strand II and the *second* Strand I). The deeper

groove is labelled as the *major groove*, and the shallower the *minor groove*. It is counter-intuitive, though, that the tilting of inner nucleosides has *negative* effect on the difference in groove depths. This can be seen clearly comparing the subfigures 1.3(b) and 1.3(d) — A-DNA has more tilted bases but the difference in groove depth is less obvious. B-DNA has less tilted bases but the groove depths are much more distinguishable.

With this on hand, we can proceed onto defining various parameters regarding the base pairs. These can be broadly divided into three types, namely base pair axis parameters, intra base pair parameters and inter base pair parameters which are shown in figures 1.7 to 1.9, with Figure 1.6 giving the reference coordinates.

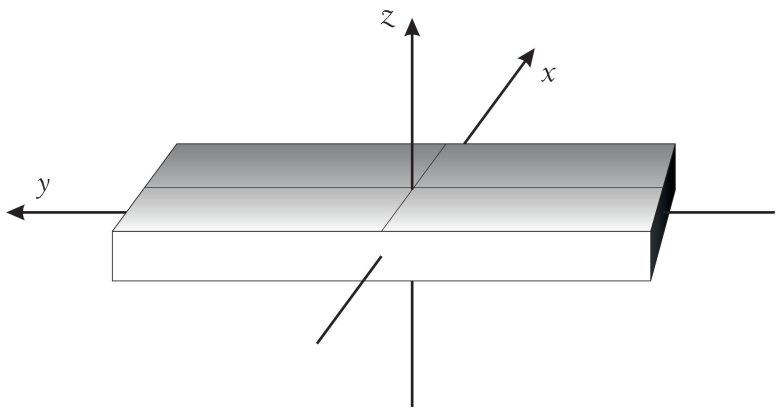
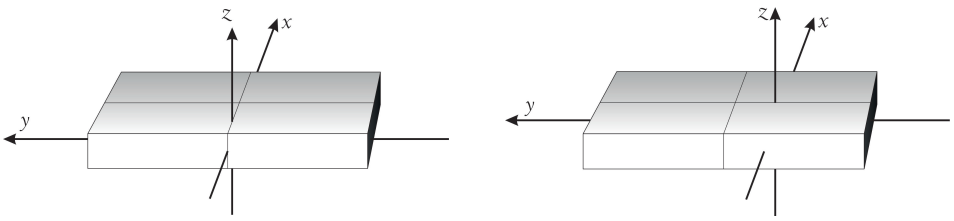
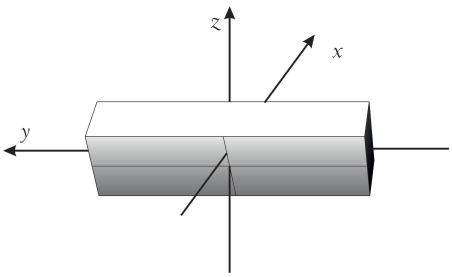


Figure 1.6: Reference coordinates

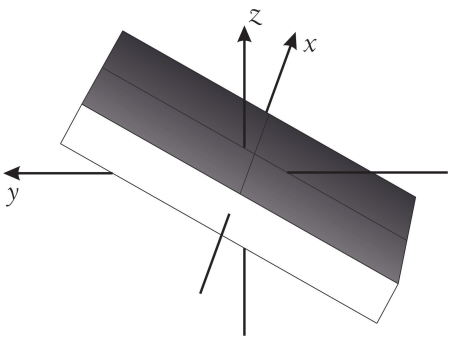


(a) y-displacement (xdisp)

(b) y-displacement (ydisp)



(c) Tip (η)



(d) Inclination (θ)

Figure 1.7: Base pair axis parameters

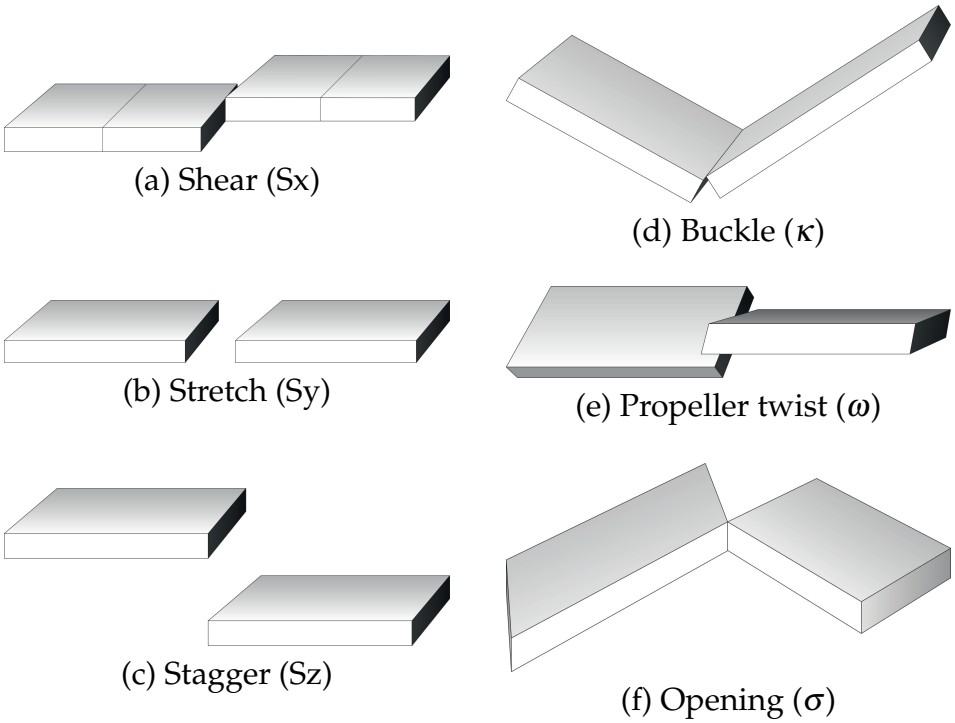


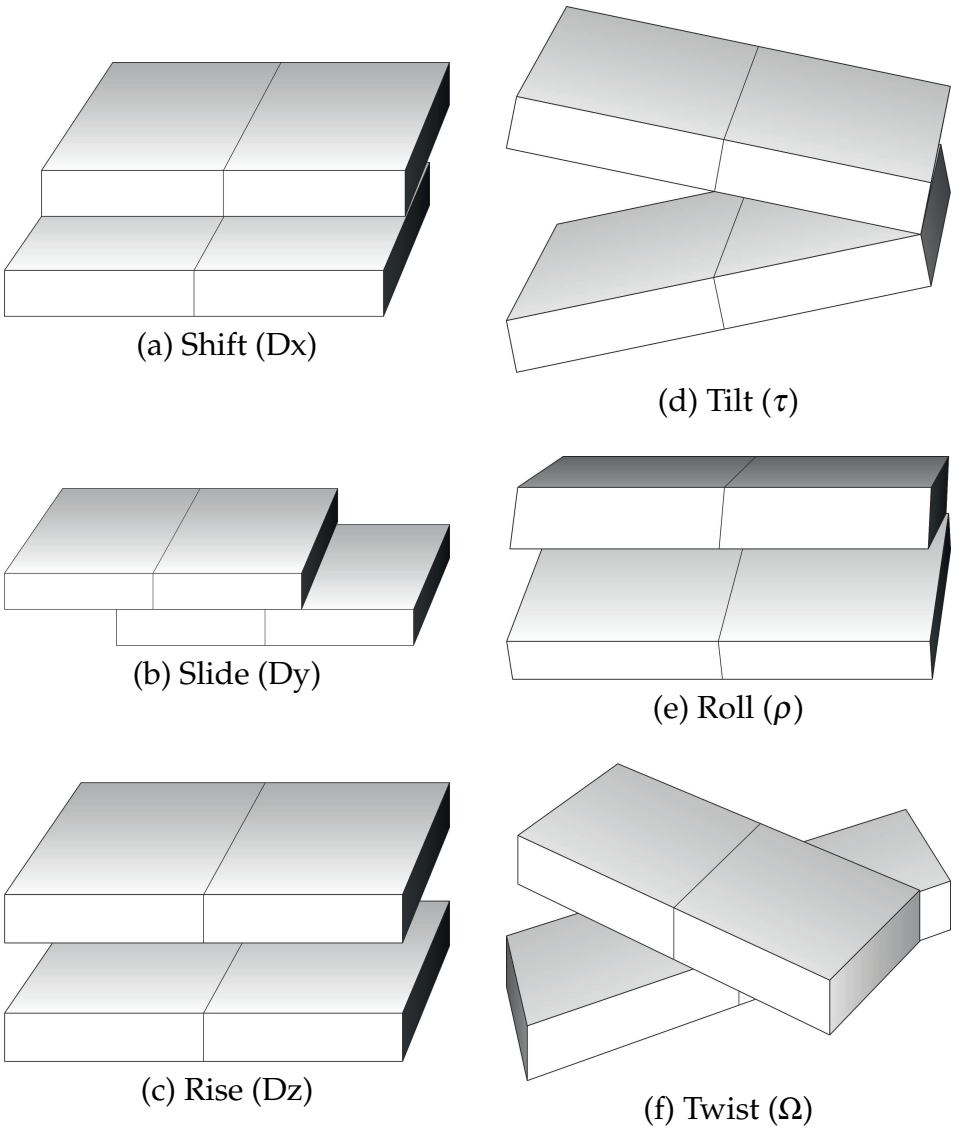
Figure 1.8: Intra base pair parameters

1.3.2 Backbone conformations

1.3.2.1 Bond lengths, bond angles, torsion angles and dihedral angles

To effectively describe the covalent bonds in a molecule, especially a complex molecule like DNA, the bond lengths and bond angles are the most fundamental yet good parameters to be used. Here we use the standard IUPAC convention — the bond length between bonded atoms A and B is denoted as b(A,B), and the angle subtended by the bonds A-B and B-C is written as θ(A,B,C). Moreover, given 4 atoms A, B, C and D, a plane could be formed by three of the atoms, say A-B-C. It can also be formed by another combination of triad, for instance, B-C-D. Then we can define an angle between the two planes, called the *torsion angle*, denoted as τ(A, B, C, D). φ(A, B, C, D), the complementary of τ(A, B, C, D), is known as the *dihedral angle* (see Fig. 1.10). Geometrically, the dihedral angle can be obtained by intersecting perpendicular lines drawn from the two planes [50].

There are some of the torsion angles within a nucleotide which are of special interest, and they are specifically labelled. $\alpha, \beta, \delta, \gamma, \varepsilon$ and ζ are designated for the torsion angles on the backbone, $\nu_0, \nu_1, \nu_2, \nu_3$ and ν_4 for the bonds within the sugar ring, and χ for the glycosidic linkage between the sugar ring and the base. Since the numbering system for purines (A and G) are different from that for pyrimidines (C and T), the true definition of χ for the two kinds of bases



(a) Shift (Dx)

(d) Tilt (τ)

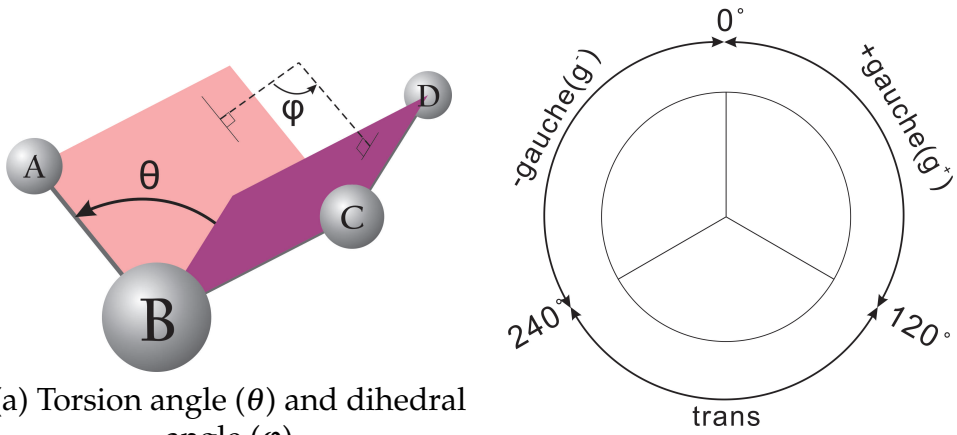
(b) Slide (Dy)

(e) Roll (ρ)

(c) Rise (Dz)

(f) Twist (Ω)

Figure 1.9: Inter base pair parameters

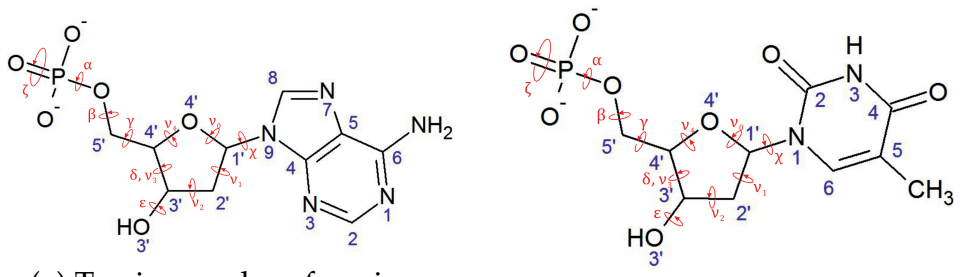


(a) Torsion angle (θ) and dihedral angle (φ)

(b) Conformational regions of torsion angle

Figure 1.10: Torsion angle and its conformational regions

are also different (cf. Figs. 1.1 and 1.11). For purines, $\chi = \tau(\mathrm{O4'}, \mathrm{C1'}, \mathbf{N9}, \mathbf{C4})$. Whereas for pyrimidines, $\chi = \tau(\mathrm{O4'}, \mathrm{C1'}, \mathbf{N1}, \mathbf{C2})$.



(a) Torsion angles of purines

(b) Torsion angles of pyrimidines

Figure 1.11: Torsion angles of purines and pyrimidines

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′–N9–C4 (Pu) O4′–C1′–N1–C2 (Py)
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′

Table 1.2: Definitions of torsion angles in nucleotides. Subscripted $(n-1)$ and $(n+1)$ indicate the preceding and subsequent nucleotide units respectively. Reproduced from [50].

1.3.2.2 Sugar puckering

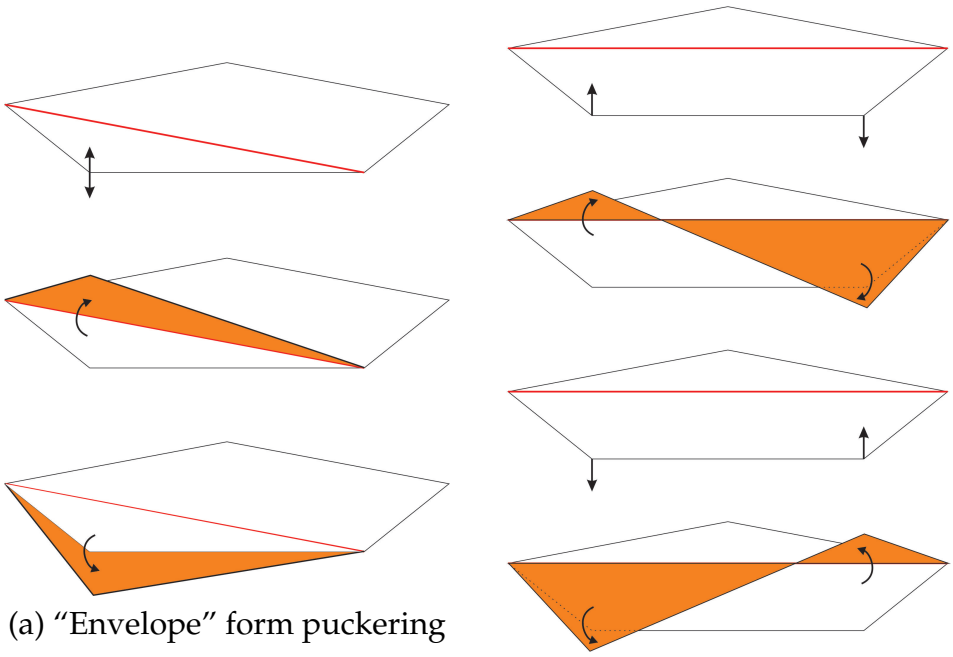
Due to the non-aromatic properties of the deoxyfuranose ring, the five members on the ring are seldom coplanar. While three of them always form a plane (by simple geometrical arguments), the remaining two may be above or below the plane. This is called "sugar puckering". Therefore a set of conventions can be defined to describe the puckering, and we adopt the definition of principle puckers from Saenger [50].

As elucidated earlier, to accurately describe the puckering forms, a reference plane must be chosen. According to Saenger, the reference plane is always chosen to be that containing atoms **C1′**, **O4′** and **C4′** (we may call it "Π" for

simplicity). Owing to the sp^3 hybridisation of C4′, the atom C5′ (on the backbone) is always either out of the plane Π or "into" Π.

There are two standard notations describing how the ring is puckered: the "envelope/twist" [10] and the "endo/exo" notations [50]. Both will be briefly explained, but the "endo/exo" notation will be used throughout this thesis as this is the most widely adopted in literature in the field.

Using the "envelope/twist" notation, the above-mentioned form of puckering is called a "twist", and is abbreviated as "T". As a remark, there is also a possibility that the *fourth* atom, either C2′ or C3′, that may also be coplanar with Π. In this case, the form is called an "envelope", with the symbol "E".

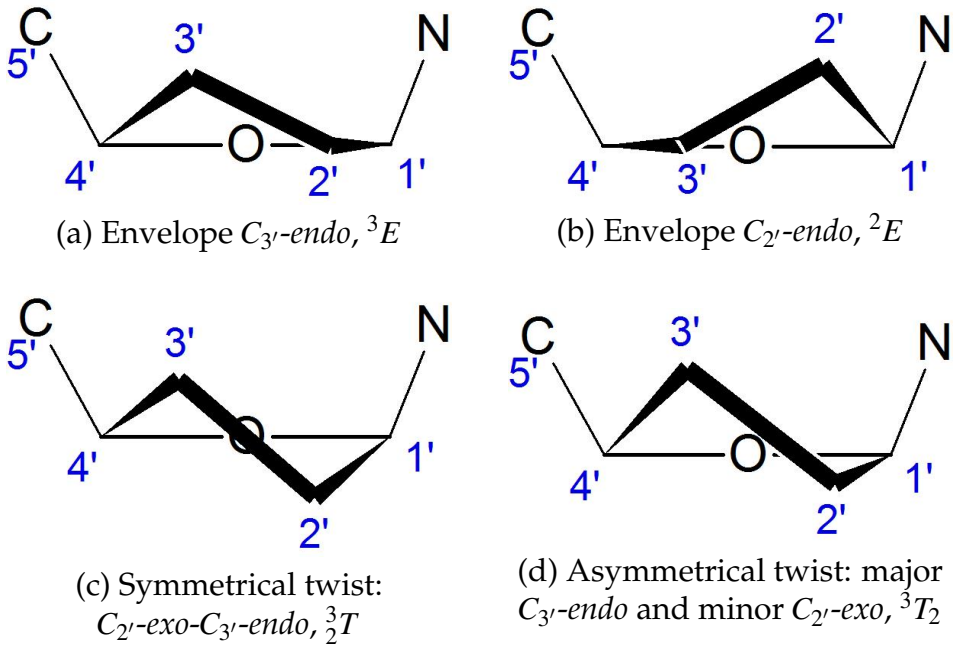


(a) "Envelope" form puckering

(b) "Twist" form puckering

Figure 1.12: Suger puckering. Reproduced from [50].

Due to different steric effects of the ring imposed by either sides, the degree of displacement of C2′ and C3′ in a T-form may not be the same. In fact, it is rare that C2′ and C3′ have the same displacement. In view of this, the larger deviation is called the "major puckering", while the smaller is the "minor puckering". Moreover, using the position of the C5′ atom (relative to Π) as a reference, the displacements of atoms in the direction of C5′ are called *endo*, and those in the opposite direction of C5′ are *exo*.



(a) Envelope $C_{3'}$-*endo*, ${}^{3}E$

(b) Envelope $C_{2'}$-*endo*, ${}^{2}E$

(c) Symmetrical twist: $C_{2'}$-*exo*-$C_{3'}$-*endo*, ${}^{3}_{2}T$

(d) Asymmetrical twist: major $C_{3'}$-*endo* and minor $C_{2'}$-*exo*, ${}^{3}T_{2}$

Figure 1.13: Definition of sugar puckering modes. Reproduced from [50].

1.3.3 Secondary structural parameters

In the previous sections, we have introduced and used some terminologies which describe the "macroscopic" properties of the DNA. They are collectively called the "secondary structural parameters", as contrasted with their counterparts explained in the last section. Here in this section we will give more details to their definitions and their calculations.

The secondary structural parameters are parameters used to describe the large-scaled observables of the helical structure. The vast majority of them were already met in Table 1.1. They are: number of nucleotides per helical turn n, rise (distance between two successive nucleotides) r, pitch of helix $p = nr$, turn (angular displacement per step) $t = \frac{360^\circ}{n}$. They are useful in describing properties of DNA molecules, which are mostly non-perfect helices.

As shown in Fig. 1.5, most DNA structures have "major" and "minor" grooves depending on the groove width. The widths and the depths of the grooves are thus defined as:

Quantity	Distance between:
Groove width	$P_{(n)} \longrightarrow P_{(n+1)}$
Minor groove depth	$P \longrightarrow N2_{\text{guanine}}$ or $P \longrightarrow O2_{\text{thymine}}$
Major groove depth	$P \longrightarrow O6_{\text{guanine}}$ or $P \longrightarrow O4_{\text{thymine}}$

1.4 Spermidine

Spermidine, along with other similar molecules such as putrescine and spermine, constitute the family of biological molecules known as "polyamines". Its structural formula is $H_3N^+ - (CH_2)_4 - N^+H_2 - (CH_2)_3 - N^+H_3$.

As its name implies, spermidine can be found inside semen. In fact, it was first isolated from semen. It can also be found at high concentrations in the brain [6], especially around the hippocampus [3, 39, 53]. It has the function of regulating biological processes, such as Ca^{2+} influx by glutamatergic N-methyl-D-aspartate (NMDA) receptor [7]. The NMDA receptor has linkage with the activation of cGMP/PKG pathway and nitric oxide synthase (NOS) which catalyses the production of nitric oxide [47], in turn decreasing the activity of the enzyme Na^+,K^+-ATPase in the cerebral cortex synaptosomes. This enzyme is crucial in the maintenance of intracellular electrolyte homeostasis [55], thus the neuronal signalling and behaviours [7, 17, 42].

Moreover, recently spermidine has been connected to the mediation of cell functions and metabolism. More importantly, it has even been attributed as a "longevity agent" because of its ability to regulate gene expression [30]. It has been shown by Ramot et al. [48] in their studies that spermidine promoted human hair shaft elongation and prolonged hair growth, hence confirming the hypothesis that polyamines play an important role in hair growth [46].

It has also been mentioned in several publications that spermidine is capable of stabilising DNA (and RNA) and modulate their replication and transcription [12, 43, 45, 51, 58]. However since little has been done on the interactions between spermidine and DNA on the molecular scale, it is still a big question to be answered. The remainder of this thesis is dedicated to the exploration of effects of spermine on DNA conformation and conformational stability.

1.5 The history of DNA simulations

Since the advent of the modern computer, simulations of molecular systems had been made possible. One way of doing them is through the so-called classical molecular dynamics, where classical mechanical and electrostatic forces are dominant.

As early as 1957 [1], molecular systems were simulated as hard spheres. Then with the rapid advancement in computer technology, systems as complicated as proteins were first simulated in 1977 [41]. As a natural result of systems becoming more complex, the demand for higher computing power also surged.

The development of various techniques such as the TIP3P water model ("3-site" water, disregarding the two lone pairs of electrons), the SHAKE algorithm (constraints the bonds involving hydrogen atoms), and the Particle Mesh Ewald (PME) summation method (deals with long-ranged electrostatic potentials), have been successful in making these highly complicated simulations much less computationally expensive while maintaining the accuracy at a high level.

The first ever reported MD simulation of DNA was by Levitt in 1983, which he simulated a Dickerson-Drew dodecamer, $d(CGCGAATTCGCG)_2$, and a tetraicosamer $d(A)_{24}$ for 90 picoseconds [37]. In the same year, Tidor et al. also performed a 60 picosecond simulation of the hexamer $d(CGCGCG)_2$ [59]. However they have shown great difficulties in common: highly approximately parametrised force fields, the net charge on DNA molecules and the omission of explicit solvent molecules and counterions [34]. Later on, Seibel et al. successfully simulated a pentamer d(CGCGA) for 106 picoseconds in 1985 [52] and Van Gunsteren et al. an octamer d(CGCAACGC) for 90 picoseconds in 1986 [62], both in the presence of a box of water molecules. These marked a milestone in the history of DNA simulation.

The history of MD simulations of DNA could possibly be divided into two eras: one from mid-1980s to mid-1990s, and the second from the mid-1990s to today. In the first decade of MD simulations the world saw the massive improvement in the quality of simulations because of the advancement in the force field calculation. Thanks to this the length and complexity of systems have greatly increased. However it was not until the second era that the breakthrough came, where the Ewald summation method – which will be discussed in detail in the next chapter – was first implemented in the evaluation of long-range potential and forces. This largely increased the computational efficiency of simulations, and hence allowed them to be brought up to the multi-nanosecond range. Recently simulation of a microsecond (obtaining 9 Tb of data with explicit solvent and 300 Gb with water molecules removed) has been reported [44].

1.6 Previous simulations of DNA/polyamine systems

The molecular simulations of DNA/polyamine systems sprouted between the late 1980s and early 1990s, which is only a few years after the start of the simulations of DNA systems alone. The initial attempts were mainly made to investigate the bending of DNA molecules due to interactions between spermine (a

naturally-existing polyamine like spermidine). It was in 1990, that Feuerstein et al. successfully showed that interactions are most favourable at the major groove of alternating Pu/Py sequences and it invokes significant bending of the DNA, whereas major groove interactions are much less favourable for homopolymers than heteropolymers and the extent of bending is minute [19].

A few years later the focus of the simulations turned from the bending of DNA to the conformational changes induced by polyamine-DNA interactions. In 1992 Haworth et al. demonstrated through simulations that a B-to-Z form transition is possible [25].

In 2001, Real [49] showed that spermine molecules provided stabilisation for the A-form of the dodecamers d(CGCGAATTCGCG) and d(CGCATATATGCG) when placed in the major groove, whereas the stabilisation effect is much less significant for the sequence d(CGCAAAAAAGCG).

Apart from the the DNA bending and conformational changes, the favourable binding sites of polyamines on DNA, i.e. specific locations at which the polyamines would reside, is also an interesting topic to be studied. In 2000 Bryson and Greenall showed that while the major groove is the preferred binding site on A-DNA for all polyamines, putrescine and cadaverine binds to the backbone of B-DNA whereas spermidine and spermine are more mobile and do not have a specifically favourable binding site [5].

More than a decade later, Shepherd [54] further confirmed that spermine favours A-form stabilisation by showing that A-to-B conformational transition is slowed down in the presence of spermine; and the more spermine added the more retarded would the transition be. However the stabilisation power is also sequence-dependent.

Computational simulations of DNA have been performed by various research groups in the world. For example, the group from Stockholm University consisting of Rupprecht, Korolev, L.yubartsev and Nordenskiöld has worked on the binding of small but mobile cations on DNA [32].

1.7 Outline of Research

The work presented in this dissertation makes use of molecular dynamics techniques to simulate the overall structures and dynamics of DNA-spermidine systems in the presence of various counterions.

In the next chapter, the theoretical backgrounds regarding molecular modelling and dynamics will be discussed. Different simulation approaches will

first be reviewed, followed by in-depth analysis of the AMBER force field. The protocol of molecular simulation and dynamics adopted by the AMBER package will then be explored. The chapter will be concluded with a detailed discussion of computational solutions to long-range forces.

The major research methodology and findings will be presented in chapters 3 and 4.

In chapter 3, the notation of DNA sequences will first be discussed in order to clarify potential confusions. Then simulations protocols which are adopted throughout the research is explained and used to study the sequences $d(A)_{20}$ and $d(A\text{-}T)_{20}$ and their interactions with varying number of spermidine molecules nearby. Magnesium ions will used as counterions in the systems with sodium as control. Structural parameters will be calculated and presented as an aid to determine possible form transitions.

Similar analytical techniques will be used in chapter 4 to investigate the effects of spermidine on two different sequences, namely d(ACCGGCGCCACA) and d(ACCGGCGCCGGT). Moreover, caesium ions will be used in place of magnesium ions as counterions, with sodium as the control.

A brief summary of this work and potential future studies will be given as an epilogue in chapter 5.

Chapter 2

Molecular Simulation and Dynamics

Investigations of DNA systems with implementation of other biological or pharmaceutical compounds (such as polyamines and proteins) include the studies of how these compounds react with the DNA macromolecule (*chemically*), how they affect the stability of DNA conformations, and how they invoke structural transitions (*physically*). It is particularly difficult to envisage experimentally the physical influences of external compounds on DNA because the changes are due to quantum mechanical effects and most of the changing processes are complete in the regime of 100ps to a few ns; even the most powerful slow-motion cameras cannot capture the processes. This is why molecular simulation, or molecular modelling, comes into play.

Molecular modelling is useful in calculating various energies and forces of structural conformations. Among various types of molecular modelling methods, molecular dynamics (MD) simulation is one of the most useful for analysing DNA conformational transitions, as it calculates not only the energetics and forces, but also the trajectories and velocities of molecules in the massive system. It is done, totally *classically*, by integrating the Newton's equations of motion $\vec{F}_i = m_i\vec{a}_i$ with the force being a conservative one, i.e. $\vec{F}_i = -\nabla_i V$.

This chapter is dedicated to explain the technicalities of molecular modelling. Starting from the AMBER force field, we will move onto the energy minimisation procedures and the MD methods. The chapter will conclude with the schemes followed to evaluate Coulomb potential which is divergent as an infinite series in the real space.

2.1 Simulation approaches

2.1.1 Quantum mechanical approaches

Force field calculations done using molecular simulations are very important before other physical quantities could be calculated; and the parameters for force field calculations can be determined theoretically by solving the time-independent many-body Schrödinger equation. However, solving the Schrödinger equation analytically is very difficult for two-electron system, and is impossible for systems any larger. Therefore, suitable approximations must be made to numerically obtain the results. There are several theories, including the Hartree and Hartree-Fock methods and the Density Functional Theory ("DFT"), which could help alleviate this problem. However since they are not in the scope of our studies we will describe them briefly here, without going into the technical details of them.

The Hartree method is an iterative scheme based on building the many-electron Hamiltonian upon the ignorance of electronic correlation and that each electron is only affected by the averaged-out influence of other electrons. It is known now as an inaccurate theory because of the violation of the generalised Pauli exclusion principle. The Hartree-Fock method is a modified version of the Hartree method by introducing the Slater determinant to account for the exchange of two fermions so that Pauli exclusion principle is obeyed. This method solves the time-independent many-body Schrödinger equation for wave functions using the variational principle, with the solution expressed in terms of linear combinations of *basis functions*. The DFT is an *exact* theory devised by Walter Kohn and Lu Jeu Sham [31] in 1965 which makes it distinctively advantageous over the two aforementioned theories.

2.1.2 Semi-empirical methods

Though the use of the density functional theory has been proven to be very powerful and accurate in producing the complicated wavefunction of the multi-particle systems, its computational cost increases significantly with the number of particles in the ensemble. As a trade-off, semi-empirical quantum chemistry methods had been developed to give a more practical simulation method.

Semi-empirical methods are approximation methods. They follow similar ideas as in the Hartree-Fock theory, but they simplify the problem by neglecting the most computationally costly terms in the Hamiltonian, while keeping the molecular orbitals to be expressed as the linear combinations of individual

atomic orbitals. The missing terms are compensated by an inclusion of other terms which are semi-empirically to reproduce observations made from experiments.

The most popular semi-empirical codes used throughout the years include: MINDO, MNDO [13] and PM3 [57]. They are implemented in the quantum chemistry package of MOPAC.

2.1.3 Empirical methods

For large molecular ensembles, such as proteins and DNA, the computational cost could still be too high even for the most powerful supercomputers. We then have to rely on empirical methods as the final cure for the problem.

In empirical methods, more approximations are taken to further simplify the system. The atoms in the molecules are treated classically as soft spheres. While the force fields governing the trajectories and velocities of the atoms are taken from the theories (such as Lennard-Jones 6-12 type potential) or approximated as classical physical phenomena (for instance, the compression and extension of a simple string—obeying the Hooke's Law), the force constants are determined by interpolation between single and double bond values using *observed* bond distances or directly taken from vibrational analyses of sp^2-hybridised atoms [11].

2.2 AMBER force field

The potential energy model of the AMBER molecular simulation package [11, 64] is a combination of several terms, which read:

$$\begin{aligned} V_{\text{total}} &= V_{\text{bonds}} + V_{\text{angles}} + V_{\text{torsions}} + V_{\text{ES}} + V_{\text{H-bonds}} \\ &= \sum_{\text{bonds}} K_b (b_{ij} - b_0)^2 + \sum_{\text{angles}} K_\theta (\theta_{ijk} - \theta_0)^2 \\ &\quad + \sum_{\text{torsions}} \frac{V_n}{2} \left[1 + \cos(n\tau_{ijkl} - \gamma)\right] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}} \right] \\ &\quad + \sum_{\text{H-bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right] \end{aligned} \tag{2.1}$$

Each of the five terms will be discussed in detail in the upcoming subsections.

2.2.1 Bond lengths and angles terms

The bond length term is the first summation in Eq. 2.1. It represents the bonding energy between all pairs of covalently bonded atoms i and j. The energy, according to Hopfinger and Pearlstein [29], can be expressed as a Hookean function, given that the bond length is not very much deviated from the equilibrium b_0. Hence

$$V_{\text{bonds}} = \sum_{\text{bonds}} K_b(b_{ij} - b_0)^2 \tag{2.2}$$

where K_b is the force constant for the specific bond between i and j which has the length b_{ij}, and b_0 is the *equilibrium* bond length of $i-j$. This Hookean model is based on the assumption that small deviations about the energy minimum could be fitted satisfactorily to a quadratic equation. Since this is a purely physical method of simulation and chemistry is disregarded, the code is unable to determine whether any two atoms are covalently bonded. Hence a reasonable way of tackling this problem is to assign a threshold distance for each bond type, beyond which the bond would be regarded as cleaved, and thus would not be take into account in the calculation of bond energies. We hereby regard all pairs of atoms, say A and B, to be "covalently bonded" if the distance between them is within the threshold value specific to the real single bond A–B. Bonds of higher orders are not considered since realistically single bonds are the longest between two atoms, i.e. $R_{A-B} > R_{A=B} > R_{A\equiv B}$. If the distance between A and B is greater than the threshold value for single bond, the chance of them being bonded is low. There is no possibility for them to be multiply-bonded.

The next summation term in Eq. 2.1 is the angular term which represents the variation of energy through the change in the covalent bond angles subtended by all possible combinations of three covalently bonded atoms $i-j-k$. This "angular energy", as given in the same paper, has the same form as the "bond length energy" which is a Hookean function [29]. Thus,

$$V_{\text{angles}} = \sum_{\text{angles}} K_\theta(\theta_{ijk} - \theta_0)^2 \tag{2.3}$$

where K_θ is the force constant for the specific bond angle subtended by the covalently bonded atoms i, j and k, and θ_0 is the *equilibrium* bond angle of $i-j-k$.

2.2.2 Torsion angles term

The third summation in Eq. 2.1 is the torsion angle term. As its name implies, it represents the energy related to the torsion angle subtended by four covalently bonded atoms $i-j-k-l$. This term is different from the previous two terms, in a sense that it must be valid throughout the 2π angle since it is related to the rotation around the bond $j-k$ [29]. This term takes the form

$$V_{\text{torsions}} = \sum_{\text{torsions}} \frac{V_n}{2}\left[1+\cos(n\tau_{ijkl}-\gamma)\right] \tag{2.4}$$

where V_n is the height of the energy barrier encountered in the course of rotation, n the periodicity, τ_{ijkl} the torsional angle subtended by the covalently bonded atoms i, j, k and l, and γ a phase angle. By taking this form, the "torsional energy" is ensured to attain its first maximum at the angle $\tau_{ijkl} = \frac{\gamma}{n}$.

2.2.3 Van der Waals and Coulomb terms

The fourth and the fifth terms in Eq. 2.1 accounts for non-bonding forces, where the fourth term includes contributions from van der Waals' and Coulomb interactions. This term reads

$$\begin{aligned} V_{\text{ES}} &= V_{\text{VdW}} + V_{\text{Coulomb}} \\ &= \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right] + \sum_{i<j}\frac{q_i q_j}{\varepsilon R_{ij}} \\ &= \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] \end{aligned} \tag{2.5}$$

, where the van der Waals term takes the form of Lennard-Jones potential (or 6-12 potential), with A_{ij} and B_{ij} being constants dependent on the atom types of i and j. R_{ij} is the distance between atoms i and j, and ε is a suitable dielectric constant. The limits of the summations are set to be $i < j$ so as to avoid double counting. q_i and q_j are the partial atomic charges on the atoms i and j.

2.2.4 Hydrogen bond term

The fifth term accounts for hydrogen bonds between atoms. Hydrogen bonds are very important non-bonding forces in biological molecules like proteins and nucleic acids. In DNA, there are two hydrogen bonds within a pair of purines (for instance, A = T) and three within a pair of pyridines (for instance, C ≡ G).

In AMBER this term takes a similar form to the van der Waals term, only that it is a 10-12 potential (rather than a 6-12 Lennard-Jones), which reads:

$$V_{\text{H-bonds}} = \sum_{i<j} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right] \tag{2.6}$$

where C_{ij} and D_{ij} are parameters dependent on the atom types of the electron donor and receptor respectively.

2.3 Molecular Simulation

The potential function obtained using the AMBER force field, as explained in the previous section, is a very important prerequisite for the dynamical simulation of a molecular system. The following subsections serve to outline briefly the simulation methods adopted in the computational tools used in this project.

2.3.1 Energy minimisation

Energy minimisation is a process in which the positions of the atoms in the systems are adjusted according a specified force field (for example, AMBER) iteratively, in order for the system to attain a minimum in the total potential energy. This method is a "local" method, i.e. potential barriers cannot be surmounted and only local minimum could be attained. Thus the initial position of the atoms determines the location of minima.

It is very crucial that energy minimisation is carried out before any molecular dynamics is taken place. This is because the initial positions of the atoms in the system, usually created assuming a perfect structure or taken from experimental results such as X-ray crystallography, might have potential defects such as clashing or unphysical bonding angles. These could contribute to a very massive potential energy and thus a large initial velocity which rips the system apart during the molecular dynamics simulations.

The minimisation scheme adopted in AMBER is described below.

2.3.1.1 Method of steepest descent

As its name implies, this method moves the atoms in the direction of the largest negative slope of the potential. Thus, mathematically,

$$\vec{X}^{(n+1)} = \vec{X}^{(n)} - \alpha \nabla V(\vec{X}^{(n)}) \tag{2.7}$$

where $\vec{X}^{(n)}$ is the conformation at step n , $\vec{X}^{(n+1)}$ the conformation after one step of adjustment, α the step length, and ∇V the gradient of the potential energy [40]. The convergence is linear and is thus slow. Hence this method is usually used primarily for the removal of any bad steric contacts in the initial configuration, and it is commonly used with the following method which is more efficient.

2.3.1.2 Method of conjugate gradient

The method of conjugate gradient takes the approximation that near the minimum the potential could be represented as a quadratic function. This method is much more efficient in searching the direction of movement in that it uses information from all previous steps. However, this method is less robust than the method of steepest descent, and requires a good initial position. Hence, this method is popularly used following steepest descent; and the two form a nice pair.

2.4 Molecular Dynamics

Molecular dynamics (MD) is a computational approach to simulate the *motions* of such molecular systems as gases, liquids and solids, i.e. their *time-dependent* positions, velocities and orientations [24]. This section is dedicated to the MD protocols and methods used in this project and how issues caused by long-ranged interactions are addressed.

2.4.1 MD protocols

The post-minimisation co-ordinates are then taken as the input co-ordinates for the MD runs. But since the minimisation process is static, initial velocities have to be assigned to individual atoms in the system — It is done through random assignment according to a normal distribution, at a reasonably low temperature. Then the system is steadily heated up to the designated tem-

perature, which could be done by slowly shifting the mean of the Gaussian distribution up to the desired temperature.

Once the target temperature is reached, equilibration process can be carried out, in order to let the temperature and the total energy stabilise.

2.4.2 MD methods

This subsection dedicates to outline the methods of molecular dynamics simulation adopted in the AMBER package (release 12).

2.4.2.1 Integration schemes for equations of motion

Since the systems are considered as fully classical, the aforementioned observables are determined solely by the familiar *Newton's equations of motion*, i.e.

$$\begin{aligned} \frac{d^2\vec{x}_i}{dt^2} &= \frac{1}{m_i}\vec{F}_i \\ \vec{F}_i &= -\nabla_{\vec{x}_i} V(\vec{X}) \end{aligned} \tag{2.8}$$

where $\vec{x}_i$ is the position vector of the i-th atom, m_i the mass of the i-th atom, $\vec{F}_i$ the *external* force acting on the i-th atom, and $\nabla_{\vec{x}_i}$ is the gradient operator with respect to the co-ordinates of the i-th atom.

Numerically, the Newton's equations of motion (eq. (2.8)) can be solved by using finite difference method. However since the evaluation of the force $\vec{F}_i$ is the most computationally expensive part throughout the simulation, integration schemes should be carefully chosen so that it will not be performed more than once per time step [61].

As a trade-off among computational cost, accuracy and numerical stability, the leapfrog scheme is popularly chosen over other integration schemes such as Runge-Kutta, Gear, Verlet and Beeman [61].

The first step of the leapfrog scheme involves the evaluation of the acceleration $\vec{a}_i$, i.e.

$$\vec{a}_i(t) = \frac{d\vec{v}_i(t)}{dt} = \frac{1}{m_i}\vec{F}_i(\{\vec{x}_i(t)\}) \tag{2.9}$$

where $\vec{v}_i(t)$ is the velocity of atom i at time t, m_i is the mass of atom i, and $\vec{F}_i(\{\vec{x}_i(t)\})$ is the force acting on atom i (which is at the position $\vec{x}_i$) at time t. The force is given, as usual, in the form of (2.8).

With the acceleration on hand, the next step is integrate the expression to arrive

at the velocity $\vec{v}_i$, which is thus done:-

$$\vec{v}_i(t_n + \frac{\Delta t}{2}) = \vec{v}_i(t_n - \frac{\Delta t}{2}) + \vec{a}_i(t_n)\Delta t + \mathcal{O}(\Delta t^3) \tag{2.10}$$

where Δt is the temporal step size, and $\mathcal{O}(\Delta t^3)$ is the usual "Big O" notation indicating the truncation of any term from third order in Δt.

In a similar sense, the position vector can be found by

$$\vec{x}_i(t_n + \Delta t) = \vec{x}_i(t_n) + \vec{v}_i(t_n + \frac{\Delta t}{2})\Delta t + \mathcal{O}(\Delta t^3) \tag{2.11}$$

In order for the leapfrog scheme to be stable, the time step, Δt, must be sufficiently small. For simulations of macromolecular systems such as nucleic acids, common values of Δt taken are of the order of 1 fs [61].

2.4.2.2 Temperature coupling

The system can be coupled to a heat bath, fixed at a reference temperature T. This is done by introducing stochastic and frictional (damping) terms in the Newton's equations of motion (eq. (2.8)), i.e.

$$\begin{cases} \dfrac{d\vec{x}_i}{dt} = \dfrac{1}{m_i}\vec{p}_i \\ \dfrac{d\vec{p}_i}{dt} = \vec{F}_i - \gamma_i\vec{p}_i + \vec{f}_i \end{cases} \tag{2.12}$$

where $\vec{x}_i$ is the positional vector of atom i, $\vec{p}_i$ is the linear momentum of the i-th atom, $\vec{F}_i$ is the force acting on the i-th atom due to the interaction potential, γ_i is a friction coefficient, and $\vec{f}_i$ is a random force with dispersion σ_i related to the friction coefficient γ_i through

$$\sigma_i{}^2 = \frac{2m_i\gamma_i k_B T}{\Delta t} \tag{2.13}$$

where Δt is the time step used in the molecular dynamics for the integrating the equation of motion. This scheme is known as the Langevin dynamics [2] and is widely used as a thermostat in molecular dynamics. Langevin dynamics is a robust and accurate formalism which correctly produces the Maxwell-Boltzmann distribution of particle velocities, in turn allowing the system to explore the entire configuration space.

2.4.2.3 Pressure coupling

In a similar way, the system can be coupled to a constant pressure bath, adding a term, which accounts for the change in pressure [4, 61], into the equations of motion,

$$\left(\frac{dP}{dt}\right)_{\text{bath}} = \frac{P_0 - P}{\tau_P} \tag{2.14}$$

where τ_P is the time constant of the pressure coupling (analogous to that of temperature coupling), and P is the pressure of the system, and is defined by

$$P = \frac{2}{3V}(E_k - \Xi) \tag{2.15}$$

where V is the total volume of a simulation box, E_k is the total kinetic energy. Ξ is known as the virial, and is defined as $\Xi = -\frac{1}{2}\sum_{\text{pairs}(i,j)}^{N} \vec{x_{ij}} \cdot \vec{F_{ij}}$, where $\vec{x_{ij}} = \vec{x_i} - \vec{x_j}$ is the displacement vector from the i-th atom to the j-th atom, and $\vec{F_{ij}}$ is the force on atom i due to atom j [4, 61].

The equation of motion is thus modified into

$$\frac{d\vec{x}}{dt} = \vec{v} - \frac{\beta(P_0 - P)}{3\tau_P}\vec{x} \tag{2.16}$$

where β is isothermal compressibility. This results also in a proportional scaling of co-ordinates x and box dimensions l (cubic box assumed, for simplicity), such that:

$$\begin{aligned} x &\longrightarrow \sqrt[3]{1 - \frac{\Delta t}{\tau_P}(P_0 - P)}\, x \\ l &\longrightarrow \sqrt[3]{1 - \frac{\Delta t}{\tau_P}(P_0 - P)}\, l \end{aligned}$$

2.4.3 Long-range Coulomb force

Whether the atoms inside the system are charged or neutral (which will still be polarised), the electrostatic potential is still a very important yet difficult term to evaluate. This is because in the absence of a "distance-dependent" dielectric, the potential takes the form $V(r) \propto \frac{1}{r}$, whose infinite series is divergent (and hence the name "long-range force"). Since computers cannot handle infinities numerically, special techniques have to be used to remove those infinities and to make computation less costly. This section dedicates to two of such powerful techniques which are most popularly used.

2.4.3.1 Truncation by a "switching function"

Truncation is an approximation method which neglects any long-range potential or force beyond the "cut-off distance (or radius)", a user-defined distance, from the atom. The rationale behind it is that the contribution of those long-range terms, though not completely zero, is so small that they can be neglected without loss of generality. Traditionally, the truncation method works by introducing a step function as the prefactor to the series, i.e.

$$\theta(r) = \begin{cases} 1 & \text{if } r < R_C, \\ 0 & \text{if } r > R_C. \end{cases} \tag{2.17}$$

We immediately see that there is a discontinuity at $r = R_C$. This could increase the kinetic energy of the atoms and hence the system's temperature.

A possible way to alleviate this problem is to modify the step function a little and smoothen the distribution about R_c, thus

$$S(r) = \begin{cases} 1 & \text{if } r < R_S, \\ \dfrac{(R_C - r)^2(R_C + 2r - 3R_S)}{(R_C - R_S)^3} & \text{if } R_S < r < R_C, \\ 0 & \text{if } r > R_C. \end{cases} \tag{2.18}$$

This "switching function" satisfies four conditions governing the behaviours of itself and its first derivative about $r = R_S$ and R_C, namely $S(R_S) = 1$, $S(R_C) = 0$, $\dfrac{dS(R_S)}{dr} = 0$ and $\dfrac{dS(R_C)}{dr} = 0$. R_S is another user-defined distance between which and R_C the interaction would be smoothened [61].

2.4.3.2 Ewald summation method

The Ewald summation method is a technique in evaluating long-range interactions exactly (as compared to the approximation approach of truncation discussed above). It is especially powerful when applied to system with PBCs.

The Coulomb interaction with PBC transforms from the second term in (2.5) into

$$V_{\text{Coulomb, PBC}} = \frac{1}{2}\sum_{i=0}^{N}\sum_{j=0}^{N}{\sum_{\vec{n}}}' \frac{q_i q_j}{\left|\vec{r_{ij}} + \vec{n}\right|} \tag{2.19}$$

where $\vec{r_{ij}} = \vec{r_i} - \vec{r_j}$ is the displacement vector from atom i to atom j, the sum over $\vec{n} = (n_x L_x, n_y L_y, n_z L_z)$ is over all the lattice points, and $(n_x L_x, n_y L_y, n_z L_z) \in \mathbb{N}$. The "prime" in the $\vec{n}$-summation signifies that the term when both $i = j$ and $\vec{n} = \vec{0}$ are satisfied, should be omitted.

The trick of Ewald summation is to split the interaction into two parts, namely the "short-range" and the "long-range"; and the "self-interaction" is removed by subtraction [23, 67]. Thus,

$$V_{\text{Coulomb, PBC}} = V^S + V^L - V^{\text{self}} \tag{2.20}$$

where $V^S = \frac{1}{2}\sum_{i=0}^{N} q_i \varphi^S_{[i]}(\vec{r}_i)$ and $V^L = \frac{1}{2}\sum_{i=0}^{N} q_i \varphi^L_{[i]}(\vec{r}_i)$. The potential terms in the summations are defined as $\varphi_{[i]}(\vec{r}_i) \equiv \sum_{\vec{n}}{\sum_{j=1}^{N}}' \iiint \frac{\rho_j(\vec{r}')}{|\vec{r}_i - \vec{r}' + \vec{n}L|} d^3\vec{r}'$. The charge distribution, ρ_j, is split into the "short-range" and the "long-range" in the similar way, by the introduction of a virtual Gaussian (counter-)charge distribution. Then

$$\begin{aligned} \rho_j(\vec{r}) &= \rho_j^S(\vec{r}) + \rho_j^L(\vec{r}) \\ &= q_i\Big\{\delta(\vec{r}-\vec{r}_i) - G_\sigma(\vec{r}-\vec{r}_i)\Big\} + q_i G_\sigma(\vec{r}-\vec{r}_i), \end{aligned} \tag{2.21}$$

where the "compensating charge" takes the form of $G_\sigma(\vec{r}) \equiv \frac{1}{(2\pi\sigma^2)^{3/2}} \exp\left[-\frac{|\vec{r}|^2}{2\sigma^2}\right]$ with σ being the standard deviation of the distribution.

The first term is called the "short-range" term because the integration of it gives the complementary error function $\text{erfc}(\vec{r}-\vec{r}_i) = 1 - \text{erf}(\vec{r}-\vec{r}_i)$ which is fast-converging and has the limit tending to zero at infinity.

The second term is "long-ranged" because integrating it results in the error function which goes to unity at infinity, hence

$$\begin{aligned} \varphi^L(\vec{r}) &= \sum_{\vec{n}}{\sum_{j=1}^{N}}' q_j \iiint \frac{G_\sigma(\vec{r}-\vec{r}_i)}{|\vec{r}_i - \vec{r}' + \vec{n}L|} d^3\vec{r}' \\ &= \sum_{\vec{n}}{\sum_{j=1}^{N}}' \frac{q_j}{|\vec{r}_j - \vec{r}' + \vec{n}L|} \text{erf}\left(\frac{|\vec{r}_j - \vec{r}' + \vec{n}L|}{\sqrt{2}\sigma}\right). \end{aligned} \tag{2.22}$$

Since the summand is converging to a definite but non-zero number, the sum can only be computed in the reciprocal space through Fourier transforming the whole expression [23]. The result (through backward Fourier transforming into the real space) is

$$\varphi^L(\vec{r}) = \frac{4\pi}{V} \sum_{\vec{k}\neq\vec{0}} \sum_{j=1}^{N} \frac{q_j}{k^2} \exp\left(i\vec{k}\cdot(\vec{r}-\vec{r}_j)\right) \exp\left(-\frac{\sigma^2 k^2}{2}\right) \tag{2.23}$$

where $\vec{k}$ is the reciprocal vector of $\vec{r}$, and V is the volume of the supercell. Putting these back to eq. (2.20), we have

$$\begin{aligned} V_{\text{Coulomb, PBC}} &= \frac{1}{2}\sum_{\vec{n}}\sum_{i=1}^{N}\sum_{j=1}^{N}{}' \frac{q_i q_j}{\left|\vec{r}_j - \vec{r}' + \vec{n}L\right|}\operatorname{erfc}\left(\frac{\left|\vec{r}_j - \vec{r}' + \vec{n}L\right|}{\sqrt{2}\sigma}\right) \\ &\quad + \frac{2\pi}{V}\sum_{\vec{k}\neq\vec{0}}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{k^2}\exp\left(-\frac{\sigma^2 k^2}{2}\right)\left|S(\vec{k})\right|^2 \\ &\quad - \frac{1}{\sqrt{2\pi}\sigma}\sum_{i=1}^{N} q_i^2 \end{aligned} \tag{2.24}$$

where $S(\vec{k}) \equiv \sum_{i=1}^{N} q_i \exp(i\vec{k}\cdot\vec{r}_i)$ is the "structure factor" which depends on the charge distribution [23], and the final term accounts for the self-energy.

Chapter 3

Simulations of d(A)$_{20}$ and d(A-T)$_{20}$

3.1 Notation of DNA sequences

In general DNA sequences will be expressed in the following form: poly-d(B)$_n$ · poly-d(b)$_n$, where "poly-d" indicates that the nucleic acid is a DNA (an RNA would have the label "poly-r"). "B" is one of the bases (A, T, C or G) and "b" is its complementary (For instance, if B = A then b = T, vice versa, etc). The subscript n indicates the length of the DNA strands in number of base pairs. The "·" indicates that the two strands are bonded by hydrogen bonds. Thus in the previous example, poly-d(B)$_n$ · poly-d(b)$_n$ would mean that the DNA duplex consists of a strand composed by n "B"-type base, while hydrogen-bonded with another strand comprising the same number of "b"-type base ("B"'s complement). It is customary to omit the second strand in notation. For example, d(A)$_{20}$ ≡ d(A)$_{20}$ · d(T)$_{20}$. This is because the Watson-Crick pairing rule [63] is understood.

In the case of heterogeneous strands, the notation will take the form poly-d(B$_1$B$_2$ ··· B$_n$) · poly-d(b$_1$b$_2$ ··· b$_n$), where b$_1$ is the complementary of B$_1$ and b$_2$ is the complementary of B$_2$, etc. There is also another possibility of heterogeneous sequences that they could be repetitive, having repetition motifs of two (or more) bases, in which case the notation would be poly-d(B$_1$B$_2$ ··· B$_n$)$_m$ · poly-d(b$_1$b$_2$ ··· b$_n$)$_m$, where m is the length of the strand, i.e. number of base pairs in the section — not the number of repetitions of the motif, which is m divided by the length of the motif.

3.2 Systems and Methodology

The simulations performed can be broadly divided into two types, according to the different subjects of study. The first type involves the studies of the effect of spermidine molecules on the conformational stability of DNA. This set of simulations are performed under the same environment, but with different numbers of spermidine molecules in the system. The second type involves the studies of the effect of different ions on DNA conformations. This set of simulations are the repetition of the aforementioned series but with different ions.

Table 3.1 shows the code of the sequences which will be discussed in this chapter, their "formal" nomenclature, the starting configuration, the types of counterions used, number of spermidine molecules inside individual systems and their initial placement.

For the codes of sequences, they all take the form of "XXY-Zzn", where "XX" is the acronym for a specific sequence (XX = "20" for $d(A)_{20}$, and XX = "AT" for $d(A\text{-}T)_{20}$), "Y" the starting configuration ("A" for an A-start and "B" for a B-start), "Zz" the counterion type chosen ("Na" for sodium and "Mg" for magnesium), and n the number of spermidine molecules.

The initial placement of spermidine is worth a more detailed explanation. The position of a spermidine molecule can be generalised into two categories — inside a DNA groove and outside the DNA structure. For the molecule buried inside the grooves, "M" means that it is in the *major* groove, while an "m" denotes the position in the *minor* groove. For the molecule to be *outside* the helix, its orientation can be further divided into two groups: aligning and bridging. Aligning means that the molecule aligns itself such that it is approximately parallel to the phosphate backbone, and is given a code "p". Bridging means that the molecule orientates such that it is somewhat perpendicular to the backbone, either having both amino end-groups near the backbone and bridging across a groove ("pMp" for bridging across a *major* groove and "pmp" for a *minor* groove), or having the end-groups near the grooves and bridging across a backbone ("Mpm" or "mpM" according to the $5' \rightarrow 3'$ sense).

For all systems, the initial configuration of each sequence was produced using the "Nucleic Acid Builder (NAB)" code, which is a tool in the AMBER package. The types and positions of the created atoms were all saved as a "PDB" files, which was then loaded into xLEAP with the parameters for the force field called "ff99SB". We chose this specific force field because this is the only force field which has the parameters for the bivalent magnesium ion.

Code	Sequence	Starting config.	Ions	# of SPD	SPD placement
20A-Na0	$d(A)_{20}$	A	Na^+	0	–
20A-Na1	$d(A)_{20}$	A	Na^+	1	pmp
20A-Na2	$d(A)_{20}$	A	Na^+	2	pMp/pMp
20A-Na3	$d(A)_{20}$	A	Na^+	3	pmp/pmp/Mpm
20A-Na4	$d(A)_{20}$	A	Na^+	4	pMp/pmp/pMp/pmp
20A-Mg0	$d(A)_{20}$	A	Mg^{2+}	0	–
20A-Mg1	$d(A)_{20}$	A	Mg^{2+}	1	pmp
20A-Mg2	$d(A)_{20}$	A	Mg^{2+}	2	pMp/pMp
20A-Mg3	$d(A)_{20}$	A	Mg^{2+}	3	pMp/pMp/pMp
20A-Mg4	$d(A)_{20}$	A	Mg^{2+}	4	mpM/pMp/m/pmp
20B-Na0	$d(A)_{20}$	B	Na^+	0	–
20B-Na1	$d(A)_{20}$	B	Na^+	1	M
20B-Na2	$d(A)_{20}$	B	Na^+	2	pMp/M
20B-Na3	$d(A)_{20}$	B	Na^+	3	pmp/M/mpM
20B-Na4	$d(A)_{20}$	B	Na^+	4	pmp/M/M/M
20B-Mg0	$d(A)_{20}$	B	Mg^{2+}	0	–
20B-Mg1	$d(A)_{20}$	B	Mg^{2+}	1	M
20B-Mg2	$d(A)_{20}$	B	Mg^{2+}	2	pMp/M
20B-Mg3	$d(A)_{20}$	B	Mg^{2+}	3	m/pmp/p
20B-Mg4	$d(A)_{20}$	B	Mg^{2+}	4	m/pMp/M/pMp
ATA-Na0	$d(A\text{-}T)_{20}$	A	Na^+	0	–
ATA-Na1	$d(A\text{-}T)_{20}$	A	Na^+	1	pmp
ATA-Na2	$d(A\text{-}T)_{20}$	A	Na^+	2	pMp/m
ATA-Na3	$d(A\text{-}T)_{20}$	A	Na^+	3	pMp/m/pMp
ATA-Na4	$d(A\text{-}T)_{20}$	A	Na^+	4	M/pmp/pMp/M
ATA-Mg0	$d(A\text{-}T)_{20}$	A	Mg^{2+}	0	–
ATA-Mg1	$d(A\text{-}T)_{20}$	A	Mg^{2+}	1	pmp
ATA-Mg2	$d(A\text{-}T)_{20}$	A	Mg^{2+}	2	pmp/pmp
ATA-Mg3	$d(A\text{-}T)_{20}$	A	Mg^{2+}	3	M/pmp/M
ATA-Mg4	$d(A\text{-}T)_{20}$	A	Mg^{2+}	4	Mpm/pMp/m/M
ATB-Na0	$d(A\text{-}T)_{20}$	B	Na^+	0	–
ATB-Na1	$d(A\text{-}T)_{20}$	B	Na^+	1	M
ATB-Na2	$d(A\text{-}T)_{20}$	B	Na^+	2	pmp/M
ATB-Na3	$d(A\text{-}T)_{20}$	B	Na^+	3	pmp/M/m
ATB-Na4	$d(A\text{-}T)_{20}$	B	Na^+	4	M/pmp/m/M
ATB-Mg0	$d(A\text{-}T)_{20}$	B	Mg^{2+}	0	–
ATB-Mg1	$d(A\text{-}T)_{20}$	B	Mg^{2+}	1	M
ATB-Mg2	$d(A\text{-}T)_{20}$	B	Mg^{2+}	2	pMp/M
ATB-Mg3	$d(A\text{-}T)_{20}$	B	Mg^{2+}	3	pMp/M/m
ATB-Mg4	$d(A\text{-}T)_{20}$	B	Mg^{2+}	4	M/M/pmp/m

Table 3.1: Summary of simulated DNA-cation systems in this chapter. Sequence for the first strand (the Watson strand) is given while that for the second (Crick) strand is assumed to be understood to be the complement.

Whenever needed, spermidine molecules were added to the system manually using XLEAP. The positions of the molecules were determined and adjusted by observation. Then the system, being still electrically negative, was neutralised by introducing suitable counterions. The process of counterion addition was totally automated, and the positions of the ions were determined by the minima of the electric field produced by the DNA and the spermidine molecules. The order of doing these were proven to be extremely crucial, as the current code only allows users to add counterions until the system comes to a neutral overall charge, but not any designated charge (in electron charges). Hence if the counterions were introduced to the system prior to the addition the spermidine molecules, the system would never become neutral. This would in turn produce a disastrous result when the system is taken to MD simulation. Because of the imposition of periodic boundary conditions on the system, any non-neutral charge would be mirror-imaged in all directions infinitely, and the total electric charge of the simulated system is thus infinite and cannot be handled properly. At last the whole neutralised system was solvated with TIP3P model water molecules using xLEAP. The buffer was set uniformly for all systems to be 8.0Å, meaning that the minimal distance between any of the atoms in the DNA molecule and the edge of the simulation cell is 8.0Å. The size of simulation box was reduced for higher simulation efficiency by taking the form of a truncated octahedron.

The following subsections will be dedicated to detailed discussions about the minimisation and MD simulation protocols used throughout the whole project. These are similar to the ones adopted previously by other researchers in the same field [49, 54].

3.2.1 Minimisation procedure

Now that the system had been constructed, molecular dynamics was ready to take place. However, since the spermidine molecules were placed by hand there were inevitably unfavourable configurations (for instance, unphysical clashing of atoms, or spermidine molecules not sitting at the local minima of the electric field, etc.). Therefore an energy minimisation was much needed to relax the system to both a physically and energetically stable condition.

In our studies, we have split the minimisation into two parts. The first part involved the minimisation of the solvent environment, in which the positions of DNA and spermidine atoms were fixed using a very large force constant of $500\,\mathrm{kcal\,mol^{-1}\,\AA^{-2}}$ while the water molecules were allowed to move freely to attain the lowest energy configuration. Moreover, a cut-off of 16Å was

adopted, meaning that beyond the 16Å threshold the electrostatic force was truncated. In this stage, 500 steps of minimisation were undertaken, with 250 steps under steepest-descent and the remaining half under conjugate gradient.

The second part of minimisation involved the minimisation of the whole system, in which the DNA and spermidine atoms were released from the previous constraints and allowed to freely move. In this round the length was increased to fivefold the previous part, with 1000 steps of steepest-descent method followed by 1500 steps of conjugate gradient.

3.2.2 Heating of system

The minimisation procedure was done statically, i.e. as though the system was at absolute zero. However at 0K nothing would be moving (except small local vibrations, a.k.a. the "zero-point motion", because of the zero-point energy, in accordance with the Heisenberg uncertainty principle [28]) and molecular *dynamics* would become totally meaningless. Therefore to reproduce both physically and physiologically realistic results, the system temperature must be raised up to temperatures similar to body temperature, or at least room temperature. For this reason we chose the simulation temperature to be 300K.

The heating procedure was performed as a molecular dynamics using the Langevin thermostat [4] to gradually bring up the temperature from 0K to 300K in 20ps (20000 1fs-timesteps).

For this particular heating protocol, we chose the collision frequency to be 1ps^{-1} in order for the Langevin thermostat to perform properly. Moreover, the SHAKE algorithm was used to constrain the hydrogen-related bonds, in turn reducing the efforts made for calculation of each timestep.

Finally, the DNA molecule was held, like in the solvent minimisation procedure, by a force constraint. But this time the force constant for the constraint was considerably smaller — only $10\,\text{kcal}\,\text{mol}^{-1}\,\text{Å}^{-2}$. The purpose of holding the DNA here was to allow the solvent to heat better without the potential (excessive) interactions with the DNA, while ensuring the system to fully prepare itself for the real molecular dynamics.

3.2.3 Molecular dynamics protocol

Molecular dynamics with constraints can be broadly divided into two families: the canonical ensemble (NVT) MD and the isothermal-isobaric (NPT) MD.

The canonical ensemble molecular dynamics, as the short-form suggests, holds

the total number of particles N, total volume V and the temperature T constant. As a result from the first law of thermodynamics, this implies that the pressure P in the ensemble has to be changed during the molecular dynamics simulations.

The isothermal-isobaric molecular dynamics conserves the number of particles N, the temperature T and the pressure P — hence the term *isobaric*. In our simulations, the control of the pressure inside the system is crucial as the pressure inside biological system is more or less a constant. As such, for the same reason as above the volume V of the system was permitted to vary with time. Thus the bond lengths were also allowed to change, according to the scheme described in subsection 2.4.2.3. The relaxation time was taken to be 2.0ps (within the range of 1.0ps and 5.0ps as suggested by AMBER [8]) and it was proven that the results from the simulations were physical and reasonable — no energetic artefacts were present throughout the simulations.

Apart from the change from NVT molecular dynamics as in the heating stage to the NPT in the current one, the major parameters were kept the same, so as to allow a seamless transition between the two stages. However, the position and the velocities of atoms were now read from the previous stage of molecular dynamics. This is essential since we were doing multi-nanosecond simulations — the runtime for the whole simulation was bound to exceed the maximum node reservation time of 48 hours within the N8 High Performance Computing facility "Polaris" (jointly operated by 8 institutions in the North of England including University of Leeds and University of York) where all the simulation runs in this project were performed. We chose to split the simulation into many shorter equal-length runs, which individually were tested to be normally-terminable within the reservation time (typically 40 hours per run). Each run was 1ns (in $1,000,000$ timesteps, i.e. 1fs per timestep) and a total of 40 runs were performed for each of the simulations to produce 40ns of molecular dynamics simulation results. At the end of the 40 runs the 40 trajectory data files were combined into one for easier analysis.

3.2.4 Analysis techniques

The first step to analyse the massive data was not directly jumping into the DNA parameters and see how they change, but rather to perform a vital check of the convergence or the stability of the runs as there are numerous sources of abnormalities which each alone could lead to fatal errors and eventually meaningless data. These anomalies include sudden changes in temperature, energies and most importantly *pressure* (which should all be more-or-less con-

stant throughout since the system is isolated).

Given that the stability tests were all passed (hence credible results from the runs), the next step was to stitch together the 40 trajectory files into one, using the "cpptraj" utility. "cpptraj" is a powerful program within AMBERTOOLS, that does not only join the trajectory files together, but also strips away the water molecules which are useless (and posing visual hindrance) in analyses and calculates root-mean-squared deviations (RMSd) of user-specified backbone atoms from the initial configuration over time which could be plotted for better visualisation. The RMSd is a very useful measure of how the conformation of the DNA has changed as the simulation progressed.

Having acquired a rough idea from the RMSd data of how drastically the DNA molecule has changed its shape, we could then proceed to the finer conformational parameters which truly reflects each of the structural (helical, intra-base pair, inter-base pair, etc.) measurements, from which we would be able to tell precisely how close the DNA is towards one form and deviated from another. This was done by inputting the unified trajectories into the codes "Curve+" [35] and "Canal" (a sister code of Curve+), which calculate the parameters at each base pair every timestep. Due to edge effects the base pairs at both ends of the section had much larger deviation than the inner ones. Hence the data for the three base pairs on either end were discarded and the rest were averaged and plotted. Moreover, since the ensemble was massive and there were numerous interactions taking place, the data were expected to be extremely noisy. Thus a running-average of 100ps was taken, without loss of generality, for every graph for easier trend-reading.

3.3 Results and discussions

In this section, results from simulations of the sequences $d(A)_{20}$ and $d(A\text{-}T)_{20}$ are presented and discussed.

To give a primary and rough idea into how much the DNA structure has deviated from the canonical A- or B-form, during the course of simulation, the RMSd graphs will first be studied. Then for more in-depth insights, we will turn to eight of the conformational parameters, which include *inter*-base pair parameters (h-rise, h-twist), *intra*-base pair parameters (xdisp, incline) and groove parameters (*major* groove depth and width, *minor* groove depth and width).

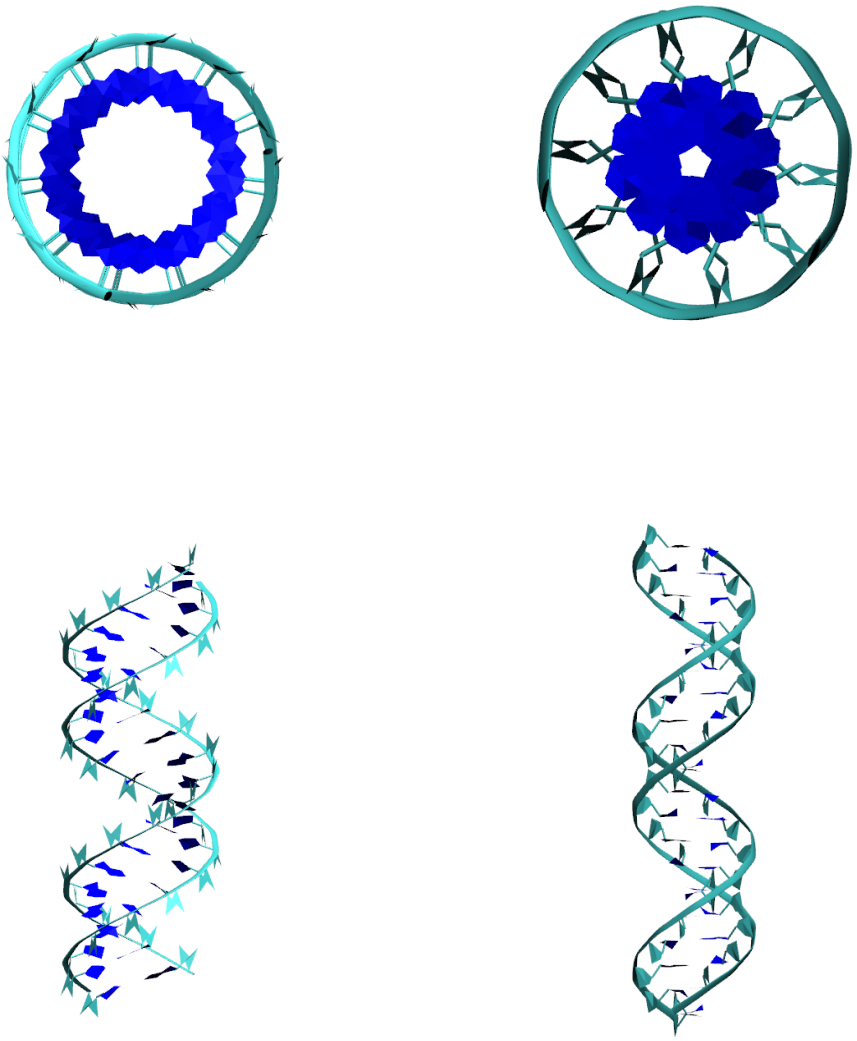


Figure 3.1: Ribbon representation of systems discussed (in canonical forms) in this chapter. Upper-left: top-down view of 20A. Upper-right: top-down view of ATB. Lower-left: side view of 20A. Lower-right: side view of ATB.

3.3.1 Spermidine-free systems

The set of spermidine-free systems consists of eight similar systems: 20A-Na0, 20B-Na0, 20A-Mg0, 20B-Mg0, ATA-Na0, ATB-Na0, ATA-Mg0 and ATB-Mg0.

Fig. 3.2 shows the snapshots of the MD simulation of the ATA-Na0 system at different time steps. It is clearly demonstrated from the top-down view that the DNA has transited from an A-form to B-form within the first nanosecond and stayed rather stably as a B-DNA for the rest of the simulation period.

The graphs of spermidine-free systems using sodium and magnesium as counterions showing instantaneous RMSd of individual system from canonical struc-

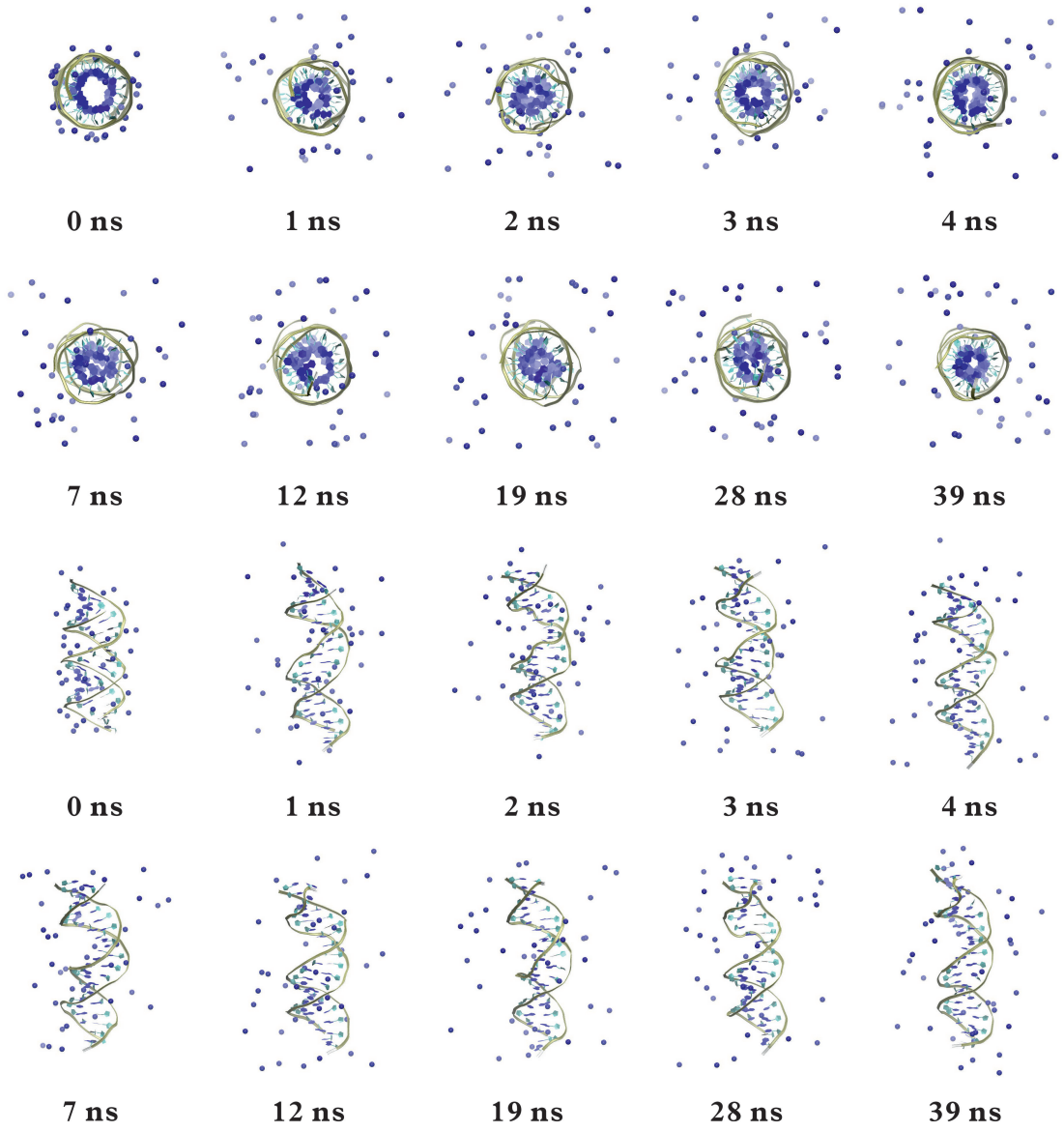


Figure 3.2: Snapshots (top-down and side view) of ATA-Na0.

tures are presented in Fig. 3.3. For the Na0 systems, it is anticipated from previous studies (and very well-established) that the DNA would eventually deviate from an A-form and approach a canonical B-form [49, 65]. And the discrepancy would be minimal as there were no spermidine molecules in the system to add "external force" on the DNA to make it change. This is rather obvious from Figure 3.3, the RMSd graphs for 20A and 20B systems, where right from the start the RMSd for A-start had been steadily larger than that for B-start, showing the system has deviated away from canonical A-form more than from B. However the RMSd for ATA and ATB are considerably closer to each other throughout than their counterparts, and they are mostly sandwiched between the 20A and 20B graphs, suggesting that the action of sodium on DNA is actually somehow sequence-specific. For sequences whose Watson and Crick strands are asymmetric (like $d(A)_{20}$) the effect is quite large and B-

form is *much* more favoured over A-form. But for symmetric sequences (like d(A-T)$_{20}$) the effect is diluted.

For the Mg0 systems, because of lack of previous data in support, it was rather difficult to back our prediction with firm experimental grounds. However from classical electrodynamics we assert that magnesium ions, being bivalent (doubly positive), should have more attraction towards the negative phosphate backbones and in turn increasing the overall stability by holding the strands in place. However it was not the case: the overall picture was quite chaotic for all the systems (20 and AT) and the general trend of the RMSd for them is all tending towards the high-end ($\approx 7\mathring{A}$), suggesting that all the systems were disturbed rather than stabilised.

Figures 3.4 to 3.7 show the histograms for base-pair parameters of four spermidine-free systems. They were produced from the time graphs (which are not shown here) and they give good insight into how structurally stable (by stability we intend to mean the extent to which the DNA system is held in place without changing its conformation) the DNA molecule is — the more bell-shaped (normally distributed) the stabler the system, the narrower the distribution the stabler the system.

From figures 3.4 and 3.6 for the sodium systems we see that all the eight graphs resemble the normal distribution fairly satisfactory, with a slight skew only; and their distributions are reasonably narrow. Moreover nearly all of the graphs have the peak around the canonical B-form values. These all suggest that the environment with the presence of sodium stably favours a B-form of DNA. It is worth mentioning that all graphs for the h-twist have an about 3° to 5° downward shift, which is probably due to the bending of the DNA molecules.

However the story is totally different for the magnesium systems, as can be observed from the remaining two figures, especially figure 3.5. Not only do the shapes of the graphs not resemble normal distributions, but also their extremely wide full width at half maximum (FWHM), especially for the incline which usually is within the range of 10 to 15Å. The FWHM for 20A-Mg0 was nearly 40Å, suggesting that the DNA system is highly unstable under the influence of magnesium.

Figures 3.8 to 3.11 show the histograms for base-pair parameters of four remaining spermidine-free systems. Here we observe that the histograms take the bell shape (with a little skew). This confirms the findings from the reading of the RMSd graphs that magnesium has much less effect on the DNA structures of the sequences with alternating AT than ones with only A on one strand

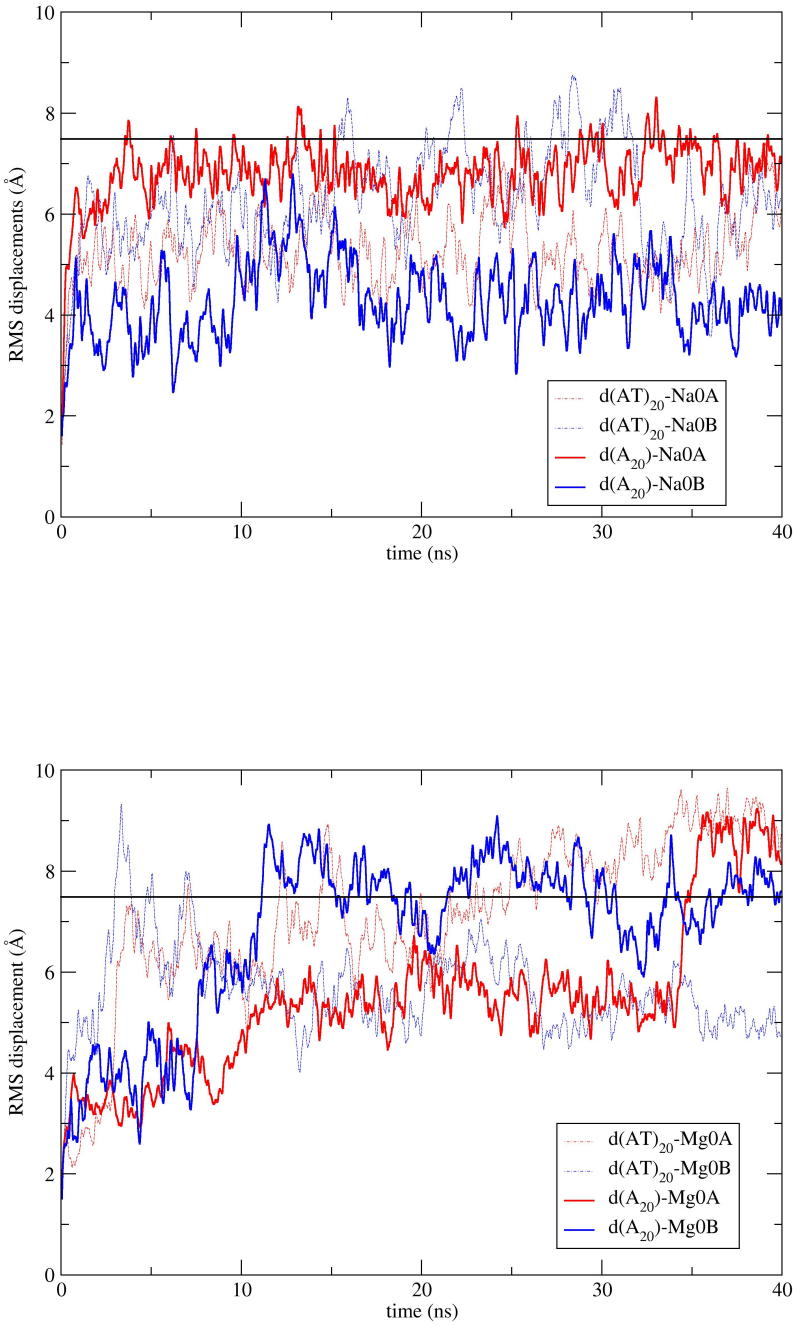


Figure 3.3: Graphs of SPD-free systems using sodium and magnesium as counterions showing instantaneous RMSd of individual system from canonical structures. Black horizontal line: RMSd of canonical A-DNA from B-DNA. (*similar treatments applied to all later RMSd graphs)

and T on the other.

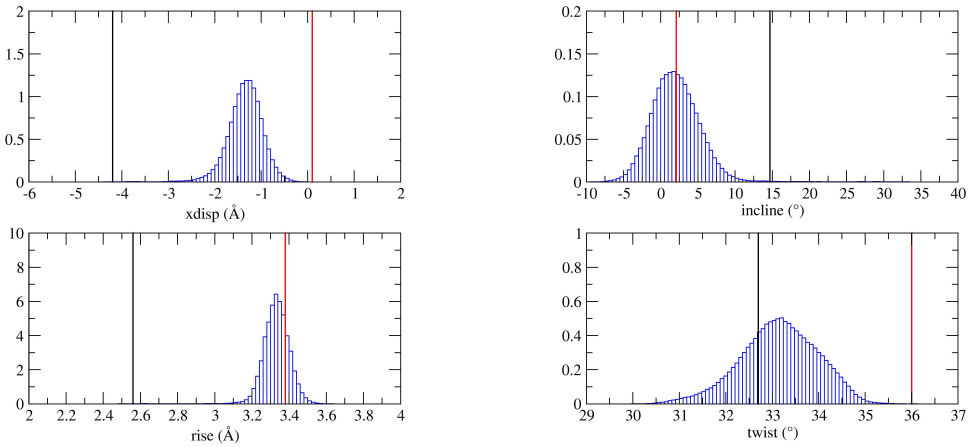


Figure 3.4: Base pair parameters histograms of 20A-Na0. Black vertical line: Typical value for canonical A-DNA. Red vertical line: Typical value for canonical B-DNA. (*similar treatments applied to all later histograms)

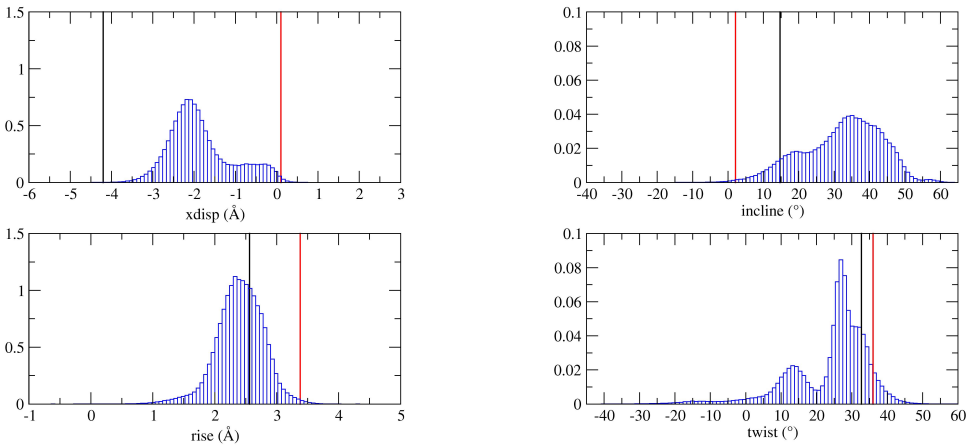


Figure 3.5: Base pair parameters histograms of 20A-Mg0.

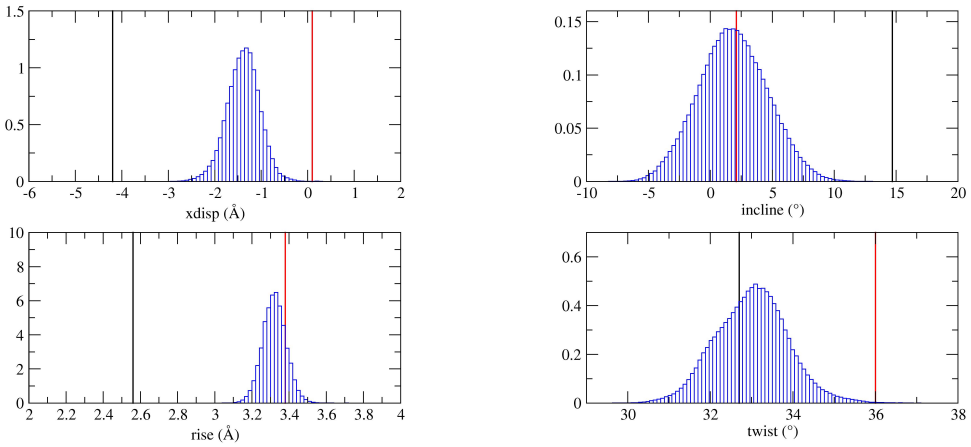


Figure 3.6: Base pair parameters histograms of 20B-Na0.

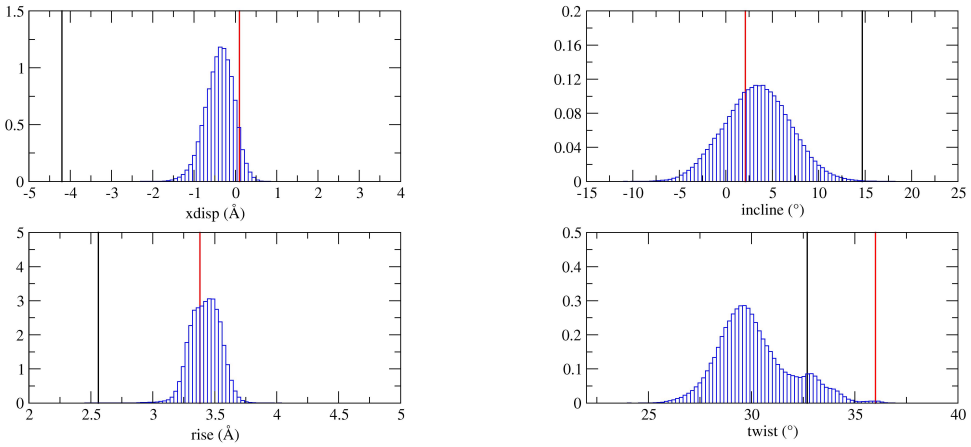


Figure 3.7: Base pair parameters histograms of 20B-Mg0.

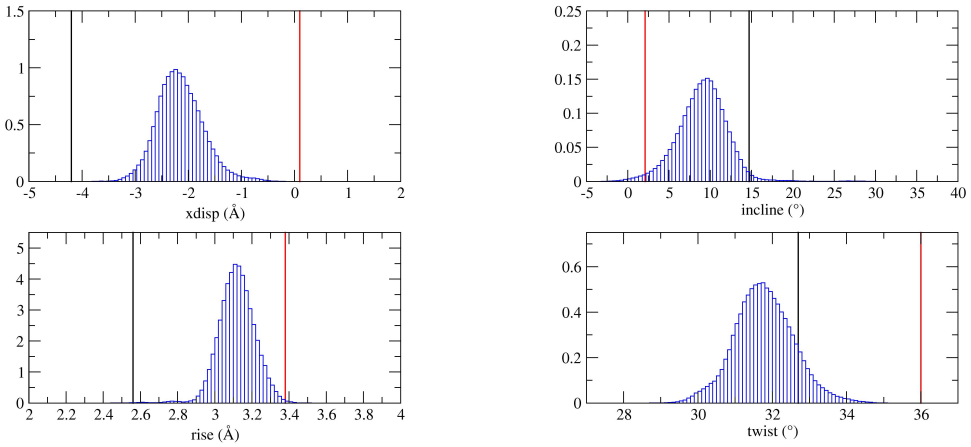


Figure 3.8: Base pair parameters histograms of ATA-Na0.

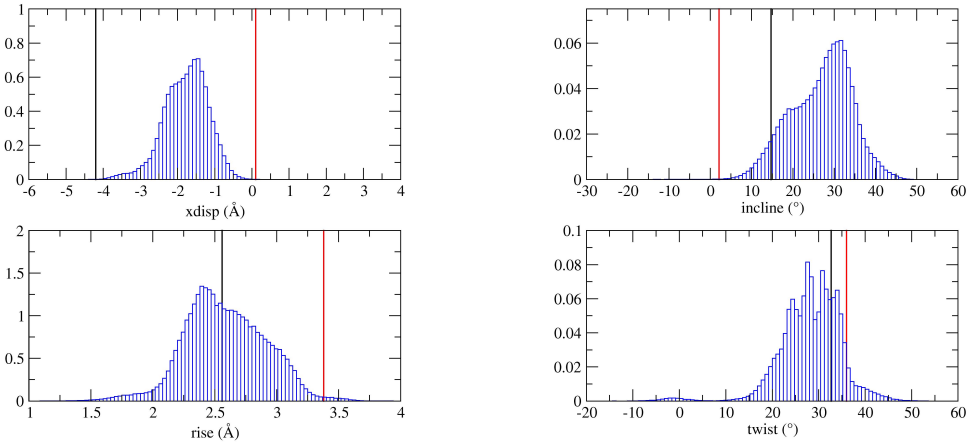


Figure 3.9: Base pair parameters histograms of ATA-Mg0.

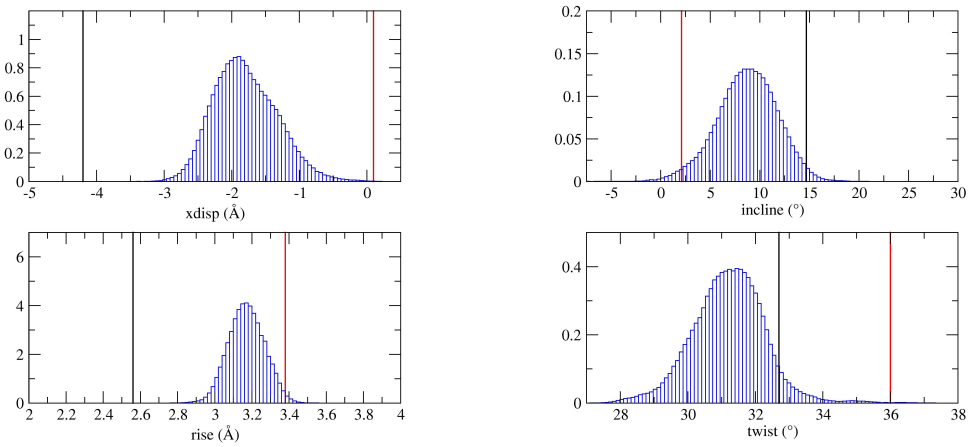


Figure 3.10: Base pair parameters histograms of ATB-Na0.

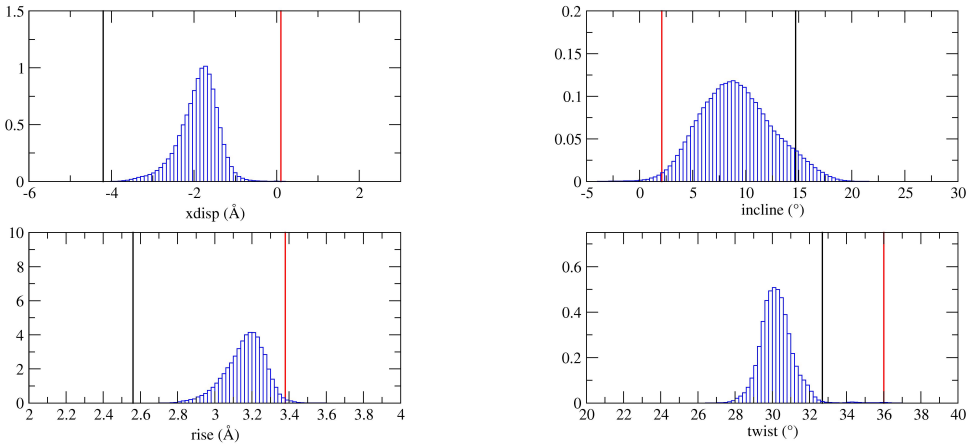


Figure 3.11: Base pair parameters histograms of ATB-Mg0.

3.3.2 Mono-spermidine systems

The set of mono-spermidine systems consists of eight similar systems: 20A-Na1, 20B-Na1, 20A-Mg1, 20B-Mg1, ATA-Na1, ATB-Na1, ATA-Mg1 and ATB-Mg1.

Fig. 3.12 shows the snapshots of the MD simulation of the ATA-Mg1 system at the time steps same as in Fig. 3.2. It is obvious from both of the top-down and side perspectives that since the start of the simulation, the system became more and more chaotic over time, with the top part of the DNA double helix unzipping and large extents of asymmetrical bending and twisting of the whole molecular structure. Since the third nanosecond it is already rather hard to tell which conformation the DNA was in.

The graphs of mono-spermidine systems using sodium and magnesium as counterions showing instantaneous RMSd of individual system from canonical structures are presented in Fig. 3.13. For the Na1 systems, since there is only one spermidine molecule inside the simulation box, it is anticipated that the behaviours of the DNA would be quite similar to those in the Na0 systems. This is proven to be a good guess, as the curves for the 20A-Na1 and 20B-Na1 systems bear great resemblance to their spermidine-free counterparts, with the RMSd of A-start hitting the high-end within the first couple of nanoseconds while that of B-start remains at the low-end ($\approx 4\mathring{A}$). Moreover, the graphs for ATA and ATB remain close to each other, but with a larger fluctuation.

For the Mg1 systems, the situation has become more complicated and unpredictable than the Mg0 system even for 20A and 20B. The RMSd curve for A-start did not shoot up to the high-range in the first few nanoseconds, but rather rose gradually and not until around the 12th nanosecond did it reach 7Å. It did not stay in the high-range for long before dropping gradually to about 6Å at the end of the simulation. As a result, the curves for 20A and 20B were rather close throughout the 40ns simulation. The curves for ATA and ATB were quite close together in the first quarter of simulation, but the gap between them slowly widened as the curve for ATB rose to and remained at the high-end and that for ATA decreased to around 3Å. This is intriguing as it is the first time that the RMSd of a B-start is larger than that of an A-start, which suggests the preference of A-form over B-form.

Figures 3.14 and 3.15 show the histograms for base pair parameters of four mono-spermidine (magnesium) systems. The histograms for the sodium systems are not included here as they resemble the normal distribution as in the previous spermidine-free sodium systems; they are stored in the attached DVD disc (Characteristics of them will be tabulated and explained in the summary

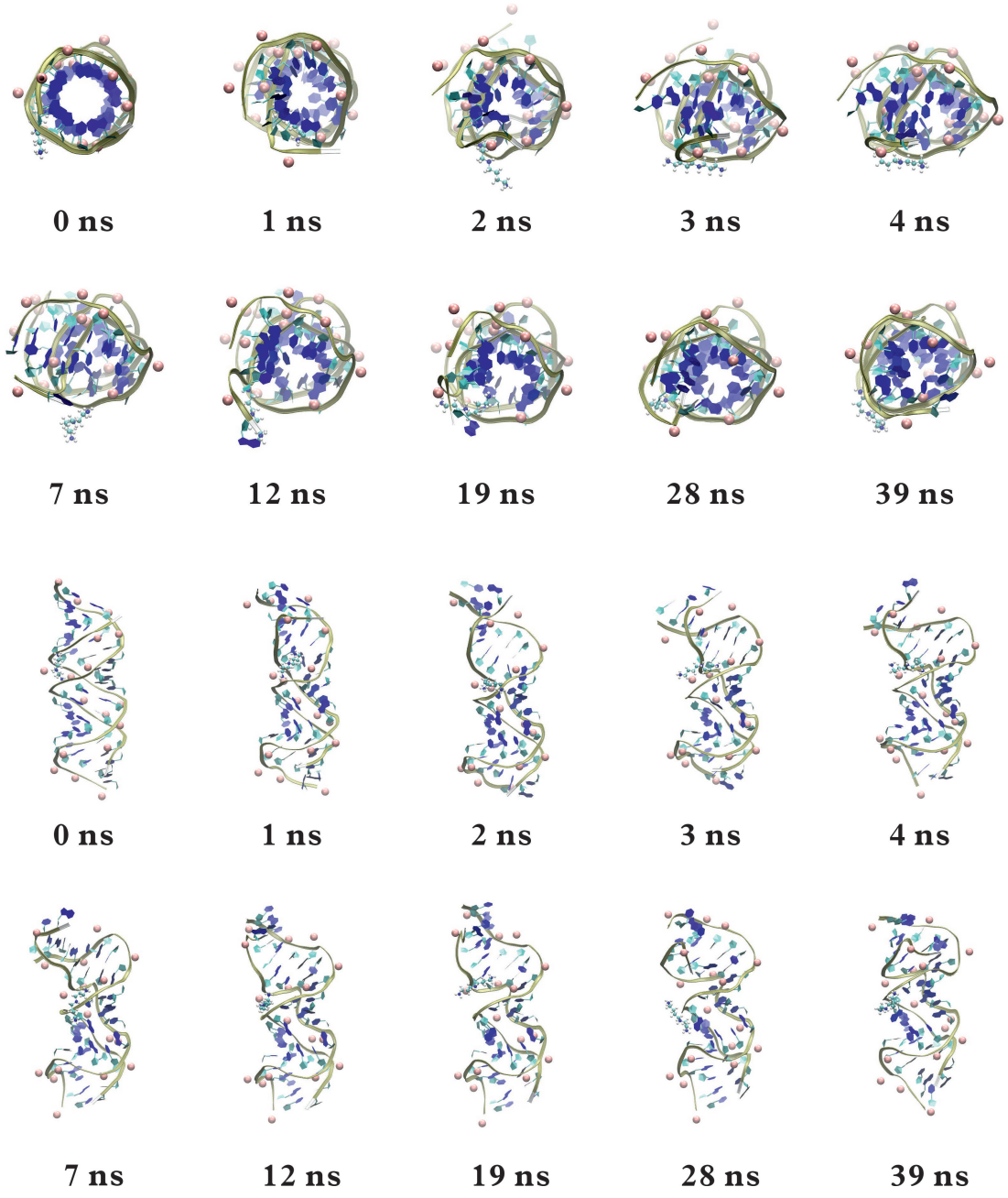


Figure 3.12: Snapshots (top-down and side view) of ATA-Mg1.

section). However, the shapes of the histograms for the 20A-Mg1 system are still very random with the width of distribution as much as 25Å for incline, for instance. This shows that one spermidine molecule (outside of the DNA) is not enough to neutralise the destabilisation effects brought by magnesium ions on A-form DNA, if it ever has the ability of doing so.

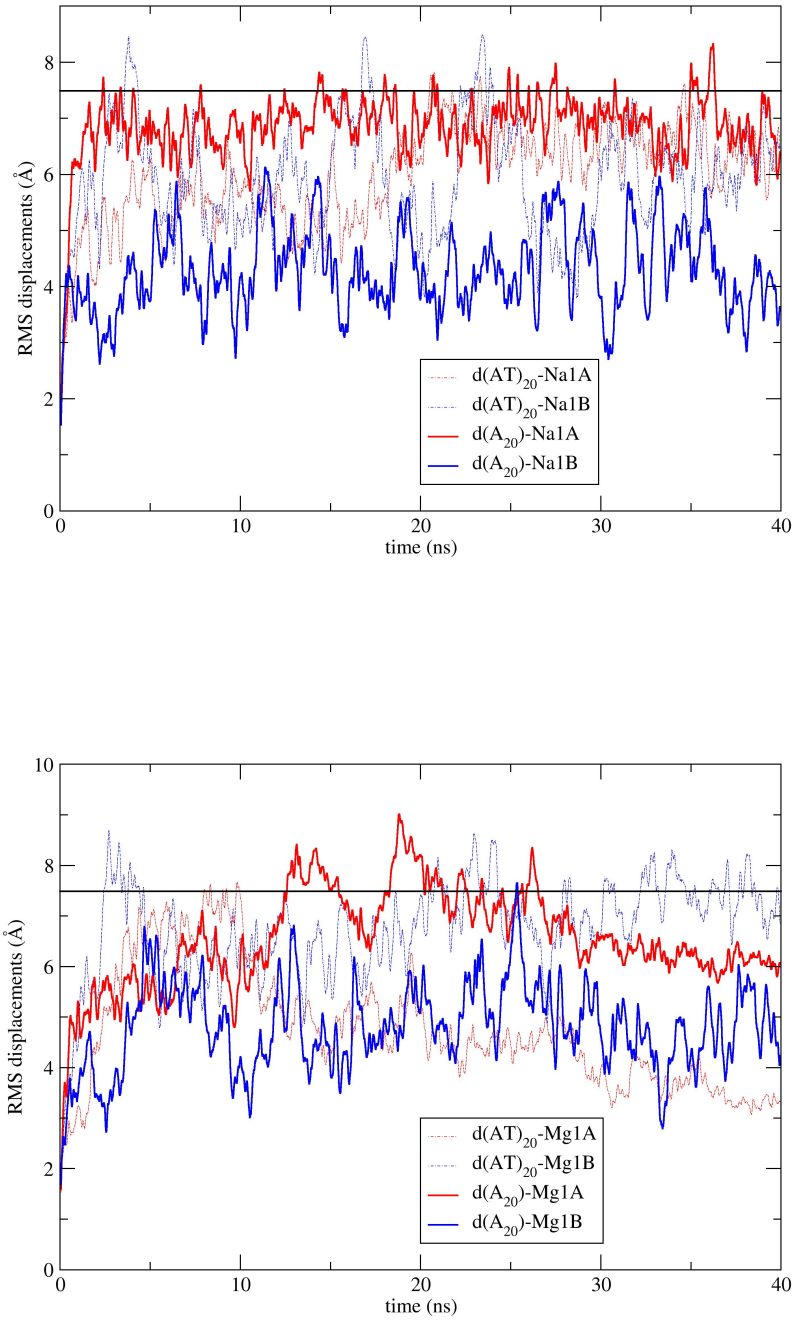


Figure 3.13: Graphs of mono-SPD systems using sodium and magnesium as counterions showing instantaneous RMSd of individual system from canonical structures

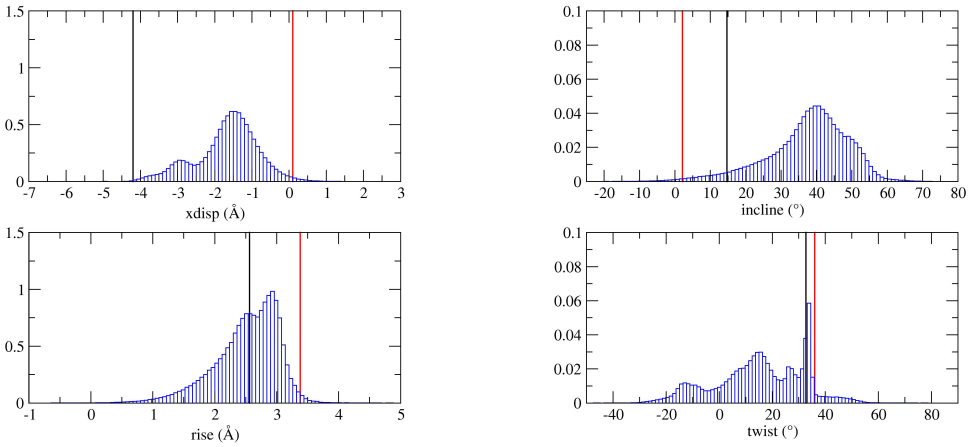


Figure 3.14: Base pair parameters histograms of 20A-Mg1.

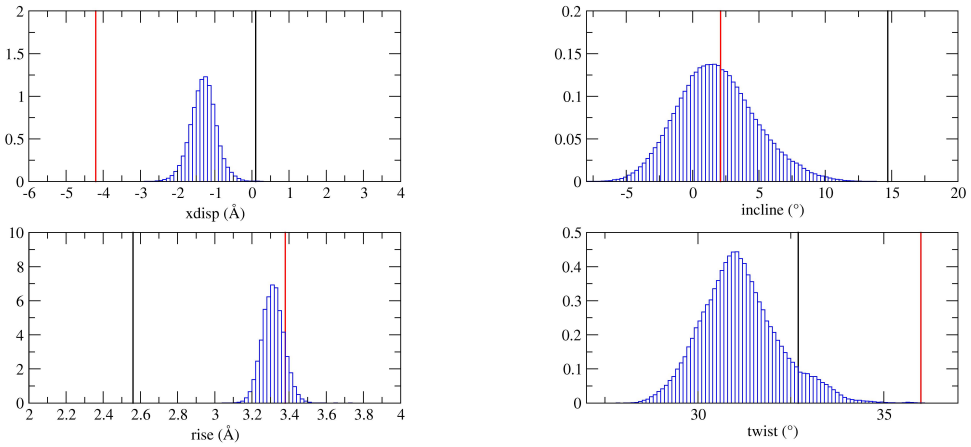


Figure 3.15: Base pair parameters histograms of 20B-Mg1.

Figures 3.16 and 3.17 show the histograms for base pair parameters of two remaining mono-spermidine magnesium systems. The near-perfect bell shape of the histograms further strengthens the previous point that the alternate placements of the bases on the strands make the DNA section less likely to be influenced by external forces. Furthermore, most of the distributions are positioned between the values of canonical A- and B-forms. This suggests that in hydrated conditions, i.e. real and physiological environments, the molecules most likely take an intermediate form between A and B. The canonical forms exist only in crystallised environments.

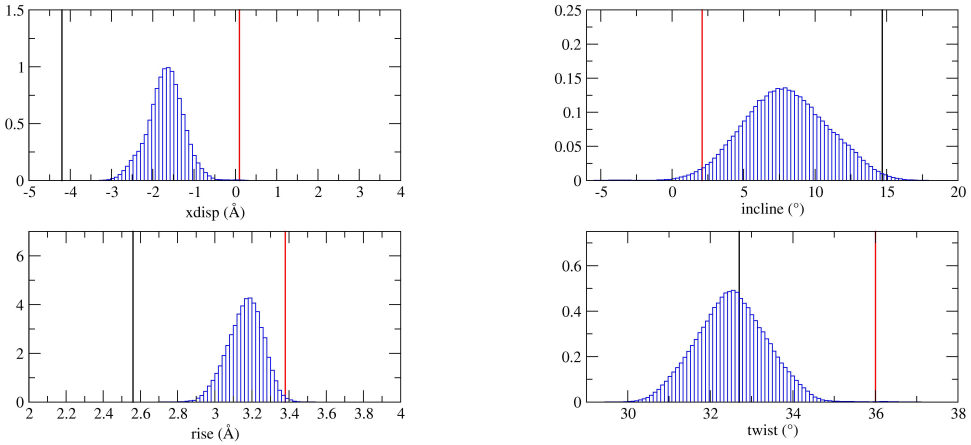


Figure 3.16: Base pair parameters histograms of ATA-Mg1.

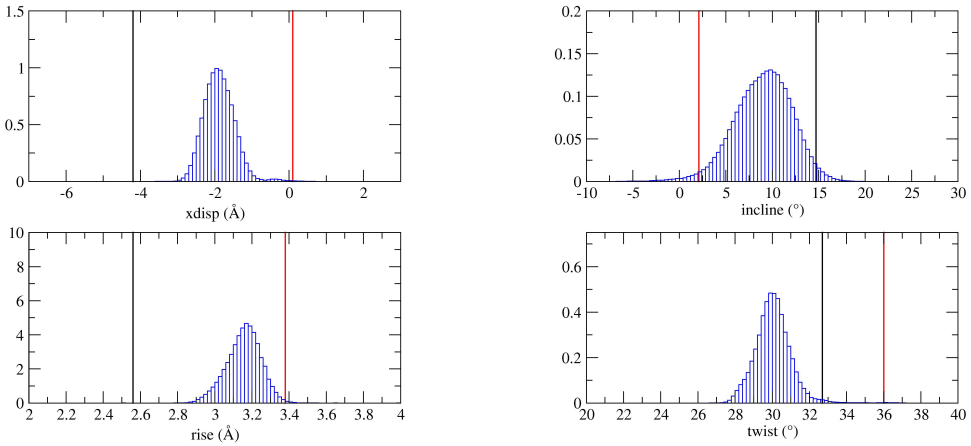


Figure 3.17: Base pair parameters histograms of ATB-Mg1.

3.3.3 Di-spermidine systems

The set of di-spermidine systems consists of eight similar systems: 20A-Na2, 20B-Na2, 20A-Mg2, 20B-Mg2, ATA-Na2, ATB-Na2, ATA-Mg2 and ATB-Mg2.

Fig. 3.18 shows the snapshots of the MD simulation of the ATA-Mg2 system. It is clear that from the very first nanosecond the top half of the DNA molecule started to unzip and the double strand was totally untwisted at the seventh nanosecond. The structure of DNA is fully destructed.

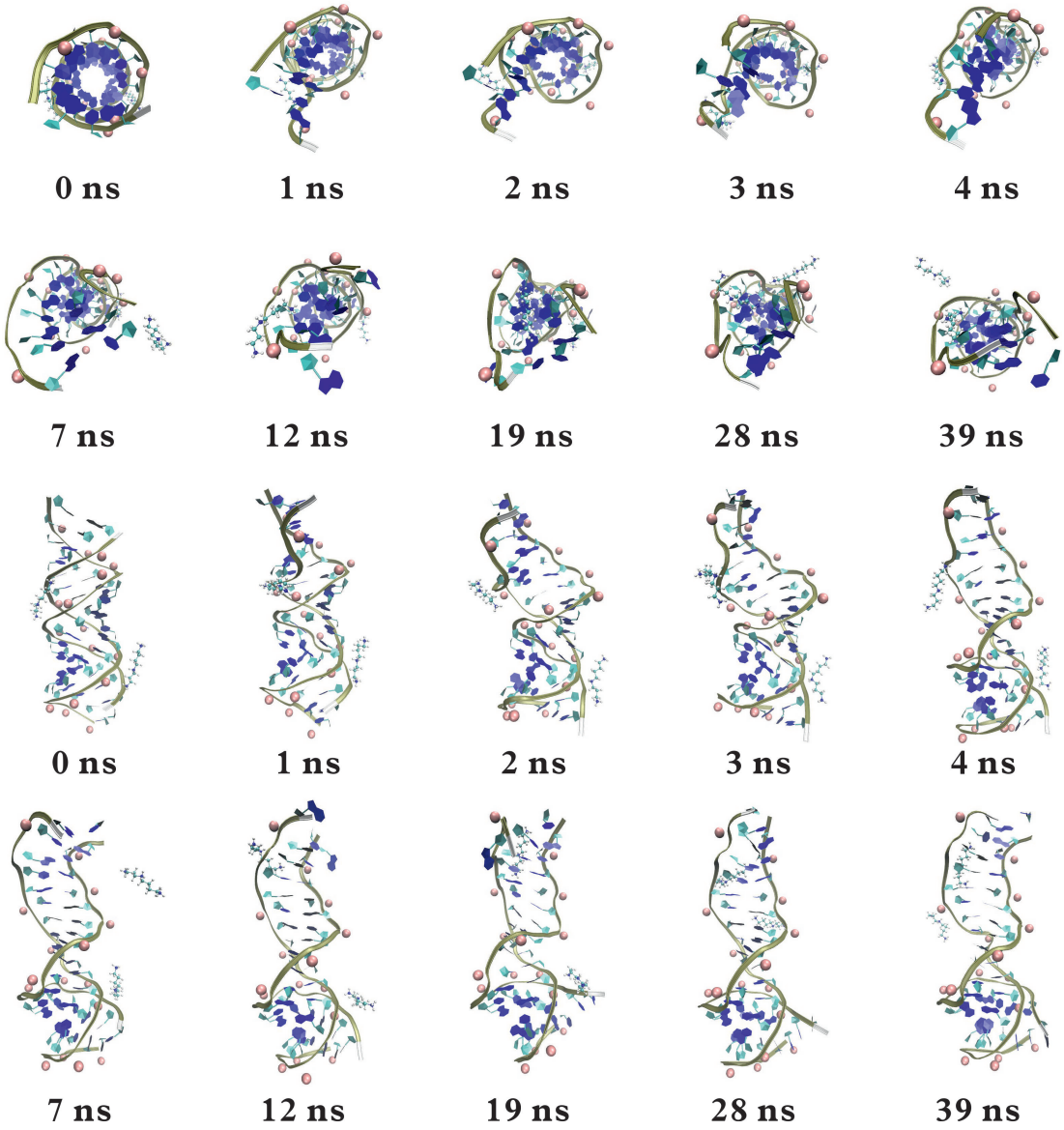


Figure 3.18: Snapshots (top-down and side view) of ATA-Mg2.

The graphs of di-spermidine systems using sodium and magnesium as counterions showing instantaneous RMSd of individual system from canonical structures are presented in Fig. 3.19. For the Na2 systems, there is not much of a change in the behaviours — RMSd for 20A shot within the first two nanoseconds up to near the 7.489Å line (RMSd of canonical A-form from B-form of

the $d(A)_{20}$ and $d(A\text{-}T)_{20}$ systems) which indicates a plausible transition from one form to another (in this case A to B) and stayed there for the rest of the simulation, while that for 20B remained at low-end about 4Å . It is worth mentioning that comparing the graphs with the spermidine-free and mono-spermidine counterparts, the fluctuation is much less for the di-spermidine systems. This suggests that spermidine may have stabilising power in the presence of sodium ions. Furthermore, the graphs for the two AT systems stick together throughout the whole simulation, signifying that the conformation right between the canonical A and B is most favourable.

For the Mg2 systems, the fluctuations in both 20A and 20B are quite strong, with the RMSd for 20A rising up to the "transition threshold" within 10ns and drops a little for a period of time and goes back up at the 30th nanosecond. That for 20B stayed at the low-to-mid range (3-6Å) most of the time near the low-end. For the AT systems, the two curves separated from each other from around the 7th nanosecond when ATA stayed near the line and ATB retained in the mid-range. This suggests that for Mg2 systems, B-form is slightly favoured against A-form.

Figures 3.20 and 3.21 show the histograms for base pair parameters of four di-spermidine systems. For the Mg2 systems, we find that the shapes of histograms for 20A and 20B differ very much — those for 20B are well-defined and normally-distributed with only slight skewness, while those for 20A are nowhere near bell-shape: spiky, double-peaked, peak-and-plateau, etc., suggesting the 20A-Mg2 system is highly disturbed.

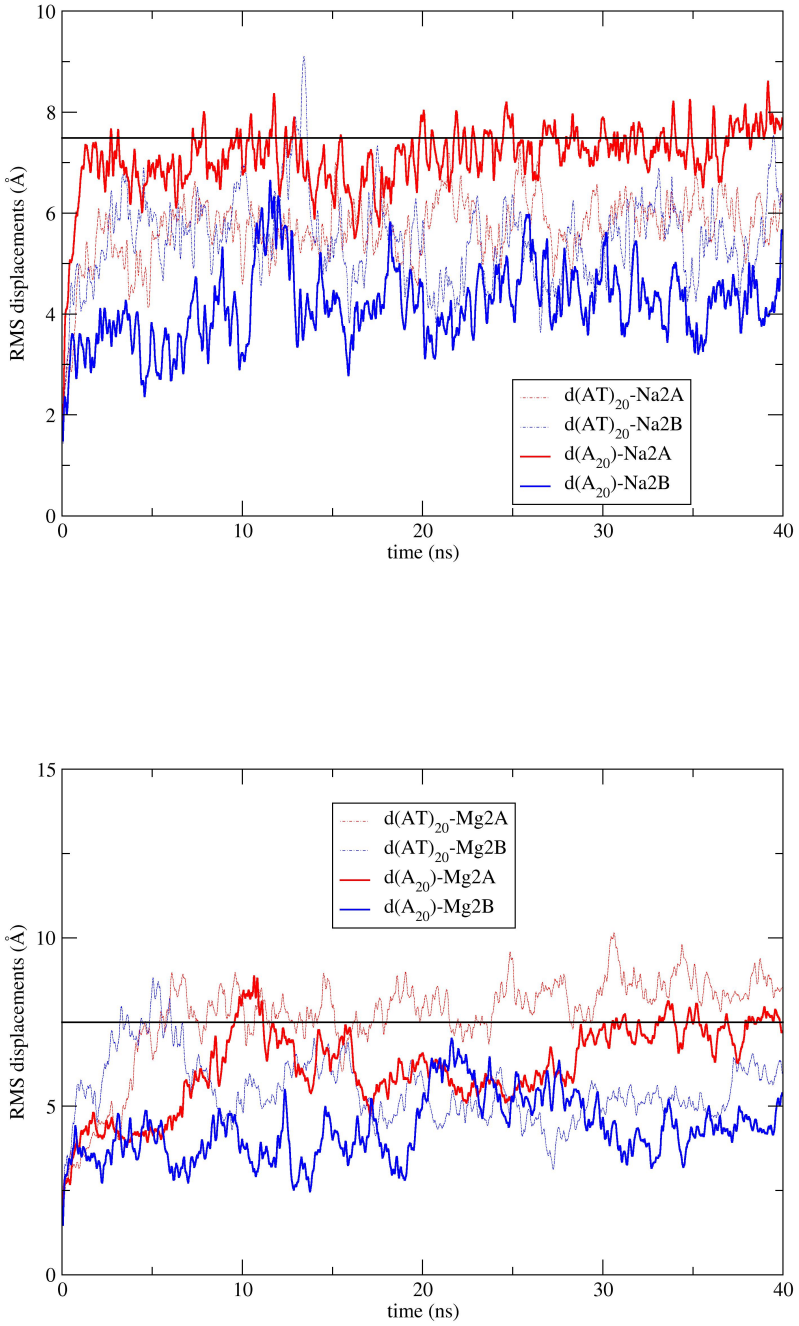


Figure 3.19: Graphs of di-SPD systems using sodium and magnesium as counterions showing instantaneous RMSd of individual system from canonical structures

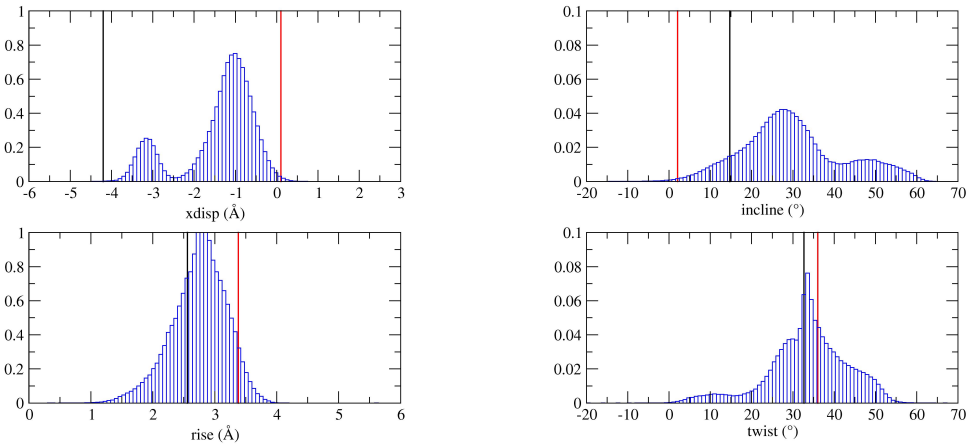


Figure 3.20: Base pair parameters histograms of 20A-Mg2.

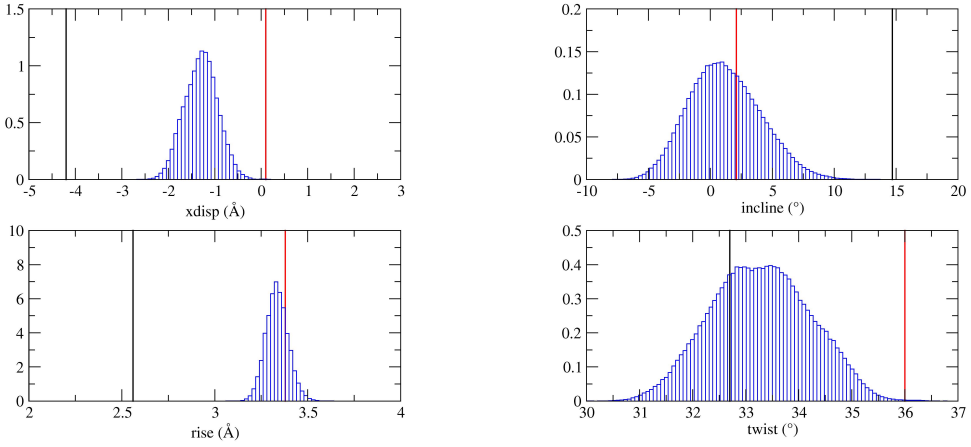


Figure 3.21: Base pair parameters histograms of 20B-Mg2.

Figures 3.22 and 3.23 show the histograms for base pair parameters of four di-spermidine AT systems. For the Mg2 systems, surprisingly, the shapes of the curves bear no resemblance with the "20" system counterparts at all, in that the distribution is much more near a normal one, though skewness and double-peak still appear. However if one looks closely to these anomalies, it is not difficult to rationalise them. The observable trends in these anomalies include: for the skewed graphs the A-side bear longer "tails", and for the double-peaked the two peaks are not of similar height (with the peak closer to the B-side taking the major peak and that to the A-side the minor). These all point to the argument that the spermidine molecules do have stabilising power, thus despite the B-form is slightly more preferred to A-form, the initial configuration of A is stabilised by the spermidine molecules so that the conformation takes longer to transform to the equilibrium structure.

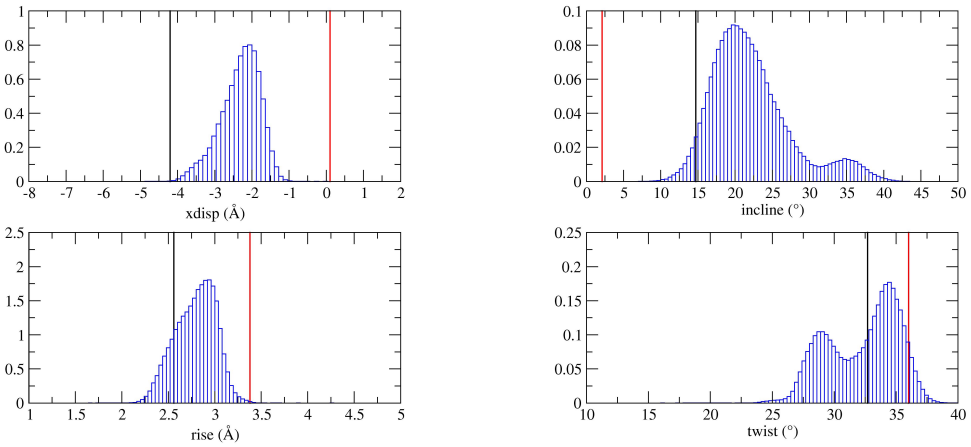


Figure 3.22: Base pair parameters histograms of ATA-Mg2.

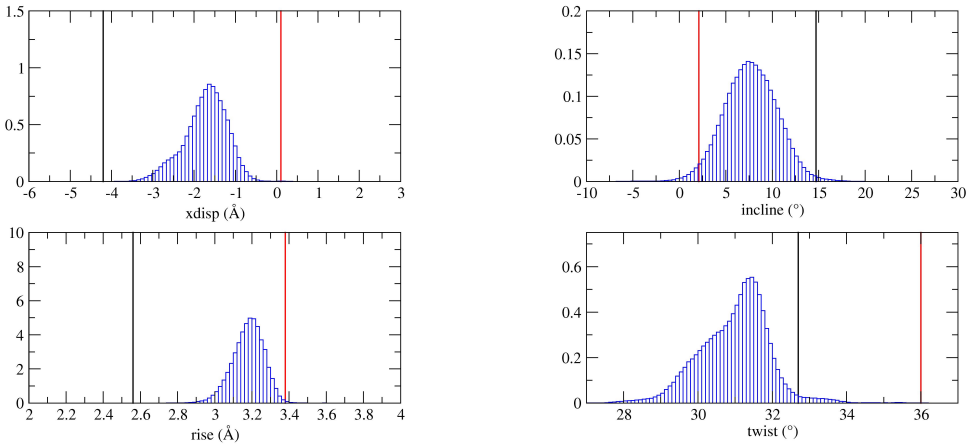


Figure 3.23: Base pair parameters histograms of ATB-Mg2.

3.3.4 Tri-spermidine systems

The set of tri-spermidine systems consists of eight similar systems: 20A-Na3, 20B-Na3, 20A-Mg3, 20B-Mg3, ATA-Na3, ATB-Na3, ATA-Mg3 and ATB-Mg3.

Figs. 3.24 and 3.25 shows the snapshots of the MD simulation of the 20B-Na3 and ATA-Mg3 systems respectively. For 20B-Na3, the variation of the structure was minimal throughout the simulation. This is predicted since B-form is very much preferred to A-form for sodium systems. However for ATA-Mg3, we still see the same kind of disruption to the structure as in the previous cases. But with 3 spermidine molecules present the extent of untwisting of the double helix is greatly reduced, that although the structure was largely disturbed (as seen from the top-down view) the top-half of the molecule is still loosely twisted.

The graphs of tri-spermidine systems using sodium and magnesium as counterions showing instantaneous RMSd of individual system from canonical structures are presented in Fig. 3.26. For the Na3 systems, the RMSd graphs look very much alike those for Na2 or even Na1. This is primary because sodium has very profound effect on leading the transformation into B-form in 20A and 20B. For the AT systems, because of the stabilising effect of the spermidine molecules, the equilibrium conformations are similar to the di-spermidine case — midway between canonical A- and B-forms.

For the Mg3 systems, it seems that spermidine is more effective in binding the DNA and restricting its motion on the "20" series than on the AT series. The RMSd curves for both 20A and 20B were kept under the "transition line" and retained inside the mid-range (≈ 6Å) throughout the simulation. For the

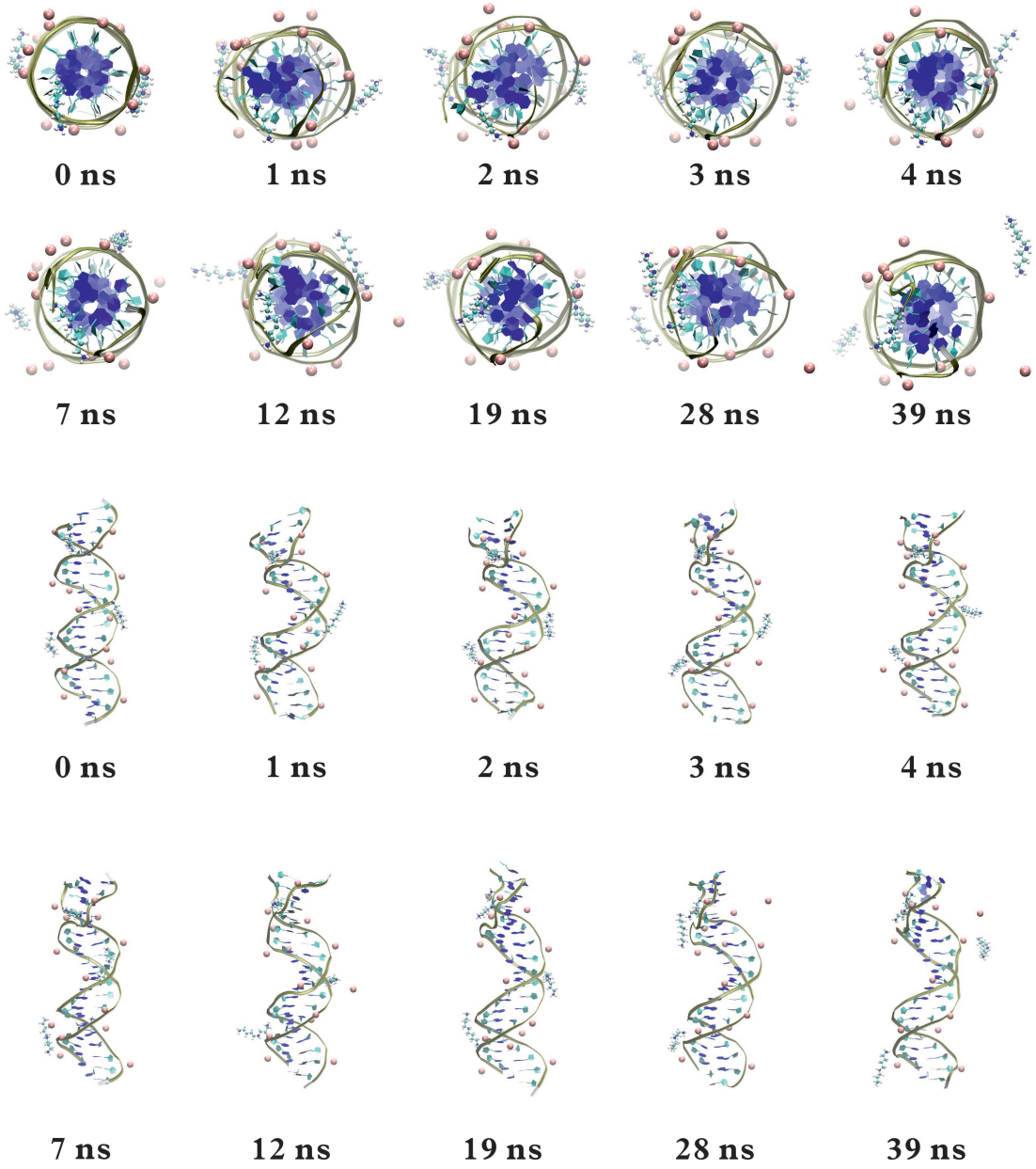


Figure 3.24: Snapshots (top-down and side view) of 20B-Mg3.

AT systems, RMSd for both reach the "transition line" within the first two nanoseconds and remained at about 7.5Å , which indicates that the conformational preference of one over another is eliminated.

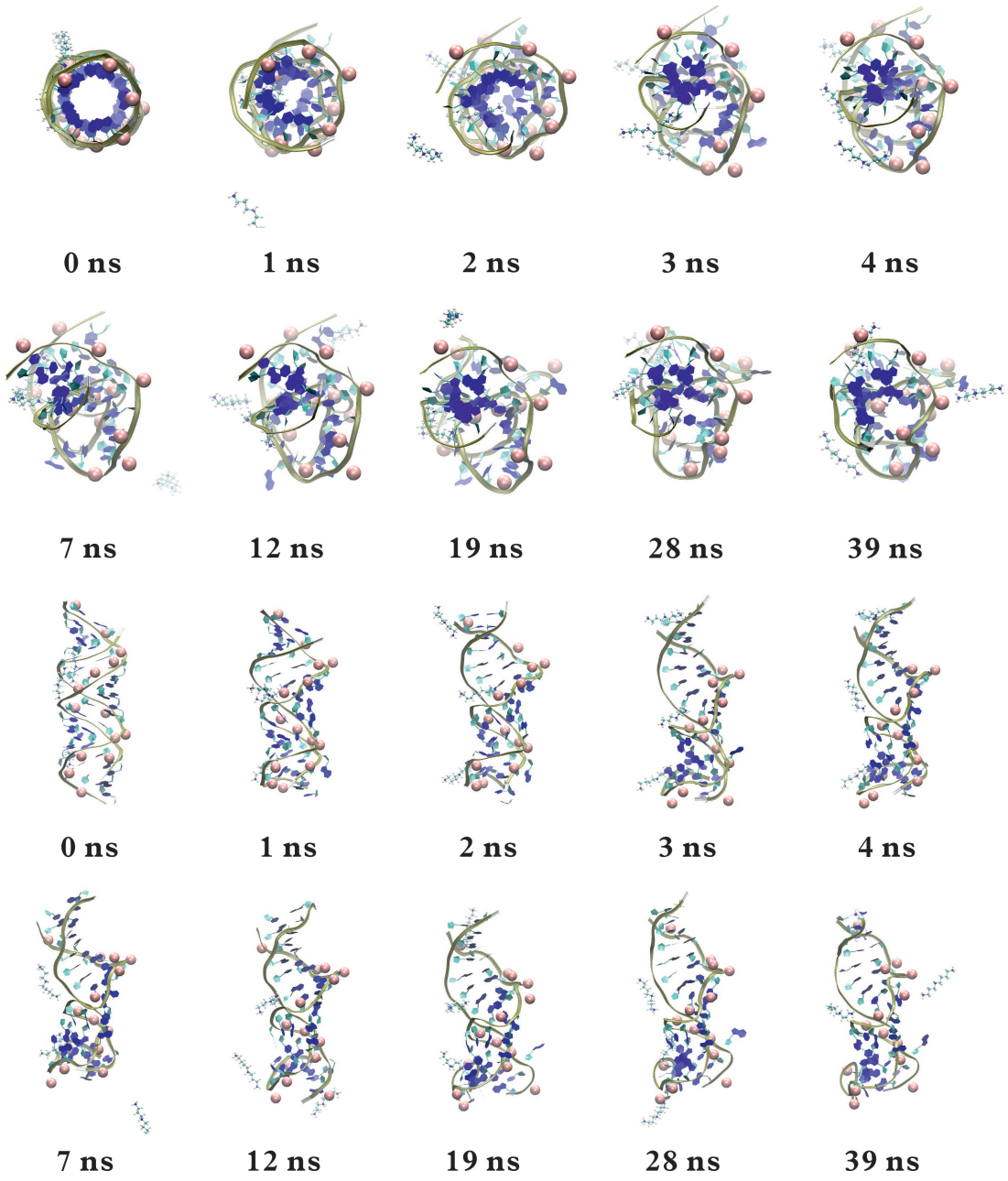


Figure 3.25: Snapshots (top-down and side view) of ATA-Mg3.

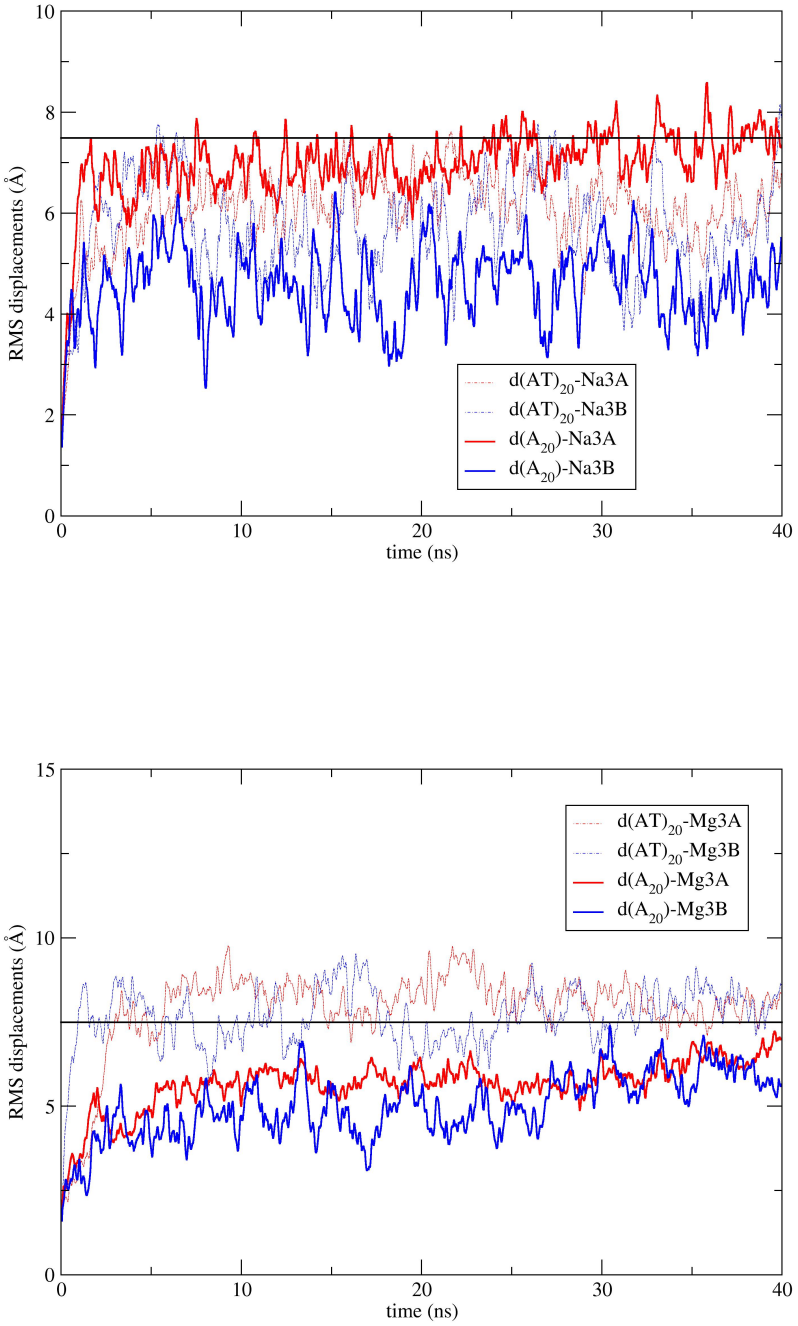


Figure 3.26: Graphs of tri-SPD systems using sodium and magnesium as counterions showing instantaneous RMSd of individual system from canonical structures

Figures 3.27 and 3.28 show the histograms for base pair parameters of two tri-spermidine magnesium "20" systems. For the AT systems, just as the Mg2 cases, the histograms of the parameters are all quite messy and chaotic, further indicating that although spermidine molecules do have stabilising power, it is due to electrostatic effect; and the stabilisation limits to the "macroscopic" helical structure — the base pair arrangements are not affected at all.

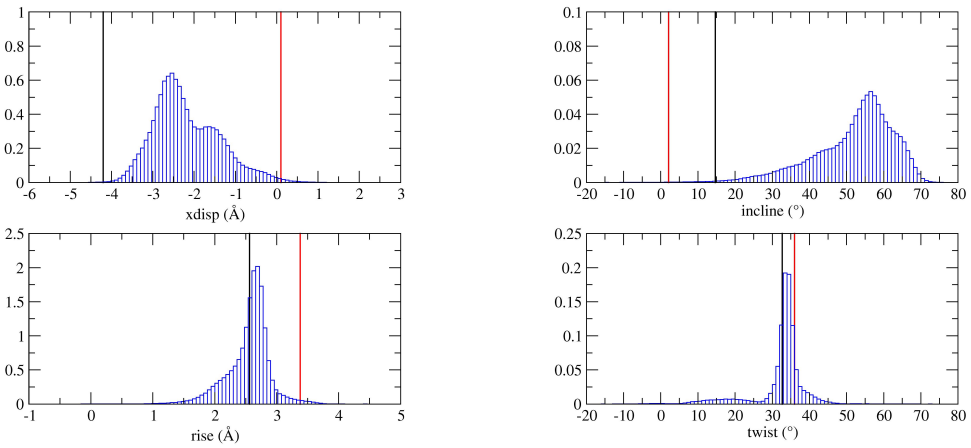


Figure 3.27: Base pair parameters histograms of 20A-Mg3.

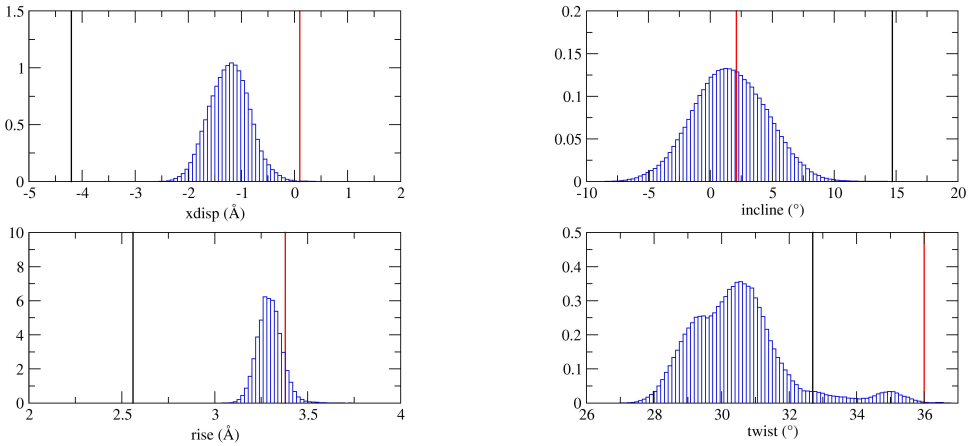


Figure 3.28: Base pair parameters histograms of 20B-Mg3.

Figures 3.29 and 3.30 show the histograms for base pair parameters of two tri-spermidine magnesium AT systems. Comparing these histograms with their AT counterparts previously, it is not hard to see that all the charts resemble the previous ones closely, with the widths narrowed a little. This is probably because of the stabilising action of spermidine.

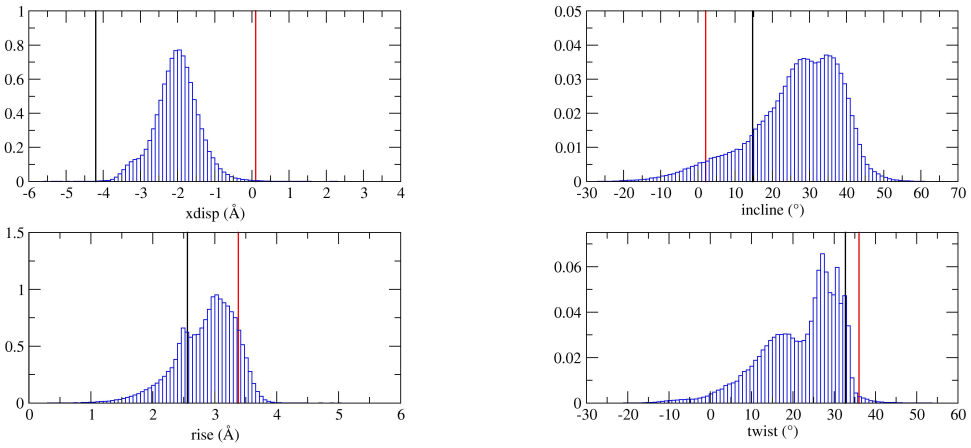


Figure 3.29: Base pair parameters histograms of ATA-Mg3.

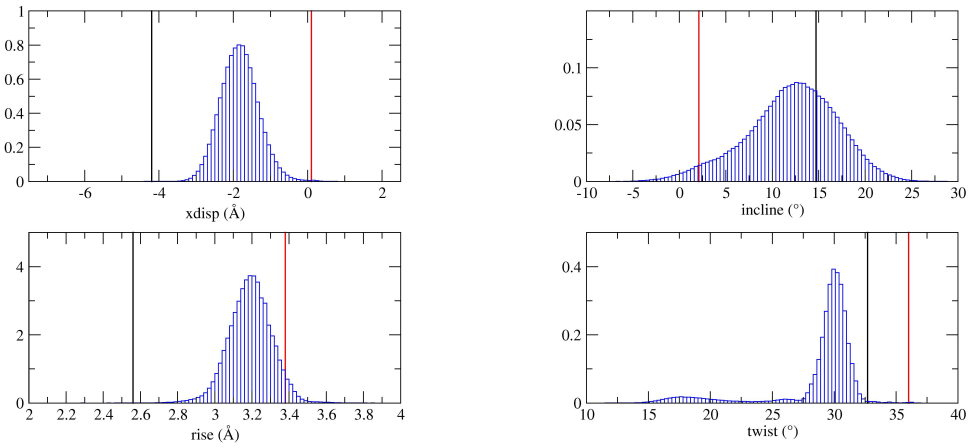


Figure 3.30: Base pair parameters histograms of ATB-Mg3.

3.3.5 Tetra-spermidine systems

This set of systems is the last in the series of five, which comprises the final eight structures: 20A-Na4, 20B-Na4, 20A-Mg4, 20B-Mg4, ATA-Na4, ATB-Na4, ATA-Mg4 and ATB-Mg4.

Fig. 3.31 shows the snapshots of the MD simulation of the ATA-Mg4 system. The extent of the unzipping of the DNA strands was even less than that in ATA-Mg3, in turn confirming that spermidine does have stabilising effects on the conformation of DNA. However, from the top-down view it is still very difficult to determine whether or not there was a transition from A to B, primarily because of the bending of the DNA which made the size of the "hole"

down the helical axis hard to determine.

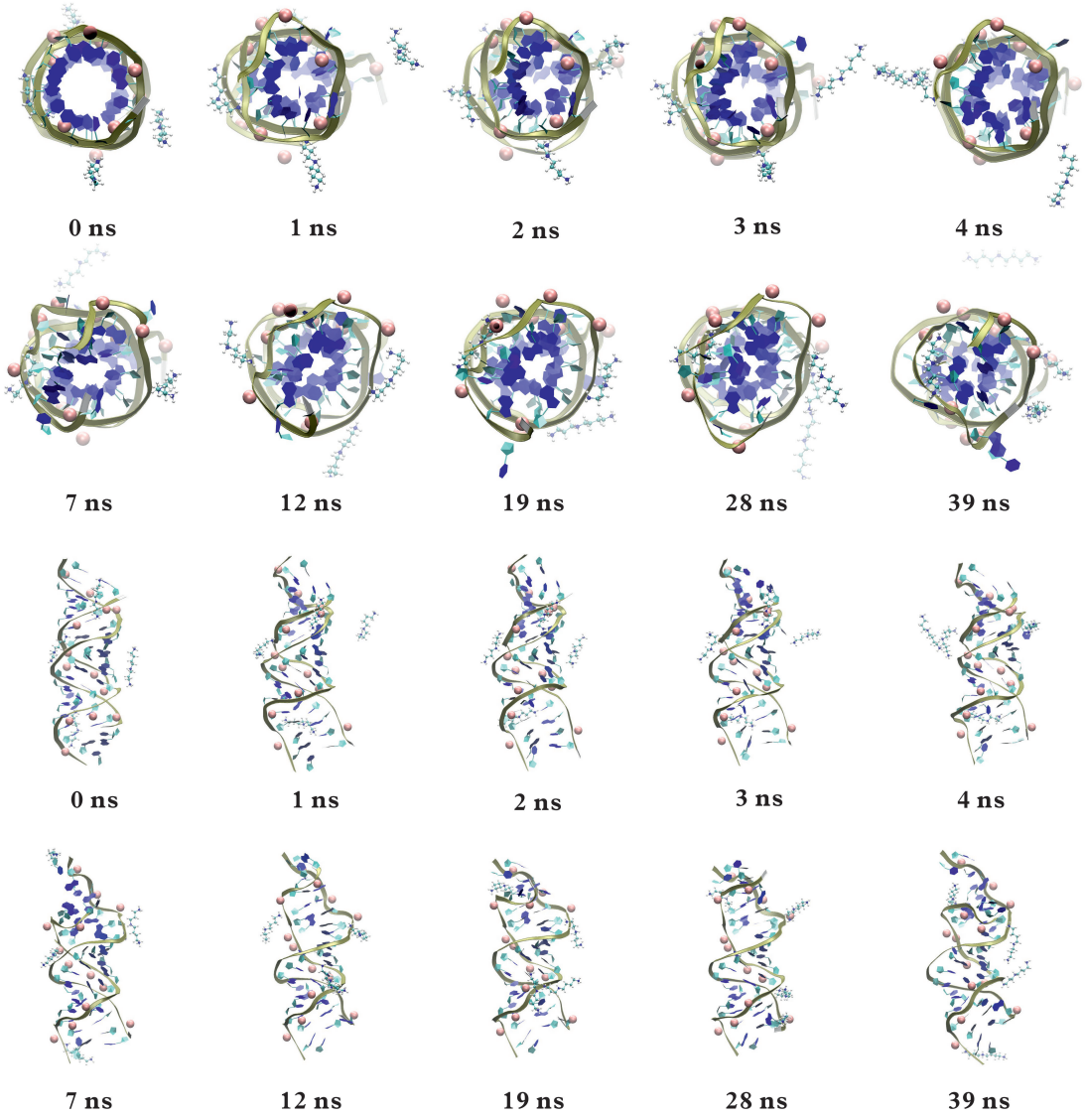


Figure 3.31: Snapshots (top-down and side view) of ATA-Mg4.

The graphs of tetra-spermidine systems using sodium and magnesium as counterions showing instantaneous RMSd of individual system from canonical structures are presented in Fig. 3.32. They resemble those for the previous systems greatly no matter in the forms of the curves or the values and positions of the curves. This can be attributed to the saturation of the effect of spermidine. However, we can still observe that the RMSd curves for 20A and 20B has been shifted down by a little when compared with the tri- and di-spermidine counterparts. This is because the addition of an extra spermidine further suppresses the motional freedom of the DNA molecules.

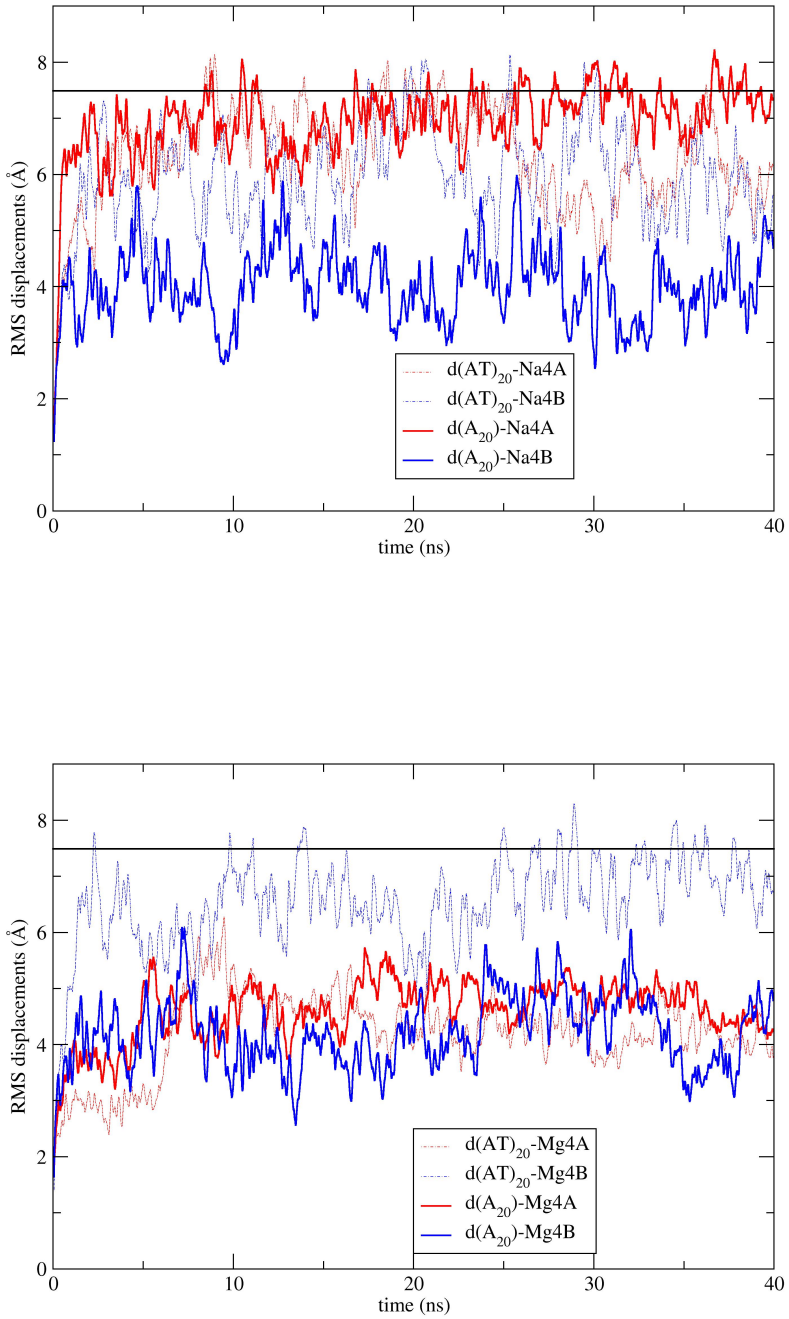


Figure 3.32: Graphs of tetra-SPD systems using sodium and magnesium as counterions showing instantaneous RMSd of individual system from canonical structures

Figures 3.33 to 3.36 show the histograms for base pair parameters of all the four tetra-spermidine magnesium systems. All of these graphs look nearly identical to the tri-spermidine ones, in turn suggesting the effect of spermidine probably saturates when there are three spermidine molecules around a 20 base-pair

section.

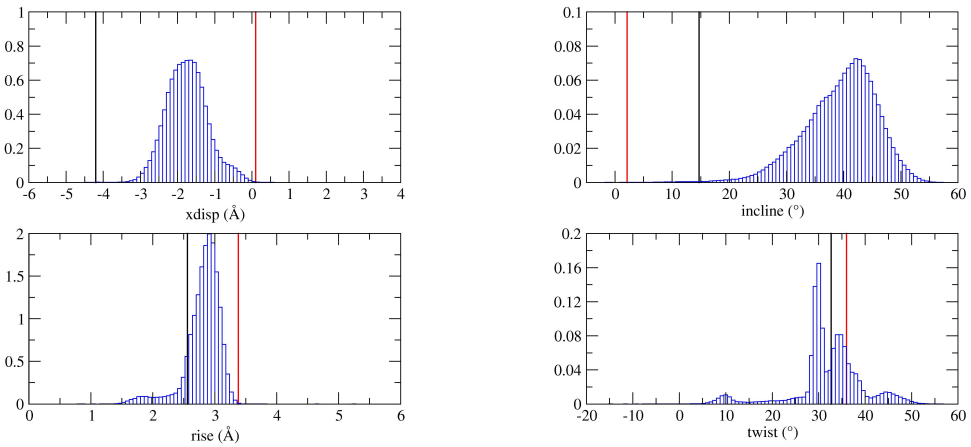


Figure 3.33: Base pair parameters histograms of 20A-Mg4.

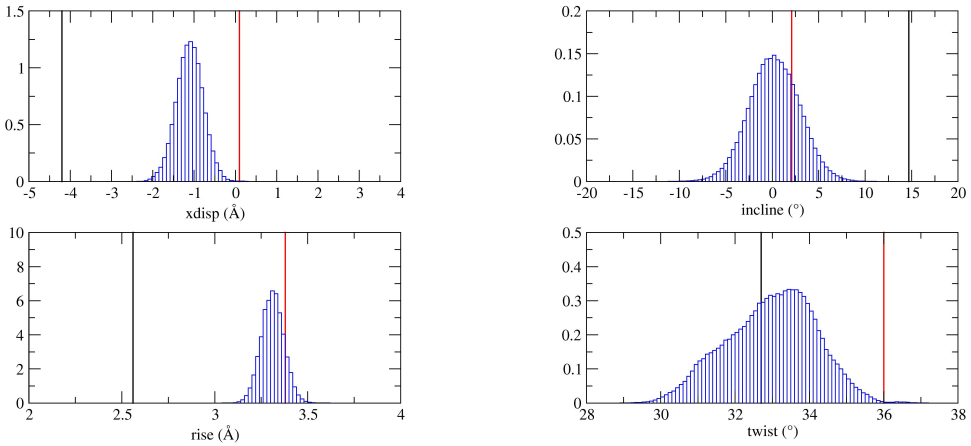


Figure 3.34: Base pair parameters histograms of 20B-Mg4.

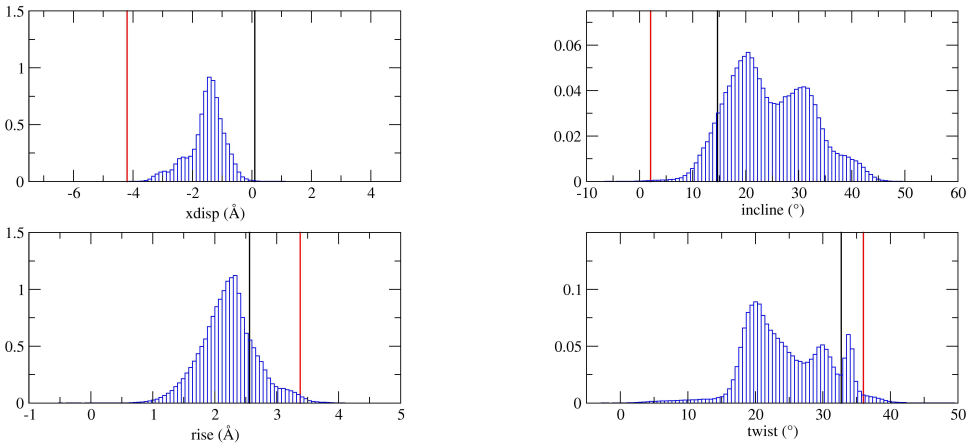


Figure 3.35: Base pair parameters histograms of ATA-Mg4.

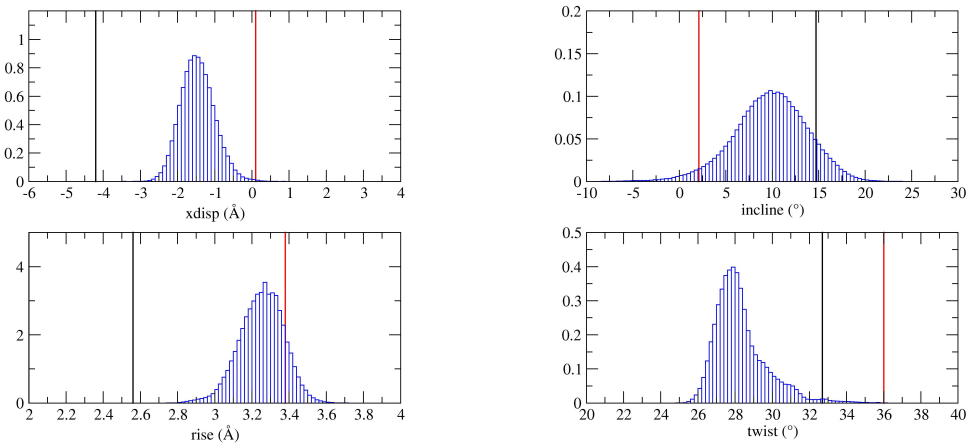


Figure 3.36: Base pair parameters histograms of ATB-Mg4.

3.3.6 Summary and discussions

In this chapter various results regarding the computer simulation of the 20 base-pair sequences $d(A)_{20}$ and $d(A\text{-}T)_{20}$ were discussed. It was found that sodium has the effect of bringing the DNA system to a near-B conformation regardless of the starting configuration, which is expected from various previous studies. However it was surprising as a discovery that magnesium, albeit having its charge double that of sodium, did not have double stabilising power that of sodium, but rather brought large disruption to the whole system. Table 3.2 gives a brief summary of the histograms for base pair parameters of all the 20 systems discussed in this chapter.

For each conformational parameter, where the histogram could be identified as either unimodal or bimodal, it will be expressed in terms of three components. For unimodal distributions the code would be "U: p, w", where U stands for unimodal, p is the abscissae of the peak and w the full width at half maximum (FWHM). For bimodal distributions the code would be "B: p_M, p_m" where B stands for bimodal, and p_M and p_m are the x values of the major and minor peaks respectively. Where the distributions are messy and chaotic, the code would have only a letter M.

From the table it is not difficult to draw a conclusion that the effect of spermidine as a source of structural stabilisation is rather uncertain, as the values for the FWHM do not drop with the increase in number of spermidine molecules in the system. Moreover, should the number of spermidine be proportional to the stability, the values of the structural parameters must converge with increasing number of spermidine. However this trend is not seen from the simulations.

Secondly, all the sodium systems behaved reasonably well, as could be seen from both the simulation snapshots presented previously and the values from Table 3.2. However the modes of the histogram distributions for the magnesium systems revealed that many of them are much more widespread than their sodium counterparts. This could be observed especially from the A-start systems, where all the bimodal and chaotic distributions reside.

This in turn inspired further investigation into the roles of counterions on DNA systems. In the next chapter we will study the effect of caesium, a monovalent but much more massive ion, on DNA.

System	xdisp (Å)	incline (°)	h-rise (Å)	h-twist (°)
20A-Na0	U: **-1.20**, 0.75	U: **2.00**, 7.20	U: **3.34**, 0.15	U: **33.24**, 1.90
20A-Na1	U: **-1.20**, 0.75	U: **2.00**, 7.20	U: **3.34**, 0.14	U: **33.64**, 2.24
20A-Na2	U: **-1.12**, 0.69	U: **2.00**, 9.74	U: **3.36**, 0.15	U: **33.22**, 2.70
20A-Na3	U: **-1.24**, 0.75	U: **1.50**, 6.35	U: **3.34**, 0.13	U: **33.40**, 1.99
20A-Na4	U: **-1.05**, 0.72	U: **-1.00**, 5.93	U: **3.34**, 0.15	U: **33.64**, 1.85
20A-Mg0	B: **-2.13**, **-0.24**	U: **35.60**, 21.97	U: **2.36**, 0.83	B: **27.45**, 13.80
20A-Mg1	B: **-1.50**, **-2.90**	U: **41.15**, 19.81	B: **2.96**, **2.54**	M
20A-Mg2	B: **-0.96**, **-3.12**	B: **27.70**, **48.40**	U: **2.82**, 0.79	U: **34.00**, 6.30
20A-Mg3	B: **-2.49**, **-1.59**	U: **57.00**, 14.85	U: **2.72**, 0.34	U: **34.00**, 3.62
20A-Mg4	U: **-1.60**, 1.26	U: **42.45**, 12.64	U: **2.94**, 0.45	M
20B-Na0	U: **-1.28**, 0.79	U: **1.40**, 6.63	U: **3.34**, 0.14	U: **33.14**, 1.97
20B-Na1	U: **-1.28**, 0.71	U: **0.20**, 6.48	U: **3.34**, 0.14	U: **33.32**, 1.93
20B-Na2	U: **-1.15**, 0.70	U: **1.10**, 6.75	U: **3.34**, 0.13	U: **33.50**, 1.72
20B-Na3	U: **-1.01**, 0.91	U: **1.10**, 6.92	U: **3.34**, 0.16	U: **33.68**, 2.12
20B-Na4	U: **-1.16**, 1.02	U: **1.24**, 7.31	U: **3.34**, 0.15	U: **33.29**, 1.97
20B-Mg0	U: **-0.32**, 0.80	U: **3.40**, 8.31	U: **3.47**, 0.32	U: **29.34**, 2.72
20B-Mg1	U: **-1.20**, 0.74	U: **1.80**, 6.70	U: **3.32**, 0.13	U: **31.10**, 2.01
20B-Mg2	U: **-1.24**, 0.85	U: **1.10**, 6.96	U: **3.34**, 0.14	U: **33.50**, 2.46
20B-Mg3	U: **-1.15**, 0.91	U: **1.40**, 7.17	U: **3.28**, 0.15	U: **30.62**, 2.66
20B-Mg4	U: **-1.04**, 0.75	U: **0.40**, 1.07	U: **3.32**, 0.14	U: **33.50**, 2.80
ATA-Na0	U: **-2.20**, 0.93	U: **9.85**, 5.98	U: **3.12**, 0.21	U: **31.80**, 1.74
ATA-Na1	U: **-1.68**, 1.00	U: **7.95**, 6.57	U: **3.18**, 0.20	U: **32.60**, 1.83
ATA-Na2	U: **-1.71**, 1.29	U: **7.40**, 5.90	U: **3.20**, 0.19	U: **31.80**, 2.31
ATA-Na3	U: **-1.56**, 0.99	U: **7.00**, 6.96	U: **3.22**, 0.20	U: **32.20**, 2.02
ATA-Na4	U: **-1.24**, 1.56	U: **4.90**, 8.67	U: **3.22**, 0.20	U: **32.32**, 2.20
ATA-Mg0	U: **-1.40**, 1.36	U: **32.10**, 16.66	U: **2.41**, 0.74	U: **28.00**, 12.42
ATA-Mg1	U: **-1.58**, 0.89	U: **8.04**, 7.09	U: **3.20**, 0.22	U: **32.60**, 1.93
ATA-Mg2	U: **-2.00**, 1.56	U: **20.00**, 9.44	U: **2.96**, 0.55	B: **34.60**, **29.20**
ATA-Mg3	U: **-1.90**, 1.11	U: **35.00**, 23.72	U: **3.06**, 1.06	B: **27.60**, **16.80**
ATA-Mg4	U: **-1.38**, 0.81	B: **20.80**, **31.30**	U: **2.36**, 0.59	M
ATB-Na0	U: **-1.87**, 1.13	U: **8.62**, 7.01	U: **3.18**, 0.23	U: **31.51**, 2.32
ATB-Na1	U: **-1.78**, 1.01	U: **8.00**, 6.43	U: **3.20**, 0.21	U: **32.10**, 2.41
ATB-Na2	U: **-1.76**, 1.12	U: **6.84**, 7.64	U: **3.20**, 0.24	U: **32.40**, 2.06
ATB-Na3	U: **-1.57**, 0.88	U: **8.00**, 7.15	U: **3.20**, 0.22	U: **32.60**, 1.93
ATB-Na4	U: **-1.78**, 1.19	U: **7.68**, 7.74	U: **3.20**, 0.23	U: **32.40**, 2.33
ATB-Mg0	U: **-1.68**, 0.87	U: **9.00**, 7.88	U: **3.20**, 0.22	U: **30.20**, 1.71
ATB-Mg1	U: **-1.90**, 0.96	U: **10.00**, 7.25	U: **3.18**, 0.19	U: **30.00**, 1.81
ATB-Mg2	U: **-1.59**, 1.05	U: **7.60**, 6.76	U: **3.20**, 0.19	U: **31.50**, 1.62
ATB-Mg3	U: **-1.80**, 1.16	U: **12.80**, 10.32	U: **3.20**, 0.25	U: **30.10**, 2.02
ATB-Mg4	U: **-1.50**, 1.06	U: **10.00**, 8.56	U: **3.28**, 0.27	U: **28.00**, 1.98
Canon. A	**0.1**	**14.7**	**2.8**	**32.5**
Canon. B	**-4.2**	**2.1**	**3.3**	**36.5**

Table 3.2: Summary of modes of histograms of DNA-cation systems in chapter 3. Values for canonical forms [56] given at the bottom of table.

Chapter 4

Simulations of d(ACCGGCGCCACA) and d(ACCGGCGCCGGT)

4.1 The choice of systems and methodology

This comparatively short chapter is dedicated to the discussions about the findings from the comprehensive studies of the two DNA dodecamers of the sequences d(ACCGGCGCCACA) and d(ACCGGCGCCGGT).

These two sequences are of particular interest because from previous studies it was shown that DNA sections which are non-palindromic are more vulnerable to mutation than the palindromic ones [60]. Palindromic sequences are sequences which read the same whether read 5′-to-3′ on the Watson strand or 5′-to-3′ on the Crick (complementary) strand. For example, the tetramer d(ACGT) is palindromic since it is read d(ACGT)·d(ACGT) when written in its full form. On the other hand, the slightly modified sequence of d(ACTG) is *non*-palindromic as its Crick strand, d(CAGT), is different from the Watson strand.

As such, the sequence which we have studied, d(ACCGGCGCCACA), which has been directly adopted from the aforementioned publication, is non-palindromic. Therefore it is the prime reason for us to have chosen this particular sequence and to study quantitatively its vulnerability to structural alteration. Also, as a control, we have also made the sequence into a palindromic one by changing the last three bases of the code from ACA to GGT. It is worthwhile to abbreviate the two sequences for the sake of better communication - d(ACCGGCGCCACA) will be aliased as “12n” while d(ACCGGCGCCGGT)

as "12p" - where n and p stand for *non-palindromic* and *palindromic* respectively.

In this set of simulations we still used sodium as one of the counterions, but for the other counterion we have chosen caesium. The prime reason for doing this is the investigation into the effects of ionic mass (rather than valence) on the structure of DNA. Caesium is chosen because it is the heaviest monovalent (Group IA) ion available for use; francium does not currently have a force field.

Moreover, in terms of spermidine placement, in this set of simulations we aim primarily to see the differences between the presence and absence of spermidine molecules. Hence the choice of number of spermidine introduced into the systems were restricted to only either 0 or 1. Wherever applicable, the spermidine was placed by hand in the major groove.

4.2 Results and discussions

In this section, results from simulations of the sequences d(ACCGGCGCCACA) and d(ACCGGCGCCGGT) are presented and discussed.

Fig. 4.1 shows the ribbon representation of systems discussed (in both canonical forms) in this chapter.

Following the same way as we have done in the previous chapter, we will analyse the data for this set of simulations first from the study of the RMSd graphs, then moving onto the conformational (base pair and groove) parameters of systems using caesium as counterions. The histograms for the sodium systems are stored in the attached DVD disc. Characteristics of them will be tabulated and explained in the summary section.

4.2.1 Spermidine-free systems

This part is dedicated to eight spermidine-free systems: 12nA-Na0, 12nB-Na0, 12nA-Cs0, 12nB-Cs0, 12nA-Cs0, 12nB-Cs0, 12pA-Cs0 and 12pB-Cs0.

Fig. 3.2 shows the snapshots of the MD simulation of the 12B-Na0 system at different time steps. It is clearly demonstrated from the top-down view that the DNA has transited from an A-form to B-form within the first nanosecond and stayed rather stably as a B-DNA for the rest of the simulation period.

The graphs of spermidine-free systems using sodium and magnesium as counterions showing instantaneous RMSd of individual system from canonical structures are presented in Fig. 4.3. First of all, the black line (RMSd between A- and B-forms of 12n and 12p) is downward shifted from 7.489Å to 5.579Å when

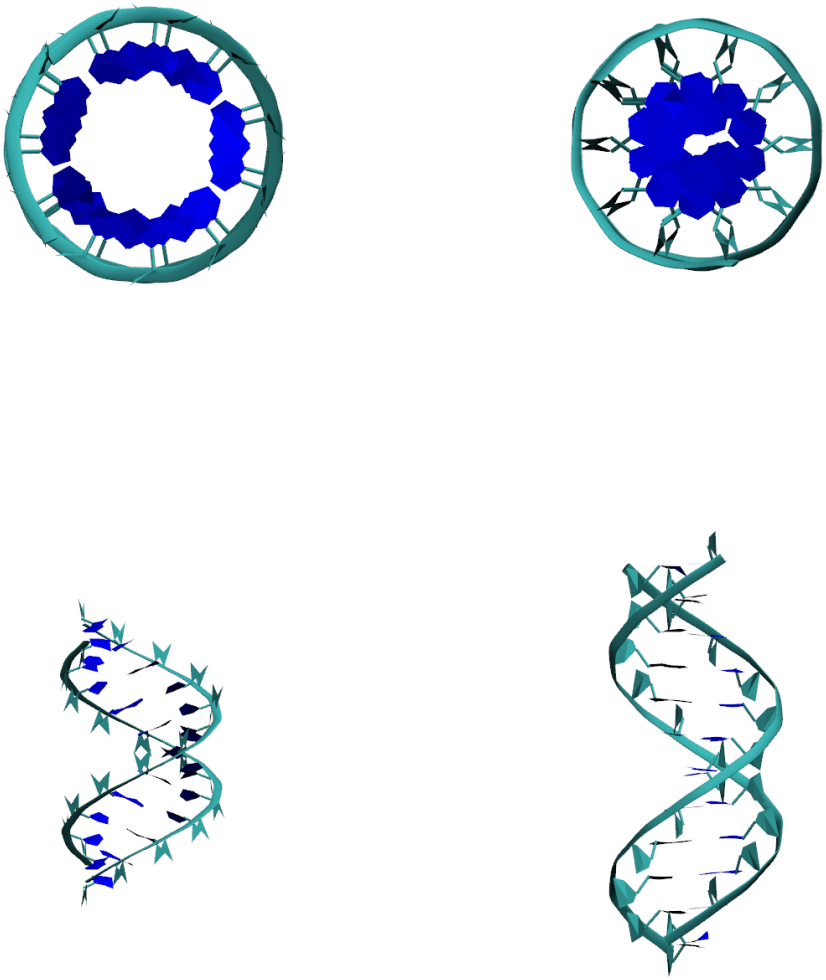


Figure 4.1: Ribbon representation of systems discussed (in canonical forms) in this chapter. Upper-left: top-down view of 12nA. Upper-right: top-down view of 12pB. Lower-left: side view of 12nA. Lower-right: side view of 12pB.

compared with the a A_{20} series. This is very much anticipated as the length of the dodecamer section is only roughly 60% that of the icosamer and thus the structural rigidity of the dodecamer should be considerably larger than its counterpart.

Now for the four systems using sodium as counterions, it is well-established that the B-form is much more favoured than the A-form [49, 65]. The RMSd graph below revealed the fact that it is not an exceptional case for this dodecamer. In fact the RMSd for the 12n systems starting from A-form rises to the 5.5Å range within the first 2ns of simulation and stayed around the black line which indicates plausible change of form, whereas that for the system start-

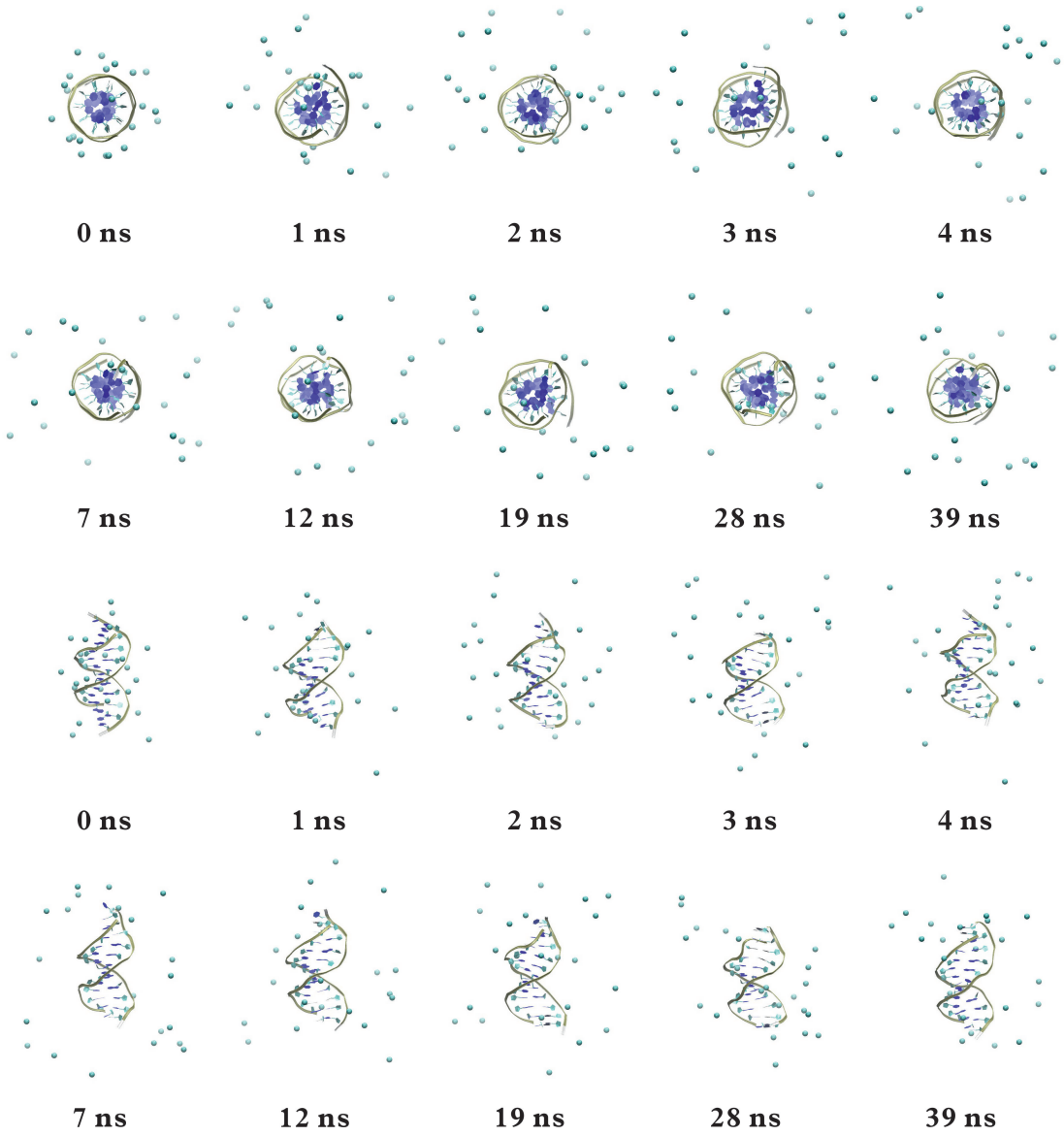


Figure 4.2: Snapshots (top-down and side view) of 12nB-Cs0.

ing with B-form remained at about 4Å , which signifies that a permanent shift from B to A is unlikely. The curves for the 12p systems further show that B-form may be even more favourable than in the 12n case, since the curve for 12pA overshoots the 5.579Å line and remained well above the line from the 15th nanosecond.

For the systems using caesium as counterions, as we have little data from previous studies, it is very hard to accurately predict the overall performance. Nevertheless it could be an educated guess that the effect of using caesium would be more diluted than sodium despite the larger atomic mass (22.989770 amu for sodium and 132.90545 amu for caesium) [38]. This is because, from a chemical point of view, caesium has a much lower first ionisation enthalpy than sodium (3.89390eV for caesium; 5.13908eV for sodium) [38]. This means

that caesium is more likely give up its outermost 6s electron to form chemical (ionic) bonds with the negative backbone, in turn implying the larger affinity with the backbone rather than repulsion, thus higher stability for the whole DNA-cation system. This is clearly shown by the curves of the 12p systems - even the unfavourable A-form resides beneath the black line - indicating that the system is so stable that neither of the forms changed permanently. However the visualisation from Figure 4.2 gives a totally different reasoning to the phenomenon: caesium is so large that the mutual repulsion led the ions to disperse within the first nanosecond. Hence the caesium ions have even less interaction with the DNA. The mode of action of caesium on DNA is thus also very different to those of sodium and magnesium, much smaller in sizes, which manipulate the motion of the DNA by attaching to the backbone. Caesium ions form a large "cage" of steep potential wells which traps the DNA molecule inside, thus stabilising it.

Figures 4.4 to 4.7 show the histograms for base pair parameters of four spermidine-free *non-palindromic* systems.

From these histograms one can observe that no matter whether the system is started from A- or B-form, or regardless of the usage of sodium or caesium as the counterions, all the histograms take a near-perfect bell-shape, only with a little skew. This indicates that the 12n systems are rather stable by themselves. Moreover, looking into the characteristics of the curves closely, it is obvious that both the shapes, the widths and the positions of the histograms of sodium systems are very similar to those of caesium systems, showing that the effects of caesium on DNA conformation may resemble those of sodium.

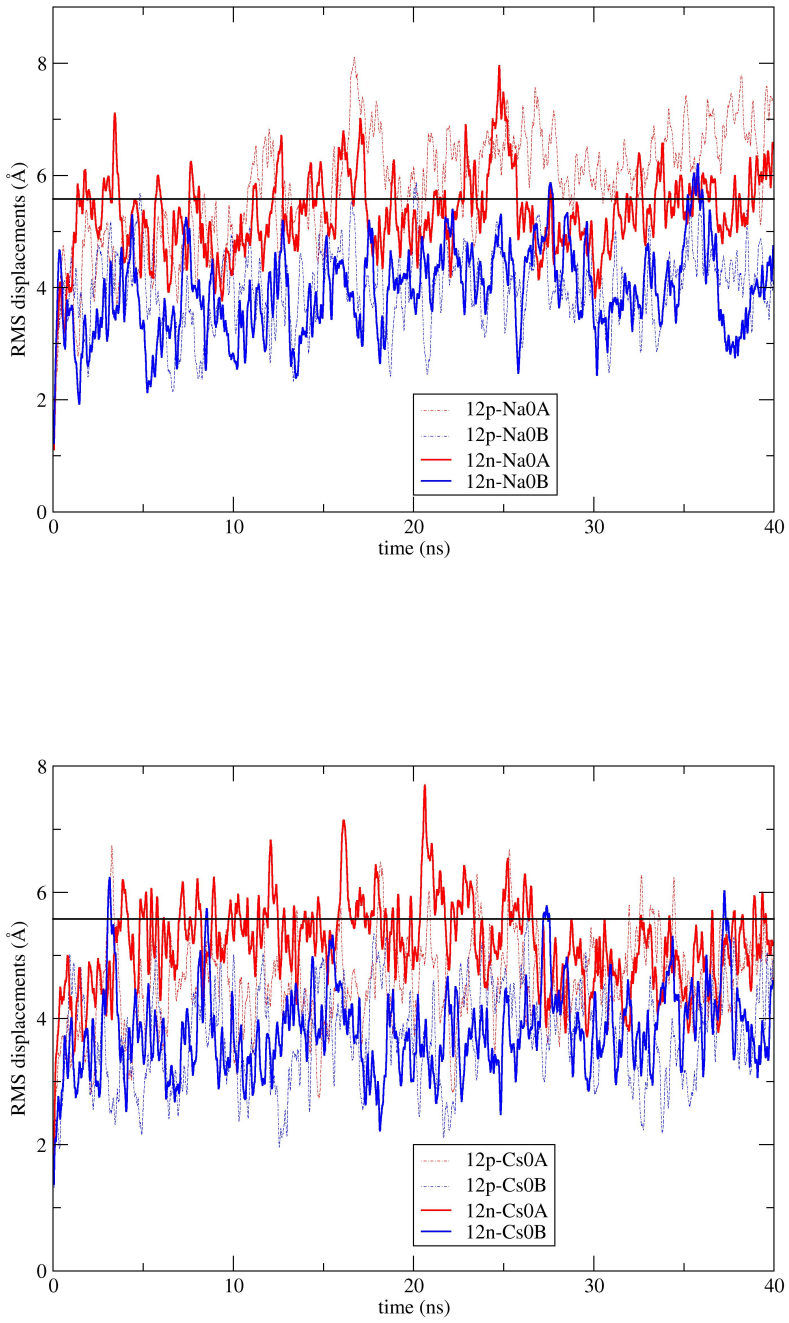


Figure 4.3: RMSd graphs of SPD-free systems using sodium and caesium as counterions. Black horizontal line: RMSd between canonical A- and B-forms.

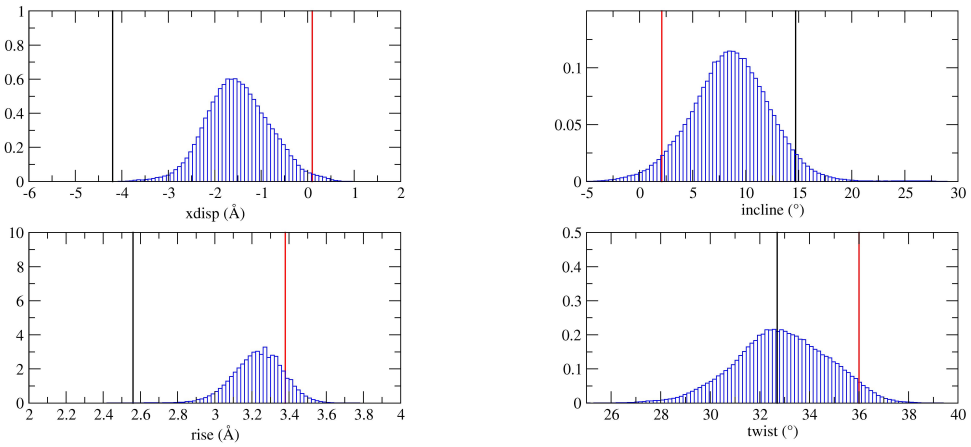


Figure 4.4: Base pair parameters histograms of 12nA-Na0.

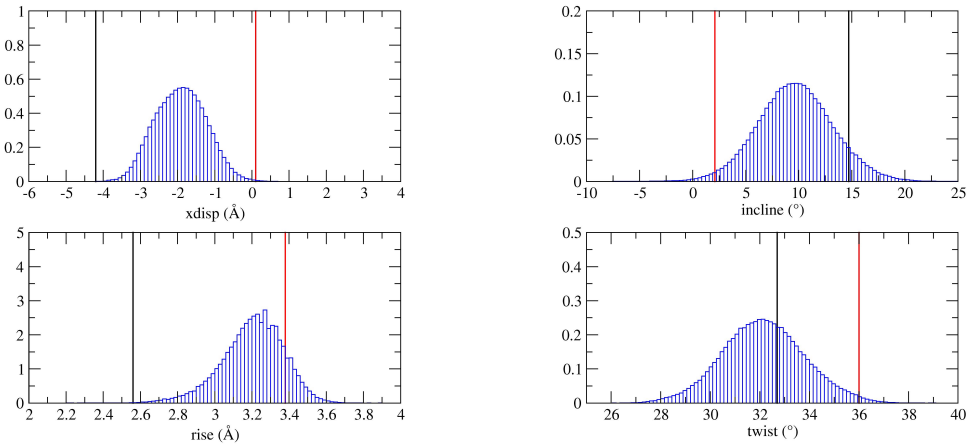


Figure 4.5: Base pair parameters histograms of 12nA-Cs0.

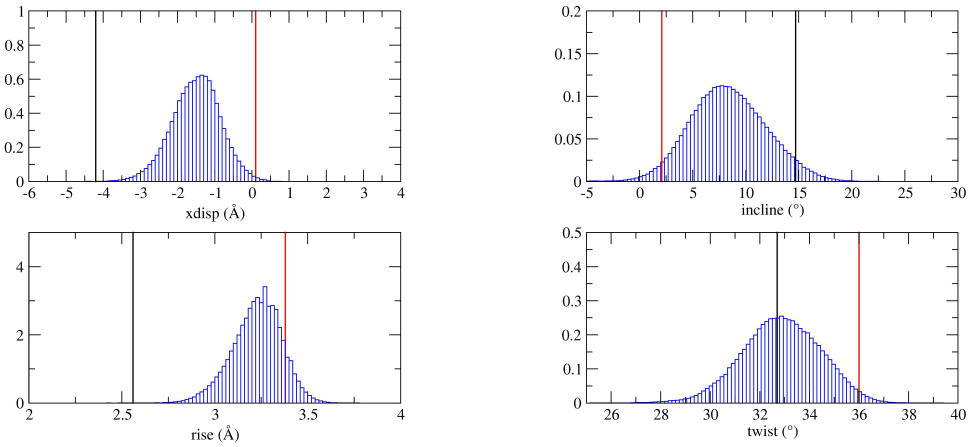


Figure 4.6: Base pair parameters histograms of 12nB-Na0.

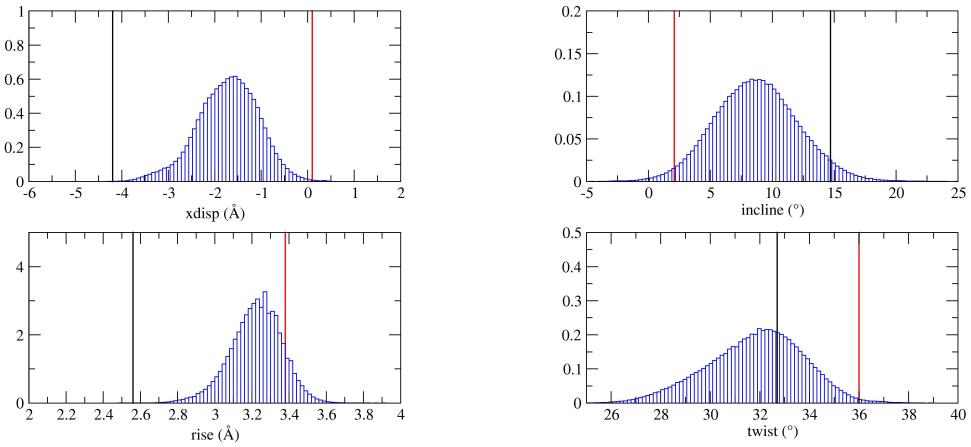


Figure 4.7: Base pair parameters histograms of 12nB-Cs0.

Figures 4.8 and 4.9 show the histograms for base pair parameters of spermidine-free *palindromic* caesium systems. We can broadly divide the four sets of graphs into two groups: A-start and B-start. The shapes of histograms for the B-start systems are highly symmetric which means the starting configuration (B-form) is highly favoured and stable, while those for the A-start systems are slightly skewed towards values of canonical B-form implying the tendency of a change in conformation from A to B.

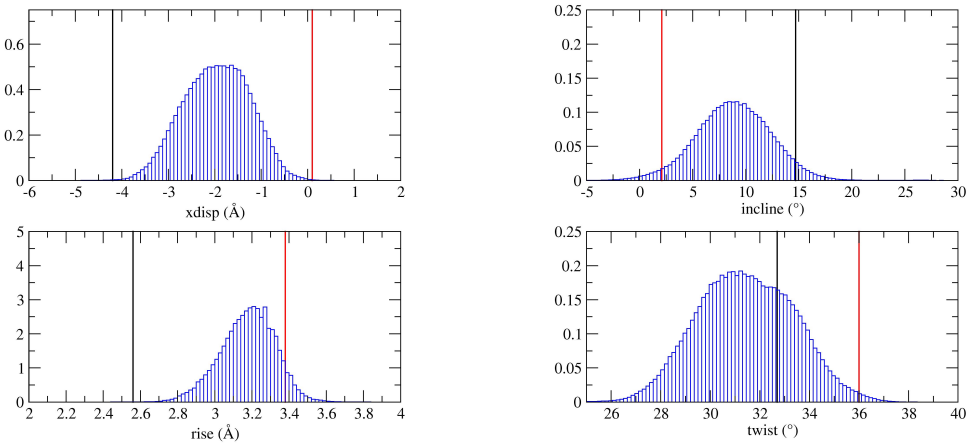


Figure 4.8: Base pair parameters histograms of 12pA-Cs0.

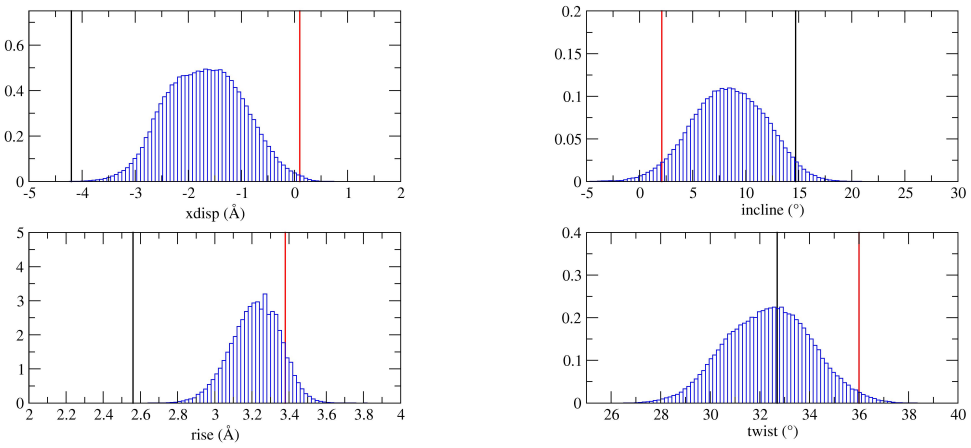


Figure 4.9: Base pair parameters histograms of 12pB-Cs0.

4.2.2 Mono-spermidine systems

In this subsection the eight remaining dodecamer systems with one spermidine in the major groove, namely 12nA-Na1, 12nB-Na1, 12nA-Cs1, 12nB-Cs1, 12nA-Cs1, 12nB-Cs1, 12pA-Cs1 and 12pB-Cs1, are analysed.

Fig. 4.10 shows the snapshots of the MD simulation of the 12pB-Cs1 system at different time steps. It is observed that the system stayed, throughout the 40ns period of simulation, in a very near-B form.

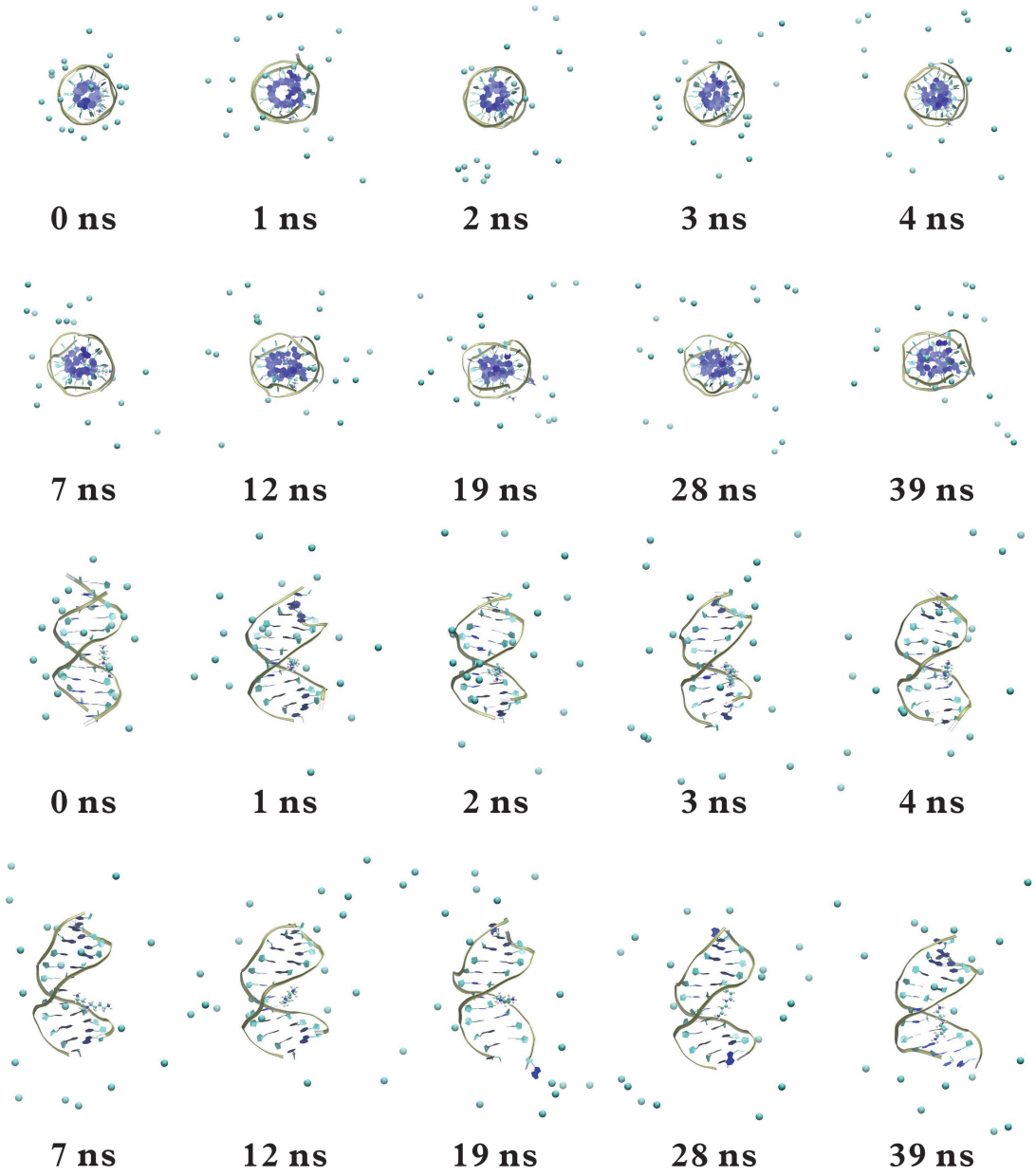


Figure 4.10: Snapshots (top-down and side view) of 12pB-Cs1.

The graphs of spermidine-free systems using sodium and magnesium as counterions showing instantaneous RMSd of individual system from canonical structures are presented in Fig. 4.11. For the systems using sodium as counterions, it

is observed that the behaviour of those with one spermidine resembles rather closely those with no spermidine molecules, i.e. the RMSd curves for A-start systems stay around the transition threshold value while those for B-start systems mostly stay below it. Moreover, with one spermidine introduced, the behaviour of the *palindromic* sequences are now much more similar to the *non*-palindromic counterparts in that the shapes of the time-relaxed RMSd curves bear great resemblance.

For the systems using caesium as counterions, the RMSd curves for all of the four systems suggest once again, that spermidine has stabilising power, as it retarded the form transition from A to B for both palindromic and non-palindromic sequences, keeping the RMSd curve well under the threshold line for more than 20 nanoseconds.

Figures 4.12 to 4.15 show the histograms for base pair parameters of four mono-spermidine *non-palindromic* systems while figures 4.16 to 4.19 show those for base pair parameters of four spermidine-free *palindromic* systems.

From these eight charts we further confirm our earlier assertion that since both sodium and caesium are monovalent ions, their effects on the DNA should be very similar. This is readily seen from the shapes of the histograms - the heights, the width and the position of the peaks are all very much alike each other.

However there is another important conclusion that could be drawn from the histograms: the B-form is much less strongly favoured over the A-form for caesium systems than sodium systems. This is rather obvious from the graphs for A-start configurations. For A-start sodium systems, several histograms such as xdisp, rise and twist are all skewed towards the "perfect B" values. In some cases a double-peak might even appear suggesting a strong tendency of an A-to-B transition. However this is not observed in the caesium cases, in which all the histograms are nicely bell-shaped and have little skewness, showing a strong stability in the initial state regardless of A-start or B-start.

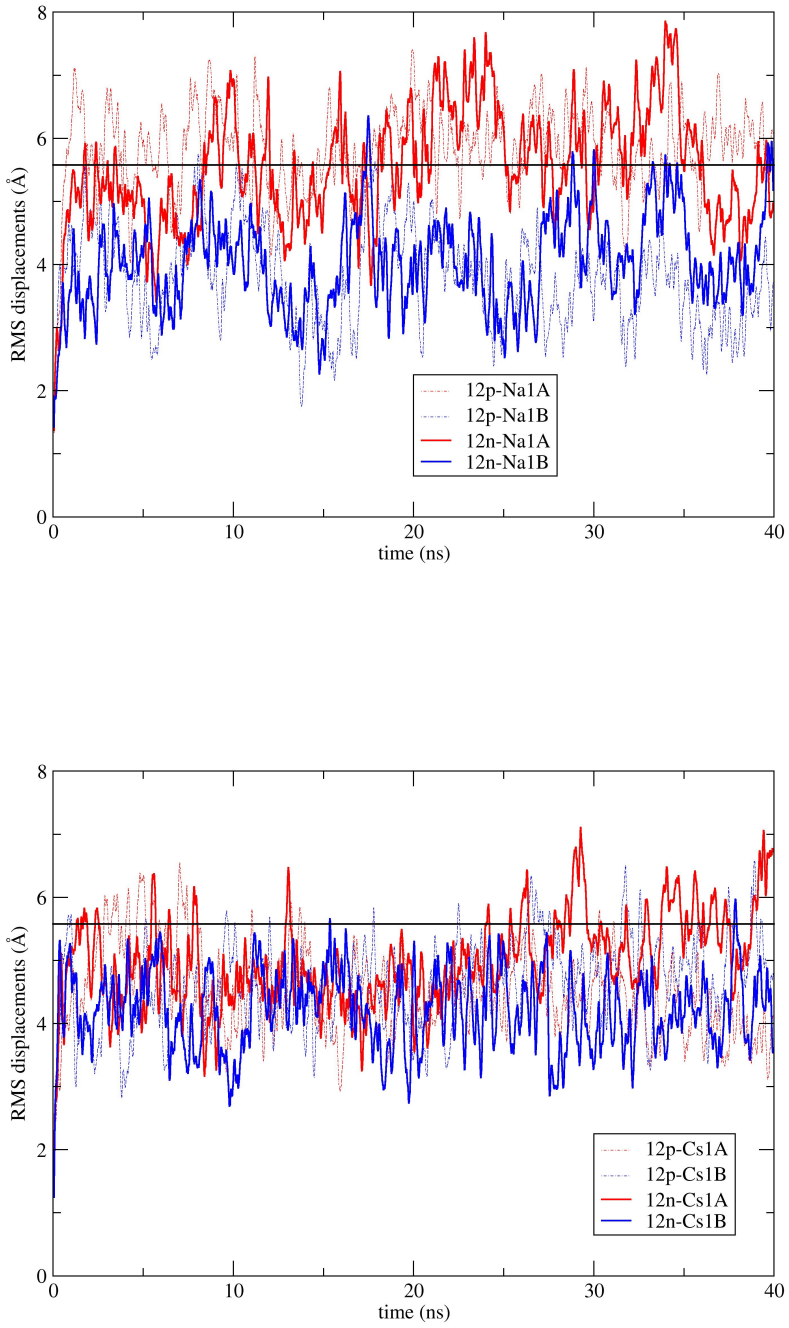


Figure 4.11: RMSd graphs of mono-SPD systems using sodium and caesium as counterions.

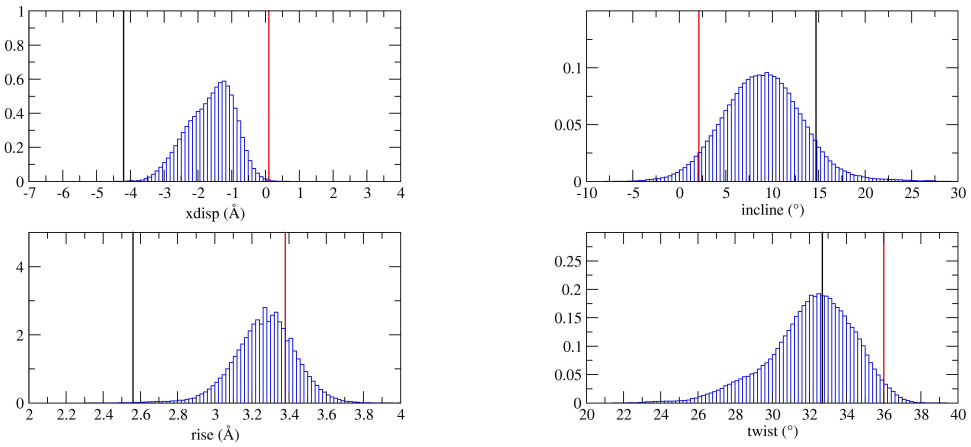


Figure 4.12: Base pair parameters histograms of 12nA-Na1.

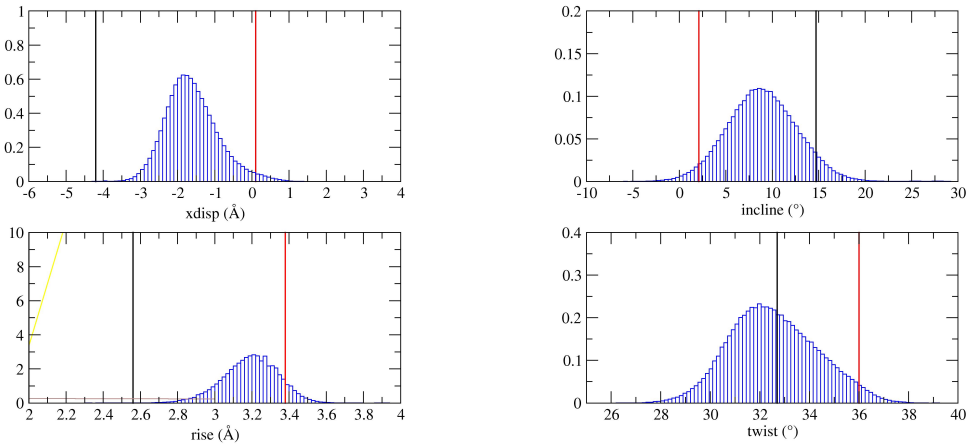


Figure 4.13: Base pair parameters histograms of 12nA-Cs1.

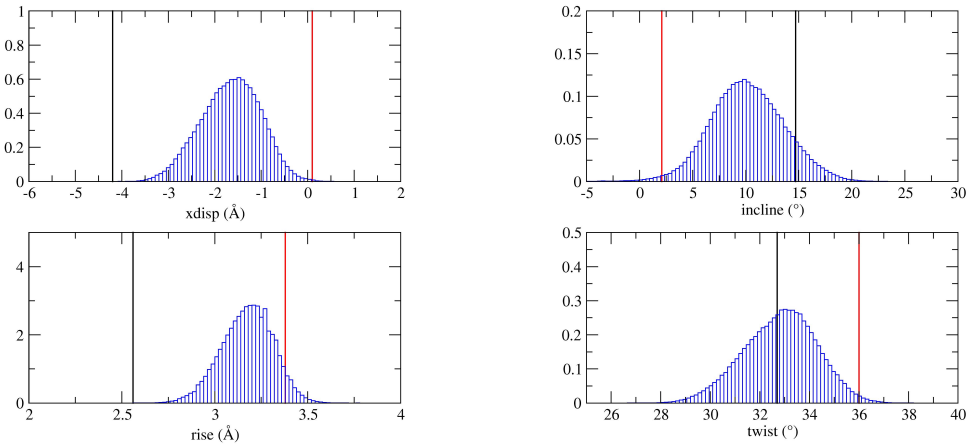


Figure 4.14: Base pair parameters histograms of 12nB-Na1.

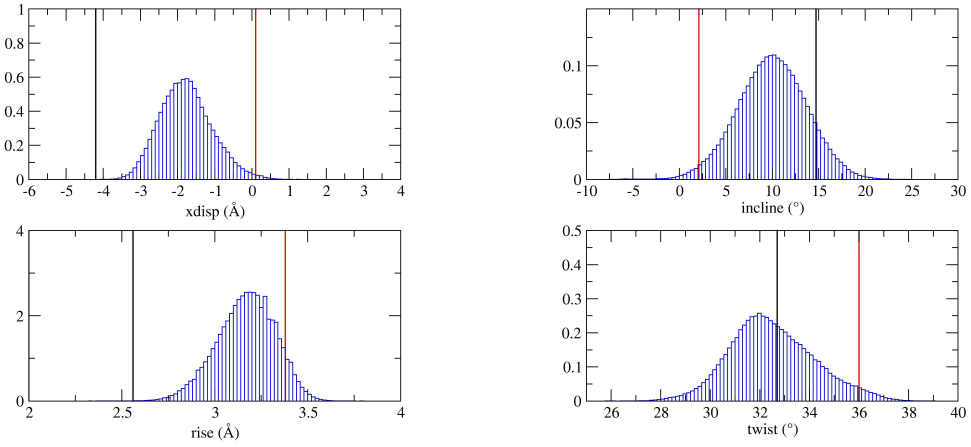


Figure 4.15: Base pair parameters histograms of 12nB-Cs1.

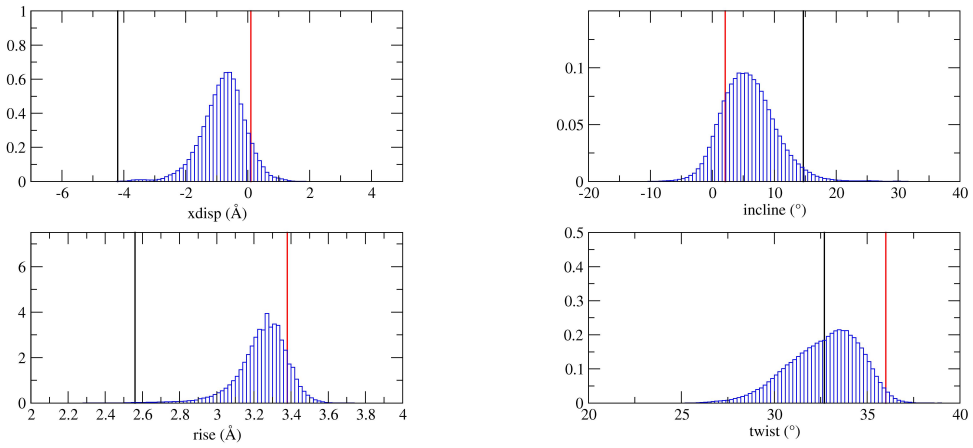


Figure 4.16: Base pair parameters histograms of 12pA-Na1.

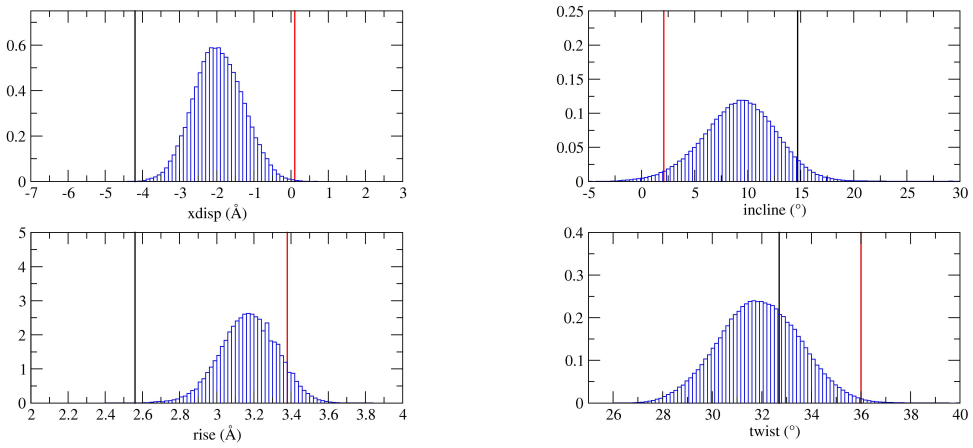


Figure 4.17: Base pair parameters histograms of 12pA-Cs1.

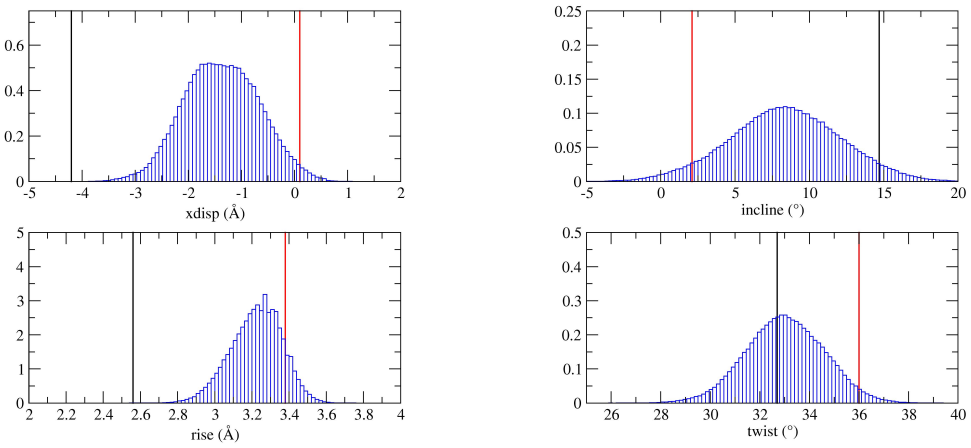


Figure 4.18: Base pair parameters histograms of 12pB-Na1.

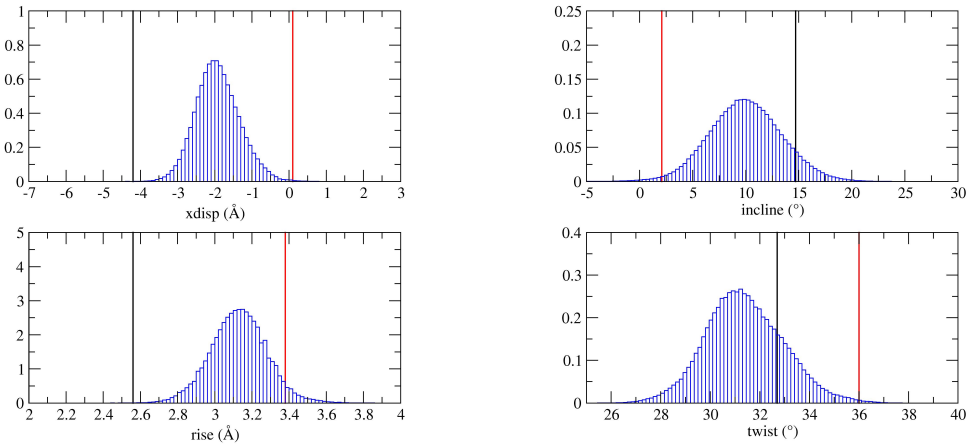


Figure 4.19: Base pair parameters histograms of 12pB-Cs1.

4.2.3 Summary and discussions

In this chapter we discussed about the results obtained from the MD simulation of the 12 base-pair sequences d(ACCGGCGCCACA) and d(ACCGGCGCCGGT). It was found that although caesium is a monovalent ion, it does not have the same effect on DNA and the mode of action of it on DNA is also quite different from that of sodium. Systems with caesium as counterions do not have such a strong tendency, as those with sodium, to adopt a near-B form. They have a slight preference to stay in the initial configuration.

The characteristics of histograms for the structural parameters of each system

will be tabulated below in the same format as in chapter 3. Since all the histograms are unimodal and resemble, to a certain extent a Gaussian distribution, the data will be presented in the format of "p, w" where p and w denote the x value of the peak and the FWHM of the histogram respectively.

System	xdisp (Å)	incline (°)	h-rise (Å)	h-twist (°)
12nA-Na0	**-1.52**, 1.56	**8.65**, 7.96	**3.28**, 0.28	**32.65**, 4.40
12nA-Na1	**-1.17**, 1.70	**9.60**, 9.78	**3.28**, 0.33	**32.60**, 4.66
12nA-Cs0	**-1.80**, 1.79	**9.95**, 7.99	**3.28**, 0.33	**32.20**, 3.75
12nA-Cs1	**-1.80**, 1.44	**8.80**, 8.54	**3.22**, 0.33	**32.05**, 4.13
12nB-Na0	**-1.30**, 1.46	**7.95**, 8.39	**3.28**, 0.27	**32.95**, 3.76
12nB-Na1	**-1.44**, 1.58	**10.05**, 7.82	**3.22**, 0.33	**33.10**, 3.44
12nB-Cs0	**-1.52**, 1.53	**8.50**, 7.79	**3.28**, 0.28	**32.05**, 4.29
12nB-Cs1	**-1.70**, 1.57	**10.40**, 8.55	**3.20**, 0.37	**32.05**, 3.61
12pA-Na0	**-1.50**, 1.84	**9.50**, 13.39	**3.28**, 0.29	**29.00**, 7.65
12pA-Na1	**-0.52**, 1.39	**4.60**, 9.63	**3.28**, 0.22	**33.60**, 4.61
12pA-Cs0	**-1.60**, 1.93	**8.65**, 8.07	**3.22**, 0.33	**31.10**, 5.11
12pA-Cs1	**-2.10**, 1.60	**9.35**, 7.75	**3.18**, 0.36	**31.75**, 3.93
12pB-Na0	**-1.82**, 1.21	**8.00**, 7.71	**3.28**, 0.29	**31.45**, 3.71
12pB-Na1	**-1.57**, 1.82	**8.50**, 8.42	**3.28**, 0.29	**32.95**, 3.72
12pB-Cs0	**-1.64**, 1.98	**8.65**, 8.79	**3.28**, 0.30	**32.95**, 4.30
12pB-Cs1	**-2.00**, 1.30	**10.05**, 7.73	**3.16**, 0.33	**31.30**, 3.51
Canon. A	**0.1**	**14.7**	**2.8**	**32.5**
Canon. B	**-4.2**	**2.1**	**3.3**	**36.5**

Table 4.1: Summary of modes of histograms of DNA-cation systems in chapter 4. Values for canonical forms [56] given at the bottom of table.

From Table 4.1 which summarised the characteristics of the histograms, we discovered that systems using caesium as counterions generally have higher xdisp values than those using sodium as counterions regardless of their starting conformation. This suggests that the conformational preference of B to A is much less for caesium systems than for sodium systems.

Secondly if we look closer to the FWHM of the xdisp and incline distributions, we see that the FWHM of both distributions for *non*-palindromic systems seems to be generally larger with spermidine added than without, whereas that for palindromic systems mostly decrease with the insertion of spermidine. This could potentially be attributed to the fact that the higher symmetry of the palindromic systems are more easily stabilised than the non-palindromic systems which are much less symmetric.

Though not performed in this project, simulations with more spermidine molecules present in the systems could be carried out. From the studies of the 20-base pair sequences in the previous chapter, we assert that the width of the bell-shaped histograms could decrease with increasing number of spermidines, as

spermidine was shown to have some stabilising power to DNA.

Chapter 5

Epilogue

5.1 Spermidine

Spermidine has been one of the main focuses of this project, and its effect on the structure of four different DNA sequences - $d(A)_{20}$, $d(A\text{-}T)_{20}$, d(ACCGGCGCCACA) and d(ACCGGCGCCGGT) was studied.

It was shown that in the presence of spermidine molecules, the transition of DNA from A-form to B-form is considerably slowed down, hence suggesting that spermidine might have stabilising power. However it is also demonstrated, rather non-intuitively, that the stabilising power brought by spermidine, is not as simple as a linear function of the number of polyamines present. This suggests that it is a highly complicated situation in which all the factors including the positions of the spermidine molecules, the positions of the counterions, the length and sequence of the DNA section, have to be considered. Furthermore, the initial placement of spermidine before simulation also plays a crucial role in the long-term effect of the spermidines on DNA. This is because there are local minima in the electrical potential between two backbone strands which could potentially trap the spermidine at the "portals", preventing it from diffusing freely into and out of the grooves.

5.2 Ions

It has been well known for long, that the choice of counterions around a DNA section has profound impact on which structural form would be preferred to others. For example, in the presence of sodium (above a certain concentration), the B-form is very much preferred to the A-form. This has been confirmed in all of our simulation studies.

Moreover, we have also shown that although it is a monovalent ion like sodium, caesium does not induce a strong structural preference of B-form over A-form. But rather it provides an extra stability to the initial configuration by restricting the movement of the DNA. This might be because caesium ions trap the DNA molecule in their steep potential wells.

It was predicted that magnesium could bring a much stronger stabilising power than sodium because of its doubly positive charge, but we discovered that it rather brought disruption to the system than stabilising it. This could be because the force field induced by magnesium is much more complex than sodium (or caesium) as the latter has one outermost electron and is thus hydrogen-like whereas magnesium has two outermost (3s) electrons which makes the situation more complicated quantum-mechanically.

5.3 Sequence-specificity

We discovered that even with the same initial conditions, such as the initial placement and number of spermidine molecules, the starting conformation of DNA and choice of counterions, the effects of the spermidine and ions can be very different. This suggests that such effects are sequence-specific.

We analysed the stabilising power with regard to the retardation of structural transition and the degree to which the structural histograms resemble Gaussian distributions - the higher extent of retardation or the more suppressed fluctuation, the more powerful the spermidine or ions are in stabilising the DNA system. We assert that the stabilising effect declines with the symmetry of the sequence, i.e. the more symmetric the sequence the less it could be stabilised by the ions. By symmetry, we mean the regularity of the sequences. For instance, palindromes are more symmetrical than non-palindromes, and repetitive sequences (with motif consisting of two or more bases) are more symmetrical than non-repetitive ones. Hence, d(ACCGGCGCCGGT) is *slightly* more symmetrical than d(ACCGGCGCCACA) because it is palindromic but the latter is not; d(A-T)$_{20}$ is *much* more symmetrical than d(A)$_{20}$ since it is both palindromic and repetitive but the latter is neither palindromic nor repetitive. To simply put, the degree of symmetry could be arranged thus: NP+NR $<$ P+NR $\ll$ NP+R $<$ P+R, where P, R, NP and NR stand for palindromic, repetitive, non-palindromic and non-repetitive respectively.

Therefore, it is justifiable that spermidine has much stronger effects on the asymmetric d(A)$_{20}$ than the highly symmetric d(A-T)$_{20}$, whereas the effect on the palindrome d(ACCGGCGCCGGT) is just a little stronger than that on the

non-palindrome d(ACCGGCGCCACA).

5.4 Future Work

The simulation of DNA-polyamine-cation systems, aiming to study the effects of biological polyamines and cations (in particular, mineral ions such as sodium and magnesium) has been proven to be successful and fruitful through this project on spermidine and previous studies on spermine. However this is for certain not the end of the story, as there remains a lot of possibilities and interesting aspects, regarding the physical interactions between polyamines and DNA in the presence of neutralising cations, which could be explored in the future.

For instance, smaller polyamines such as putrescine ($H_3N^+ - (CH_2)_4 - N^+H_3$) and cadaverine ($H_3N^+ - (CH_2)_5 - N^+H_3$) could be introduced to the systems for further simulation to see how other straight-chain polyamines have reactions with DNA and how they affect the conformation of DNA.

Furthermore, in the future studies, the DNA sections could be divided into groups in accordance to their degree of symmetry. This way the effects of the polyamines on DNA could hopefully be investigated more systematically.

As computational costs become cheaper as a result of technological advancement, even more complex systems consisting of longer DNA sequences, different solution environments and more uncommon cations could be considered to be simulated to give broader insights of different solutes and solutions on the structure of DNA.

Last but not least, theoretical (computational) simulations of any systems can only give a rather brief overview of how things are related to each other; and they have certain limitations which cannot be surmounted. Therefore it is also vital that the results obtained from such research be taken into the laboratory to test experimentally. By doing so we could hopefully have more all-rounded understanding about the roles of polyamines in the cell.

There are numerous possible further investigations related to this project as this is a broad field of studies. Some of them are listed below:

1. Choose one specific sequence (e.g. ATB) from this thesis and test with different counterions down the group (e.g. Li^+, Na^+, K^+, Rb^+ and Cs^+).

2. Choose one specific sequence (e.g. ATB) from this thesis and test with different counterions across the period (e.g. Na^+, Mg^{2+} and Al^{3+}).

3. Choose one specific sequence (e.g. ATB) from this thesis and compare effects of different polyamines (e.g. spermine vs spermidine).

Bibliography

[1] B.J. Alder and T.E. Wainwright. Phase transitions for a hard sphere system. *J. Chem. Phys.*, 27:1208–1209, 1957.

[2] M.P. Allen and D.J. Tildesley. *Computer Simulation of Liquids*. Oxford Science Publications, 2008.

[3] D.J. Anderson, J. Crossland, and G.G. Shaw. The actions of spermidine and spermine on the central nervous system. *Neuropharmacology*, 14:571–577, 1975.

[4] H.J.C. Berensen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, and J.R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[5] K. Bryson and R.J. Greenall. Binding sites of the polyamines putrescine, cadaverine, spermidine and spermine on A- and B-DNA located by simulated annealing. *J. Biomol. Struct. Dyn.*, 18(3):393–411, 2000.

[6] C. Carter. *The Neuropharmacology of Polyamines*. American Press, 1994.

[7] F.B. Carvalho, C.F. Mello, P.C. Marisco, R. Tonello, B.A. Girardi, J. Ferreira, M.S. Oliveira, and M.A. Rubin. Spermidine decreases Na^+, K^+-ATPase activity through NMDA receptor and protein kinase g activation in the hippocampus of rats. *Eur. J. Pharmacol.*, 684:79–86, 2012.

[8] D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman. *Amber 12: Reference Manual*. University of California, San Diego, 2012.

[9] E. Chargaff, R. Lipshitz, and C. Green. Composition of the deoxypentose nucleic acids of four genera of sea-urchin. *J. Biol. Chem.*, 195:155–160, 1952.

[10] IUPAC-IUB Commission. Abbreviations and symbols for the descriptions of conformations of polynucleotide chains. *European Journal of Biochemistry*, 131:9–15, 1983.

[11] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1998.

[12] N.E. Davidson, H.A. Hahm, Woster P.M. McCloskey, D.E., and R.A. Casero Jr. Clinical aspects of cell death in breast cancer: the polyamine pathway as a new target for treatment. *Endocr Relat Cancer*, 6:69–73, 1999.

[13] M.J.S. Dewar and W. Thiel. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.*, 99(15):4899–4907, 1977.

[14] R.E. Dickerson. DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acid Res.*, 26(8):1906–1926, 1998.

[15] R.E. Dickerson and T.K. Chiu. Helix bending as a factor in protein/DNA recognition. *Biopolymers*, 44(4):361–403, 1997.

[16] S. Diekmann. Definitions and nomenclature of nucleic acid structure parameters. *The EMBO Journal*, 8(1):1–4, 1989.

[17] E.A. dos Reis, L.S. de Oliveira, M.L. Lamers, C.A. Netto, and A.T. Wyse. Arginine administration inhibits hippocampal Na^+, K^+-ATPase activity by polyamines in the rat renal medullary cells of the thick ascending limb of Henle's loop. *J. Endocrinol.*, 127:377–382, 2002.

[18] D. Elson and E. Chargaff. On the deoxyribonucleic acid content of sea urchin gametes. *Experientia*, 8(4):143–145, 1952.

[19] B.G. Feuerstein, N. Pattabiraman, and L.J. Marton. Molecular mechanics of the interactions of spermine with DNA: DNA bending as a result of ligand binding. *Nucleic Acids Res.*, 18(5):1271–1282, 1990.

[20] R.E. Franklin and R.G. Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171:740–741, 1953.

[21] W. Fuller and W.H.F. Wilkins. The molecular configuration of deoxyribonucleic acid: IV. X-ray diffraction study of the A form. *J. Mol. Biol.*, 12 (1):60–76, 1965.

[22] S. Furberg. On the structure of nucleic acids. *Acta Chem. Scand.*, 6:634–640, 1952.

[23] Lee H. and Cai W. Ewald summation for Coulomb interactions in a periodic cell. 2009.

[24] J.M. Haile. *Molecular Dynamics Simulation: Elementary Methods*. John Wiley & Sons, Inc., 1992.

[25] I.S. Haworth, D. Rodger, and W.G. Richards. A molecular dynamics simulation of a polyamine-induced conformational change of DNA. A possible mechanism for the B to Z transition. *J. Biomol. Struct. Dyn.*, 10(1):195–211, 1992.

[26] S.R. Holbrook, A.H.J. Wang, A. Rich, and S.H. Kim. Local mobility of nucleic acids as determined from crystallographic data: Z-form DNA. *J. Mol. Biol.*, 187(3):429–440, 1986.

[27] K. Hoogsteen. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica*, 16:907–916, 1963.

[28] J.R. Hook and H.E. Hall. *Solid State Physics*. Manchester Physics Series. John Wiley & Sons, second edition edition, 1991.

[29] A.J. Hopfinger and R.A. Pearlstein. Molecular mechanics force-field parametrization procedures. *J. Comp. Chem.*, 5(5):486–499, 1984.

[30] M. Kaeberlein. Spermidine surprise for a long life. *Nat. Cell. Biol.*, 11: 1277–1278, 2009.

[31] W. Kohn and L.J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133–A1139, 1965.

[32] N. Korolev, A.P. Lyubartsev, A. Rupprecht, and L. Nordenskiöld. Competitive binding of Mg^{2+}, Ca^{2+}, Na^{+}, and K^{+} ions to DNA in oriented DNA fibers: Experimental and monte carlo simulation results. *Biophysical Journal*, 77:2736–2749, 1999.

[33] R. Langridge, D.A. Marvin, W.E. Seeds, and H.R. Wilson. The molecular configuration of deoxyribonucleic acid: II. molecular models and their fourier transforms. *J. Mol. Biol.*, 2(1):38–62, 1960.

[34] C.A. Laughton and S.A. Harris. The atomistic simulation of DNA. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(4):590–600, 2011.

[35] R. Lavery, M. Moakher, J.H. Maddocks, D. Petkeviciute, and K. Zakrzewska. Curves+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.*, 37:5917–5929, 2009.

[36] A.G.W. Leslie, S. Arnott, R. Chandrasekaran, and R.L. Ratliff. Polymorphism of DNA double helices. *J. Mol. Biol.*, 143(1):49–72, 1980.

[37] M. Levitt. Computer simulation of DNA double-helix dynamics. *Cold Spring Harb Symp Quant Biol*, 47(1):251–262, 1983.

[38] David R. Lide. *CRC Handbook of Chemistry and Physics, Internet Version 2005*. <http://www.hbcnetbase.com>, CRC Press, Boca Raton, FL, 2005.

[39] P. Liu, N. Gupta, Y. Jing, and H. Zhang. Age-related changes in polyamines in memory-associated brain structures in rats. *Neuroscience*, 155:789–796, 2008.

[40] J.A. McCammon and S.C. Harvey. *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, 1987.

[41] J.A. McCammon, B.R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267:585–590, 1977.

[42] A.E. Moseley, M.T. Williams, T.L. Schaefer, C.S. Bohanan, J.C. Neumann, M.M. Behbehani, C.V. Vorhees, and J.B. Lingrel. Deficiency in Na, K-ATPase alpha isoform genes alters spatial learning, motor activity, and anxiety in mice. *J. Neurosci.*, 27:616–626, 2007.

[43] S.M. Oredsson. Polyamine dependence of normal cell-cycle progression. *Biochem. Soc. Trans.*, 31:366–370, 2003.

[44] M. Pasi, J.H. Maddocks, D. Beveridge, T.C. Bishop, D.A. Case, T.E. Cheatham III, P.D. Dans, B. Jayaram, F. Lankas, C. Laughton, J. Mitchell, R. Osman, M. Orozco, A. Pérez, D. Petkevičiūtė, N. Spackova, J. Sponer, K. Zakrzewska, and R. Lavery. μ-DNA: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acid Research*, 42(19):12272–12283, 2014.

[45] A.E. Pegg. Recent advances in the biochemistry of polyamines in eukaryotes. *Biochem. J.*, 234:249–262, 1986.

[46] M. Pietila, J.J. Parkkinen, L. Alhonen, and J. Janne. Relation of skin polyamines to the hairless phenotype in transgenic mice overexpressing spermidine/spermine N-acetyltransferase. *J. Invest. Dermatol.*, 116:801–805, 2001.

[47] H. Prast and A. Philippu. Nitric oxide as modulator of neuronal function. *Prog. Neurobiol.*, 64:51–68, 2001.

[48] Y. Ramot, S. Tiede, T. Bíró, M.H. Abu Bakar, K. Sugawara, M.P. Philpott, W. Harrison, M. Pietilä, and R. Paus. Spermidine promotes human hair growth and is a novel modulator of human epithelial stem cell functions. *PLoS ONE*, 6(7):e22564, 2011.

[49] A.N. Real. *Molecular Dynamics Simulations of AT-rich DNA and DNA-spermine Complexes*. PhD thesis, University of York, 2001.

[50] W. Saenger. *Principles of Nucleic Acid Structure*. Springer-Verlag, 1984.

[51] F. Scorcioni, A. Corti, P. Davalli, S. Astancolle, and S. Bettuzzi. Manipulation of the expression of regulatory genes of polyamine metabolism results in specific alterations of the cell-cycle progression. *Biochem. J.*, 354: 217–223, 2001.

[52] G.L. Seibel, U.C. Singh, and P.A. Kollman. A molecular dynamics simulation of double-helical B-DNA including counterions and water. *Proc. Natl. Acad. Sci. USA*, 82:6537–6540, 1985.

[53] N. Seiler and T. Schmidt-Glenewinkel. Regional distribution of putrescine, spermidine and spermine in relation to the distribution of RNA and DNA in the rat nervous system. *J. Neurochem.*, 24:791–795, 1975.

[54] J.W. Shepherd. Classical and quantum simulations of DNA/spermine systems. Master's thesis, University of York, 2015.

[55] J.C. Skou and M. Esmann. The Na, K-ATPase. *J. Bioenerg. Biomembr.*, 24: 249–261, 1992.

[56] J.R. Stagno, B. Ma, J. Li, A.S. Altieri, R.A. Byrd, and X. Ji. Crystal structure of a plectonemic RNA supercoil. *Nature Communications*, 3(901), 2012.

[57] J.J.S. Stewart. Optimization of parameters for semiempirical methods. I. Method. *J. Comp. Chem.*, 10(2):209–220, 1989.

[58] K. Sugiura, J.Z. Min, T. Toyo'oka, and S. Inagaki. Rapid, sensitive and simultaneous determination of fluorescence-labeled polyamines in human hair by high-pressure liquid chromatography coupled with electrospray-ionization time-of-flight mass spectrometry. *J Chromatogr A*, 1205:94–102, 2008.

[59] B. Tidor, K.K. Irikura, B.R. Brooks, and M. Karplus. Dynamics of DNA oligomers. *J Biomol Struct Dyn*, 1:231–252, 1983.

[60] Y. Timsit, E. Westhof, R.P.P. Fuchs, and D. Moras. Unusual helical packing in crystals of DNA bearing a mutation hot spot. *Nature*, 341:459–462, 1989.

[61] W.F. van Gunsteren and H.J.C. Berendsen. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.*, 29:992–1023, 1990.

[62] W.F. van Gunsteren, H.J. Berendsen, R.G. Geurtsen, and H.R. Zwinderman. A molecular dynamics computer simulation of an eight-base-pair DNA fragment in aqueous solution: comparison with experimental two-dimentional NMR data. *Ann. N. Y. Acad. Sci.*, 482:287–303, 1986.

[63] J.D. Watson and F.H.C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[64] S.J. Weiner, P.A. Kollman, D.T. Nguyen, and D.A. Case. An all atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.*, 1986.

[65] R.K. Williams. *Molecular conformational studies of deoxyribonucleic acid by potential energy minimisation with normal mode analysis*. PhD thesis, University of Keele, 1990.

[66] G.R. Wyatt. The nucleic acids of some insect viruses. *J. Gen. Physiol.*, 36, 1952.

[67] B.Y. Yee, Y.L. Mong, and T.L. Chan. *in progress*, 2014.