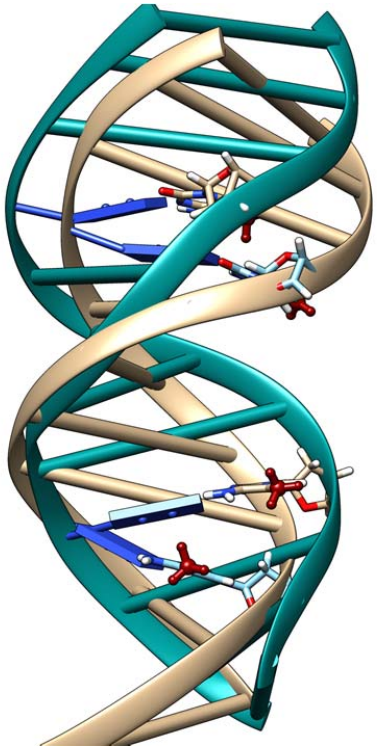
Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Department for Plant Biology

Computational studies on DNA: insights into structure and dynamics



Charan Raju Kanna

Independent degree project in biology/Master Thesis, 45 hp, A2E, Master's in Biotechnology, 2015.

Computational studies on DNA: insights into structure and dynamics

Author's name: *Charan Raju Kanna*

Supervisor: Dr. Lynn Kamerlin, Department of Cell and Molecular Biology, Uppsala University.

Internal Supervisor: Dr. Lars Hennig, Department of Plant Biology, SLU.

Assistant Supervisor: Alexandra Carvalho, Department of Cell and Molecular Biology, Uppsala University.

Examiner: Dr. Örjan Carlborg, Department of Clinical Sciences, Division of Computational Genetics.

Credits: 45

Level: A2E

Course title: Independent project in biology – master's thesis

Course code: EX0596

Programme/education: Master's in Biotechnology

Place of publication: Uppsala

Year of publication: 2015

Cover picture: Charan Raju Kanna

Online publication: http://stud.epsilon.slu.se

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Department for Plant Biology

Table of Contents

Abstract:

The best-characterized epigenetic modification is DNA methylation, which was found to shape significantly gene expression in mammalian cells, though bacteria and plants display less sequence specificity. The cytosine methylation of DNA sequence here was studied by the molecular dynamic (MD) method. We found that the optimized structures are not significantly altered between unmethylated, methylated sequences. However, sampling of the molecule dynamical landscape via long MD simulations showed that the altered sequences could adopt different conformations (revealed in the twist and roll angles) from the unmodified DNA, which are dependent on the §sequence environment. We propose that this conformational diversity may play a role on protein recognition.

The G-rich sequences are studied with MD simulations to check the propensity of the sequence to adopt to A-form or B-form of DNA. The results showed that the DNA sequence adopts most of the time in A-form and also spend sometime in between A-form and B-form. The groove widths showed the conformation in B-form. By analyzing the DNA in different contexts using computational methods we were able to test limits that force field methods offer for these systems. We propose that this conformational diversity may play a role in protein recognition.

Finally, we have studied the conformational changes of DNA sequence with uracil as a base. Sometimes in the replication process uracil can be formed and the mismatch base pair in the DNA sequence is removed by the enzyme UDG (Uracil-DNA sequence). In the present work MD simulations were used to study the structure, dynamics and base pair kinetics of DNA sequence. The opening and closing kinetics of the T-A base pair were studied. A fast base pair opening was observed and also additionally base flipping was spotted. A very fast opening of the bases may be responsible for the fast closing of the bases. In order to characterize the complete study of base pair kinetics in relation to T-A versus U-A base pairs is currently under way.

Keywords: Epigenetic modifications, DNA methylation, G-rich sequence, Uracil, Molecular dynamics, demethylation.

Abbreviations: DDD: Dickerson-Drew Dodecamer, DNA: Deoxyribonucleic acid, DNMT: DNA methyltransferases, MD: Molecular dynamics, LINES: Long interspred transposable elements, SINES: Short interspred transposable elements, TET: Ten-Eleven Translocation, MBD: Methyl binding domain, PBC: Periodic boundary conditions. UDG's: Uracil-DNA glycosylase.

1. Introduction

This work comprises three different studies on the dynamical structure of DNA. The first part is about DNA methylation and how this small change affects the overall DNA structure. The second part is about the propensity of sequences with a high GC content to adopt a particular DNA form. Finally, the last part is about DNA sequences containing uracil and how this affects their dynamics.

1.1 DNA Methylation:

Epigenetics is a recent and fast growing research area that deals with the study of changes in organisms caused by modification of gene expression. Epigenetic changes are persistent and even heritable changes that occur in gene activity and expression and that do not involve any change in DNA sequence. The gene expression can be altered in different tissues, diseases and during developmental stages. The epigenetic information in the altered gene can be passed to the offspring through the male germ line (Soubry et al., 2014) and maternal germ line (Muers, 2011) by the epigenetic mechanisms. These include DNA methylation, histone modifications and transcription of non-coding RNA's. The epigenotypes that are metabolic disease-related, have been proposed to also be transmitted by germ-line (Lavebratt et al., 2012).

DNA methylation is one of the main epigenetic modifications that play an important role in development, carcinogenesis and gene silencing (Temiz et al., 2012). Methylation occurs most frequently in cytosines at 5^{th} position. The introduction of the methyl group is catalyzed by DNA methyltransferases that transfer this group from the S-adenosyl-L-methionine to the DNA (Guz et al., 2010). The DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b maintain the methylation patterns during cell division especially in mammals (Rottach et al., 2009).

Regions with more CpG content are called "CpG islands". In the mammalian genome, approximately about 60 -70 % of CpG sites are methylated (Rottach et al., 2009). Gene silencing by DNA methylation can occur in two ways: i) transcriptional factors are prevented from binding to the promoters by methylation; ii) Methyl Binding Domain proteins (MBD) recognize the methylated sites and bind to them, which then further prevents the binding of RNA polymerase to the promoter (Ballestar and Wolffe, 2001).

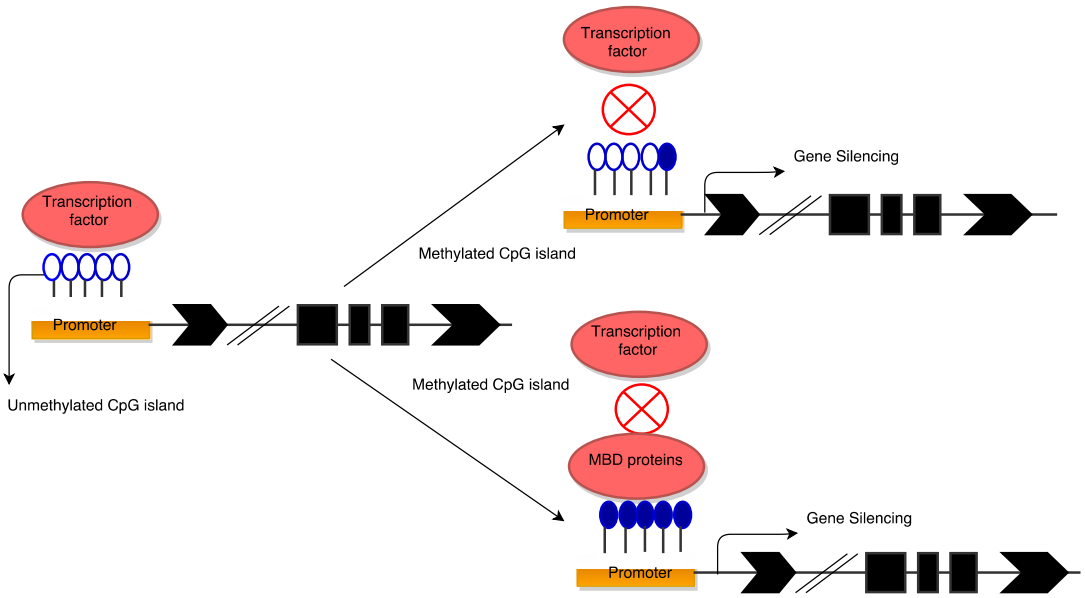


Figure 1: Gene silencing occurs either by directly preventing the binding of transcriptional factors or MBD proteins bind to the methylated sites (Urdinguio et al., 2009).

Another recently discovered epigenetic modification is hydroxymethylation, which is present in significant amounts in genome. Hydroxymethylation is an intermediate product formed in the demethylation process. Recent studies showed that 5-hydroxymethylation (5hmc) plays an important role in many diseases including Alzheimer's disease and high levels of 5hmc are present in brain and low levels are present in liver, lung and heart (Szulwach et al., 2011). TET (Ten-Eleven Translocation) proteins are responsible for the conversion of methylcytosine to hydroxymethylcytosine in the oxidation process (Haseeb et al., 2014). 5hmc is involved in the gene regulation patterns, as intermediate of demethylation process and also modulates the binding of proteins responsible for transcription regulation (Branco et al., 2012). The hydroxymethylation pattern shows effect on DNA flexibility and stability (Wanunu et al., 2010) and this epigenetic regulation is critical in aging, neurodevelopment and in human diseases (Szulwach et al., 2011).

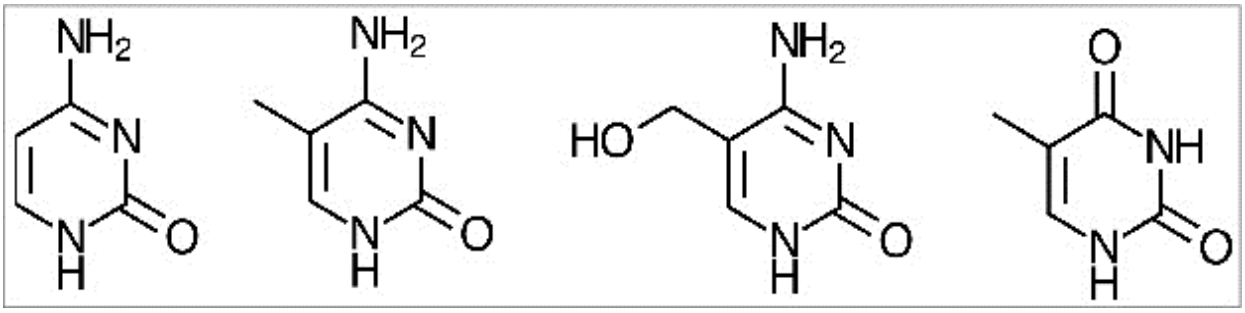


Figure 2: From left to right structures of Cytosine (C), 5-methylcytosine (5mc) and 5-hydroxymethylcytosine (5hmc) and thymine (Carvalho et al., 2014).

The epigenetic modifications show structural consequences on cytosine such as cytosine methylation, hydroxymethylated cytosine, formylcytosine and carboxycytosine. Some of the structures were shown in the above figure 2. The methylated cytosines mainly present in the genomic repetitive elements, intergenic regions and also in transposable elements such as Long Interspersed Transposable Elements (LINES) and Short Interspersed Transposable Elements (SINES) (Robertson, 2005). The DNA methylation patterns are essential for the normal functioning of organisms like insects and plants and for mammalian development. In somatic cells the loss of DNA methylation patterns leads to loss of growth control. Illegitimate recombination can occur when methylation is absent in the highly repetitive sequences (Robertson, 2005). Methylation plays an important role in disease by altering the gene expression. Many of the CpG sites are methylated in the intergenic regions and genomic repetitive sequences. Hypomethylation of the CpG sites takes place in cancer, where the chromatin becomes less densely packed and DNA can be transcribed (Bishop and Ferguson, 2015). In genomic repetitive elements and transposable elements the hypomethylation patterns leads to genome instability, DNA breakage and loss of imprinting (Bishop and Ferguson, 2015). Demethylation of promoters can lead to cancer (Suter et al., 2003) and also hypermethylation of promoters leading to gene silencing can result in cancer (Herman and Baylin, 2003).

A proper understanding DNA conformation is crucial since DNA is a very flexible polymer. Proteins recognize and bind DNA sequences either by directly interacting with the bases (direct readout) or by indirect readout, which is based on the sequence dependent conformational properties, such as bending and deformability (Napoli et al., 2006). DNA flexibility is dependent on torsional flexibility and bending and these parameters have important biological consequences. They play a significant role in packing of the molecule. The sequence dependent conformations of DNA can be described by six conformational parameters, which are shown in figure 5. Shift, slide and roll are translational parameters and tilt, roll and twist are rotational parameters (Olson et al., 1998). Out of these parameters twist is an important parameter for binding of protein to DNA (Fujii et al., 2007). The bending property is sequence dependent and can reflect the changes in the backbone geometry and the stability of the base steps (Travers, 2004). The twisting of the molecule is related to bending and unbending of DNA, which may have energetical conformations that occurs at base pair level (Travers, 2004). The mechanical stability of DNA can be altered by

methylation and the mechanical properties also lead to structural transitions of the molecule, which play a major role in biological functioning (Travers and Thompson, 2004).

The methylation patterns may affect the binding specificity of proteins or other ligands with DNA. In the present study, the cytosine methylation patterns were studied to understand the structural properties of DNA. The increased flexibility at the methylated regions could be important for the recognition of methylated cytosines by transcriptional factors, methylated CpG-binding proteins (MeCPs) involved in gene silencing (Carvalho et al., 2014).

1.2. G-Rich Sequences:

In bacteria, the levels of metals are important for the cells optimal growth and these levels are controlled by the metal-responsive machinery and intracellular metalloenzymes (Iwig et al., 2008). The bacterial metal signal responsive binding proteins are well-known proteins involved in the control of gene expression (the metal ions can increase or decrease the DNA binding affinity of the transcriptional factors). When the metal ions are present in excess concentration, they are expelled from the bacterial cell, which results in detoxification (Blaha et al., 2011).

RncR and CsoR proteins are recently discovered in bacterial families, which are signal responsive transcriptional factors that respond to a variety of environmental stresses. RcnR was discovered in *Escherichia coli* and CsoR was discovered in *Mycobacterium tuberculosis*.

In *E.coli,* the first biochemical member identified was RcnR, a transcriptional regulator of gene rcnA. RcnR controls the expression of metal efflux protein RcnA only with Nickel and Cobalt metal ions (Iwig et al., 2008). These RcnA proteins regulate the transcription of the gene encoding the Co and RcnA Ni. The RcnA proteins are resistant to Ni and Co ions and the expression of these proteins are induced by these two metal ions. According to Iwig and coworkers the RncR tetramer recognizes two TACT-G6-N-AGTA motifs of the rcnA-rcnR intergenic region, which has A-form DNA characteristics. The recognition is most likely through minor groove contacts leading to DNA wrapping around the protein (Iwig and Chivers, 2009). In this work, we studied this sequence desperation to MD simulations to check the propensity of the sequence whether to adopt the A-form or B-form.

1.3. Uracil:

Uracil is a unusual base found in DNA that can be formed by misincorpration of dUMP instead of dTMP during DNA synthesis and also can be formed by occasional deamination of cytosine to uracil (Doseth et al., 2011) which is shown in the figure 4. The misincorpration of dUMP leads to U:A pairing which is most common route to DNA uracil (Doseth et al., 2011). For every cell generation it is estimated that $\sim 10^4$ uracil residues are present in the human genome (Mosbaugh and Bennett, 1994). Spontaneous deamination has been estimated at a rate of 70-200 events per

day for a cell (Hagen et al., 2008). Deamination can occur spontaneously in the genome or by enzyme activation induced deaminase in Ig loci. Enzymatic deamination of cytosine can occur when S-adenosylmethionine (SAM) is low or when methyltransferases (MT) is high, it is observed in several tumors with mutagenic U-G which is harmful and whereas U-A pair is non-mutagenic. The mutagenic and recombinogenic pathways of regulating the processing of uracil are not well understood.

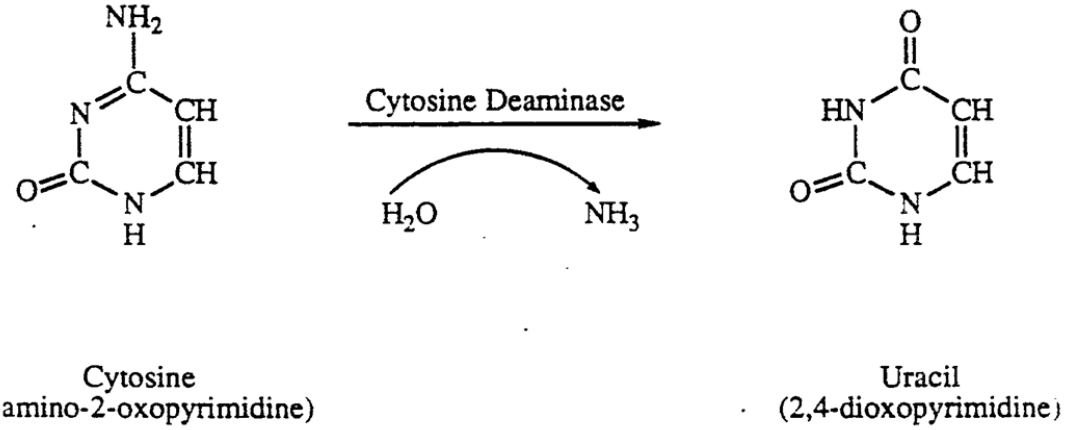


Figure 4: Spontaneous deamination of cytosine converts into uracil by cytosine deaminase (Senter et al., 1990).

Uracil DNA glycosylases (UDG's) are the class of enzymes expressed by mammalian cells are helpful in removal of uracil from DNA. These enzymes are capable of cleaving the *N*-glycosidic bond between deoxyribose and uracil leaving a substrate for base excision repair pathway (Doseth et al., 2011). The enzymes (UDG's), which are responsible for the repair pathway are classified into four super families. Family-1 enzymes are responsible for repairing uracil in ssDNA and dsDNA, they recognize the uracil in extrahelical conformation along with protein and water interactions. Family-2 enzymes are specific to mismatch and recognize guanine on the complimentary strand. The structures are not available yet for Family-3 and Family-4 still identifies the common active site but lacks a conservation of catalytic residues (Pearl, 2000). The process of identification of DNA lesions by the UDGs is particularly important. After the identification of the lesions, the first step is to extract the damaged base from the DNA into enzyme active site where the bond is cleaved between the base and deoxyribose sugar of the DNA backbone and then releasing the uracil base and abasic site (Parker et al., 2007). A template directed DNA polymerase inserts the missing nucleotide and with the help of DNA ligase, the backbone is closed. Cytosine deamination can also be reversed by uracil reamination to form the cytosine. This would be highly mutagenic because during replication sometimes DNA polymerase has the inability to distinguish between TTP and dUTP in template directed DNA synthesis (Pearl, 2000). Stacking interactions helps in maintaining the stability for the rest of the bases when the

repair base is pushed to the enzyme active site (Krokan et al., 2002). In the process of complete extraction of the base from the DNA i.e. base flipping requires the action of repair enzymes, partial base opening or base pair breathing, these are naturally occurring events in DNA dynamics which are influenced by thermal fluctuations (Krokan et al., 2002). To better understand the DNA-damage recognition mechanism the fundamental step is, characterization of structure and dynamics of these naturally occurring base pair opening events (Fadda and Pomès, 2011). Base pair opening is required in many fundamental processes such as transcription and recombination (Dornberger et al., 1999). The base pair opening kinetics were studied by NMR imino proton exchange experiments (Priyakumar and MacKerell, 2006). From these data it was suggested that the methylation of C5 of pyrimidines affects the base pair opening kinetics. The base pair opening kinetics in G-tracts is faster than the A-tracts which means the base pairing of T-A has a greater lifetime than in G-C (Dornberger et al., 1999). The Watson and Crick base pair lifetimes of GC and AT are observed approximately to be 10-50 ms and 1-5 ms respectively, whereas in the open base pairing it is observed to be in nanoseconds (Priyakumar and MacKerell, 2006). By using the experimental data we can extrapolate the base pair opening kinetics and the closing kinetics, otherwise it also can be estimated directly by molecular simulation techniques (Fadda and Pomès, 2011). In recent years computational techniques had proved to be more valuable in DNA recognition and repair mechanisms (Priyakumar and MacKerell, 2006). In the present work our aim is to study the structural and dynamics of the base pairs T-A in relation to U-A. By using molecular dynamic simulations our goal is to study the differences in DNA with U relative to T and how this will show the effect on the dynamics, structure and open base pair parameters.

2. Aim:

The aim of the work was to understand the structural and dynamic behavior of DNA by using molecular dynamic simulations.

3. Background:

The first molecular dynamic (MD) simulations of model liquids were performed by Metrapolis (Metrapolis et al., 1953), Alder and Wainwright (Alder and Wainwright, 1957). Since then MD simulations had a continuous progress and were applied in a large variety of scientific areas. In the present work all the classical MD simulations were done by using GROMACS software (Van der Spoel et al., 2005). MD simulations allow studying the time-dependent behavior of a system by producing a dynamical trajectory. The initial conditions of the system i.e; the positions and velocities of each particle should be known. The initial velocities are generated randomly by Maxwell distribution and then the MD simulations are performed by integrating the Newton's equation of motion as described below.

$$m_i \frac{d^2 ri}{dt^2} = f_i = -\frac{\partial}{\partial r_i} U(r_1 r_2, \dots r_N) \qquad (1)$$

Where $U(r_1 r_2, \dots r_N)$ is the potential energy of the system with N number of particles. The force field used in this study is amber99parmbsc0 and it was specially designed for nucleic acids (Pérez et al., 2007). This force field version has an improvement in the description of α/γ concerted rotations in nucleic acids when compared to previous versions (Pérez et al., 2007). The AMBER (Case D et al., 2014) group of force fields is used because they have satisfactorily describe the conformational transitional changes in DNA molecules and also did well in extreme environments (Cheatham and Kollman, 1996). The expression of force field is described as below.

$$U = \sum_{bonds} \frac{1}{2} K_b \ (r - r_0)^2 + \sum_{angles} \frac{1}{2} K_a \ (\theta - \theta_0)^2 + \sum_{torsions} \frac{V_n}{2} \left[1 + cos(n\phi - \delta)\right] + \sum_{improper} V_{imp} + \sum_{LJ} 4\varepsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}}\right) + \sum_{elec} \frac{q_i q_j}{r_{ij}} \qquad (2)$$

The first four terms describe the intramolecular terms in which the first two terms are related to bond stretching and angle bending, the next two terms are related to dihedrals and improper torsions and the last two terms are related to describe the repulsive and Van der Waals interactions

and the Coulombic interactions. The reliability and accuracy depends upon the quality of the force field that describes the inter- and intramolecular interactions.

Periodic boundary conditions (PBC) are a set of boundary conditions, which are used to represent a large system by a unit cell. PBC will minimize the edge effects into a finite system. The PBCs are maintained to avoid artifacts of the system. The PBC conditions are used with minimum image convection means that the interactions that are nearer to the system are only considered. A 1.0 nm cut off is used. For log range effects the PBCs can be used in combination with the Particle mesh Ewald method (PME) for calculating the electrostatic interactions in the system (Darden et al., 1993). The LINCS algorithm was used to constraint the bond lengths and more stable and faster than SHAKE (Hess et al., 1997). A 3-point TIP3P water model (Jorgensen et al., 1983) was used in this work.

The temperature coupling of the system should be maintained constant through out the simulation in order to avoid drift during equilibration, integration errors and heating due to external forces. The direct use of molecular dynamics gives rise to NVE ensemble (constant number, constant volume and constant energy ensemble). The NVT ensemble or canonical ensemble at constant temperature is used to calculate the most quantities. The Velocity-rescaling thermostat is used for temperature coupling. The v-rescale temperature coupling produces a correct canonical ensemble and also has an advantage of first order decay of temperature deviations without oscillation (Bussi et al., 2007) . The expression of the v-rescaling temperature coupling is described as below.

$$\mathrm{d}k = (K_0 - K)\frac{dt}{\tau_T} + 2\sqrt{\frac{KK_0}{N_f}}\frac{\mathrm{d}W}{\sqrt{\tau_T}} \qquad (3)$$

Where K is the kinetic energy, N_f is the number of degrees of freedom and d is a Wiener process. Within this ensemble the conserved energy quantity is written into the energy file and log file.

In the same process the system pressure coupling should be maintained constant throughout the system in order to avoid the fluctuations in pressure or volume. The pressure is maintained constant by using NPT ensemble with Parinello-Rahman barostat (Parrinello and Rahman, 1981). In early stages of equilibration Berendsen barostat is used as it converges quickly, however this barostat can lead to unrealistic pressure oscillations. For this reason Parinello-Rahman barostat was used, which gives the true NPT ensemble.

$$\frac{db^2}{dt^2} = VW^{-1}b'^{-1}\left(P - P_{ref}\right) \qquad (4)$$

where V is the volume of the box, W is the matrix parameter, which determines the strength of the coupling. The matrices P and P_{ref} are current and reference pressures.

Base pair parameters:

The structure of DNA molecule is with succession of base pairs. The relative orientations of these bases for short and long axes and also the position with these planes allow determining the base pair structure of DNA. The base pair step parameters of local step parameters will plays an important role in studying the dynamics of DNA.

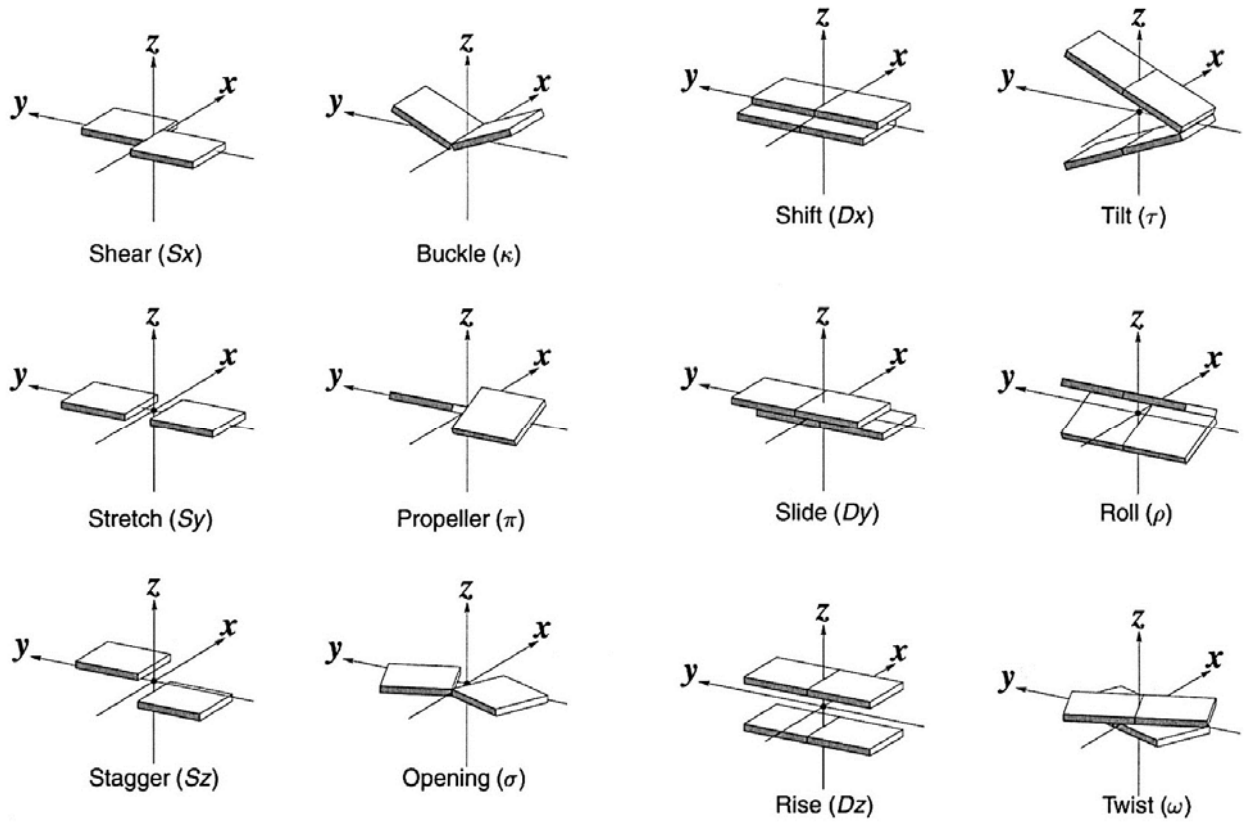


Figure 5: The schematic representations of base parameters. The first two vertical rows represent local base pair parameters (Shear, Buckle, Stretch, Propeller, Stagger and Opening) and other two represents the base pair step parameters (Shift, Tilt, Slide, Roll, Rise and Twist). (Lu and Olson, 2003).

4. Methodology:

4.1: DNA Methylation:

4.1.1: Crystal structure:

The initial coordinates were obtained from the crystal structure of the B-DNA dodecamer (Pdb: 1BNA (5'-D(*CP*GP*CP*GP*AP*AP*TP*TP*CP*GP*CP*G)-3'), that is deposited in the protein data bank. The Dickerson drew structure is well known structure with 1.9 Å resolution and with R-value of 17.8 % (Drew et al., 1981).

4.1.2: Model Building:

In the Dickerson drew structure the guanine base was replaced manually at the 4^{th} position instead of cytosine and at 9^{th} position cytosine is replaced by guanine. The methyl group was added manually at 5^{th} position of 4^{th} cytosine and at 5^{th} position of 4^{th} position of cytosine in opposite strand. MD simulations were performed for both the wild type and methylated sequence. The final structure used in the experiment is (5'-D(*CP*GP*CP*5mcP*AP*AP*TP*TP*GP*GP*CP*G)-3').

4.1.3: Molecular Dynamic Simulations:

The MD simulations were done by using Gromacs program of version 4.6.5 (Van der Spoel et al., 2013). The MD simulations were done by using amber99_parmsbc0 force field (Pérez et al., 2007) and the TIP3P water solvent model (Jorgensen et al., 1983) system. All the bonded parameters and non-bonded parameters were defined in the topology file. To build our system we simulated a simple aqueous system. By using editconf command a cubic box is defined and genbox command fills the solvent molecules to the box. Tip3p, a 3-point water model system is used to fill the solvent molecules.

Introduction of Counterions:

The solvated system is ready, the ionic concentration and total charge of the system should be set to zero. Replacement of solvent molecule is done by Na^+ and Cl^- ions until the total charge becomes zero. The adding of ions is done by the tool genion. Grompp, a Gromacs pre-processor tool is used to generate the necessary input files contains the atomic description of the system. For this an input em.mdp is created which contains the parameters such as number of steps, electrostatics and vanderwall options. The genion will add the ions until the system attains a zero charge. The 0.15 M represents the physiological ionic concentration of the system.

Energy Minimization:

Now we have solvated and neutralized system. The structure should be relaxed by the process called energy minimization before beginning the real dynamics. Once again grompp is used to generate the necessary input files. In this step the following files are generated em.log (contains the log information), em.edr (contains the information of the energies), em.gro (energy minimized structure file), em.trr (trajectory file) that gives the information of the system about the minimized structure.

The two important steps to determine if EM was successful are the potential energy E_{pot} should be negative and maximum force F_{max} should not be greater than 1000 kJ mol^{-1} nm^{-1}.

Equilibration:

After energy minimization step the structure is reasonable with geometry and solvent orientation. The system is equilibrated with the solvent molecules and ions around the DNA. The temperature of the system should be maintained constant through out the simulation, otherwise the system may collapse if we use unrestrained dynamics. After attaining a desired temperature, pressure is applied to the system until it reaches to a proper density. The position restraint file generated in the starting is used in this step. The restraint file will apply a restraining force on heavy atoms and will allow equilibrating the solvent around the DNA. Equilibration is done in two steps. The first phase is called NVT ensemble (constant number of particles, volume and temperature). This ensemble is also called isothermal-isochoric or canonical ensemble. Once again grompp is used to generate the necessary files. The temperature is controlled by v-rescale algorithm. LINCS (Linear Constraint Solver) algorithm is used to constraint all bond distances. The Maxwell distribution is used to assign the different initial structures, so that different simulations can be started from the same starting structure. A 50ps simulation is done in this ensemble to attain a desired temperature. A 0.002fs time step is used. The following files will be generated, a new structure files with new coordinates, a file with energies, a file with log information about the run and a trajectory file. A graph is plotted to check whether the temperature reached a desired temperature or not.

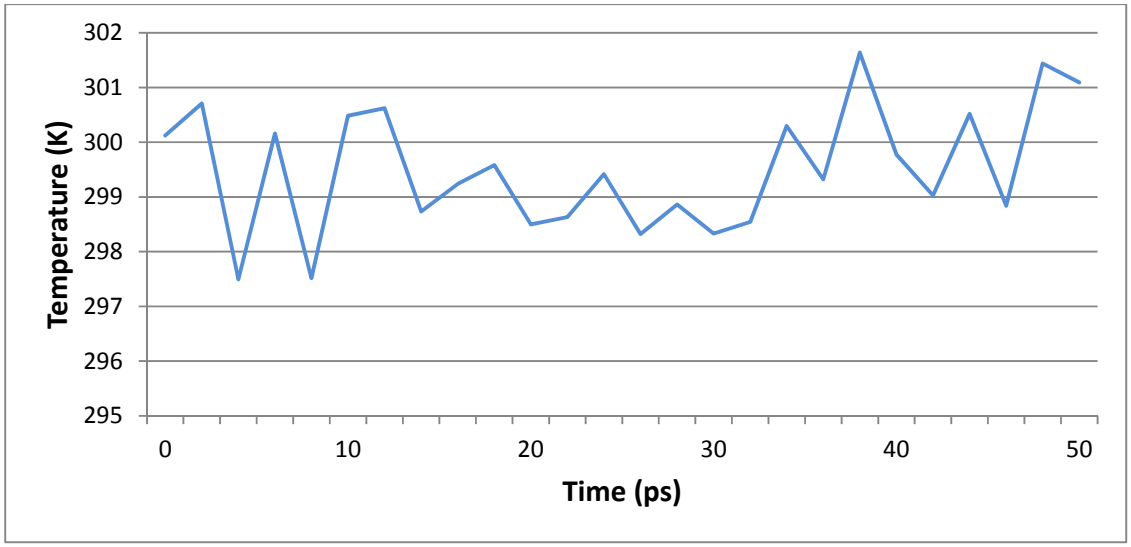


Figure 6: The above graph is plotted between time on X-axis and temperature on Y-axis. The above graph shows that the temperature reaches to a desired value (300K) quickly and remains stable throughout the simulation.

In the second phase stabilizing the pressure is done by using NPT ensemble (constant number of particles, pressure and temperature). This ensemble is also called isothermal-isobaric ensemble. The pressure is controlled by Berendsen thermostat. The process is same as before in first phase. A 500ps simulation is done to attain a stable pressure. A graph is plotted to analyze the pressure progression.

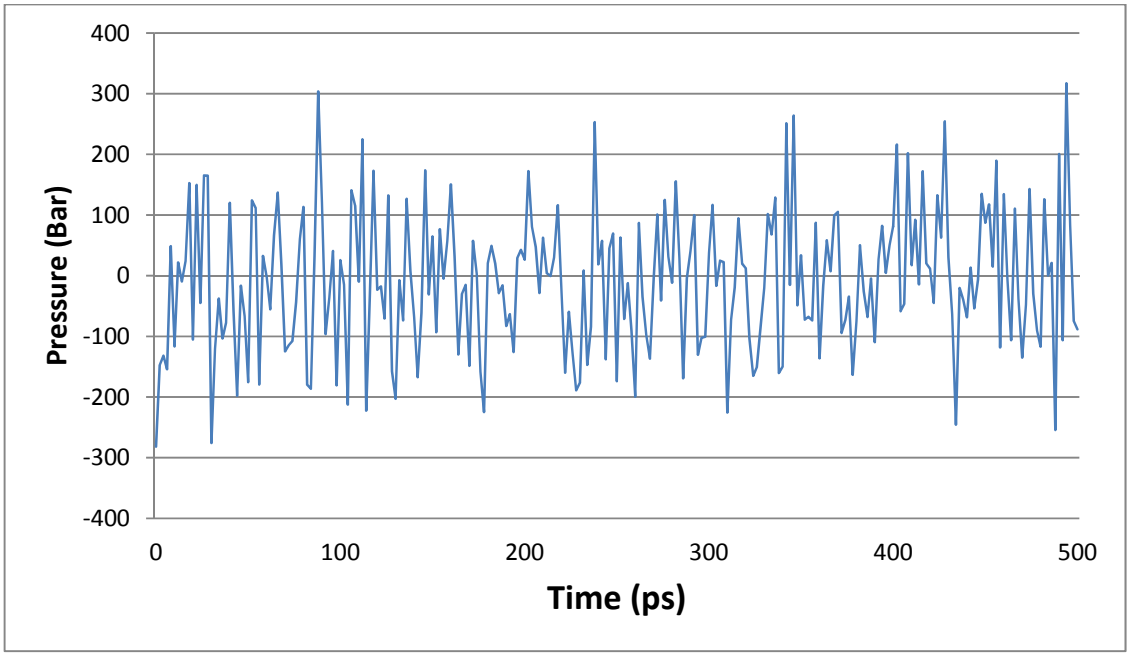


Figure 7: The above is plotted between time on X-axis and pressure on Y-axis. The above graph shows that the pressure reaches to a desired value and remains stable throughout the simulation.

Production:

The system is well equilibrated with a desired temperature and pressure and ready for the production run. GROMPP a pre processor tool is used to generate necessary input files before starting the production run.

In this step Parinello-Rahman barostat is used instead of Berendsen barostat because Berendsen is good in early stages of equilibrium that can lead to unrealistic pressure oscillations. So Parinello-Rahman barostat is used instead of Berendsen barostat. A 100ns production run was performed for replica. A total time of 300ns production was performed for each type of sequence. In total 600ns run was performed for both types of sequences. Now the simulated trajectory is analyzed to understand the dynamics of DNA. Analysis was performed by using X-3 DNA software and histograms were plotted by using R (R core team, 2012).

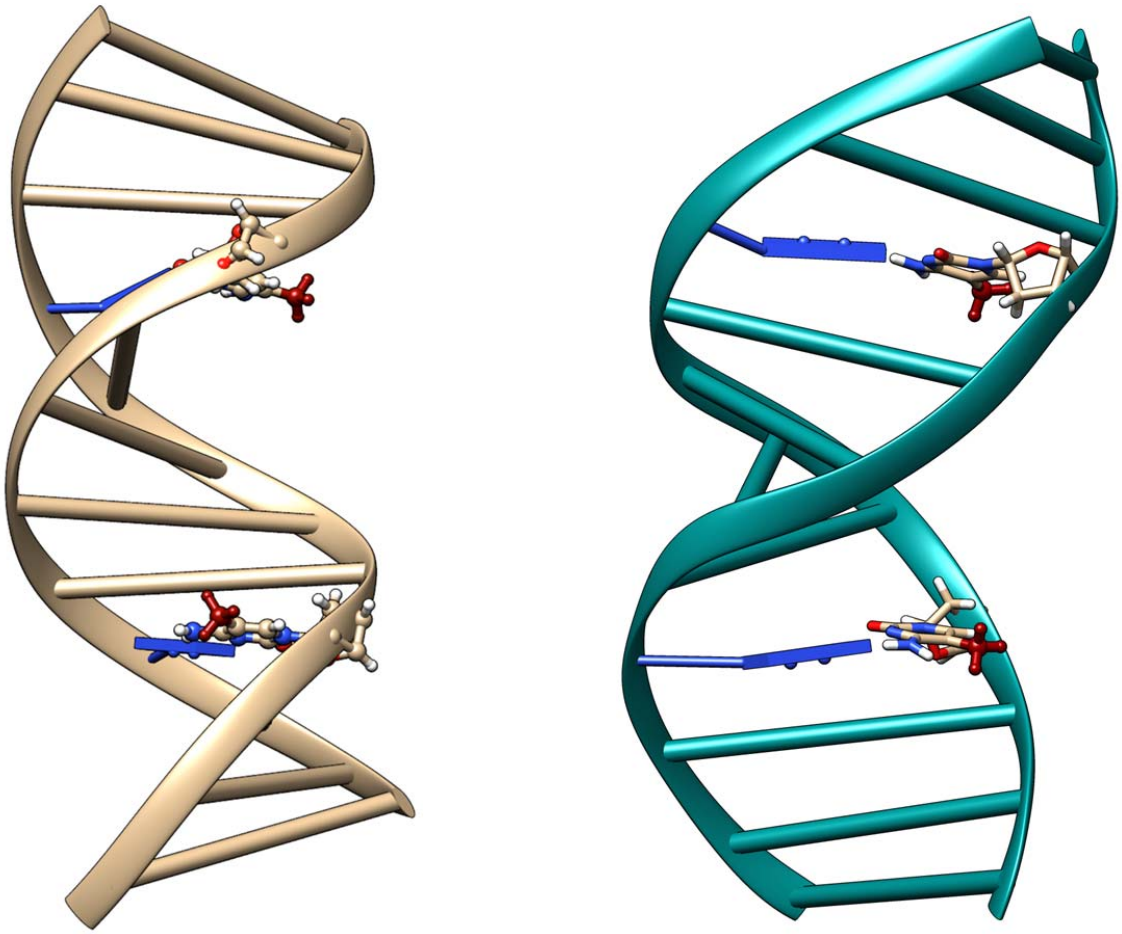


Figure 8: Schematic representation of DNA behaviour during the 100ns simulation. The left picture shows the starting structure of DNA at 1ps and the right one shows the structure of DNA at the last frame of simulation run. The methyl groups are shown in the red colour in both structures.

4.2. G-Rich Sequences:

4.2.1: G-Rich sequence with 15 base pairs:

Method

The crystal structure of the sequence TACT-GGGGGGG-AGTA was build into A-and B-DNA forms by using 3DNA software (Zheng et al., 2009). MD simulations were performed for a total time of 100ns for each replica, using the Gromacs program with the amber99sb_parmbsc0 force field. The system was solvated by using TIP3P water model system and neutralized with Na^+ and Cl^- counter ions until a 0.15 M concentration was achieved. The structure was first subjected to an energy minimization using the steepest descent algorithm, and then equilibrated in two phases; the first phase in a NVT ensemble and afterwards in a NPT ensemble. First phase of equilibration is NVT phase, a 50ps simulation was done in order to raise the temperature to 300k. The temperature of the system is maintained at 300K throughout the run. Second phase of equilibration is NPT ensemble in order to maintain a constant pressure. The ensemble is also called the isothermal-isobaric ensemble. In this phase a simulation run of 500ps was done in order to maintain a constant pressure throughout the system. After attaining a stable temperature and pressure, simulation of the system was done for 100ns. Analysis of trajectories was performed using X3DNA software (Zheng et al., 2009).

4.2.2: G-Rich sequence with 23 base pairs:

Method

According to the part 1, simulation of the sequence $d(TACTGGGGGGGAGTA)_2$ was done for 100ns. But this sequence was small, only comprising the binding sequence. So we decided to simulate a larger sequence. By doing this we are trying to minimize the end effect artifact from the simulation with the free DNA. Our extended model system comprises the sequence, $d(AGGTACTGGGGGGGAGTAGAATC)_2$. The same protocol as above is followed to perform MD simulations for a total time of 100ns for each replica. A total time of 300ns simulation were performed.

4.3. Uracil:

Method

The crystal structure of the sequence d(CAGAAAAATTTTTCTG)$_2$ was build into B-DNA form by using 3DNA software (Zheng et al., 2009). Molecular dynamics (MD) simulations were performed for a total time of 10ns for unrestrained sequence, using the Gromacs program with the Amber99sb_parmbsc0 force field. This force field was chosen because it is designed specially for nucleic acids. The system was solvated by using TIP3P water model system and neutralized with Na^+ and Cl^- counter ions until a 0.15 M concentration is achieved. The structure was first subjected to an energy minimization using the steepest descent algorithm, and then equilibrated in two phases; In the first NVT phase a 50ps simulation was done in order to raise the temperature to 300k and maintained throughout the run and in the second NPT phase, a simulation run of 500ps is done in order to maintain a constant pressure throughout the system. After attaining a stable temperature and pressure, simulation of the system is done for 10ns. Base pair opening was generated by a harmonic constraint distance of r_o = 2.90 Å and a force constant k = 5000 kJ.mol^{-1}.nm^{-2} in between the O_2 of thymine and NH_2 of adenine. The simulation was performed for 10ns. Base flipping was observed. Now the force constant of the constraint is lowered to 1000 kJ.mol^{-1}.nm^{-2} and simulation was performed for 10ns simulation, base flipping was not observed in this case.

5.Results and Discussions:

5.1: DNA Methylation:

The MD simulations were performed with the objective of better understanding the structure and dynamics of DNA. The well-known Dickerson-Drew Dodecamer structure is modified at the 4th base and 9th base of the structure. The end effects are not taken into consideration because of their flexible nature. So, the first base pair and last base pair were not considered while plotting the histograms.

5.1.1: Base pair step parameters:

DNA molecule is highly flexible and polymorphic in nature. The conformational space for the DNA is limited which depends upon the base pair parameters. The above figure1 shows the frequency of obtained values for the base pair step parameters for methylated sequence in relation with the unmethylated sequence. In the base pair step parameters, there are small differences observed in roll and tilt and twist, which result in bending and unwinding of DNA. The local unwinding was observed at the methylated regions. A 3° difference was observed at the GG-Cc in relation to the unmodified DNA which results in unwinding of DNA. The differences are only observed at the modified base pair steps. The central part of the system (AA-TT, AT-AT and TT-AA) is stable in all the parameters, which has less scope for conformational space. The differences are observed near the methylation region in roll and tilt. In all the base pair steps twist shows a similar value in relation to the unmodified sequence. The twist relates to winding and unwinding of the DNA, which is an important parameter. Differences were observed in roll parameter at the methylated sites which is around 3° in the base pair steps Cc-GG and GG-Cc when compared to unmethylated sequence. In tilt parameter no significant differences were observed expect a 1° degree difference in the G and C base pair steps. The differences for the parameters are even though smaller which behaves almost same as the unmodified sequence. But the modified sequence has bending near the methylated regions. The modified DNA sequence has methyl group on cytosine base at 4th position of one strand and whereas the methyl groups of thymine 7 and 8 are present on the other strand. The methyl group of modified base can alter the hydration patterns by creating a hydrophobic region that can lead to untwisting of DNA. The changes in the solvation patterns are accommodated by the DNA unwinding. The table 1 below shows the values of the base pair step parameters and the graph in figure 10 shows the histograms of base pair step parameters of both unmethylated and methylated sequence. The figure 9 below shows that the DNA bending and unwinding near the methylation regions. Bending was observed near the methylated region at 4th base and unwinding was observed near the 16th base.

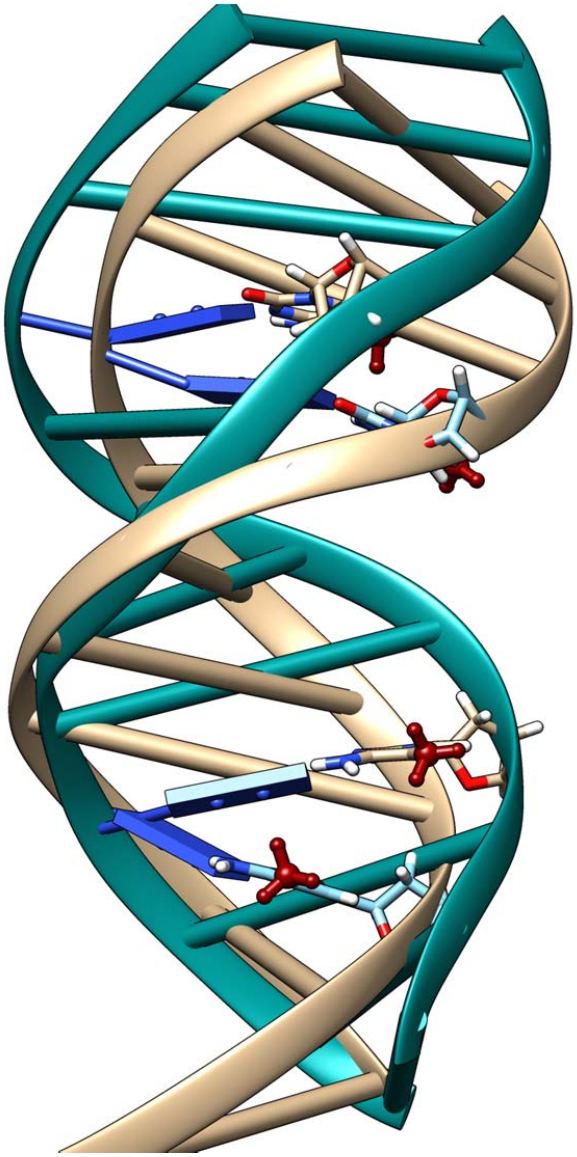


Figure 9: Schematic representation of behavioural changes caused by the methylation on the DNA structure in the simulation run. In the above figure tan colour DNA is the starting structure of the simulation and the cyan colour is the last frame of the 100ns simulation run. The methylated region is shown in red colour.

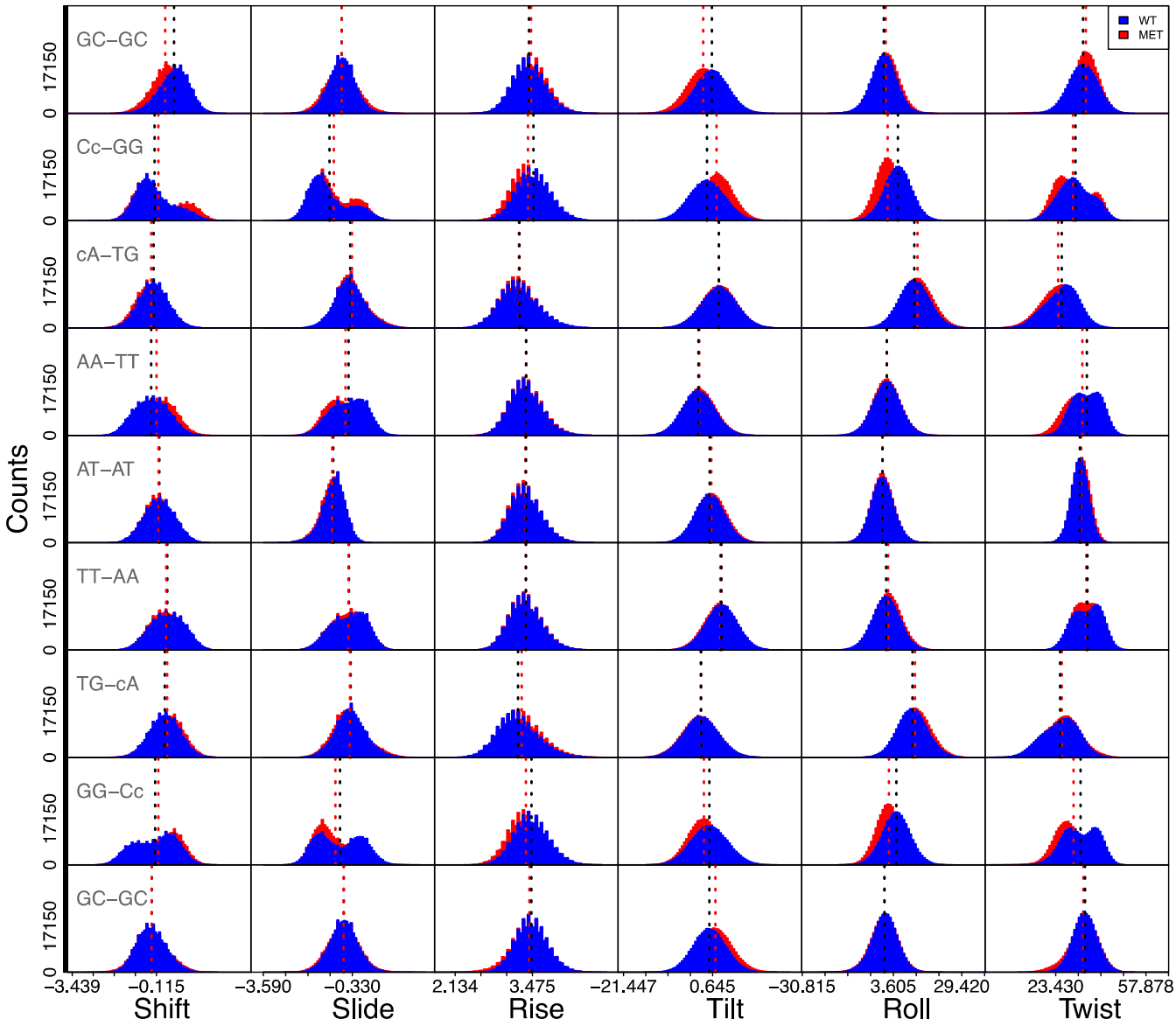


Figure 10: Frequency of values for base pair step parameters of both unmethylated and methylated structure of DNA.

		Shift(A^o)	*Slide(A^o)*	*Rise(A^o)*	*Tilt(^{o})*	*Roll(^{o})*	*Twist(^{o})*
GC-GC	WT	0.58 ± 0.6	−0.72 ± 0.49	3.4 ± 0.29	0.49 ± 4.2	−0.04 ± 4.9	3.9 ± 5.6
	MET	0.23 ± 0.56	0.72 ± 0.58	3.5 ± 0.29	1.7 ± 4.2	0.72 ± 5.0	34.9± 5.1
Cc-GG	WT	−0.19 ±0.75	−1.2 ± 0.76	3.5 ± 0.30	−0.73 ± 4.4	5.3 ± 5.2	31.0 ± 6.5
	MET	−0.039 ± 0.9	−1.0 ± 0.72	3.4 ± 0.30	1.7 ± 4.0	1.5 ± 4.7	30.0 ± 7.1
cA-TG	WT	−0.23 ± 0.59	−0.41 ± 0.58	3.2 ± 0.35	2.2 ± 4.4	11.5 ± 5.9	25.7 ± 6.6
	MET	−0.32 ± 0.6	−0.35 ± 0.58	3.2 ± 0.33	2.2 ± 4.4	12.7 ± 5.8	24.4 ± 6.7
AA-TT	WT	−0.32 ± 0.67	−0.47 ±0.64	3.4 ± 0.29	−2.8 ± 4.0	1.1 ± 5.3	35.3 ± 5.4
	MET	−0.11 ± 0.71	−0.58 ±0.65	3.4 ± 0.30	−2.6 ± 4.0	1.1 ± 5.2	33.6 ± 6.3
AT-AT	WT	0.003 ± 0.6	−1.0 ± 0.43	3.4 ± 0.28	−0.05 ± 3.7	−0.48 ± 4.4	32.6 ± 3.3
	MET	−0.032±0.57	−1.1 ± 0.43	3.4 ± 0.28	0.35 ± 3.8	−0.46±4.3	33.1 ± 3.4
TT-AA	WT	0.31 ± 0.67	−0.45 ± 0.64	3.4 ± 0.29	2.9 ± 4.0	0.86 ± 5.3	35.5 ± 5.7
	MET	0.27 ± 0.65	−0.48 ± 0.63	3.4 ± 0.29	2.9 ± 4.0	1.7 ± 5.2	35.1 ± 5.2
TG-cA	WT	0.22 ± 0.61	−0.39 ± 0.58	3.2 ± 0.37	−2.2 ± 4.4	10.8 ± 5.8	25.1 ± 7.2
	MET	0.31 ± 0.64	−0.42 ± 0.64	3.3 ± 0.36	−2.3 ± 4.5	11.8 ± 5.9	25.7 ± 7.1
GG-Cc	WT	−0.17 ± 0.83	−0.78 ± 0.83	3.3 ± 0.31	−0.12 ± 4.5	4.7 ± 5.4	32.9 ± 6.6
	MET	−0.04± 0.86	−0.96 ± 0.75		−1.5 ± 4.0	1.9 ± 4.9	30.2 ± 6.9
GC-GC	WT	−0.3 ± 0.58	−0.64 ± 0.52	3.5 ± 0.28	−0.12 ± 4.1	0.27 ± 5.0	34.6 ± 4.7
	MET	−0.31 ± 0.63	−0.65± 0.57	3.4 ± 0.31	1.4 ± 4.3	0.2 ± 5.1	34.0 ± 5.6

Table 1: Values of base pair parameters for unmethylated DNA and methylated DNA(d $(CGCCAATTGGCG)_2$.

5.1.2: Local Base Step Parameters:

The local base pair parameters also play an important in conformational dynamics of the DNA structure. In the present work the local base pair parameters does not have effects on structure as base pair step parameters had. The central of the system has with A and T bases are stable which shows similar values in both methylated and unmethylated sequence. This central part does not show any effect on the dynamics and structure of DNA. The differences were observed in buckle, propeller. In the buckle parameter a 5° difference was observed in the first base steps G-C, C-G and the last base step C-G. A 2° difference was observed in the base step G-C. In propeller a 4° difference was observed in C-G base step and a 2° difference was observed in base steps in G-c and G-C. These differences in buckle and propeller were observed near the methylated regions. The remaining parameters of local base step parameters show almost similar values for both methylated and unmethylated sequence. The graph in figure 11 below shows the histograms of local base step parameters for both methylated and unmethylated sequence and table 2 shows the obtained values of local base step parameters.

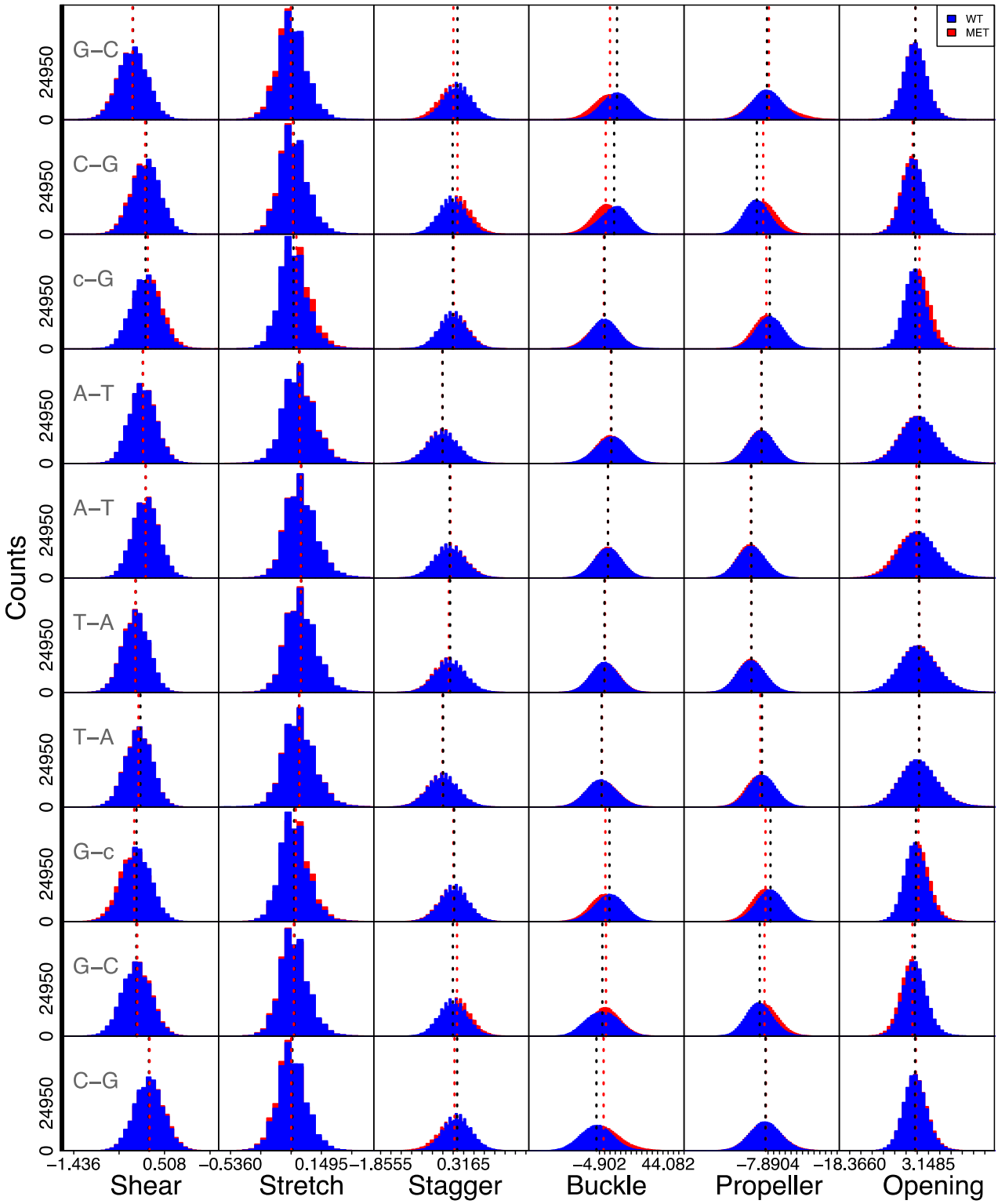


Figure 11: Frequency of values for local base step parameters of unmethylated and methylated structure of DNA.

		Shear(A^o)	*Stretch(A^o)*	*Stagger(A^o)*	*Buckle(^{o})*	*Propeller(^{o})*	*Opening(^{o})*
G-C	**WT**	−0.17 ± 0.31	−0.07 ± 0.1	0.16 ±0.37	7.7 ± 10.8	−8.4 ± 8.5	−0.42 ± 3.2
	MET	−0.18 ± 0.3	−0.08 ± 0.1	0.05 ±0.39	2.7 ± 11.5	−7.4 ± 10.6	−0.41 ± 3.2
C-G	**WT**	0.12 ± 0.3	0.06 ± 0.1	0.03 ±0.36	5.7 ± 10.5	−14.2 ± 7.4	−0.83 ± 3.3
	MET	0.093 ± 0.3	0.07 ± 0.1	0.16 ±0.36	−0.34 ± 9.8	−10.6 ± 7.7	−1.2 ± 3.2
c-G	**WT**	0.1 ± 0.3	−0.06 ± 0.11	0.04 ±0.36	−1.2 ± 9.9	−6.8 ± 7.8	−0.45 ± 3.1
	MET	0.15 ± 0.31	−0.04 ± 0 .11	0.06 ±0.37	−1.8 ± 10.3	−8.8 ± 7.6	0.67 ± 3.2
A-T	**WT**	0.041 ± 0.28□	0.02 ± 0.12	−0.23 ±0.4	3.6 ± 10.8	−11.5 ±7.7□	0.74 ± 5.2
	MET		−0.02±0.12	−0.24 ±0.4	3.2 ± 10.4	−11.8 ± 7.6	0.62± 5.3
A-T	**WT**	0.11 ± 0.28□	−0.002±0.12□	−0.05±0.4□	1.2 ± 9.6	−17.3 ± 7.9	0.59 ± 5.3
	MET	0.1 ± 0.27	−0.01±0.12	−0.03± 0.4	1.2 ± 9.4	−17.6 ± 7.6	−0.17 ± 5.4
T-A	**WT**	−0.11 ± 0.28	−0.002 ±0.12	−0.04 ±0.4	−1.3 ± 9.6	−17.3 ± 7.8	0.55 ± 5.3
	MET	−0.12 ± 0.27	−0.005±0.12	−0.07± 0.4	−1.0 ± 9.7	−17.4 ± 7.6	0.66 ± 5.2
T-A	**WT**	−0.009 ± 0.51	−0.02 ± 0.13	−0.22±0.41	−3.4 ± 10.6	−11.2 ± 7.8	0.58 ± 5.3
	MET	−0.049± 0.28	−0.02± 0.12	−0.23±0.41	−3.1±10.9	−12.2 ± 7.9	0.6 ± 5.4
G-c	**WT**	−0.09 ± 0.31□	−0.06 ± 0.11□	0.07 ±0.37□	2.4 ± 10.6□	−6.5 ± 8.0	−0.3 ± 3.1
	MET		0.04 ±0.11	0.052±0.38		−9.3 ± 7.9	0.4 ± 3.2
G-C	**WT**	−0.096 ± 0.3	−0.05 ±0.11	0.04± 0.36	−2.8 ± 11.3	−12.6 ± 7.7	−0.59 ± 3.3
	MET	−0.077±0.31	−0.06±0.11	0.14 ± 0.36	−0.2 ± 10.2	−9.9 ± 7.9	−1.2 ± 3.3
C-G	**WT**	0.17 ± 0.31	−0.07± 0.11	0.15 ± 0.37	−7.0 ± 11.4	−9.3 ± 8.8	−0.51± 3.3
	MET	0.19 ± 0.31	−0.08± 0.11	0.08 ± 0.41	−1.9 ± 12.5	−9.1 ± 9.0	−0.31 ± 3.3

Table 2: Values of local base pair parameters for unmethylated DNA and methylated DNA(d $(CGCCAATTGGCG)_2$.

5.1.3: RMSD:

The RMSD (Root mean square deviation) shows the deviation of the atoms from their mean position. The deviation of the atoms from their mean positions is not so much. The graph is plotted between time in picoseconds and distance in nanometers. In the graph it shows the system is stable in the entire simulation time. The figure 12 shows the RMSD values for unmethylated DNA and the three colours indicate the replicates, which shows similar type of values in each individual run. The figure 13 shows the RMSD values of methylated DNA. The values are also stable in relation with unmodified DNA. The system is stable in both unmethylated and methylated simulation runs. The end bases are not considered while analyzing.

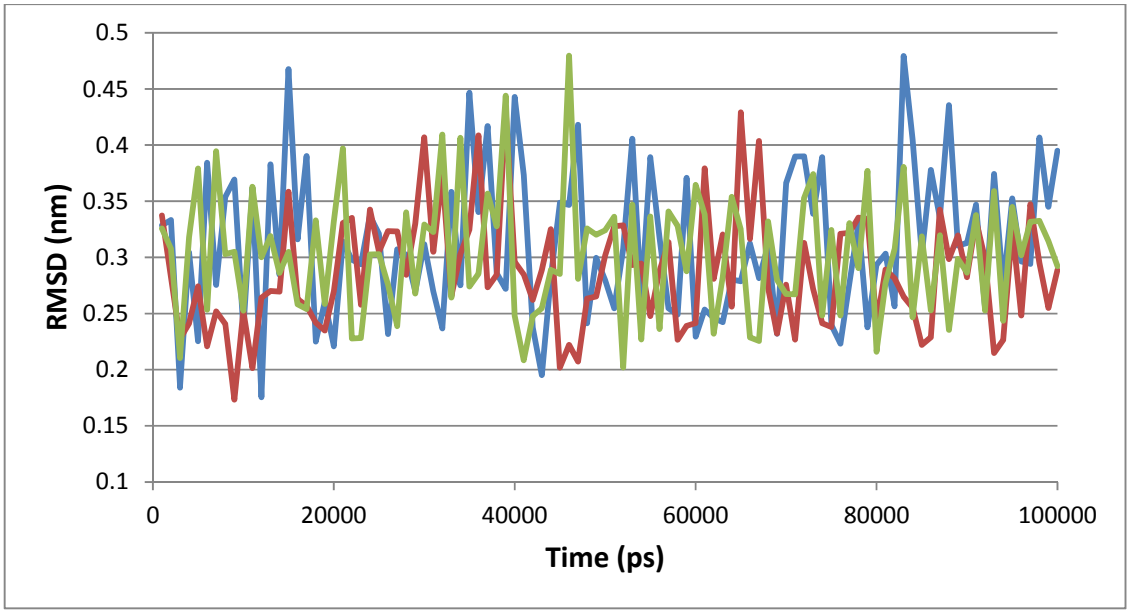


Figure 12: The above graph shows the RMSD of the wild type DNA during the 100ns run. The three different colours green, red and blue shows that the replicates of each run.

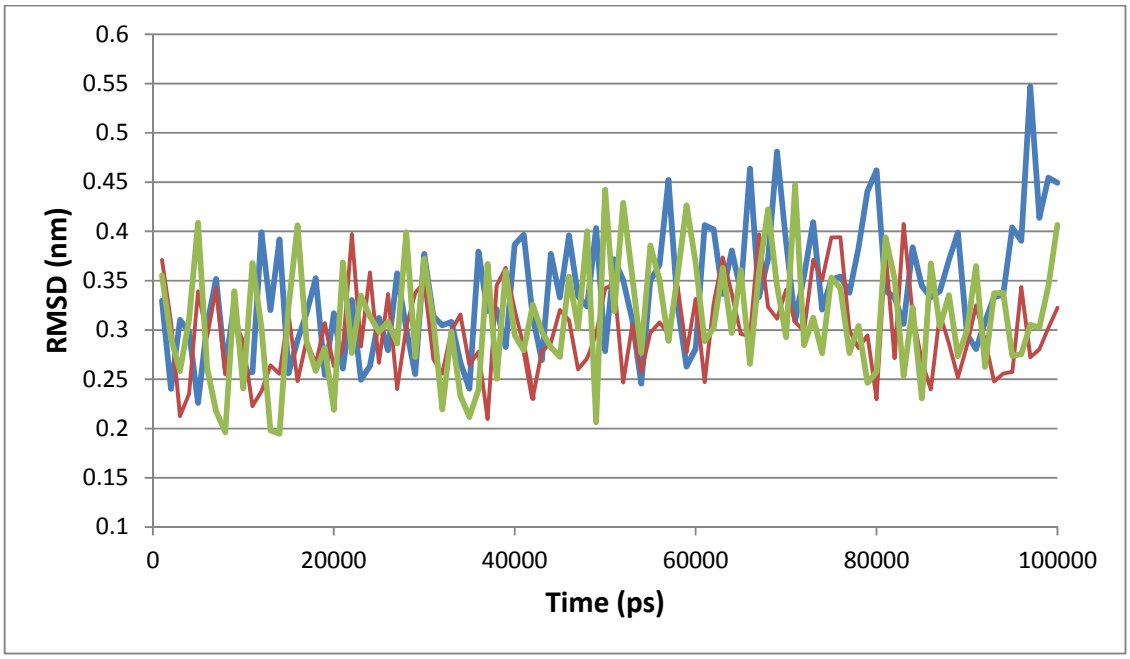


Figure 13: The above graph shows the RMSD of methylated DNA during the 100ns run. The three different colours green, red and blue shows that the replicates of each run.

5.2: G- Rich Sequences:

5.1.1: G-Rich sequence with 15 base pairs:

The conformational differences between the A and B-DNA forms of the double helical DNA sequences are well known and are conventionally characterized by a variety of conformational parameters, such as base pair parameters, major and minor groove widths, phase angle of sugar puckering (P), (Lu et al., 2000). Out of the six base pair step parameters three of them shows differences in the twist, roll and slide.

Table 3: Representative parameters distinguishing A and B type steps in high- resolution X-ray crystal structures and canonical fiber models.

	A-DNA		B-DNA	
Parameter	crystal	Fiber	crystal	Fiber
Base pair parameters				
Twist (°)	$31.1_{(4.0)}$	30.3	$36.0_{(6.8)}$	36.0
Roll (°)	$8.0_{(3.9)}$	12.4	$0.6_{(5.2)}$	1.7
Slide (Å)	$-1.53_{(0.34)}$	-1.40	$0.23_{(0.81)}$	0.45
Helical Parameters				
Inclination (°)	$14.6_{(7.3)}$	22.6	$2.1_{(9.3)}$	2.8
x- displacement (Å)	$-4.2_{(1.2)}$	-4.5	$0.1_{(1.3)}$	0.5
Helical Twist (deg.)	$32.5_{(3.8)}$	32.7	$36.5_{(6.6)}$	36.0
Helical rise (Å)	$2.8_{(0.4)}$	2.6	$3.3_{(0.20)}$	3.4
Groove dimensions (Å)				
Major-groove width	$12.9_{(2.6)}$	11.1	$17.4_{(1.3)}$	17.2
Minor-groove width	$15.8_{(0.5)}$	16.7	$10.8_{(1.4)}$	11.7

Above table shows the characteristics of DNA transformations whether into A or B form (Lu et al., 2000).

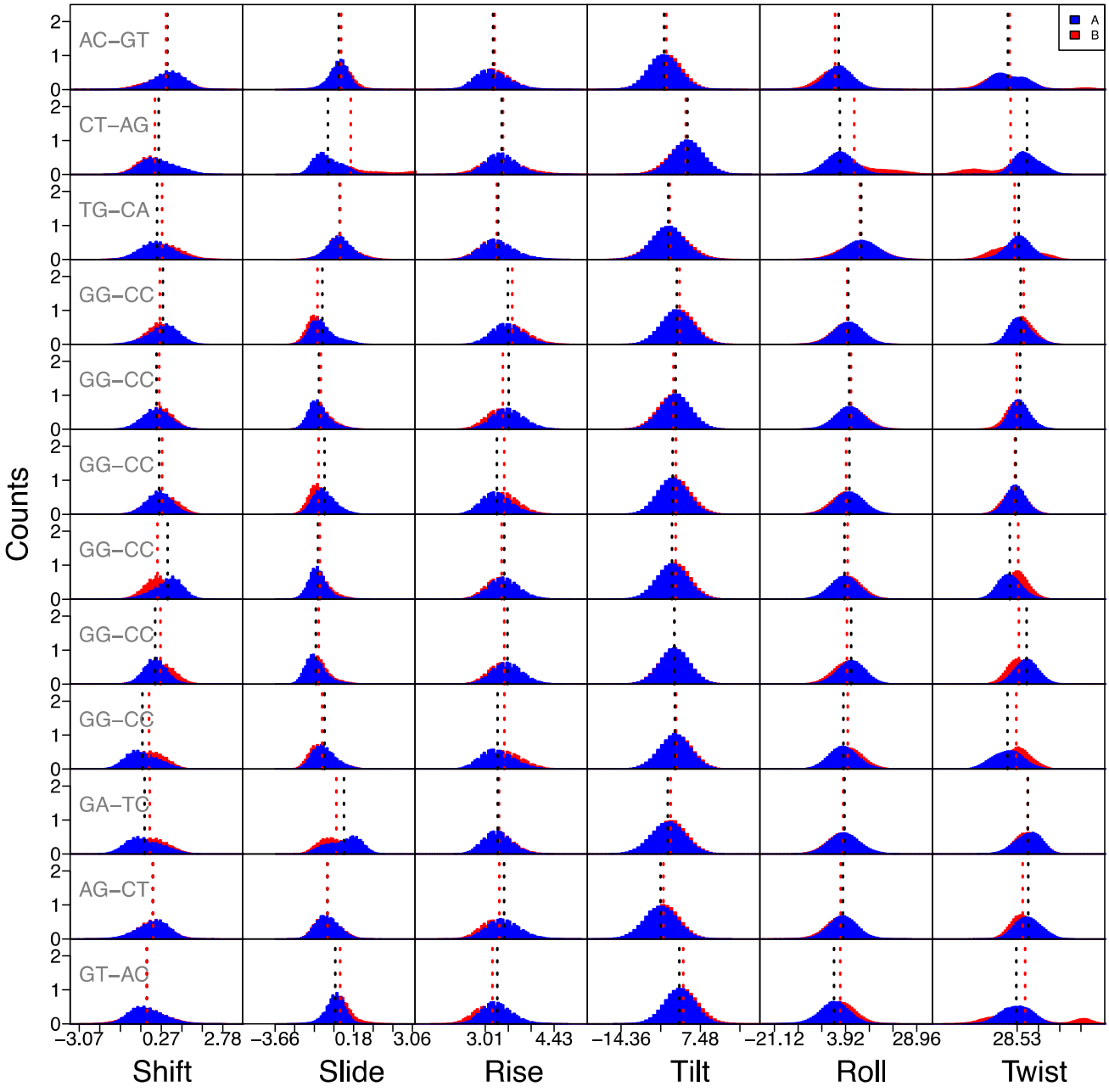


Figure 14: Frequency of values for base pair step parameters of both A and B forms of DNA.

		Shift(A^o)	*Slide(A^o)*	*Rise(A^o)*	*Tilt(^{o})*	*Roll(^{o})*	*Twist(^{o})*
AC-GT	**A**	0.52 ± 0.67	−0.54 ± 0.5	3.2 ± 0.35	−2.5 ± 4.3	1.50 ± 5.1	24.7 ± 7.0
	B	0.46 ± 0.76	−0.43 ± 0.6	3.2 ± 0.35	−1.9 ± 4.6	0.27 ± 6.0	25.5 ± 11.1
CT-AG	**A**	0.16 ± 0.73	−1.1 ± 0.66	3.3 ± 0.33	3.9 ± 4.4	1.9 ± 5.3	32.6 ± 5.4
	B	0.01 ± 0.70	0.04 ± 1.80	3.4 ± 0.44	3.5 ± 4.7	7.0 ± 9.1	25.9 ± 11.0
TG-CA	**A**	0.09 ± 0.68	−0.51 ± 0.61	3.3 ± 0.36	−1.30 ± 4.5	9.4 ± 6.1	29.2 ± 5.4
	B	0.30 ± 0.78	−0.48 ±0.68	3.2 ± 0.36	−0.96 ± 5.1	8.9 ± 6.8	27.5 ± 8.0
GG-CC	**A**	0.33 ± 0.63	−1.3 ± 0.60	3.5 ± 0.32	1.0 ± 4.3	4.9 ± 5.2	29.9 ± 4.7
	B	0.21 ± 0.55	−1.6 ± 0.54	3.6 ± 0.34	1.8 ± 4.3	4.5 ± 5.4	31.2 ± 4.4
GG-CC	**A**	0.07 ± 0.57	−1.5 ± 0.53	3.5 ± 0.33	0.67 ± 4.2	5.2 ± 5.2	29.7 ± 4.2
	B	0.19 ± 0.54	−1.4 ± 0.60	3.4 ± 0.34	0.11 ± 4.3	5.8 ± 5.4	28.3 ± 4.5
GG-CC	**A**	0.18 ± 0.52	−1.2 ± 0.56	3.2 ± 0.31	−0.12 ± 4.1	5.2 ± 5.2	27.9 ± 4.2
	B	0.30 ± 0.60	−1.5 ± 0.50	3.4 ± 0.32	0.76 ± 4.2	4.1 ± 5.2	27.6 ± 4.8
GG-CC	**A**	0.53 ± 0.59	−1.6 ± 0.48	3.4 ± 0.31	−0.3 ± 4.2	3.5 ± 5.0	25.5 ± 4.8
	B	0.11 ± 0.53	−1.5 ± 0.54	3.3 ± 0.32	0.66 ± 4.1	4.6 ± 5.1	29.0 ± 4.5
GG-CC	**A**	0.02 ± 0.47	−1.7 ± 0.52	3.5 ± 0.31	0.36 ± 4.1	5.9 ± 4.9	32.5 ± 4.8
	B	0.23 ± 0.57	−1.5 ± 0.55	3.4 ± 0.32	0.33 ± 4.2	4.4 ± 5.2	29.2 ± 4.9
GG-CC	**A**	−0.50 ± 0.65	−1.2 ± 0.58	3.3 ± 0.35	0.45 ± 4.3	3.1 ± 5.1	24.5 ± 6.2
	B	−0.23 ± 0.63	−1.4 ± 0.63	3.4 ± 0.37	0.8 ± 4.4	4.7 ± 5.4	28.2 ± 6.3
GA-TC	**A**	−0.42 ± 0.72	−0.29 ± 0.79	3.3 ± 0.3	−1.5 ± 4.7	3.5 ± 5.5	32.9 ± 5.5
	B	−0.2 0± 0.70	−0.67 ± 0.78	3.3 ± 0.31	−0.71 ± 4.5	3.1 ± 5.6	32.9 ± 5.2
AG-CT	**A**	−0.08 ± 0.64	−1.1 ± 0.56	3.4 ± 0.35	−3.5 ± 4.5	3.0 ± 5.1	33.1 ± 5.3
	B	−0.07 ± 0.64	−1.1 ± 0.59	3.3 ± 0.37	−2.7 ± 4.4	2.2 ± 5.4	30.8 ± 5.0
GT-AC	**A**	−0.32 ± 0.69	−0.72 ± 0.48	3.3 ± 0.32	1.7 ± 4.2	−0.18 ± 5.4	28.2 ± 6.2
	B	−0.32 ± 0.77	−0.48 ± 0.59	3.2 ± 0.37	2.8 ± 4.5	2.1 ± 5.5	31.8 ± 13.2

Table 4: Values of base pair parameters for A and B forms of DNA (d (TACTGGGGGGGAGTA)$_2$)

We compared our results with the values in the table 3 above in order to assign the DNA structures to the A or B forms. In most base pair steps the values are closer to the A-form. More specifically, and in relation to the twist angle, all the reported values are closer to an A-form 31.1° (Lu et al., 2000). In the first base pair step (AC-GT) the values are also related to A- form even though it has low experimental value around 25° which is close to the crystallographic value

31.1°. The average crystallographic value from the set of crystals of the parameter roll of A-form is 8.0° and B-form is 0.6°. According to the values of the parameter roll, the values are related to A-form and also the dimer steps are in between A-form and B-form. The values in the first (AC-GT) base pair step 1.5° for A-DNA and 0.2° for B-DNA and last (GT-AC) base pair step -0.18° for A-DNA and 2.1° for B-DNA are related to B-form of DNA. Most of the dimer steps of roll are in between A-form and B-form, but closer to the A-form, which has a value around 5. But according to (Blackburn, 2006) the value of roll are 6.3° for A-DNA and 0.6° for B-DNA. The crystal values of slide for A-DNA are -0.15 Å and 0.23 Å for B-DNA. According to the slide it is closely related to A-form. All the values are related to A-form. The B→A transition in slide occurs in all base steps and easily converted to A-form. All the base pair steps are easily converted into A-form of DNA expect AC-GT that is resistant in all the three parameters which maintains a value related to. The graph in figure 14 below shows the histograms of base pair step parameters for both A-form and B-form of DNA and table 4 shows the obtained values of base pair step parameters. So, we can conclude that our sequence is related to A-form according to the graph.

5.1.2: G-Rich sequence with 23 base pairs:

According to the above section, simulation of the sequence $d(TACTGGGGGGGAGTA)_2$ was done for 100ns. But this sequence was small, only comprising the binding sequence. So we decided to simulate a larger sequence. By doing this we are trying to minimize the end effect artifact from the simulation with the free DNA. Our extended model system comprises the sequence, $d(AGGTACTGGGGGGGAGTAGAATC)_2$. We follow a similar protocol as before. The table below shows the parameters of the conformations, we analyzed the values related to major groove, minor grooves and also base pair step parameters such as roll and twist. Based on these values we can determine that our model system is related to A-form. All the values of twist, roll and slide are closely related to A-form. In the parameter twist all the base pair step values are similar to the crystal values 31.1° expect in the AA-TT base pair step which is 34.0° in A-form and 34.5° in B-form, which are closely related to B-form. But in roll the base steps (GT-AC),(AC-GT),(CT-AG),(AG-CT),(AA-TT) and (AT-AT) the values are related to B-form. These values are slightly above the crystal value (0.6°) of B-DNA but in reality these are closer to the B-form than the A-form. All the experimental values of slide shows similar values to the crystal value (-1.53 Å) of A-form expect (AG-TC),(AA-TT) and (GA-TC) which has slightly lower value close to -0.6 Å in both types of DNA and in the base pair step TA-TA the experimental value of B-DNA is closely related to B-form of DNA. We compare our results with the values in the table 3 above in order to know the conformations of DNA. The graph in figure 15 below shows the histograms of base pair step parameters for both A-form and B-form of DNA sequence and table 5 shows the obtained values of base pair step parameters.

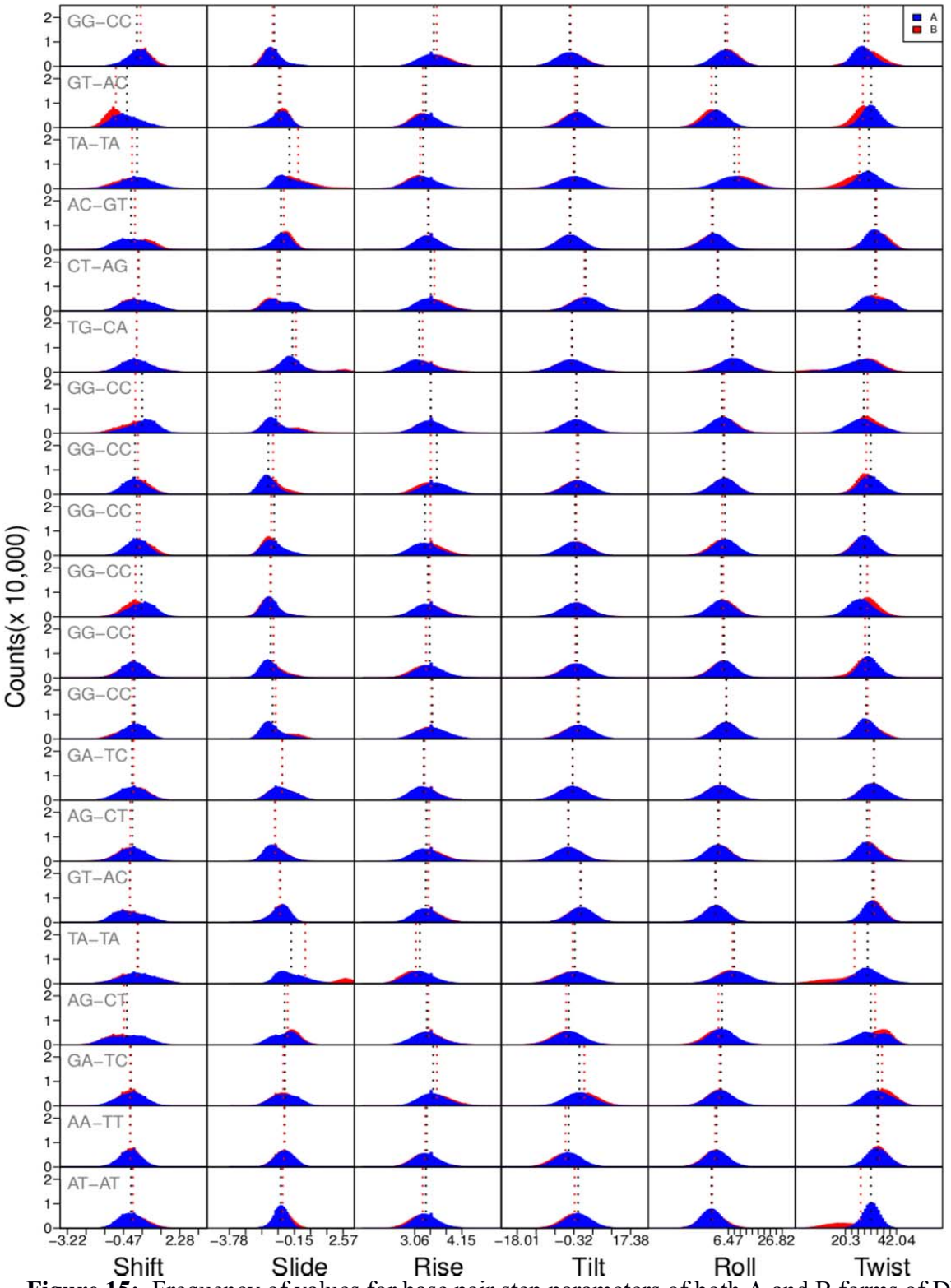


Figure 15: Frequency of values for base pair step parameters of both A and B forms of DNA.

		Shift(A^o)	*Slide(A^o)*	*Rise(A^o)*	*Tilt(^{o})*	*Roll(^{o})*	*Twist(^{o})*
GG-CC	**A**	0.19 ± 0.60	−1.30 ± 0.60	3.50 ± 0.34	−1.60 ± 4.20	5.20 ± 5.30	28.4 ± 5.0
	B	0.37 ± 0.52	−1.40 ± 0.60	3.60 ± 0.35	−1.30 ± 4.40	6.10 ± 5.10	30.0 ± 6.1
GT-AC	**A**	−0.30 ± 0.69	−1.0 ± 0.58	3.3 ± 0.30	0.71 ± 3.9	1.20 ± 5.10	31.30 ± 4.2
	B	−0.85 ± 0.53	−0.9 ± 0.49	3.3 ± 0.28	0.08 ± 3.9	−0.7 ± 4.9	27.80 ± 4.4
TA-TA	**A**	0.19 ± 0.81	−0.43 ± 0.74	3.30 ± 0.34	−0.15 ± 4.9	9.10 ± 7.0	30.3 ± 5.6
	B	−0.05 ± 0.78	0.06 ± 0.91	3.20 ± 0.31	−0.48 ± 5.0	11.0 ± 6.7	26.3 ± 6.8
AC-GT	**A**	−0.10 ± 0.74	−0.90 ± 0.57	3.40 ± 0.29	−1.5 ± 4.0	−0.07 ± 5.6	33.0 ± 4.5
	B	0.09 ± 0.81	−0.74 ± 0.54	3.40 ± 0.29	−1.6 ± 4.0	−0.57 ± 5.8	33.6 ± 5.1
CT-AG	**A**	0.27 ± 0.75	−0.97 ± 0.71	3.4 ± 0.33	3.4 ± 4.3	2.1 ± 5.3	33.0 ± 5.7
	B	0.23 ± 0.74	−1.10 ± 0.75	3.5 ± 0.34	2.9 ± 4.3	1.8 ± 5.3	33.5 ± 5.3
TG-CA	**A**	0.17 ± 0.72	−0.25 ± 0.82	3.2 ± 0.35	−0.98 ± 4.9	8.3 ± 6.3	26.1 ± 7.50
	B	0.18 ± 0.74	−0.07 ± 1.10	3.2 ± 0.37	−0.85 ± 5.0	8.2 ± 6.8	26.3 ± 10.3
GG-CC	**A**	0.44 ± 0.71	−1.20 ± 0.70	3.4 ± 0.32	0.44 ± 4.4	3.8 ± 5.4	28.1 ± 6.5
	B	0.11 ± 0.74	−0.97 ± 0.93	3.4 ± 0.34	0.37 ± 4.6	4.6 ± 5.7	29.7 ± 6.1
GG-CC	**A**	0.09 ± 0.59	−1.6 ± 0.51	3.6 ± 0.35	1.0 ± 4.3	4.5 ± 5.5	31.1 ± 5.0
	B	0.24 ± 0.60	−1.3 ± 0.61	3.4 ± 0.34	0.5 ± 4.2	4.7 ± 5.4	29.2 ± 4.6
GG-CC	**A**	0.18 ± 0.55	−1.3 ± 0.64	3.3 ± 0.32	0.08 ± 4.4	5.0 ± 5.5	28.5 ± 4.6
	B	0.32 ± 0.61	−1.4 ± 0.54	3.4 ± 0.33	0.42 ± 4.2	4.0 ± 5.4	28.3 ± 4.8
GG-CC	**A**	0.41 ± 0.63	−1.5 ± 0.52	3.4 ± 0.32	0.33 ± 4.2	3.5 ± 5.2	26.7 ± 5.2
	B	0.12 ± 0.57	−1.5 ± 0.54	3.4 ± 0.32	0.50 ± 4.2	4.3 ± 5.1	29.6 ± 4.9
GG-CC	**A**	0.0059 ± 0.55	−1.5 ± 0.57	3.4 ± 0.32	0.63 ± 4.1	4.6 ± 5.1	30.4 ± 4.3
	B	−0.021 ± 0.59	−1.3 ± 0.62	3.3 ± 0.33	0.15 ± 4.2	4.1 ± 5.2	28.7 ± 5.1
GG-CC	**A**	0.056 ± 0.59	−1.4 ± 0.64	3.5 ± 0.35	1.2 ± 4.3	5.7 ± 5.3	29.1 ± 4.7
	B	−0.04 ± 0.65	−1.2 ± 0.77	3.4 ± 0.34	0.8 ± 4.4	5.6 ± 5.2	29.8 ± 5.1
GA-TC	**A**	−0.05 ± 0.67	−0.83 ± 0.67	3.3 ± 0.30	−0.72 ± 4.3	3.1 ± 5.9	32.6 ± 5.3
	B	0.015 ± 0.66	−0.83 ± 0.68	3.3 ± 0.31	−0.69 ± 4.3	3.0 ± 5.7	32.4 ± 5.3
AG-CT	**A**	−0.04 ± 0.65	−1.2 ± 0.57	3.3 ± 0.33	−2.1 ± 4.2	2.1 ± 5.2	29.6 ± 4.9
	B	−0.14 ± 0.64	−1.2 ± 0.57	3.4 ± 0.34	−2.0 ± 4.3	2.5 ± 5.2	30.6 ± 4.7
GT-AC	**A**	−0.15 ± 0.76	−0.95 ± 0.54	3.3 ± 0.29	1.9 ± 3.9	1.1 ± 5.2	31.9 ± 4.3
	B	−0.17 ± 0.72	−0.95 ± 0.54	3.4 ± 0.31	1.7 ± 3.9	1.0 ± 5.2	32.6 ± 4.1
TA-TA	**A**	0.24 ± 0.85	−0.33 ± 0.78	3.2 ± 0.33	−0.03 ± 4.9	9.0 ± 6.9	29.8 ± 6.3
	B	0.20 ± 0.87	0.46 ± 1.40	3.1 ± 0.32	−0.71 ± 5.0	8.2 ± 6.5	24.2 ± 9.4
AG-TC	**A**	−0.31 ± 0.88	−0.69 ± 0.68	3.3 ± 0.33	−1.9 ± 4.5	3.8 ± 5.4	31.1 ± 6.6
	B	−0.45 ± 0.83	−0.53 ± 0.66	3.4 ± 0.32	−2.7 ± 4.5	2.3 ± 5.7	33.0 ± 6.0
GA-TC	**A**	−0.11 ± 0.66	−0.67 ± 0.70	3.5 ± 0.32	1.4 ± 4.4	3.3 ± 5.6	34.2 ± 5.6
	B	−0.15 ± 0.58	−0.77 ± 0.66	3.6 ± 0.35	2.9 ± 4.5	2.8 ± 5.3	35.9 ± 5.2
AA-TT	**A**	−0.14 ± 0.54	−0.69 ± 0.53	3.3 ± 0.30	−1.9 ± 4.1	1.50 ± 5.2	34.0 ± 4.7
	B	−0.14 ± 0.49	−0.73 ± 0.53	3.3 ± 0.31	−3.0 ± 4.3	0.87 ± 5.2	34.5 ± 4.4
AT-AT	**A**	−0.10 ± 0.59	−0.91 ± 0.41	3.3 ± 0.29	0.990 ± 4.0	−0.66 ± 4.5	31.2 ± 3.6
	B	−0.01 ± 0.65	−0.80 ± 0.47	3.2 ± 0.30	−0.06 ± 4.1	−0.33 ± 5.0	26.8 ± 8.2

Table 5: Values of base pair parameters for A and B forms of DNA ($d(AGGTACTGGGGGGGAGTAGAATC)_2$).

According to the minor groove width the values are in between A-form and B form when compared to crystallographic value and coming to major groove it is related to B-form in both runs. The histograms were shown in figure 16 and obtained values are shown in below table 6.

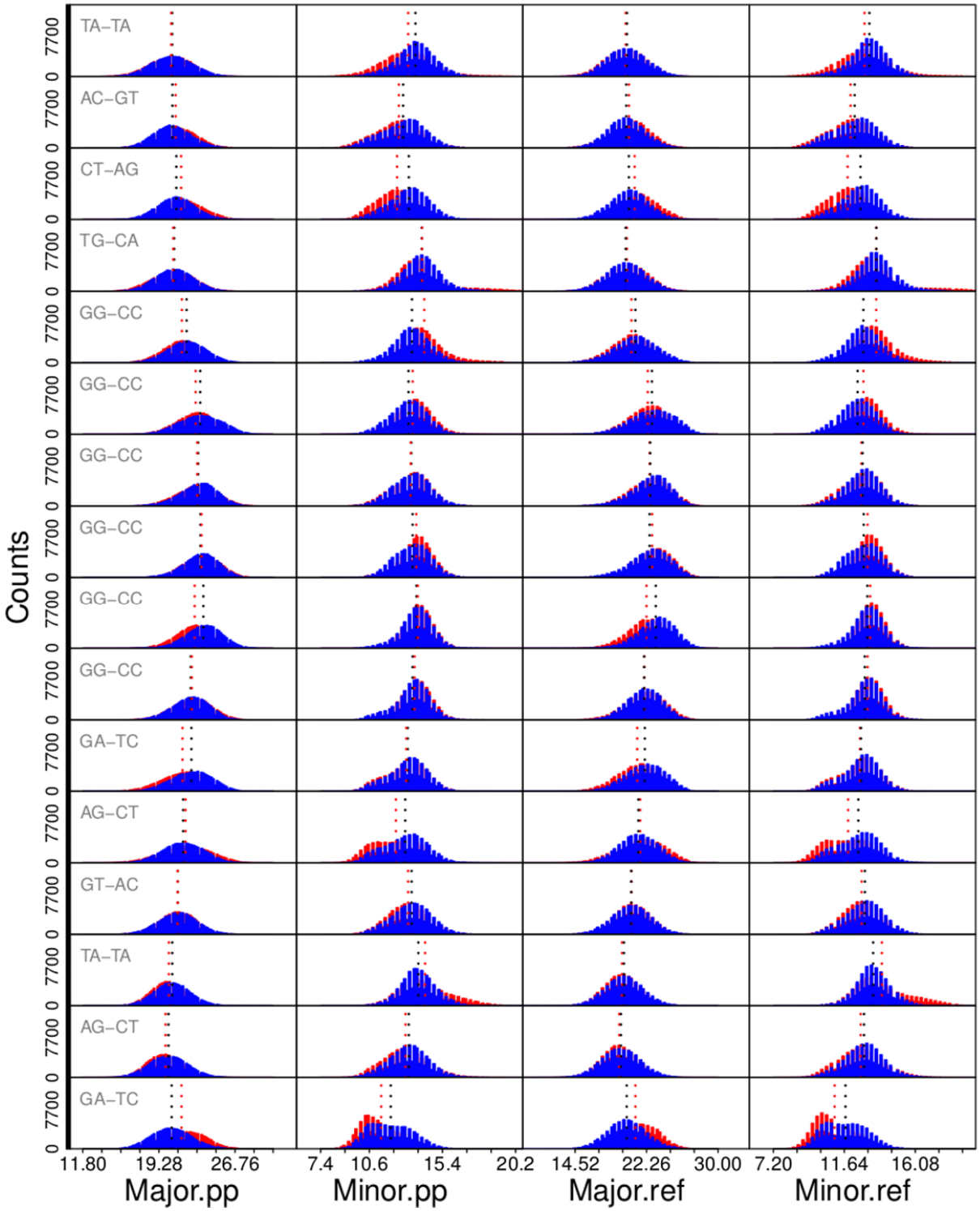


Figure 16: Frequency of values for grove widths of A and B forms of DNA.

		Major.pp	*Minor.pp*	*Major.pp*	*Minor.pp*
TA-TA	**A**	20.6 ± 2.0	13.6 ± 1.3	20.1 ± 2.1	13.2 ± 1.2
	B	20.5 ± 2.1	13.1 ± 1.7	20.0 ± 2.2	12.9 ± 1.6
AC-GT	**A**	20.6 ± 1.9	12.8 ± 1.5	20.1 ± 2.1	12.3 ± 1.4
	B	20.9 ± 2.0	12.5 ± 1.5	20.4 ± 2.2	12.0 ± 1.4
CT-AG	**A**	21.0 ± 1.9	13.2 ± 1.3	20.3 ± 2.2	12.7 ± 1.2
	B	21.5 ± 2.1	12.4 ± 1.3	21.0 ± 2.3	11.8 ± 1.3
TG-CA	**A**	20.8 ± 1.8	14.0 ± 1.3	20.0 ± 2.2	13.6 ± 1.2
	B	20.7 ± 2.0	14.1 ± 1.8	20.1 ± 2.2	13.6 ± 1.8
GG-CC	**A**	22.0 ± 1.9	13.4 ± 1.3	21.0 ± 2.4	12.8 ± 1.2
	B	21.5 ± 1.8	14.2 ± 1.5	20.6 ± 2.2	13.6 ± 1.3
GG-CC	**A**	23.3 ± 2.1	13.2 ± 1.2	22.8 ± 2.4	12.5 ± 1.1
	B	22.9 ± 1.9	13.4 ± 1.3	22.4 ± 2.1	12.8 ± 1.2
GG-CC	**A**	23.1 ± 1.9	13.4 ± 1.3	22.7 ± 2.1	12.8 ± 1.1
	B	23.0 ± 2.0	13.3 ± 1.3	22.6 ± 2.2	12.7 ± 1.2
GG-CC	**A**	23.4 ± 1.8	13.4 ± 1.2	22.6 ± 2.2	12.9 ± 1.2
	B	23.5 ± 1.8	13.7 ± 1.1	22.9 ± 2.2	13.1 ± 1.1
GG-CC	**A**	23.6 ± 1.8	13.7 ± 1.1	23.2 ± 2.1	13.1 ± 1.1
	B	22.8 ± 1.8	13.8 ± 1.0	22.3 ± 2.1	13.3 ± 0.1
GG-CC	**A**	22.4 ± 1.8	13.4 ± 1.2	22.0 ± 2.0	12.9 ± 1.1
	B	22.5 ± 1.9	13.6 ± 1.1	22.0 ± 2.1	13.1 ± 1.0
GA-TC	**A**	22.5 ± 2.0	13.1 ± 1.3	22.1 ± 2.2	12.7 ± 1.2
	B	21.6 ± 2.1	13.0 ± 1.3	21.2 ± 2.2	12.6 ± 1.3
AG-CT	**A**	21.7 ± 2.0	13.0 ± 1.4	21.4 ± 2.1	12.5 ± 1.3
	B	21.9 ± 2.4	12.3 ± 1.5	21.6 ± 2.5	11.9 ± 1.5
GT-AC	**A**	21.1 ± 1.9	13.4 ± 1.3	20.6 ± 2.1	12.9 ± 1.2
	B	21.1 ± 1.8	13.1 ± 1.3	20.6 ± 2.0	12.7 ± 1.2
TA-TA	**A**	20.6 ± 1.8	13.8 ± 1.2	19.8 ± 2.1	13.4 ± 1.1
	B	20.3 ± 1.7	14.3 ± 1.7	19.6 ± 1.9	14.0 ± 1.7
AG-TC	**A**	20.2 ± 1.9	13.2 ± 1.3	19.5 ± 2.1	12.9 ± 1.2
	B	19.9 ± 1.8	13.0 ± 1.5	19.3 ± 1.9	12.7 ± 1.4
GA-TC	**A**	20.5 ± 2.0	12.0 ± 1.4	20.1 ± 2.1	11.7 ± 1.3
	B	21.5 ± 2.2	11.4 ± 1.4	21.0 ± 2.3	11.0 ± 1.3

Table 6: Values of groove width parameters for A and B forms of DNA d((AGGTACTGGGGGGGAGTAGAATC)$_2$).

5.3: Uracil:

The kinetics of sequence dependent base pair openings has been investigated by using NMR imino proton exchange experiments (Priyakumar and MacKerell, 2006). The kinetics of base pair opening was a difficult task in computational field because the opening events will occur at milliseconds timescale (Dornberger et al., 1999). The closing kinetics usually takes place in nanoseconds that are difficult to study experimentally but can be studied by using computational

techniques. The base pair kinetics for the sequence d(CAGAAAAATTTTTCTG)$_2$ were studied. The base pair openings were studied between NH_2- O_2 atoms because it is structural parameter that captures the base pair event directly. An early study used the same coordinates while generating the open base pairs that shows consistent with NMR data (Keepers et al., 1982). Base pair opening was generated by a harmonic constraint distance of r_o = 2.90 Å and a force constant k = 5000 kJ.mol^{-1}.nm^{-2} in between the O_2 of thymine and NH_2 of adenine. The simulation was performed for 10ns. Base opening occurs by breaking the hydrogen bond between the thymine and adenine. We observed base pair opening immediately and also partial base flipping. Base flipping study is based on opening reaction coordinate, which is in open state held together by single H-bond between the bases. Base flipping occurs when H-bond breaks that hold the open conformation state. The base pair eventually return into Watson-crick conformation from partial flipped out base.

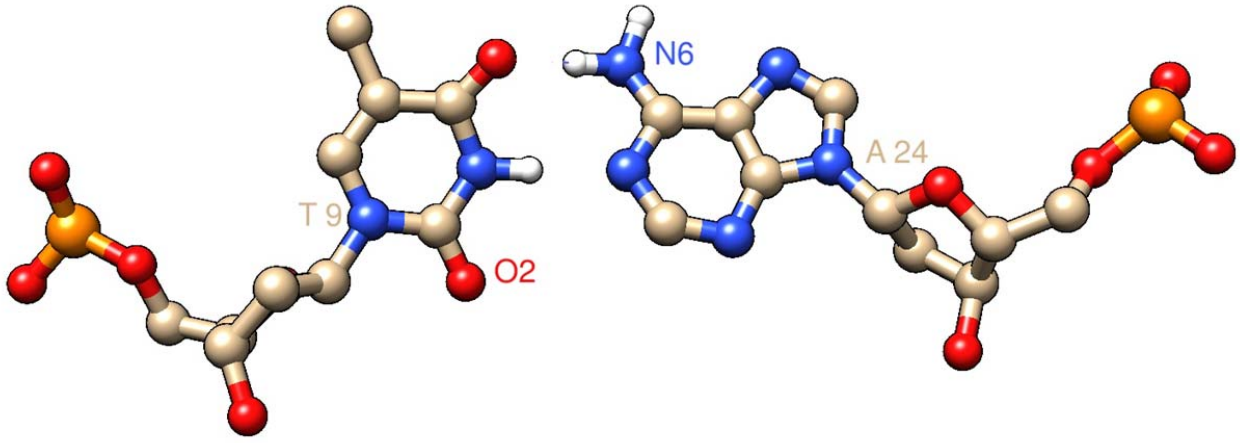


Figure 17: The above figure shows the closed conformation of the bases thymine and adenine at 970ps. Blue atoms indicate nitrogen and red atoms indicate oxygen.

Now the force constant of the restraint is lowered to 1000 kJ.mol^{-1}.nm^{-2}. Unrestrained molecular dynamic simulation is done for 10ns with maintaining the distance of r_o = 2.90 Å between the two atoms. The kinetic decay of open base pair i.e; closing kinetics was studied until the base pair gets a closed conformation. For the proton exchange the base pair state should be in open configuration. Base pair closing is observed at 970ps, which is shown in the above figure 17. A closed base conformation was defined by considering the distance between O_2 of thymine and NH_2 of adenine was greater than 4.5Å (Fadda and Pomès, 2011).

6. Overview and Conclusions:

6.1: DNA Methylation:

It has been recognized that DNA methylation has a great importance for maintaining the genomic stability. The chemical modifications of the bases such as DNA methylation play an important role in cell events responsible for development and gene silencing, which leads to many diseases such as cancer (Temiz et al., 2012). The study of the cytosine modifications helps to understand the conformational behavioural studies such as flexibility and deformability of DNA. These conformational properties are sequence dependent.

In this study, the investigation of base pair dynamics that has influence on DNA structure was monitored by using MD simulations. This study shows how the flexibility is dependent on base pair step parameters (mainly in twist and roll). The central part of the sequence with AA-TT, AT-AT and TT-AA regions are identical in all the parameters show consistent with the result obtained in a recent study (Carvalho et al., 2014). The result obtained with the base step AA-TT shows consistent with recent studies (Travers, 2004) (Carvalho et al., 2014) shows no variability, which means it has less scope for conformational space. The data obtained in the present study with sequence D(5'(CGCCAATTGGCG)-3') shows consistent with the results obtained in a recent study (Carvalho et al., 2014) with the sequence D(5'(CGCGAATTCGCG)-3'). The methylation in the 4^{th} base and the methyl groups of thymine in the other strand that shows the changes in hydration patterns is also in good agreement. In the methylated sequence there are differences observed in the roll parameter near the modification site at Cc-GG and GG-Cc and also in tilt parameter at G and C steps, which shows the methylation has an effect on the conformational state such as flexibility of DNA. The study by (Carvalho et al., 2015) with different approaches of QM/MM and DFT methods also shows the differences in twist, tilt and roll parameters, which are in good agreement with present study. The conformations are also influenced by neighbouring steps, the work done (Lavery et al., 2010) by shows that the C5 methyl groups had influence on energetic difference when interacts with the neighbouring base steps and to the interaction of methyl group on CG:CG itself. Another study also shows that nearest neighbouring effects were observed on base pair parameters. In the study the modified sequence has influence on adopting different conformations and also can have impact on recognition of methylated cytosines by the proteins.

6.2: G-Rich Sequences:

By analyzing the DNA in different contexts using computational methods we were able to test limits that force field methods offer for these systems. The force field used in this study is amber99sb_parmbsc0. Molecular dynamics were performed to study the distributions of base pair parameters and groove widths. As per the results most of the values in all the parameters are closely related to A-form when compared to crystallographic values (Lu et al., 2000). The

transition of B-A form is easy and shows similar values for both runs. The MD simulation of A-form DNA sequence remains in the same form after 100ns simulation time. The conformational changes occur in the sequence of B-form, which adapts to A-form during the simulation time. But in both parts of individual runs the sequence sometimes spends the simulation time in between the A and B-form. Major groove shows that the conformations is related to B-form where as in minor groove it shows that it is in between A and B-form. In A-form of DNA minor groove should be larger than minor groove. But in the study major groove is larger than the minor groove that favors B-DNA form. The sequence undergoes large fluctuations during the simulation towards the minor and major groove. The limitations of the force fields may be a problem for this study, even though the structure was constructed by using X3DNA software. In the base pair parameters all the values are not in good agreement to experimental values that adopts A-form. For ex. Roll shows a difference in the base steps that adopts B-form in those particular base steps for both sequences. However the important point to emphasize here is, the MD simulations when coupled to force-filed parameters shows a transition from B-A form when the study of base pair parameters is considered.

6.3: Uracil:

The aim of the work is to study the effect of the mismatch base pair U-A and in relation with natural base pair T-A to study the structure and dynamics of the DNA sequence. But in the present study the kinetics of T-A are only studied by using MD simulations in $d(CAGAAAAATTTTTCTG)_2$. A spontaneous base pair opening is observed immediately in the simulation time. In the 1st picosecond of the simulation time the base pair opening is observed. The base pair opening occurs only when two hydrogen bonds is replaced by one hydrogen bond between the Watson-Crick base pairing. The open structure was studied which allowed observing the spontaneous base flipping. The base flipping was observed only when the force constant is k = 5000 kJ.mol^{-1}.nm^{-2} and no base flipping were observed when the force constant is lowered. The open state conformation can develop into a complete flipped out base (extra-helical base). In the present study the base pair kinetics is not in good agreement with the experimental data. The kinetic decay i.e: closing kinetics of the sequence was studied and it occurs when the force constant is lowered to 1000 kJ.mol^{-1}.nm^{-2}. No base flipping was observed during this time. The closing of the base pair is observed very soon at approximately at 1ns during 10ns total simulation time. Generally the closing kinetics of base C-G and T-A were occurred in nanosecond time scale and were not in good agreement with NMR imino proton exchange study by (Priyakumar and MacKerell, 2006). A very fast opening of the bases may be responsible for the fast closing of the bases. The study of (Fadda and Pomès, 2011) shows that the shortest and longest life time of T-A base pairs obtained are 5.7 and 11.7 ns respectively at 300 K (or 27 °C) shows significant results with the NMR studies by (Priyakumar and MacKerell, 2006) at 35°C. In order to characterize the complete study of base pair kinetics in relation to T-A versus U-A base pairs is currently under way.

7. Acknowledgements:

I would like to thank my supervisors Dr. Lynn Kamerlin and Alexandra Carvalho for their support, patience and open to discussions anytime whenever I faced any problems in my project. In spite of their busy schedule, they guided me throughout to understand the background of the work, guided me patiently whenever I faced any hurdle. They helped me a lot with scientific report writing and with the help I was able to learn within a very short span of time.

8. References:

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

Ballestar, E., and Wolffe, A.P. (2001). Methyl-CpG-binding proteins. Targeting specific gene repression. Eur. J. Biochem. FEBS *268*, 1–6.

Bishop, K.S., and Ferguson, L.R. (2015). The Interaction between Epigenetics, Nutrition and the Development of Cancer. Nutrients *7*, 922–947.

Blaha, D., Arous, S., Blériot, C., Dorel, C., Mandrand-Berthelot, M.-A., and Rodrigue, A. (2011). The Escherichia coli metallo-regulator RcnR represses rcnA and rcnR transcription through binding on a shared operator site: Insights into regulatory specificity towards nickel and cobalt. Biochimie *93*, 434–439.

Branco, M.R., Ficz, G., and Reik, W. (2012). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat. Rev. Genet. *13*, 7–13.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. J. Chem. Phys. *126*, 014101.

Carvalho, A.T.P., Gouveia, L., Kanna, C.R., Wärmländer, S.K.T.S., Platts, J.A., and Kamerlin, S.C.L. (2014). Understanding the structural and dynamic consequences of DNA epigenetic modifications: Computational insights into cytosine methylation and hydroxymethylation. Epigenetics Off. J. DNA Methylation Soc. *9*, 1604–1612.

Carvalho ATP, Gouveia ML, Raju Kanna C., Wärmländer, S.K.T.S., Platts, J.A., and Kamerlin, S.C.L. (2015) Theoretical modelling of epigenetically modified DNA sequences [version 1; referees: 1 approved, 1 approved with reservations] *F1000Research* 2015, **4**:52 (doi: 10.12688/f1000research.6148.1).

Cheatham, T.E., and Kollman, P.A. (1996). Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. J. Mol. Biol. *259*, 434–444.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N□log(N) method for Ewald sums in large systems. J. Chem. Phys. *98*, 10089–10092.

D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui,R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu and P.A. Kollman (2014), AMBER 14, University of California, San Francisco.

Dornberger, U., Leijon, M., and Fritzsche, H. (1999). High base pair opening rates in tracts of GC base pairs. J. Biol. Chem. *274*, 6957–6962.

Doseth, B., Visnes, T., Wallenius, A., Ericsson, I., Sarno, A., Pettersen, H.S., Flatberg, A., Catterall, T., Slupphaug, G., Krokan, H.E., et al. (2011). Uracil-DNA Glycosylase in Base Excision Repair and Adaptive Immunity SPECIES DIFFERENCES BETWEEN MAN AND MOUSE. J. Biol. Chem. *286*, 16669–16680.

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

Fadda, E., and Pomès, R. (2011). On the molecular basis of uracil recognition in DNA: comparative study of T-A versus U-A structure, dynamics and open base pair kinetics. Nucleic Acids Res. *39*, 767–780.

Fujii, S., Kono, H., Takenaka, S., Go, N., and Sarai, A. (2007). Sequence-dependent DNA deformability studied using molecular dynamics simulations. Nucleic Acids Res. *35*, 6063–6074.

Guz, J., Foksiński, M., and Oliński, R. (2010). [Mechanism of DNA methylation and demethylation--its role in control of genes expression]. Postepy Biochem. *56*, 7–15.

Hagen, L., Kavli, B., Sousa, M.M., Torseth, K., Liabakk, N.B., Sundheim, O., Peňa-Diaz, J., Otterlei, M., Hørning, O., Jensen, O.N., et al. (2008). Cell cycle-specific UNG2 phosphorylations regulate protein turnover, activity and association with RPA. EMBO J. *27*, 51–61.

Haseeb, A., Makki, M.S., and Haqqi, T.M. (2014). Modulation of Ten-Eleven Translocation 1 (TET1), Isocitrate Dehydrogenase (IDH) Expression, α-Ketoglutarate (α-KG), and DNA Hydroxymethylation Levels by Interleukin-1β in Primary Human Chondrocytes. J. Biol. Chem. *289*, 6877–6885.

Herman, J.G., and Baylin, S.B. (2003). Gene Silencing in Cancer in Association with Promoter Hypermethylation. N. Engl. J. Med. *349*, 2042–2054.

Iacobuzio-Donahue, C.A. (2009). Epigenetic Changes in Cancer. Annu. Rev. Pathol. Mech. Dis. *4*, 229–249.

Iwig, J.S., Leitch, S., Herbst, R.W., Maroney, M.J., and Chivers, P.T. (2008). Ni(II) and Co(II) sensing by Escherichia coli RcnR. J. Am. Chem. Soc. *130*, 7592–7606.

Iwig, J.S., and Chivers, P.T. (2009). DNA Recognition and Wrapping by Escherichia coli RcnR. J. Mol. Biol. *393*, 514–526.

Iwig, J.S., and Chivers, P.T. (2010). Coordinating intracellular nickel-metal-site structure-function relationships and the NikR and RcnR repressors. Nat. Prod. Rep. *27*, 658–667.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926–935.

Keepers, J.W., Kollman, P.A., Weiner, P.K., and James, T.L. (1982). Molecular mechanical studies of DNA flexibility: Coupled backbone torsion angles and base-pair openings. Proc. Natl. Acad. Sci. U. S. A. *79*, 5537–5541.

Krokan, H.E., Drabløs, F., and Slupphaug, G. (2002). Uracil in DNA--occurrence, consequences and repair. Oncogene *21*, 8935–8948.

Lavebratt, C., Almgren, M., and Ekström, T.J. (2012). Epigenetic regulation in obesity. Int. J. Obes. *36*, 757–765.

Lavery, R., Zakrzewska, K., Beveridge, D., Bishop, T.C.,Case, D.A., Cheatham, T., Dixit, S., Jayaram, B., Lankas, F., Laughton, C., et al. (2010). A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. Nucleic Acids Res. 38, 299–313.

Lee, J.-H., and Pardi, A. (2007). Thermodynamics and kinetics for base-pair opening in the P1 duplex of the Tetrahymena group I ribozyme. Nucleic Acids Res. *35*, 2965–2974.

Lu, X.-J., Shakked, Z., and Olson, W.K. (2000). A-form Conformational Motifs in Ligand-bound DNA Structures. J. Mol. Biol. *300*, 819–840.

Lu, X.-J., and Olson, W.K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three□dimensional nucleic acid structures. Nucleic Acids Res. *31*, 5108–5121.

Mosbaugh, D.W., and Bennett, S.E. (1994). Uracil-Excision DNA Repair1. In Progress in Nucleic Acid Research and Molecular Biology, W.E.C. and K. Moldave, ed. (Academic Press), pp. 315–370.

Muers, M. (2011). Epigenetics: Methylation from mother. Nat. Rev. Genet. *12*, 6–6.

N. Metropolis, A.W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953

Napoli, A.A., Lawson, C.L., Ebright, R.H., and Berman, H.M. (2006). INDIRECT READOUT OF DNA SEQUENCE AT THE PRIMARY-KINK SITE IN THE CAP-DNA COMPLEX: RECOGNITION OF PYRIMIDINE-PURINE AND PURINE-PURINE STEPS. J. Mol. Biol. *357*, 173–183.

Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M., and Zhurkin, V.B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc. Natl. Acad. Sci. U. S. A. *95*, 11163–11168.

Parker, J.B., Bianchet, M.A., Krosky, D.J., Friedman, J.I., Amzel, L.M., and Stivers, J.T. (2007). Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. Nature *449*, 433–437.

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. *52*, 7182–7190.

Pearl, L.H. (2000). Structure and function in the uracil-DNA glycosylase superfamily. Mutat. Res. *460*, 165–181.

Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E., Laughton, C.A., and Orozco, M. (2007). Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. Biophys. J. *92*, 3817–3829.

Priyakumar, U.D., and MacKerell, A.D. (2006). NMR Imino Proton Exchange Experiments on Duplex DNA Primarily Monitor the Opening of Purine Bases. J. Am. Chem. Soc. *128*, 678–679.

R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012.

Robertson, K.D. (2005). DNA methylation and human disease. Nat. Rev. Genet. *6*, 597–610.

Rottach, A., Leonhardt, H., and Spada, F. (2009). DNA methylation-mediated epigenetic control. J. Cell. Biochem. *108*, 43–51.

Senter, P.D., Su, P.C.-D., Marquardt, H., Hayden, M.S., and Linsley, P. (1990). Thermally stable cytosine deaminase.

Soubry, A., Hoyo, C., Jirtle, R.L., and Murphy, S.K. (2014). A paternal environmental legacy: evidence for epigenetic inheritance through the male germ line. BioEssays News Rev. Mol. Cell. Dev. Biol. *36*, 359–371.

Suter, C.M., Martin, D.I., and Ward, R.L. (2003). Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue. Int. J. Colorectal Dis. *19*, 95–101.

Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I., et al. (2011). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nat. Neurosci. *14*, 1607–1616.

Temiz, N.A., Donohue, D.E., Bacolla, A., Luke, B.T., and Collins, J.R. (2012). The Role of Methylation in the Intrinsic Dynamics of B- and Z-DNA. PLoS ONE *7*, e35558.

Travers, A.A. (2004). The structural basis of DNA flexibility. Philos. Transact. A Math. Phys. Eng. Sci. *362*, 1423–1438.

Travers, A.A., and Thompson, J.M.T. (2004). An introduction to the mechanics of DNA. Philos. Transact. A Math. Phys. Eng. Sci. *362*, 1265–1279.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., and Berendsen, H.J.C. (2005). GROMACS: Fast, flexible, and free. J. Comput. Chem. *26*, 1701–1718.

Van Der Spoel, D., Lindahl, E., Hess, B., and the GROMACS development team, *GROMACS User Manual version 4.6.5.*

Urdinguio, R.G., Sanchez-Mut, J.V., and Esteller, M. (2009). Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. Lancet Neurol. *8*, 1056–1072.

Wanunu, M., Cohen-Karni, D., Johnson, R.R., Fields, L., Benner, J., Peterman, N., Zheng, Y., Klein, M.L., and Drndic, M. (2010). Discrimination of Methylcytosine from Hydroxymethylcytosine in DNA Molecules. J. Am. Chem. Soc. *133*, 486–492.

Yusufaly, T.I., Li, Y., and Olson, W.K. (2013). 5-Methylation of cytosine in CG:CG base-pair steps: a physicochemical mechanism for the epigenetic control of DNA nanomechanics. J. Phys. Chem. B *117*, 16436–16442.