Summer 2017

# Clustering of Epigenetic Methylation and Oxidative Damage: Effects on Duplex DNA Solution Structure, Thermodynamic Stability and Local Dynamics

David R. Gruber
*Western Washington University*, gruberd@wwu.edu

**Clustering of Epigenetic Methylation and Oxidative Damage: Effects on Duplex DNA Solution Structure, Thermodynamic Stability and Local Dynamics**

By

David R. Gruber

Accepted in Partial Completion
Of the Requirements for the Degree
Master of Chemistry

Kathleen L. Kitto, Dean of the Graduate School

ADVISORY COMMITTEE

Chair, Dr. Sergey Smirnov

Dr. Paul C. Spiegel

Dr. Spencer Anthony-Cahill

# MASTER'S THESIS

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work, and does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Signature: David R. Gruber
Date: 7/21/2017

# Clustering of Epigenetic Methylation and Oxidative Damage: Effects on Duplex DNA Solution Structure, Thermodynamic Stability and Local Dynamics

A Thesis

Presented to

The Faculty of

Western Washington University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

By

David R. Gruber

July 2017

# Abstract

All known living organisms use DNA to store genetic templates used for development, proper function and reproduction. The structural integrity of DNA is therefore of extreme importance and cellular machinery continuously regulates our DNA either through addition of covalent molecules to regulate the transcription of genes or the removal of DNA lesions propagating from the exposure to reactive molecules. One of the most common DNA lesions**,** 8-oxoguanine (8OG), is a prominent, pro-mutagenic DNA adduct present at a baseline level from consistent generation of reactive oxygen species through oxidative metabolism or at greater concentrations through exposure to ionizing radiation and other toxins. Its mutagenic potential is attributed to its ability in the *syn*- conformation, to mimic thymine during DNA replication, resulting in a mispair with adenine. In contrast, 5-methylcytosine (5MC), occurs from the covalent addition of a methyl group to a cytosine base by a DNA methyltransferase. 5MC acts as an epigenetic gene regulator, often found densely packed within CpG islands upstream of transcriptionally inactive genes. It can be estimated that each human diploid cell contains hundreds of CpG dinucleotides undergoing active methylation while also harboring 8OG. Previous results obtained by Kasymov et al, showed reduced endonuclease activity by hOGG1 for substrates containing 5MC adjacent and cross strand from 8OG. In addition, the work presented by Maltseva et al, conveyed that the enzymatic methylation rates by maintenance DNA methyltransferases were severely impacted when 8OG is adjacent to the methylation target. These results prompted us to investigate the clustering of these two modifications in greater detail.

We present the results of solution NMR structure determination, thermodynamic stability analysis and molecular dynamics simulations on the DNA sequence 5’-d(CGCGAATTCGCG)-3’ with

clustered 5MC and 8OG in CpG dinucleotides. NMR spectroscopy and restrained molecular dynamics were used to refine the structure of 11 DNA duplexes containing different methylation and oxidation patterns. The results reveal that 8OG induces local unwinding 5' to itself and $^{31}P$ chemical shifts indicate an increase in the BII phosphate backbone conformation 3' relative to 8OG. Melting temperatures of the duplexes was shown to decrease with the addition of 8OG in all contexts. Surprisingly, the addition of 5MC in two separate instances led to lower $T_m$ values of already oxidized DNA samples. 1D-$^{1}H$ NMR linewidths indicate 8OG increases the base dynamics while incorporation of 5MC leads to a stabilizing effect. Our results indicate that addition of 8OG to a fully-methylated CpG induces a sequence dependent stabilizing effect. Molecular dynamics trajectories were analyzed for BI/BII phosphate conformation populations, conformational flexibility and local dynamics. Comparison of helical geometries and backbone angles indicated that our MD simulations accurately and reliably reproduced our NMR structures within one standard deviation. Principal component analysis was carried out to highlight the most dominant modes of motion for CpG sites with clustered 5MC and 8OG. Particularly, we report significant differences in concerted atomic displacements, with the 8OG:5MC base pair displaying the greatest dynamic effects.

# Acknowledgments

There are numerous individuals I would like to thank for their contributions to this work. Specifically, our collaborators at the institute of chemical biology and fundamental medicine (Novosibirsk, Russia) in the Dr. Dmitry Zharkov lab. I would also like to thank Dr. Hla Win-Piazza for her assistance with my NMR experiments.

I would like to thank my thesis committee, Professors Sergey Smirnov, P. Clint Spiegel, and Spencer Anthony-Cahill, for their support throughout my project as well as essential edits during the preparation of this thesis.

This work was a collaboration within the Smirnov lab and much of this work would not be possible without countless hours spent with my colleagues. Foremost, I would like to extend gratitude towards Stas Fedetchkin and Jacob Brockerman for mentoring me as an undergraduate and laying the foundation for this project. In addition, Joanna Hoppins and Heather Miears were invaluable for discussing issues that arose with many of the experiments presented in this work.

Above all else, I am extremely grateful for the continued support from my mentor, Dr. Sergey Smirnov. Thank you for allowing me to conduct research in your lab. Your guidance shaped me into a better researcher.

# Table of Contents

# List of Figures

# List of Tables and Equations

# *Chapter I – Introduction to DNA and DNA modifications*

***Canonical Deoxyribonucleic acid*.** Deoxyribonucleic acid (DNA) is a molecule that contains vital information for the proper development and function of all known living organisms and some viruses. DNA is classified as a nucleic acid and is a type of biopolymer, comprised of a string of monomers called nucleotides. There are four nucleotides in DNA each defined by a specific nitrogenous base: the purines, guanine (G) and adenine (A) and the pyrimidines, thymine (T) and cytosine (C) (Figure 1.1a).

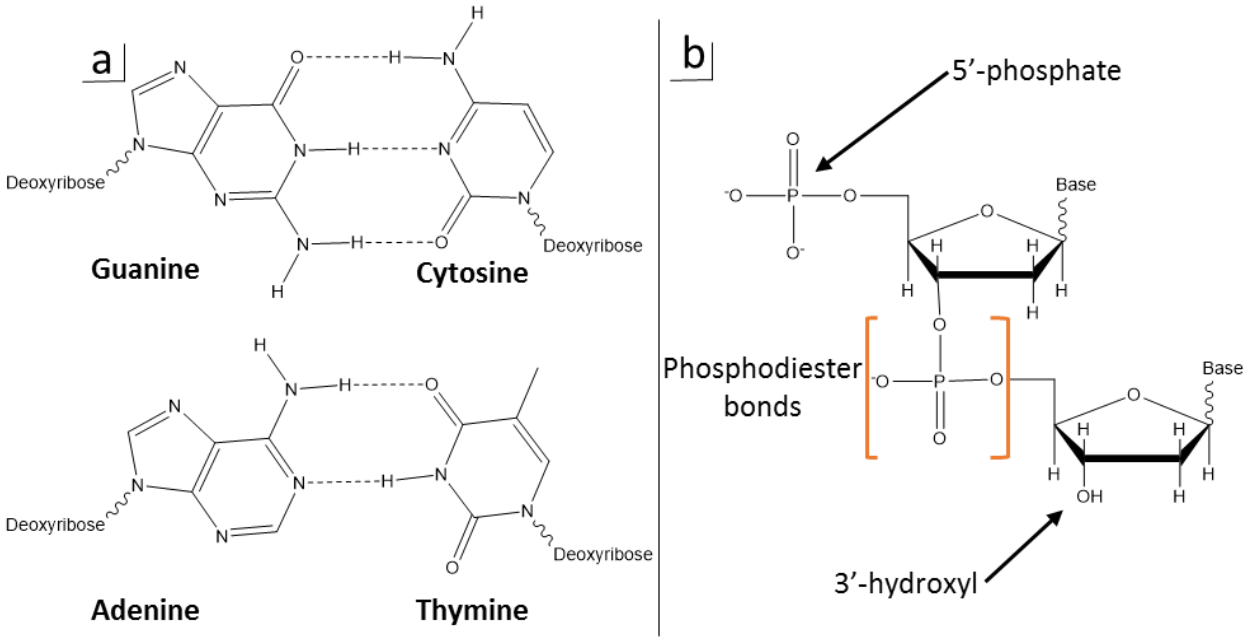



**Figure 1.1 (a) The four nitrogenous bases in DNA shown hydrogen-bonded to their canonical base pair partner. (b) Depiction of the DNA backbone and deoxyribose sugars. Highlighted are the 5'-phosphate and 3'hydroxyl as well as the phosphodiester bond that connect two sugar molecules.**

The chemical makeup of these nucleotides is broken into three fundamental components: a pentose sugar ring called 2'-deoxyribose, a phosphate group and a nitrogenous base. The DNA backbone of each strand is comprised of identical, repeating sugar and phosphate groups. Each

nucleotide unit is covalently linked to the next through a phosphodiester bond between the 5'-phosphoryl group of one sugar and the 3'-hydroxyl group of the next (Figure 1.1b). This orientation gives the strand directionality where the starting nucleotide will have a protruding 5'-phosphate and the terminus of the chain will contain a free 3'-hydroxyl. The nitrogen-containing base defining the nucleotide is attached to the 1'-position of the sugar through an *N*-glycosidic bond (Figure 1.1a).

The tertiary structure of DNA, as its simplest definition, is a double helix. The helix contains two antiparallel nucleotide strands twisting around each other to form a spiral. The two strands associate favorably due to two stabilizing forces derived from individual bases. First, the nitrogenous base of each nucleotide within the strands are oriented towards each other forming specific hydrogen bonds, creating a base pair. This interaction is highly specific, with typical Watson-Crick hydrogen bonding occurring between a purine from one strand and a pyrimidine from the opposing strand (Watson and Crick, 1953). Guanine pairs with cytosine forming three hydrogen bonds and thymine pairs with adenine forming two hydrogen bonds (Figure 1.1a). Second, π- π stacking between the aromatic rings on each nucleobase stabilize the individual strands (Matta et al., 2006).

The macromolecular structure of double stranded DNA can exist in numerous conformations with B-DNA, A-DNA and Z-DNA being the prominent structures. The strands can twist in either direction, giving rise to either right- or left-handed helices. Due to the twisting of the strands and the consistent positioning of the *N*-glycosidic bond of each nucleotide (oblique with respect to

one another) two grooves form, denoted the major and minor groove. Both B-DNA and A-DNA exhibit right-handedness with A-DNA being more compressed, causing the major groove to narrow and the minor groove to widen (Figure 1.2). Z-DNA twists in the opposing direction from A- and B-form DNA, exhibiting a left-handed conformation. DNA conformation and stability is directly associated with electrostatic interactions, hydration level and the addition of chemical modifications. In physiological conditions, the prevalent conformation is B-DNA, however, transient Z-DNA conformations have been documented during DNA transcription as the helix is unwound and in negative supercoiling (Lafer et al., 1983). In addition, certain chemical modifications such as cytosine methylation have been demonstrated to enhance Z-DNA characteristics for helices *in vivo (*Zacharias et al., 1990). The third major conformer, A-DNA, is found in living cells only in enzyme-DNA complexes where otherwise buried atoms are exposed for enzymatic attack. This is due to the high energetic barrier of the B-DNA to A-DNA transition in hydrated environments (Lu et al., 2000). However, A-form double helices are the predominant secondary structure of RNA.

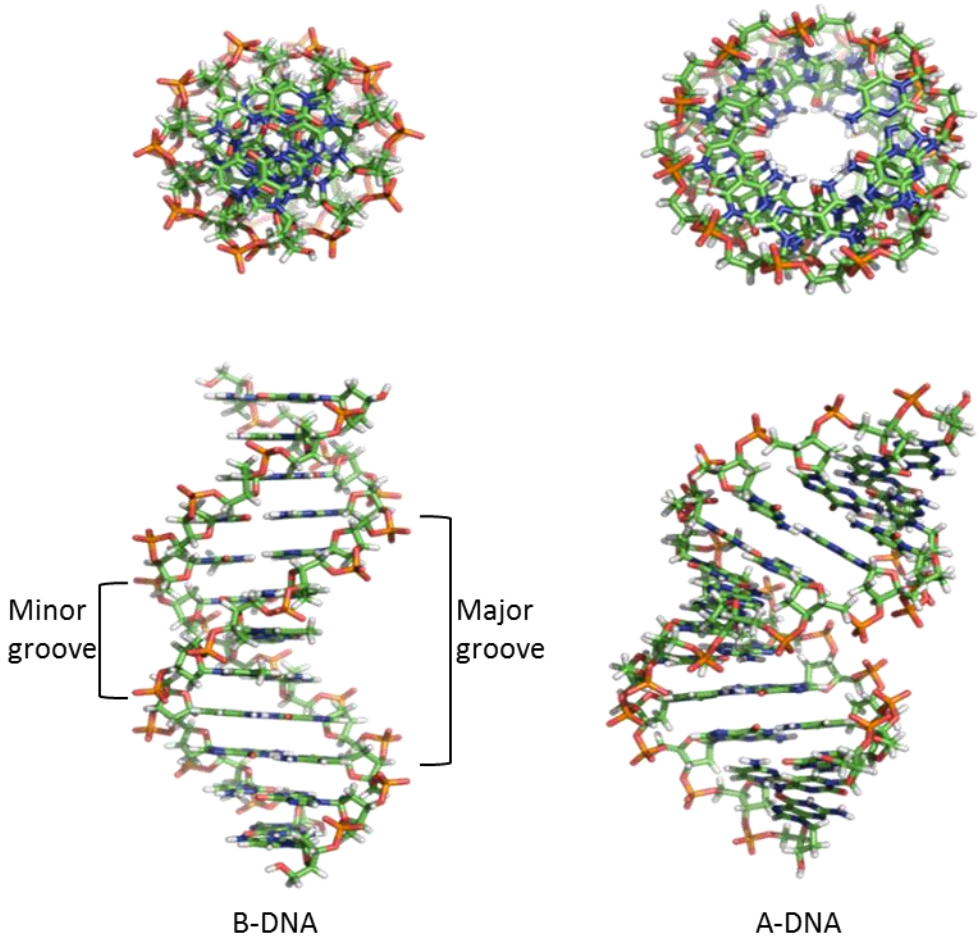



**Figure 1.2. 3D representations of "perfect" B- and A- DNA shown top-down (top) and side-view (bottom). Both are right handed helices with A-DNA exhibiting a narrow major groove and widened minor groove with respect to B-DNA. Structures were generated using nucleic acid builder from AmberTools 15.**

***DNA torsion angles*.** The torsion angles of the DNA backbone are flexible and defined by six covalent bonds (α, β, γ, δ, ε and ζ) (Figure 1.4). In B-DNA, the backbone conformation is frequently described by two sets of torsion angles. First, the phosphate can adopt two major conformations, called BI and BII (Figure 1.3) (Schneider et al., 1997). These conformations are dictated by the torsion angle values of ε (C4'-C3'-O3'-P) and ζ (C3'-O3'-P-O5'). Each conformation corresponds to a range of values determined from the difference between ε and ζ (ε – ζ). For BI, the range is -

160° to -20° with an average value being approximately -90°. For BII, the range is 20° to 200° with an average value of approximately 90°. Transitions between BI and BII sub-states have been frequently observed to occur on the nanosecond timescale (Winger et al., 1998; Isaacs and Spielmann, 2001).

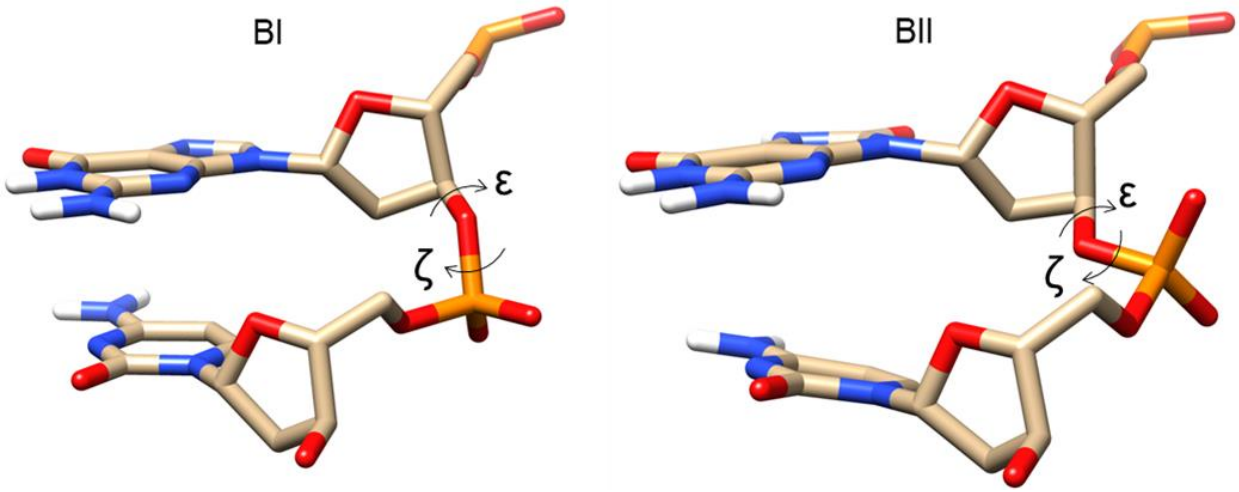



**Figure 1.3. Illustration of the BI (left) and BII (right) phosphodiester conformations. BI and BII conformations are defined by the difference between ε and ζ (ε – ζ).**

Second, fluctuations in α (O3'-P'-O5'-C5') and γ (O5'-C5'-C4'-C3') angles are used to define the canonical and non-canonical backbone states. NMR and X-ray crystal structures solved on free, unmodified B-DNA show almost exclusive preference for the canonical α/γ conformation: *g-*/*g+* (Schneider et al., 1997). However, non-canonical α/γ angles present themselves in DNA-protein complexes (Castagné et al., 2000; Olson et al., 1998), as nucleotide flipping is directly associated with changes in the α/γ torsion angles. An investigation of crystal structures for DNA-protein complexes revealed that up to 10% of 1245 dinucleotide steps analyzed display non-canonical α/γ conformations. (Várnai et al., 2002).

Five torsion angles define the pucker of the ribose sugar ring which can exist in ten possible configurations with each configuration occupying a range of 36° (pseudorotation range is 360°/10). In unmodified B-DNA, the sugar primarily adopts the C2'-endo (144°-180°) configuration. Analysis of internal sugar pucker distributions in crystal structures reveal that the uncompressed and flexible structure of B-DNA allows a broad range to be defined (60°-210°) (Dickerson and Ng, 2001). Finally, the *N*-glycosidic bond connecting the sugar and base of each nucleotide is defined by the angle χ. The relative orientation of the base and sugar are grouped into two conformations, *anti* and *syn*, with the former being most prominent in A- and B-DNA.

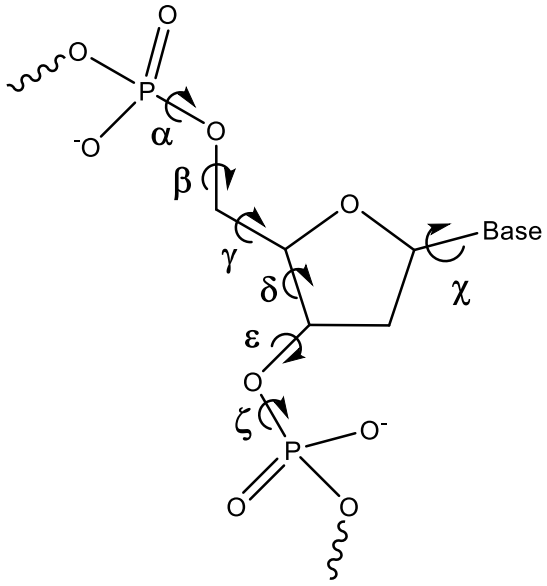



**Figure 1.4. The six dihedral angles (α, β, γ, δ, ε and ζ) that define the DNA backbone. The angle χ, which defines the *N*-glycosidic bond, dictates the orientation of the base.**

***Helical geometries***. The local structure of the DNA molecule is often defined by base-pair and base-pair step geometries measured by their relative position and orientation to cross-strand and successive bases, respectively (Figure 1.3). In B-DNA, the helical parameters are highly

sequence specific and polymorphic (Dans et al., 2014). Specifically, the tetranucleotide (four sequential base pairs) sequence is tightly correlated with the dynamics of the helical parameters and backbone sub-state transitions for the internal dinucleotide steps. Specifically, fluctuations in the helical parameters shift, slide and twist for tetranucleotide sequences are often linked directly to the BI/BII phosphate transition equilibrium (discussed in detail in subsequent sections). Furthermore, dinucleotide steps comprised of purine-purine or pyrimidine-purine exhibit correlation between BII backbone conformation and the formation of a base-backbone hydrogen bond (Dans et al., 2014). This hydrogen bond stabilizes a low twist value when the dinucleotide step is flanked by a 3' purine. These dinucleotide steps promote a high twist/low twist helical transition. The cause of this transition is believed to be initiated by penetration of ions into the minor groove.

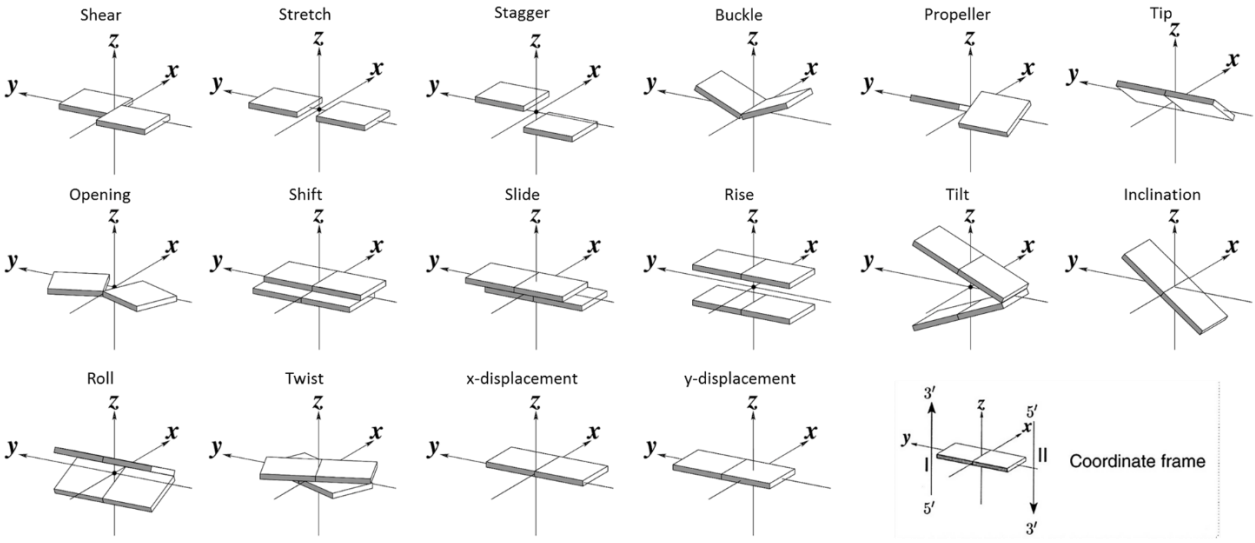



**Figure 1.5. Representation of base-pair and base-pair step helicoidal geometries. Adapted with permission (Lu et al., 2003).**

***Oxidative Stress***. Oxidative stress is characterized as an imbalance in a system between oxidants and antioxidants. In living systems, the primary source of this imbalance is the production of reactive oxygen species (ROS) either through endogenous or exogenous routes. ROS are molecules involving oxygen that are reactive with biological molecules such as DNA, proteins and lipids. Endogenously, ROS manifest from a response to inflammation and during oxidative phosphorylation in the form of superoxide ($^{\bullet}O_2^-$), hydrogen peroxide ($H_2O_2$) and hydroxyl radical ($^{\bullet}OH$) (Apel and Hirt, 2004). Two distinct enzymes of the electron transport chain are responsible for production of superoxide radicals: at complex I (NADH dehydrogenase) and at complex III (ubiquinone-cytochrome *c* reductase), with complex III generating the bulk (Turrens, 1997). These partially reduced species can leak from the mitochondria where they are free to react with essential biological molecules. Experiments completed *in vitro* reveal that the upper limit of molecular oxygen converted to superoxide anions by the mitochondria is approximately 3% when exposed to an environment of excess oxygen (Boveris and Chance, 1973). However, exposure to UV or ionizing radiation, prooxidants and transition metals have the potential to dramatically increase the cellular concentration of ROS species (Azzam et al., 2012; Shah et al., 2001)

Oxidation of nuclear and mitochondrial DNA, as well as free nucleotide pools, is well documented, with more than 20 oxidatively damaged DNA bases successfully identified (Cooke et al., 2003). Without removal or repair, these damaged nucleotides can lead to an increase in mutations or cell death. Mutations caused by oxidative damage to DNA have been linked to a multitude of degenerative diseases including cancer, cardiovascular disease, nervous and

immune-system decline as well as brain dysfunction (Ames et al., 1993). Guanine bases are the most susceptible to oxidation as they have the lowest redox potential of the four nucleotides (Jovanovic and Simic, 2001).

***Oxidation of guanine bases***. 8-oxo-7,8-dihydroguanine (8OG) is recognized as one of the most prominent forms of DNA damage resulting from ROS oxidation of a guanine base (Floyd, 1990). 8OG differs from native guanine nucleotides by two atoms: the introduction of an oxygen at carbon 8 (C8) and a hydrogen at nitrogen 7 (N7), effectively converting a hydrogen bond acceptor, N7, into a hydrogen bond donor (Figure 1.5).

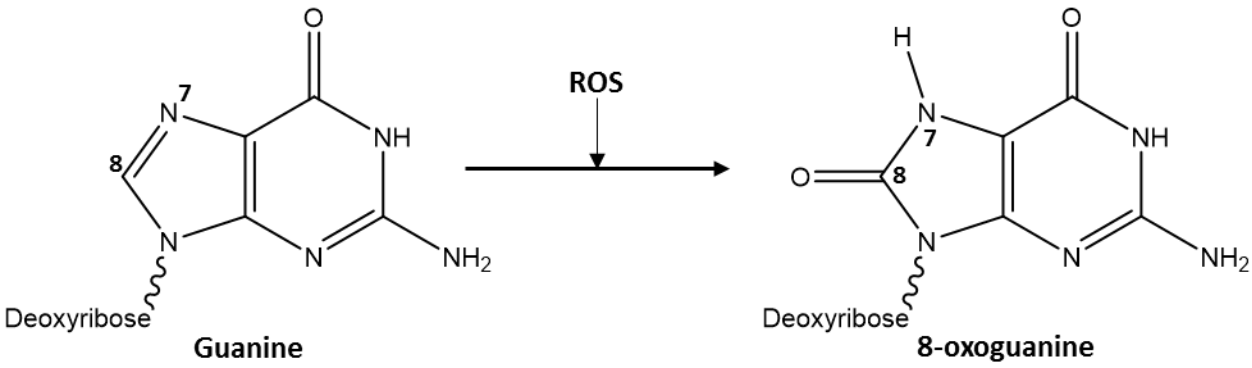



**Figure 1.6. Native guanine (left) and 8-oxoguanine (right) after ROS oxidation of the guanine base.**

This DNA lesion has been demonstrated to be both cytotoxic and mutagenic (Kozmin et al., 2005). Its mutagenic potential has been well characterized due to its ability to form a stable base pairing in both *anti*- and *syn*- conformations. In the former conformation, 8OG retains its ability to correctly form Watson-Crick hydrogen bonds with cytosine. However, when assuming the *syn*

conformation, 8OG can form a Hoogsteen base pair with adenine, effectively acting as a mimic to thymine (Figure 1.6). While many forms of DNA damage reduce or remove the function of proper DNA replication, the structural features of 8OG allow it to be erroneously bypassed by high-fidelity polymerases (Hsu et al., 2004). As a result, 8OG can mismatch with adenine during DNA replication forming an A·8OG intermediate. Consequently, 8OG becomes a direct source of G:C → T:A transversion mutations (Shibutani et al., 1997). Immunoaffinity-based urinalysis detected the formation and excision of approximately 100,000 8OG lesions per rat cell per day (Park et al., 1992). Furthermore, analysis of 2,500 somatic point mutations in the p53 gene of human tumors uncovered G→T transversion hotspots and established it as one of the most represented mutations (Hollstein et al., 1996).

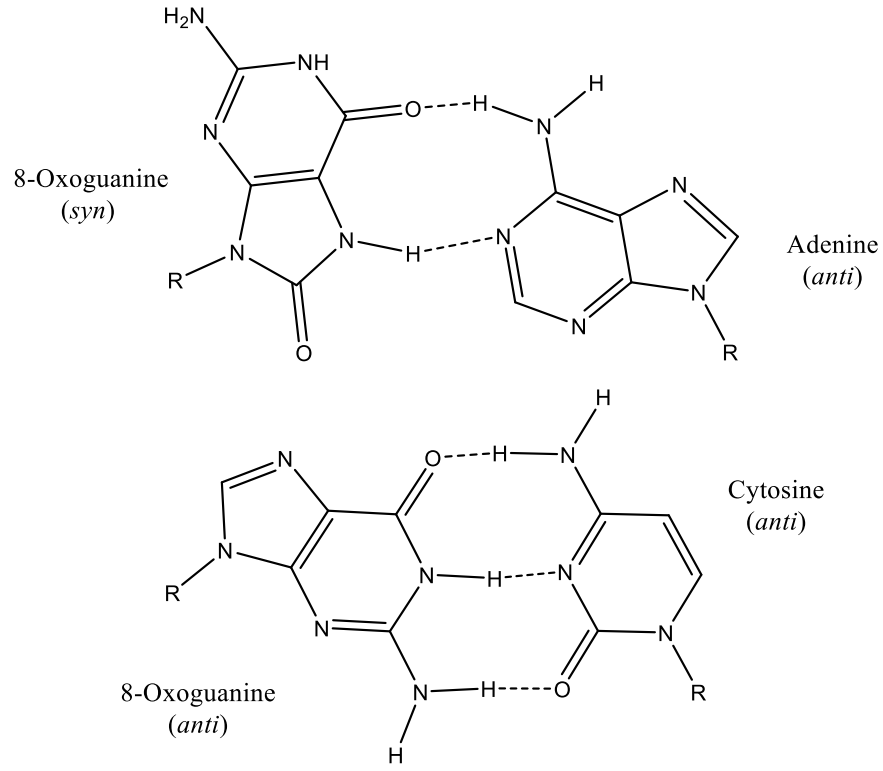



**Figure 1.7. (Top) 8OG in the *syn* conformation highlighting the two stable hydrogen bonds that make up the Hoogsteen base pair with adenine. (Bottom) 8OG correctly forming Watson-Crick hydrogen bonds with cytosine.**

***8OG repair mechanisms*.** Due to the various endogenous routes giving rise to 8OG, an extensive cellular repair pathway exists functioning to maintain the integrity of the genome. 8OG is repaired through the base excision repair (BER) pathway (Michaels and Miller, 1992). This cellular defense mechanism is catalyzed by three enzymes and has been well characterized in bacteria and humans. In bacteria, MutT, MutM and MutY constitute the BER repair pathway, with the corresponding human enzymes being MTH1, OGG1 and MUTYH. In the first step, MutT (or MTH1 in humans), acts to cleanse the free nucleotide pool by hydrolyzing 8O-dGTP, rendering it incapable of incorporation into the genome by DNA polymerases (Setoyama et al., 2011; Nakabeppu et al., 2006). The two remaining steps utilize DNA glycosylases that attempt to

recognize and excise their target nucleotide leaving an apurinic/apyrimidinic site that is identified and subsequently repaired by DNA repair machinery. The substrate for MutM (or OGG1 in humans) is the 8OG:C base pair, where after recognition, 8OG is everted from the DNA duplex and inserted deep into the catalytic pocket allowing hydrolysis of the *N*-glycosidic bond (David et al., 2007). The third enzyme, MutY (MUTYH in humans) acts on the 8OG:A mismatch as its substrate, inducing extraction by catalyzing hydrolysis of the *N*-glycosidic bond between the adenine base and its sugar (David et al., 2007). The initiation and subsequent excision of the 8OG nucleotide rely heavily on the ability to rotate the DNA backbone and formation of stable contacts with neighboring and cross-strand bases.

***DNA methylation*.** DNA methylation is a biological mechanism involving the enzymatic addition of a methyl group to a DNA base. In eukaryotes, this chemical modification is essential and is an established epigenetic gene regulator. Additionally, DNA methylation has been demonstrated to be an important component in cell differentiation, genomic imprinting and X-chromosome inactivation. For instance, experiments disrupting DNA methylation caused divergent regulation in developmental processes such as X-chromosome inactivation and genomic imprinting (Li, 2002). Disruption and errors in the methylation process have been linked to numerous degenerative diseases, including cancer (Robertson and Jones, 2000). The most frequent methylation process occurs at the 5-carbon of a cytosine base, resulting in 5-methylcytosine (5MC) (Figure 1.7). Single base resolution maps from human embryonic stem cells and fetal fibroblasts reveal 5MC accounts for ~1.5% of genomic bases (Lister et al., 2009). In mammals, 5MC most commonly occurs at a CpG dinucleotide (shorthand for cytosine-phosphate-guanine),

with approximately 70-80% of CpG dinucleotides methylated (Ehrlich et al., 1982*)*. Further, methylation of CpG dinucleotides in humans is highly symmetrical, with 99% of dinucleotides being methylated on both strands (Lister et al., 2009). DNA methylation is regulated by various enzymes, with more than 95% of genomic methylation controlled by two DNA methyltransferases (DNMTs): DNMT1, a maintenance DNMT, and DNMT3b, a *de novo* DNMT (Rhee et al., 2002; Balada et al., 2008). Both enzymes catalyze a reaction between cytosine and S-adenosyl-L-methionine (SAM) to produce 5MC. DNMT1 is responsible for maintaining DNA methylation patterns post DNA replication (Jeltsch, 2006). Following DNA replication, CpG sites will only contain a single methylated cytosine. Consequently, hemimethylated CpG sites are the preferred substrate for DNMT1, with an approximate 24-fold preference over unmethylated CpG sites (Hermann et al., 2004). In contrast, DNMT3b does not hold preference over hemimethylated CpG and is responsible for methylation during embryogenesis and early development sites (Okano et al., 1998).

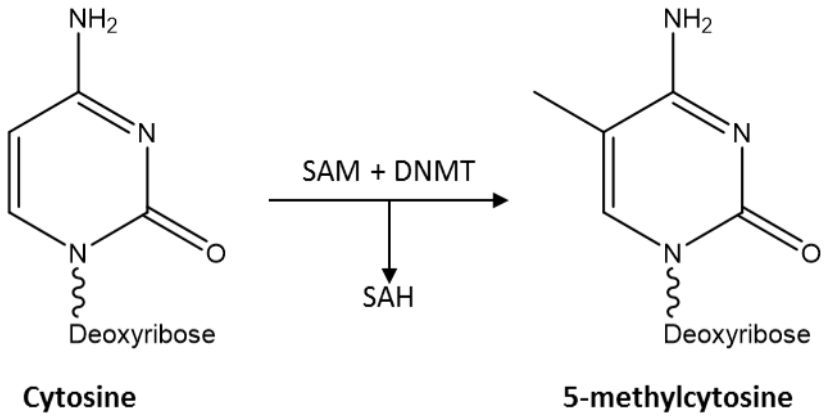



**Figure 1.8. Cytosine methylation occurs when a DNMT catalyzes a reaction between SAM and cytosine. As seen on the right, the methyl group is attached to the 5-carbon of the cytosine base.**

From the earliest nearest neighbor base sequence analysis, it has been apparent that in vertebrates, the frequency of genomic CpG distribution is lower than expected from statistical prediction (Josse et al., 1961; Swartz et al., 1962). Initial DNA sequencing of the human genome detected an average CG content of 41%, therefore, the expected frequency of CpG dinucleotides would be ~4.2% (Lander et al., 2001; Venter et al., 2001). However, the observed frequency is underrepresented, showing the frequency to be ~1%. This discrepancy can in part be rationalized due to spontaneous deamination of 5MC to thymine *in vivo* (Shen et al., 1994)

CpG dinucleotides are commonly found in clusters of several hundred base pairs and are referred to as CpG islands. In humans, CpG islands are distributed at the 5'-untranslated regions, upstream of over 70% of genes. Most often, CpG islands are unmethylated; which is particularly apparent in genes which are ubiquitously expressed and required for cellular function (Zhu et al., 2008). Consequently, the methylation status of a CpG island is associated with the regulation of expression of the downstream gene (Jones et al., 1999). Without proper CpG island hypomethylation, expression of essential genes such as tumor suppressor genes can become

compromised (Esteller, 2005). Alternatively, genes natively suppressed through CpG island hypermethylation can become aberrantly over-expressed.

***Enzymatic implications of clustering 8-oxoguanine and 5-methylcytosine*.** The enzymatic reaction between OGG1 and 8OG creates multiple, stable reaction intermediates (Kuznetsov et al., 2005). Accordingly, the kinetics of the OGG1 mediated reaction (Figure 1.8) cannot be described by a classical Michaelis-Menten scheme.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} EP \xrightarrow{k_3} E + P$$

**Figure 1.9. Kinetic scheme for catalysis by OGG1. $k_2$ is the rate constant for cleavage of the N-glycosidic bond and $k_3$ is the rate-limiting β-elimination step.**

The enzymatic rates of 8OG removal by OGG1 have been determined with numerous methylated CpG substrates (*trans*-, *cis*- and fully-methylated). OGG1 effectively removed 8OG from all methylation contexts (Kasymov et al., 2013). However, when the CpG dinucleotide was *cis*-methylated, that is, 8OG and 5MC are adjacent; the rate of excision ($k_2$) was reduced by half while the rate of product release ($k_3$) showed no significant deviation from unmethylated control. In contrast, when OGG1 was introduced to DNA substrate with 5MC opposite to 8OG (both trans- and fully-methylated contexts), the rate of $k_2$ was unchanged while $k_3$ was reduced by half. The rates for the fully-methylated substrate are consistent with the *trans*-methylated substrate.

The introduction of 8OG into a CpG site has been determined to drastically affect the *in vitro* methylation rates of murine DNA methyltransferase DNMT3a (Maltseva et al., 2009) and bacterial HpaII methylase (Weitzman et al., 1994). The introduction of 8OG immediately downstream of the target cytosine on a hemimethylated CpG substrate reduced the rate of methylation 25-fold. On the other hand, when 8OG was inserted directly across from the target cytosine, a 1.8-fold acceleration in methylation rates was observed. In human DNMT1, a similar but less drastic effect was observed; with up to a 13-fold reduction in methylation rates in the presence of downstream 8OG (Turk et al., 1995). It is hypothesized that the downstream 8OG inhibits selective binding from steric repulsion between O8 of 8OG and key binding residues of the methyltransferases. (Maltseva et al., 2009)

Under physiological conditions, the steady state level of 8OG in the human genome is estimated to be 0.5-5 8OG per $10^6$ guanine nucleotides (ESCODD, 2002). This level may dramatically increase with exposure to environmental oxidative stress and as a result of many diseases. As CpG dinucleotides represent a frequency of ~1% of all nucleotide pairs and the magnitude of each human diploid cell is estimated to be ~$6x10^9$ nucleotide pairs, each cell would, therefore, contain $6x10^7$ CpG dinucleotides. Consequently, each diploid cell can be estimated to have several hundred CpG dinucleotides that harbor 8OG lesions. In mammals, up to 80% of all CpG sites are methylated, making the introduction of 8OG to a methylated CpG a biologically significant occurrence.

***Importance of DNA structure and dynamics for enzymatic recognition*.** The capacity for proteins to selectively recognize DNA sequences is the foundation for biological regulation and repair. Sequence-specific recognition was found to occur through hydrogen bonding on the amino acid-DNA base interface (Seeman et al., 1976). It is acknowledged that proteins can distinguish base-sequence identity in the major groove utilizing at minimum two hydrogen bond interactions (Seeman et al., 1976). Analysis of 3D structures of protein—DNA interfaces within the last decade revealed that binding specificity is often reliant on insertion of positively charged side chains into the narrower minor groove, which interact with the negatively charged phosphate groups (Rohs et al., 2009). Although protein—DNA interfaces most often rely on positively charged contact surfaces, some proteins use helix-dipole interactions to stabilize phosphates along the DNA minor groove. Figure 1.9a displays the full crystal structure solved by (Bruner et al., 2000) of OGG1 bound to its 8OG-containing DNA substrate. Four residues (Cys 253, Gln 315, Phe 219 and Gly 42) make up the catalytic pocket, which sample the everted nucleotide (Figure 1.9b). The phosphate backbone is stabilized by a lone basic residue (His 270). Unlike many other DNA-protein complexes, helix-dipole interactions form the bulk of the protein—phosphate contacts. Tyr 203 is inserted into the middle of the duplex acting as a wedge, helping increase the total bend and kink of the DNA substrate. Finally, Asn 149 directly stabilizes the cytosine cross-strand from the everted 8OG.

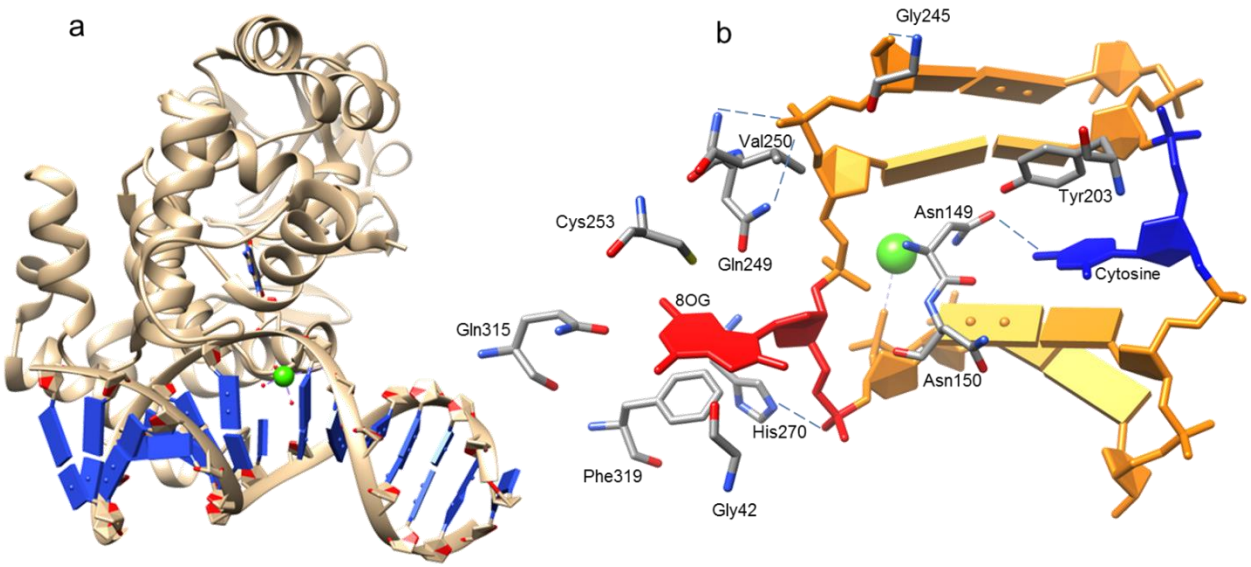



**Figure 1.10. (a) Full crystal structure (2.1 Å resolution) of OGG1 bound to 8OG containing substrate (PDB ID 1EBM). 8OG is seen everted out of the DNA duplex and into the catalytic pocket of OGG1 (b) Depiction of the OGG1—DNA interface. 8OG is in red, 5MC in blue. Hydrogen bonds are shown with blue lines. Helix-dipole interactions include Gly 245, Val 250 and Gln 249 Both images were produced using the UCSF Chimera package.**

DNA regulatory glycosylases and methyltransferases seek out their respective substrates through facilitated diffusion, in a Brownian manner, sliding along the DNA backbone and sampling stretches of DNA during a single binding event (Hedglin et al., 2014; Jeltsch, 2006). Therefore, the hydrogen bonding stability, solvation, ion-directed contacts, electrostatics, steric hindrance, base-pair/-step geometries and conformational dynamics of the DNA duplex at the local level are all important determinants for enzyme substrate discrimination. For instance, the base pair and backbone dynamics are integral features for recognition of uracil in DNA by uracil DNA glycosylase (Parker et al., 2007), which are hypothesized to play critical roles for damage recognition by similar DNA glycosylases.

The introduction of 8OG has been shown to decrease the thermal stability of the DNA duplex, with only minor local structural perturbations located on the backbone (Singh et al., 2011). Nevertheless, 8OG is efficiently detected and repaired through the BER pathway. While the reaction intermediates of base excision have been well characterized (Kim and Wilson III, 2012), it remains unclear how DNA *N*-glycosylases are able to locate the 8OG nucleotides within the vast sea of undamaged G bases. Solid-state NMR experiments previously determined that structural alterations from cytosine methylation are somewhat more pronounced. This includes narrowing of the minor groove and reduction in the amplitudes of motions in the sugar-phosphate backbone (Geahigan et al., 2000). Furthermore, it has been demonstrated that the thermal stability of the DNA duplex is increased with the addition of 5MC (Rodríguez López et al., 2010). It is apparent that DNA binding proteins are able to establish specificity using a unique variety of stabilizing interactions, all dependent on the local and global conformations of the DNA substrate. As DNA is a highly flexible molecule, these conformations fluctuate, on different timescales. Motions such as backbone transitions and fluctuations in helicoidal geometries are fast and happen on the picosecond to nanosecond timescale (Trieb et al., 2004; Pérez et al., 2007). In contrast, fluctuations in base-pair openings, which are critical for base flipping, are much slower, occurring on the millisecond timescale (Coman and Russu, 2005). Fast motions in DNA on the femtosecond to nanosecond timescale have been probed by a multitude of experiments including FTIR spectroscopy (Rüdisser et al., 1997) and fast field cycling NMR relaxometry (Roberts et al., 2004). Millisecond timescales can be probed with NMR, evaluating imino proton dynamics (Coman and Russu, 2005; Bhattacharya et al., 2002). Results from these experiments have all been corroborated with molecular dynamics simulations on DNA sampling

motions from nanosecond to more recently millisecond timescales (Dans et al., 2016; Galindo-Murillo et al., 2014).

***Research aims***. Recognizing the apparent alteration of kinetic rates for 8OG base extraction by OGG1 (Kasymov et al., 2013) on *trans*-, *cis*- and fully-methylated CpG substrates, as well as the perturbation of methylation efficiency of DNA methyltransferases (Weitzman et al., 1994; Maltseva et al., 2009), presents the need to investigate clustering of these modifications in more detail. Specifically, the lack of information regarding the thermal stability, structural perturbations and all-atomic motions of DNA duplexes containing clustered 8OG and 5MC prompted this investigation. We probed for structural perturbations as well as base dynamics with solution NMR spectroscopy and evaluated the thermodynamic stability through UV spectrophotometry of duplexes with CpG sites containing all biologically relevant methylation and oxidation patterns. Finally, to investigate dynamics differences on the ps-ns timescale, we evaluated the all atomic motions of the duplexes using free molecular dynamic simulations.

# *Chapter II – Methodology*

***Sample Design.*** In total, twelve synthetically constructed DNA duplexes based on the Drew-Dickerson dodecamer (DDD) (Drew et al., 1981), a *de facto* A/B DNA standard were analyzed. Each sample contains a different oxidation, methylation or oxidation+methylation pattern. The list of samples is presented in Table 2.1. This table includes the naming convention that will be used for each sample throughout the remainder of the manuscript along with the associated PDB deposition ID solution NMR structure determination. Oligonucleotides $^{m}$C3, $^{m}$C9 and $^{m}$C3/$^{m}$C9 were purchased from Midland Certified Reagent Co. whereas the remaining were synthesized using commercially available phosphoramidites (Glen Research, Sterling, VA) in the lab of our collaborator Dr. Dmitry Zharkov (Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia). The samples were designed to gain insight for clustering of oxidation and methylation. Due to the palindromic nature of the DNA sequence, the designed modification exists twice for each sample. Sample logic is presented in Figure 2.1.

| Sample name | methylation with respect to 8OG | PDB ID | Sequence |
|---|---|---|---|
| DDD* | | 1BNA | d($C_1$ $G_2$ $C_3$ $G_4$ $A_5$ $A_6$ $T_7$ $T_8$ $C_9$ $G_{10}$ $C_{11}$ $G_{12}$) |
| $^{oxo}$G4 | - | 5IV1 | d(CGC(**$^{oxo}$G$_4$**)AATTCGCG) |
| $^{oxo}$G10 | - | 5IZP | d(CGCGAATTC(**$^{oxo}$G$_{10}$**)CG) |
| $^{m}$C3 | - | 5L06 | d(CG(**$^{m}$C$_3$**)GAATTCGCG) |
| $^{m}$C9 | - | 5L2G | d(CGCGAATT(**$^{m}$C$_9$**)GCG) |
| $^{m}$C3/$^{m}$C9 | - | 5TMI | d(CG(**$^{m}$C$_3$**)GAATT(**$^{m}$C$_9$**)GCG) |
| $^{m}$C3/$^{oxo}$G10 | trans | 5UZ1 | d(CG(**$^{m}$C$_3$**)GAATTC(**$^{oxo}$G$_{10}$**)CG) |
| $^{m}$C9/$^{oxo}$G4 | trans | 5TRN | d(CGC(**$^{oxo}$G$_4$**)AATT(**$^{m}$C$_9$**)GCG) |
| $^{m}$C3/$^{oxo}$G4 | cis | | d(CG(**$^{m}$C$_3$**)(**$^{oxo}$G$_4$**)AATTCGCG) |
| $^{m}$C9/$^{oxo}$G10 | cis | 5UZ3 | d(CGCGAATT(**$^{m}$C$_9$**)(**$^{oxo}$G$_{10}$**)CG) |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | full | | d(CG(**$^{m}$C$_3$**)(**$^{oxo}$G$_4$**)AATT(**$^{m}$C$_9$**)GCG) |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | full | 5UZ2 | d(CG(**$^{m}$C$_3$**)GAATT(**$^{m}$C$_9$**)(**$^{oxo}$G$_{10}$**)CG) |

**Table 2.1. Sequences of oligonucleotides used in this study.**

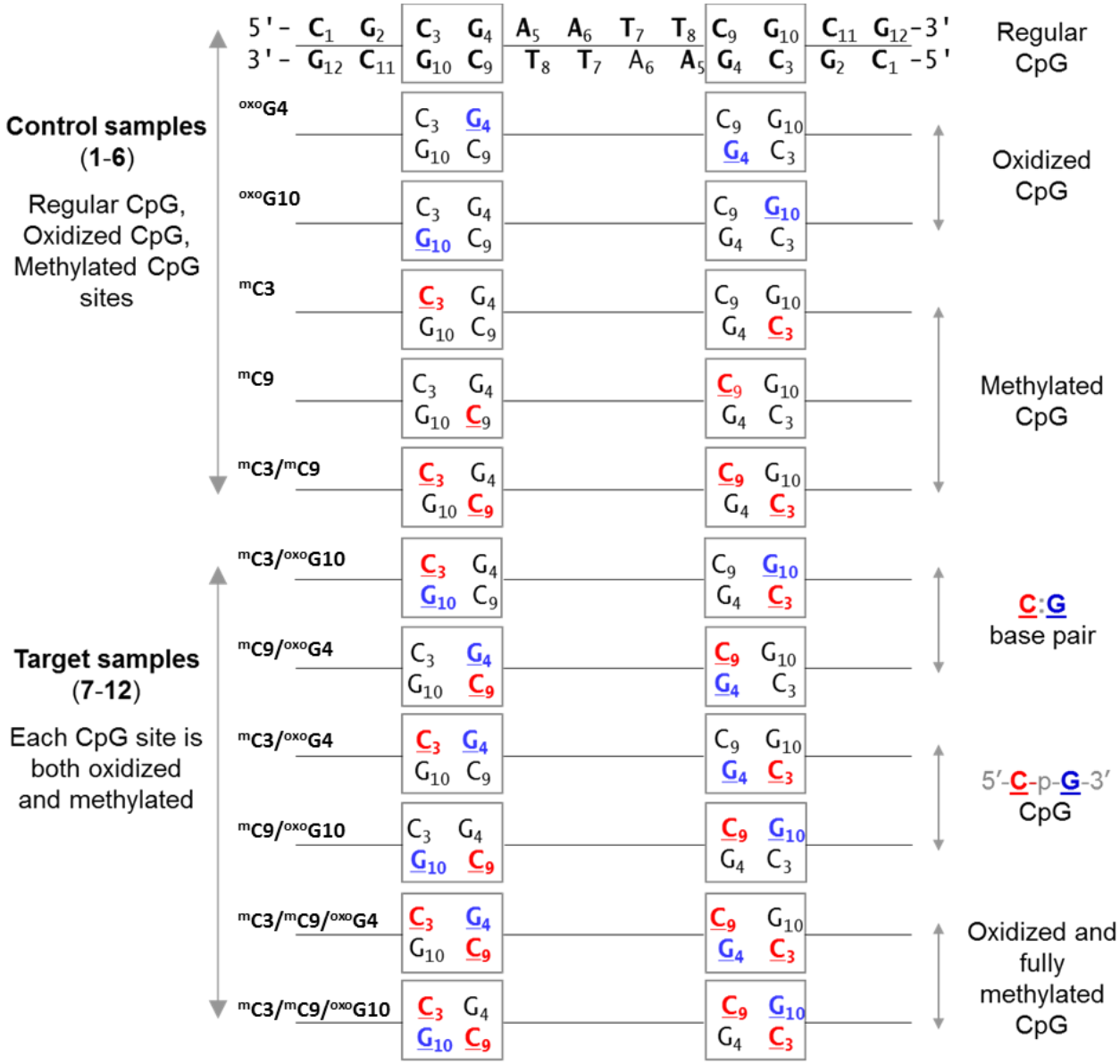



**Figure 2.1. Sample logic highlighting the methylation and oxidation pattern for each control and target sample. Blue is oxidation; red is methylation.**

***Thermal denaturation.*** Melting temperatures for all sampled duplexes were determined by UV-Vis spectrophotometry in the lab of Dr. Dmitry Zharkov.

***NMR spectrometers.*** NMR data were recorded on Bruker AVANCE III and Avance III HD spectrometers operating at proton Larmor frequencies of 500, 600, 700 and 850 MHz and equipped with 5 mm probes (Prodigy TCI cryoprobes and room temperature SmartProbes).

***NMR sample preparation*.** DNA samples were prepared from lyophilized solid with a buffer comprised of 10 mM sodium phosphate (pH 6.8), 50 mM NaCl, 1 mM EDTA and 0.3-1.1 mM oligonucleotide. The pH of the solution was adjusted to 6.8 with NaOH and HCl without regard for the DHO effect. The aqueous solvent consisted of either $H_2O$ + $D_2O$ at a 9:1 ratio or 99.9% $D_2O$. When required, solvent exchange between $D_2O$ and $H_2O$ was performed on either an Air Products CSW-202 Displex Cryogenic Refrigerator or a Labconco FreeZone 2.5 freeze dryer. Approximately 500 µL DNA samples were prepped for lyophilization by flash freezing in liquid nitrogen. Freeze drying was completed overnight or until visual inspection revealed no remaining ice in the sample tube. Following lyophilization, rehydrated DNA samples were annealed to ensure proper hybridization. A microfuge tube containing the DNA sample was incubated on a heat block at 95 °C for 5 minutes. Following incubation, the sample was left to cool to room temperature prior to data collection.

***DNA purification***. Trace impurities from chemical synthesis were commonly found through initial 1D-$^1$H NMR experiments. To remove residual impurities, samples were loaded onto a HiPrep 16/60 Sephacryl S-100 HR size exclusion column (GE Healthcare Life Sciences) and fractionated at a flow rate of 0.5 mL/min using fast protein liquid chromatography (FPLC—ÄKTA Prime Plus). The buffer used during FPLC was 10 mM sodium phosphate (pH 6.8), 50 mM NaCl and 1 mM EDTA. Following FPLC, fractions containing DNA were pooled and concentrated to receive adequate NMR signature. Concentration was achieved using pre-rinsed (to remove trace

glycerin) 3 kDa MW cutoff 15-mL Amicon centrifugal filters. Centrifugation was carried out at 4000 x g in 8 min intervals. DNA concentration was determined at 260 nm using micro-volume UV-Vis spectrophotometry on a BioTek Epoch Microplate Spectrophotometer.

***1-Dimensional NMR spectroscopy*.** 1D $^{1}H$ spectra for temperature profiles were recorded for each sample starting from 333 K to 278 K in 5 K decrements. Phase correction, baseline correction and NMR line width analysis was conducted using MestReNova version 10.0. The resonance line position of $H_2O$ was used as chemical shift reference for all proton spectra with the values corrected for each temperature (Gottlieb et al., 1997). For the 1D $^{1}H$ data recorded at 500 MHz, 600 MHz and 700 MHz experiments, 65536 (64K), 98304 (96K) and 131072 (128K) data points respectively were collected to deliver a digital resolution of 0.11 - 0.15 Hz per data point; For each spectrum, 1024 scans were acquired to obtain sufficient sensitivity (S/N). Solvent peak-suppression was performed using excitation sculpting with the pulse program 'zgesgp'. The homogeneity of the magnetic field was optimized at each temperature point by shimming along x, y and z axes. In total, 132 individual experiments were conducted. With each experiment taking approximately 2 hours to complete, manual adjustment at each temperature point was impractical. Therefore, a TopSpin AU macro was written to automate the process (Appendix Script 1).

***2-Dimensional NMR spectroscopy*.** The standard set of $^{1}H$-$^{1}H$ and $^{1}H$-$^{31}P$ 2D data was collected and utilized for the solution NMR structure determination (Hare et al., 1983). A combination of TOCSY and NOESY (70/90, 140, 200 and 260 ms mixing times) allowed us to perform the NMR resonance assignment. For the solution structure determination, the NOE distance restraints and backbone conformational restraints (BI vs. BII ranges) from $^{1}H$-$^{31}P$ HETCOR were utilized. Watson-

Crick hydrogen bonding was verified using $^{1}H$-$^{1}H$ NOESY spectra recorded in 10% $D_2O$. NMR data were processed with NMRPipe (Delaglio et al., 1995) and visualized using NMRDraw from the NMRPipe package and NMRViewJ (Johnson, 2004). Figures of NMR spectra were produced using MestReNova. NOESY cross-peak assignment was performed using the standard procedure (Hare et al., 1983).

***Nuclear Overhauser effect H-H distance calibration***. Generation of distance restraints were derived from NOESY cross-peak intensities. NOE peak volumes were obtained through integration in NMRViewJ. The CYANA 2.1 (Güntert, 2004) program was used to calibrate conformational distances. The fixed cytosine H5-H6 distance (2.5 Å) was used as internal reference to determine the quality of the calibration. Following calibration of each of the four mixing time NOESY experiments, cross-peaks present to sufficient intensity in > 1 NOESY spectra had their associated distances averaged. Cross peaks not present in multiple NOESY spectra were kept as non-averaged distances.

***Conformational restraints for simulated annealing***. Calibrated NOE derived distances were binned into four groups based on distance: < 3.0 Å, 3.0-5.0 Å, > 5.0 Å, and a fourth group contained distances involving methyl groups as their NOE peak volumes are saturated with three proton resonances. Each grouping was assigned an upper and lower distance tolerance. Distances less than 3.0 Å were given a tolerance of 0.50 Å, those between 3.0-5.0 Å were assigned 0.70 Å and 0.90 Å for distances greater than 5.0 Å. Methyl groups were treated as heteroatoms. Grouped distances were transformed into AMBER input format with an in-house python script. Watson-crick hydrogen bonding distances were obtained from (Saenger, 1984) and given a tolerance of 0.10 Å. Dihedral angle values were taken from (Arnott and Hukins, 1972). Epsilon

and zeta torsion angles for 3'-phosphates of 8OG nucleotides were restrained in BII conformation (ε-ζ > 20°) when $^{1}H$-$^{31}P$ resonances indicated BII phosphate conformation. Angles defining sugar rings were not restrained.

***AMBER driven simulated annealing.*** NMR refinement was performed using the sander program in AMBER 14 (Case et al., 2005; Pearlman et al., 1995) in which minimization and simulated annealing were carried out in the generalized Born (Still et al., 1990) electrostatic continuum. A-form and B-form starting structures for each system were built using nucleic acid builder (NAB) from AmberTools 15 and parameterized using the AMBER ffOL15 force field (Zgarbova et al., 2015). The libraries for parameters and topology were generated in-house for 8OG (Cheng et al., 2005) and 5MC (Lankaš et al., 2002) nucleotides. 8OG and 5MC nucleotides were manually edited in the starting structure PDB files and the resulting structure was subjected to 1000 cycles of energy minimization with 500 steps using the steepest decent algorithm. The force constant for lower and upper bounds of 32 kcal $mol^{-1}$ $deg^{-2}$ was set to all restraints excluding the backbone angles epsilon and zeta which were set to 512 kcal $mol^{-1}$ $deg^{-2}$. A total of 18 (9 for each A- and B-form starting structures) restrained molecular dynamics simulations were performed with varying target temperatures (580 K, 600 K and 620 K) and simulation times (205 ps, 215 ps and 225 ps). Temperature was controlled using the Berendsen coupling algorithm (Berendsen et al., 1984). The temperature of the system was increased rapidly from 0 K to the target temperature and held for 5 ps with a coupling constant of 0.4 ps. The weight of the restraints was gradually increased from 0.1 to 1.0 over the first 3 ps. The system remained at the target temperature (coupling constant of 4 ps) for 90 ps, 100 ps or 110 ps dependent for increasing total simulation times. The temperature bath was gradually cooled from the target temperature to 100 K over

100 ps with a coupling constant of 4 ps. To complete the simulated annealing the system was then rapidly cooled to 0 K over 10 ps while tightening the coupling constant from 1.0 to 0.05 ps. The refined structures were subjected to a final 1000 cycles of energy minimization. The ensemble of structures was determined based on lowest conformational penalty/energy values. A representative structure was built by averaging the lowest energy models and subjecting the resulting structure to energy minimization. Ensemble visualizations and RMSD calculations were carried out in VMD (Humphrey et al., 1996). Ensemble models were built using the UCSF Chimera package (Pettersen et al., 2004).

## *Free molecular dynamics simulations*

***System initialization***. The starting structure for each system was the B-form initial structure used for NMR refinement. Each structure was parameterized with the AMBER ffOL15 force field: a recently developed force field that combines ff99bsc0 with three angle modifications ($\varepsilon/\zeta_{OL1}$, $\chi_{OL4}$, and $\beta_{OL1}$). Each system was neutralized with 22 $Na^+$ ions using CPPTRAJ (Roe et al., 2013) from AmberTools 15. Approximately 5000 explicit solvent molecules were added using the SPC/E water model (Berendsen et al., 1987) in truncated octahedral boxes. Additional $Na^+$ and $Cl^-$ ions were added for a final excess salt concentration of approximately 150 mM using Joung and Cheatham parameters (Joung and Cheatham III., 2008). Initial placement of excess ions was randomized using CPPTRAJ by swapping water and ion positions such that no ion was closer than 4 Å to another ion and that all ions were at least 6 Å away from the DNA duplex.

***System minimization and equilibration***. Each system followed the same minimization and equilibration protocol. Initially, the DNA atoms were held fixed. To eliminate Van der Waals

clashes, the solvent molecules and ions were subjected to 500 steps of the steepest descent minimization. Subsequently, 500 steps of conjugate gradient minimization, with a force constant of 500 kcal/mol-Å$^2$ applied to the solute molecule. Next, the whole system was minimized with 1000 steps of the steepest descent followed by 1500 steps of conjugate gradient minimization without restraints. Heating was completed over 20 ps at constant volume from 0 K to 300 K with weak positional restraints (10 kcal/mol-Å$^2$) applied to the DNA molecule. A 2 fs time step was used and a weak coupling thermostat were used to control temperature, with a collision frequency of 1.0 ps$^{-1}$. SHAKE (Ryckaert, 1977) was used to constrain bonds involving hydrogen atoms, with a tolerance of 0.002. A 9 Å cutoff was used for non-bonded interactions and Particle Mesh Ewald (PME) was used to handle long range electrostatics. Finally, 100 ps of MD was run at 300K with no restraints and constant pressure to relax the density of water. All other parameters were retained from the previous equilibration.

***Production molecular dynamics***. Production MD simulations were run for 600 ns using graphics processing (GPU) code with the PMEMD.cuda implementation of SANDER from Amber14 on an NVIDIA GTX 970 GPU. Simulations were held at constant pressure (1 atm) periodic boundaries and 300K using Berendsen coupling constants of 5.0 ps. Long-range interactions were calculated with PME. A 2 fs integration time step was used. Simulation coordinates were recorded every 1 ps. Twelve 600 ns simulations accumulated approximately 800 gigabytes of raw data before any post production processing occurred.

***Trajectory analysis.*** The first stage of trajectories analysis was performed with CPPTRAJ, where, the structural stability of the simulation was examined in terms of RMSD vs simulation time. RMSD figures were generated and running averages were calculated using the XMGRACE

program (http://plasma-gate.weizmann.ac.il/Grace/). RMSF per nucleotide was calculated using CPPTRAJ and plotted with GNUPLOT (Williams et al., 2012). After, the first 45 ns of each simulation was removed as equilibration giving 555 ns total simulation time for each sample to analyze. This step also worked to remove water molecules from the trajectory. Next, the raw angle data for backbone and sugar pucker torsions was retrieved using CPPTRAJ. Raw ε and ζ angle data was further processed and the BI/BII phosphate conformation ratio was calculated using an in-house python script. Average residence times and transition occurrences were calculated with respective in-house python scripting. The dihedral angles α and γ ratios were calculated using pytraj. Sugar puckers angles were binned, normalized and graphed using XMGRACE. The truncated trajectory was further analyzed to extract raw base-pair and base-pair step geometries followed by an in-house python script. The CpG site (C3-G22pG4-C21) was analyzed by PCA for each sample using CPPTRAJ. Only heavy atoms were used in the analysis by removing all other nucleotides outside of the target CpG. Global rotational and translational movements were removed by subjecting each frame of the trajectory to a RMS-fit against the overall average CpG coordinates. Psuedo-trajectories were built by projecting the averaged CpG coordinates over each principal component. Visualization of the pseudo-trajectory was accomplished using the UCSF Chimera package (Pettersen et al., 2004). Porcupine plots were generated using the normal mode wizard plugin for VMD.

# *Chapter III – NMR solution structure, thermal stability and base dynamics of DNA duplexes*

## *Introduction to NMR*

Solution nuclear magnetic resonance (NMR) is a technique used to gain site-specific insight at the atomic resolution on biomolecules. Since 1967, NMR has been used to study proteins and nucleic acids. Trends from the protein data bank (PDB) (Bernstein et al., 2000) indicate that the number of biomolecular structures solved using NMR increased substantially within the last two decades. The reliability of NMR for determining 3D structures was significantly increased with the development and rapid improvements of heteronuclear (e.g., $^{13}C$, $^{15}N$, $^{2}H$) NMR applications (Bax, 1994). One of the major advantages of solution NMR is that it allows detection of behaviors of biomolecules that are more synonymous to *in vivo* conditions. NMR relies on the magnetic moment proportional to the spin of atomic nuclei. Upon delivery of a constant external magnetic field by the spectrometer, two spin energy states exist. "spin +1/2" nuclei are lower energy and are aligned with the magnetic field while the higher energy "spin -1/2" nuclei oppose it. Then, the sample is irradiated by a radio frequency pulse sequence that corresponds to the energy difference (proportional to the magnetic moment) between the two spins, inducing an energy transfer between the +1/2 spin -1/2 spin states. Free induction decay, or the nuclear magnetic resonance signal, is generated from the precession of the excited spins around the magnetic field axis. This signal is recorded as electric current in detection coils and decomposed into a spectrum of frequencies through Fourier transformation (Cavanagh et al., 1995). Structural information about the sample of interest can be obtained from various nuclear spin interactions affecting the spectral features, which include:

***Chemical shift.*** The chemical shift is defined as the resonance frequency of the nucleus and is dependent on the type of nucleus that is resonating and its covalent and local magnetic environment. Therefore, chemical shift values help to uncover the geometry of the dihedral angles.

***J-coupling interactions.*** J-couplings arise from through-bond ("scalar") interactions between two nuclear spins which propagate via the shared electrons of the covalent bonds. These interactions help to delineate the covalent structure of the biomolecule and assign NMR chemical shift values to specific atoms. TOCSY and COSY and their heteronuclear analogs (e.g. HETCOR) are typical NMR experiments reporting J-coupling interactions.

***Dipole-dipole interactions.*** Dipole-dipole coupling refers to the magnetic interaction of two spins that are in close proximity. The internuclear interaction between two spins is proportional to $\frac{1}{r^6}$, where *r* is the distance between two nuclei undergoing cross-relaxation (nuclear Overhauser effect, NOE). Thus, these interactions are highly valuable in obtaining internuclear distance information as restraints for structural determination. NOESY and ROSY are typical NMR experiments reporting dipolar-coupling interactions.

## *Results*

***NMR characterization of the DNA duplexes***. Our lab has determined solution structures of 11 duplex DNA samples derived from the Drew–Dickerson dodecamer sequence listed in Table 2.1. This thesis will focus on the NMR refinement statistics for all samples presented in Figure 2.1, specifically looking for effects of clustering 5MC and 8OG. The composition of the 2D $^{1}H$ and $^{31}P$ resonance lines on the recorded NMR spectra indicates that each sample adopts a single A/B-DNA conformation. The peak patterns for the 2D NOESY spectra recorded in 99.9% $D_2O$ for all the samples are analogous to the NMR signature of $d(CGCGAATTCGCG)_2$ (Hare et al., 1983).

***Exchangeable proton resonances***. The 2D NOESY spectra recorded in 90% $H_2O$ confirmed Watson-Crick type hydrogen bonding within all the base pairs for all samples. Shown in Figure 3.1 are the internal G-C and A-T imino resonances for $^{m}C3/^{oxo}G4$ (spectra for all other samples was omitted). The terminal GC pairs were either low intensity or absent due to fast exchange with $H_2O$. The imino protons were assigned by referencing cytosine H5/H6 and adenine H2 base protons in the $D_2O$ NOESY. Additionally, a low intensity NOE between G4 H1 and A5 H2 allows discrimination between adenine 5 and adenine 6 resonance signatures as well as allow independent assignment of guanine 4 (Figure 3.1). Watson-Crick base pairing for GC pairs was substantiated from NOE connectivities between guanine H1 protons and cytosine H41/H42 protons. Moreover, canonical hydrogen bonding for TA pairs was verified by NOE connectivities between thymine H3 and adenine H2 protons.

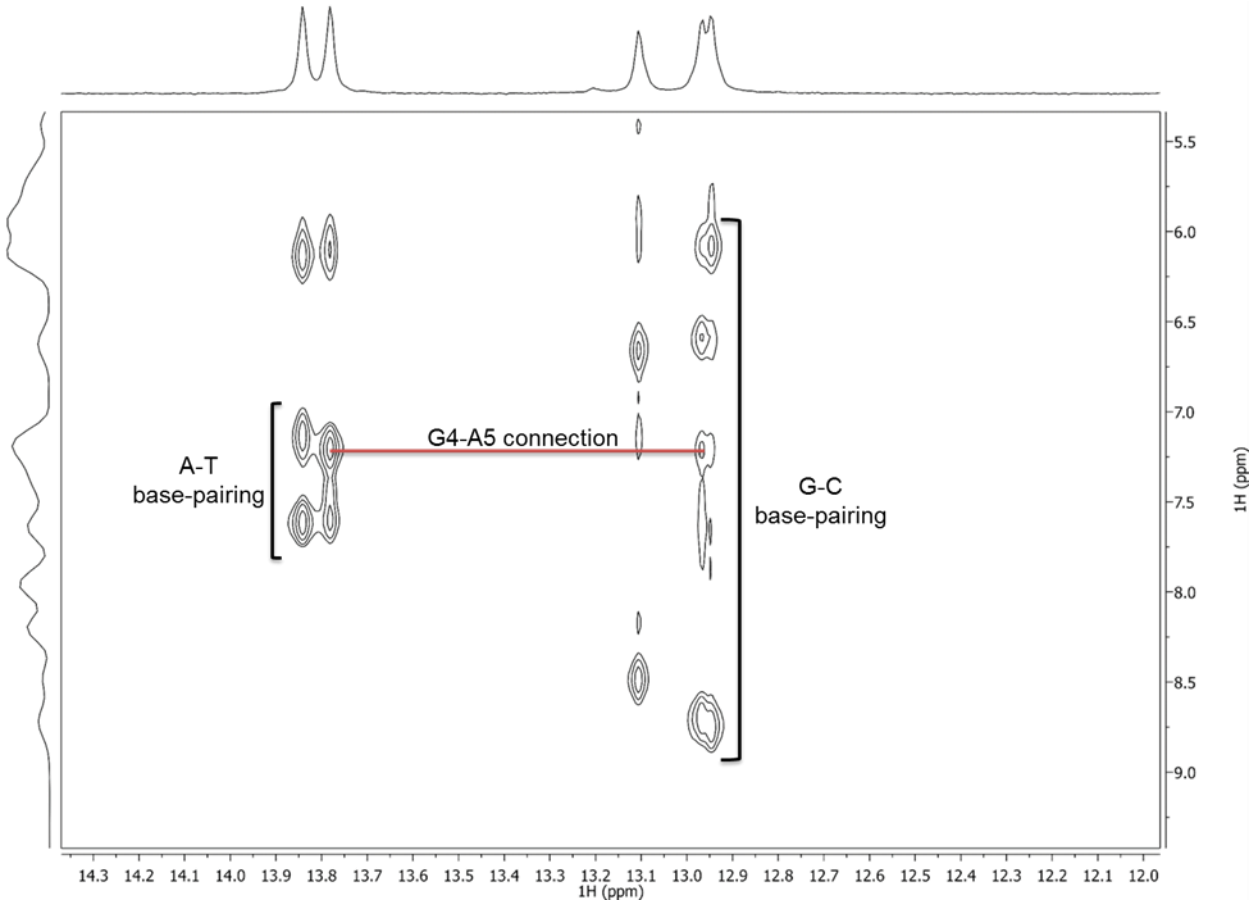



**Figure 3.1. $^{1}$H-$^{1}$H NOESY in 10% $D_2O$. The experiment was conducted on $^{m}$C3/$^{oxo}$G4 at 600 MHz at 279 K with a 200ms mixing time. G-C and A-T NOE patterns indicate Watson-Crick base pairing. Red bar highlights the G4-A5 NOE.**

***Non-exchangeable proton resonances***. The 2D NOESY spectra recorded in 99.9% $D_2O$ (Figure 3.2a) delivered a sufficient density of $^{1}$H-$^{1}$H NOE restraints for every sample and target CpG site within the samples (Table 3.1). NOE assignment was initiated with the identification of cytosine H5-H6 NOEs and mapping the connections between non-exchangeable H8/H6 base protons and sugar H1' protons (Figure 3.2b), in a region deemed the "walk". Beginning with the cytosine H5-H6 NOE, H8/H6 base protons to H1' sugar NOEs were identified in the sequential manner (*n*)H8/H6-(*n*)H1'-(*n*+1)H8/H6-(*n*+1)H1', where *n* is the nucleotide number. Sequential base-H1' connectivities involving 8OG were disrupted for all samples containing 8OG due to the loss of the

aromatic H8 proton. The remaining NOEs involving base-sugar protons were then assigned by moving stepwise upfield along the base proton signal and assessed based on NMR statistics for DNA in the Biological Magnetic Resonance Data Bank (BMRB). All remaining cross peaks were then assigned based off the initial base-sugar assignments. Adenine A5 showed cross-peaks involving its H2 to H1' protons belonging to A5 and A6. The presence of these NOEs further indicates the DNA duplex samples adopt right-handed conformations. All intra-residue H5-H6 cross-peaks were high intensity, dominating the "walk" region and at lower mixing times these NOEs displayed COSY-like patterns. Of the base-sugar proton interactions, the H8/H6-H2' cross peaks were the highest intensity indicating all bases were in the anti-conformation. All 8OG-containing samples presented the characteristic spectral features: absence of purine H8 NOESY cross-peaks, downfield shift of H2" (> 2.33 ppm), and presence of peaks involving HN7. Samples containing 5MC displayed the following unique attributes: absence of the cytosine H5-H6 NOE in the "walk region", an upfield shift on the base H6 $^{1}$H signal and additional high intensity methyl NOEs present upfield.

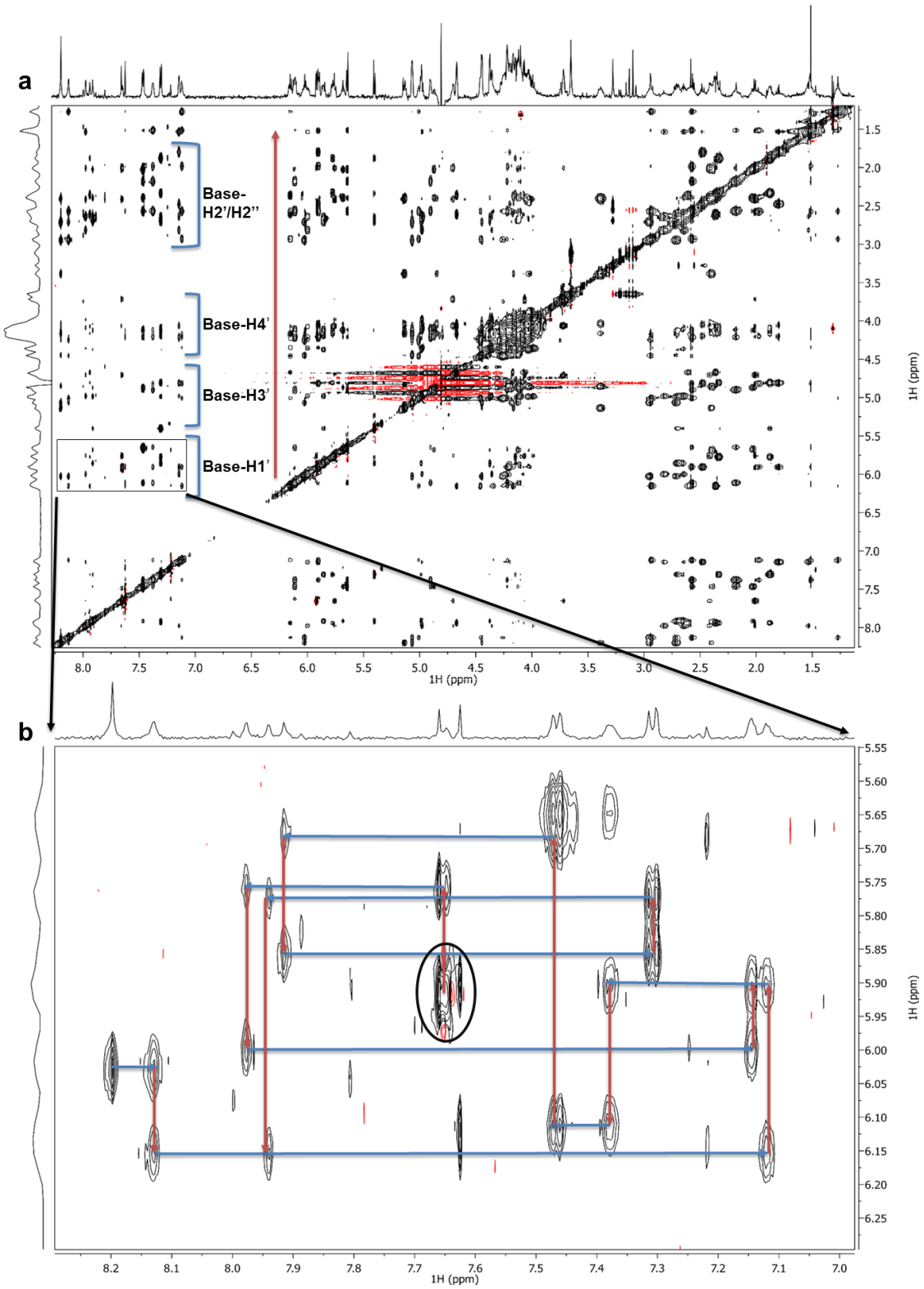



**Figure 3.2. (a) $^{1}$H-$^{1}$H NOESY of $^{m}$C3/$^{oxo}$G4 in 99.9% $D_2O$ at 298 K. Data collected at 600 MHz. Red arrow indicates the order of analysis during NOE assignment. (b) The "walk" base—sugar proton connectivities. Blue arrows indicate inter-base—sugar NOEs. Red arrows show intra-sugar—Base NOEs. Circled is the cytosine 1 H5-H6 NOE.**

| Sample | NOEs for entire sample | NOEs for target CpG sites |
|---|---|---|
| $^{oxo}$G4 | 379 | 140 |
| $^{oxo}$G10 | 462 | 188 |
| $^{m}$C3 | 404 | 154 |
| $^{m}$C9 | 328 | 138 |
| $^{m}$C3/$^{m}$C9 | 329 | 112 |
| $^{m}$C3/$^{oxo}$G10 | 404 | 134 |
| $^{m}$C9/$^{oxo}$G4 | 372 | 124 |
| $^{m}$C3/$^{oxo}$G4 | 336 | 108 |
| $^{m}$C9/$^{oxo}$G10 | 348 | 120 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | 376 | 130 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | 416 | 172 |

**Table 3.1. Number of assignable NOEs from $^{1}$H-$^{1}$H NOESY recorded in 99.9 % $D_2O$ for each sample**

***Insights into backbone conformation populations with $^{31}$P NMR***. The backbone conformation was probed with 1D $^{31}$P and 2D $^{31}$P-$^{1}$H NMR for all samples. Figure 3.3 displays the $^{31}$P-$^{1}$H HETCOR for $^{m}$C3/$^{m}$C9/$^{oxo}$G4 (spectra for remaining samples excluded). Assignment of the phosphorous signals between the sugar protons H3'/H4' and the backbone P was established with reference to the $D_2O$ NOESY and TOCSY (omitted) spectra *via J*-coupling. Both 1D (not shown) and 2D $^{31}$P-$^{1}$H NMR spectra show the majority of $^{31}$P resonances clustered within ~0.5 ppm (Table 3.2, Figure 3.3) for all samples as previously reported for the unmodified sequence (Ott and Eckstein, 1985). The $^{31}$P peaks from the 8OG phosphate groups were shifted downfield by ~0.15-0.44 ppm from the cluster of other $^{31}$P lines in every 8OG-containing sample (Table 3.2). The downfield chemical

shift value of the 8OG $^{31}$P resonance indicates a significant population of BII backbone conformation (Nikonowicz et al., 1989; Thiviyanathan et al., 2003) unlike the BI conformation of all other nucleotides. This prompted our restraints for the 8OG epsilon and zeta angles, and all refined structures exhibit the BII conformation for nucleotides containing 8OG.

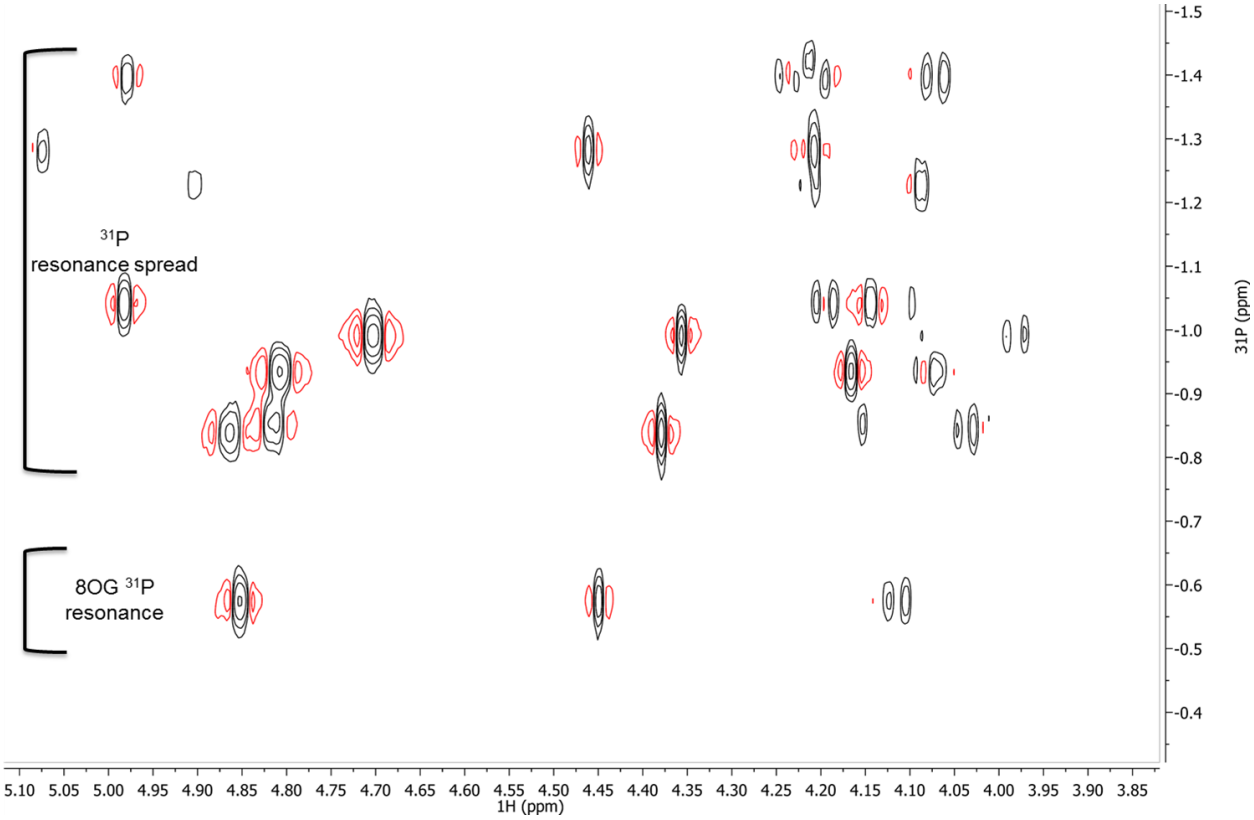



**Figure 3.3. 2D $^{1}$H-$^{31}$P NMR spectrum on $^{m}$C3/$^{m}$C9/$^{oxo}$G4. The spectrum was recorded at 295 K on a spectrometer operating at 600 MHz. The shifted $^{31}$P signature from 8OG is observed downfield from the major spread of phosphorous resonances.**

| Sample | $^{31}$P spread excluding 8OG (ppm) | 8OG $^{31}$P chemical shift from spread (ppm) |
|---|---|---|
| $^{oxo}$G4 | 0.57 | 0.15 |
| $^{oxo}$G10 | 0.47 | 0.35 |
| $^{m}$C3/$^{oxo}$G10 | 0.50 | 0.37 |
| $^{m}$C9/$^{oxo}$G4 | 0.57 | 0.24 |
| $^{m}$C3/$^{oxo}$G4 | 0.74 | 0.14 |
| $^{m}$C9/$^{oxo}$G10 | 0.52 | 0.44 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | 0.81 | 0.27 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | 0.54 | 0.44 |

**Table 3.2. $^{31}$P chemical shift spread for all samples with an oxidized guanine. A downfield shift of at least 0.14 ppm for all 8OG $^{31}$P resonances.**

***Solution NMR structures of control and target duplexes***. For each sample, an ensemble of lowest-energy structures was derived from rMD simulations and all eleven NMR ensembles are displayed in Figure 3.4a and 3.4b. The overall statistics for the final structural refinements for all samples is provided in Table 3.3. The greatest NOE restraint violation observed for any model was 0.35 Å and dihedral violations did not exceed 17.6°. Within each refined ensemble, the largest pairwise backbone RMSD value did not exceed 1.06 Å. Overall, our structures are highly homologous to one other, with an average backbone RMSD of 1.07 Å between the representative models, as well as to known DDD structures (PDB 1BNA, 355D and 1NAJ) with a backbone RMSD range of 1.26–3.8 Å across all samples. All samples containing 8OG show their backbone conformation 3' to 8OG in the BII conformation in accordance with $^{31}$P NMR resonances.

Analysis of the base-pair step geometries showed structural differences between target and control samples were minor and generally localized at and near their respective modification. Specifically, incorporation of 8OG at either G4 or G10 induces local unwinding (as indicated by twist and tilt) at the C3pG4 and C9pG10 base pair steps. Previous reports on DNA structures containing 8OG present mixed results in regard to local unwinding. For instance, 8OG has been shown to cause local unwinding at the site of modification (Thiviyanathan et al., 2003) while others were unable to detect significant local distortions (Crenshaw et al., 2011). All other base-pair steps did not show noteworthy structural differences. A general trend can be identified where 8OG lowers the twist (Figure 3.5a) value by up to 10° versus samples containing solely cytosine methylation (samples $^{m}$C3, $^{m}$C9 and $^{m}$C3/$^{m}$C9). The twist value for $^{m}$C3/$^{oxo}$G4 is divergent from the trend, showing an elevated twist value by several degrees above the rest. In addition, values of tilt (Figure 3.5b) within the target CpG base pairs are generally decreased for the C3pG4 step and increased for the C9pG10 step upon clustering 8OG and 5MC, with respect to methylated only samples.

| Sample | Number of models in ensemble | Greatest backbone RMSD within ensemble (Å) | Backbone RMSD range between representative models (Å) | Greatest NOE Violation (Å) | Greatest angle violation (deg.) |
|---|---|---|---|---|---|
| $^{oxo}$G4 | 13 | 0.66 | 1.20-1.86 | 0.07 | 7.49 |
| $^{oxo}$G10 | 14 | 0.87 | 1.13-2.05 | 0.10 | 10.19 |
| $^{m}$C3 | 8 | 0.89 | 0.62-2.22 | 0.03 | 17.6 |
| $^{m}$C9 | 8 | 0.69 | 0.62-2.34 | 0.17 | 13.5 |
| $^{m}$C3/$^{m}$C9 | 10 | 0.71 | 0.89-2.10 | 0.04 | 13.2 |
| $^{m}$C3/$^{oxo}$G10 | 11 | 0.80 | 1.07-2.23 | 0.08 | 15.1 |
| $^{m}$C9/$^{oxo}$G4 | 11 | 0.85 | 1.52-2.35 | 0.14 | 12.5 |
| $^{m}$C3/$^{oxo}$G4 | 10 | 0.80 | 1.08-2.26 | 0.19 | 11.3 |
| $^{m}$C9/$^{oxo}$G10 | 9 | 1.06 | 0.88-2.36 | 0.07 | 13.7 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | 10 | 0.68 | 1.17-1.96 | 0.35 | 8.9 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | 13 | 0.84 | 0.90-1.82 | 0.09 | 15.9 |

**Table 3.3. Statistics for final structural refinements on all control and target samples.**

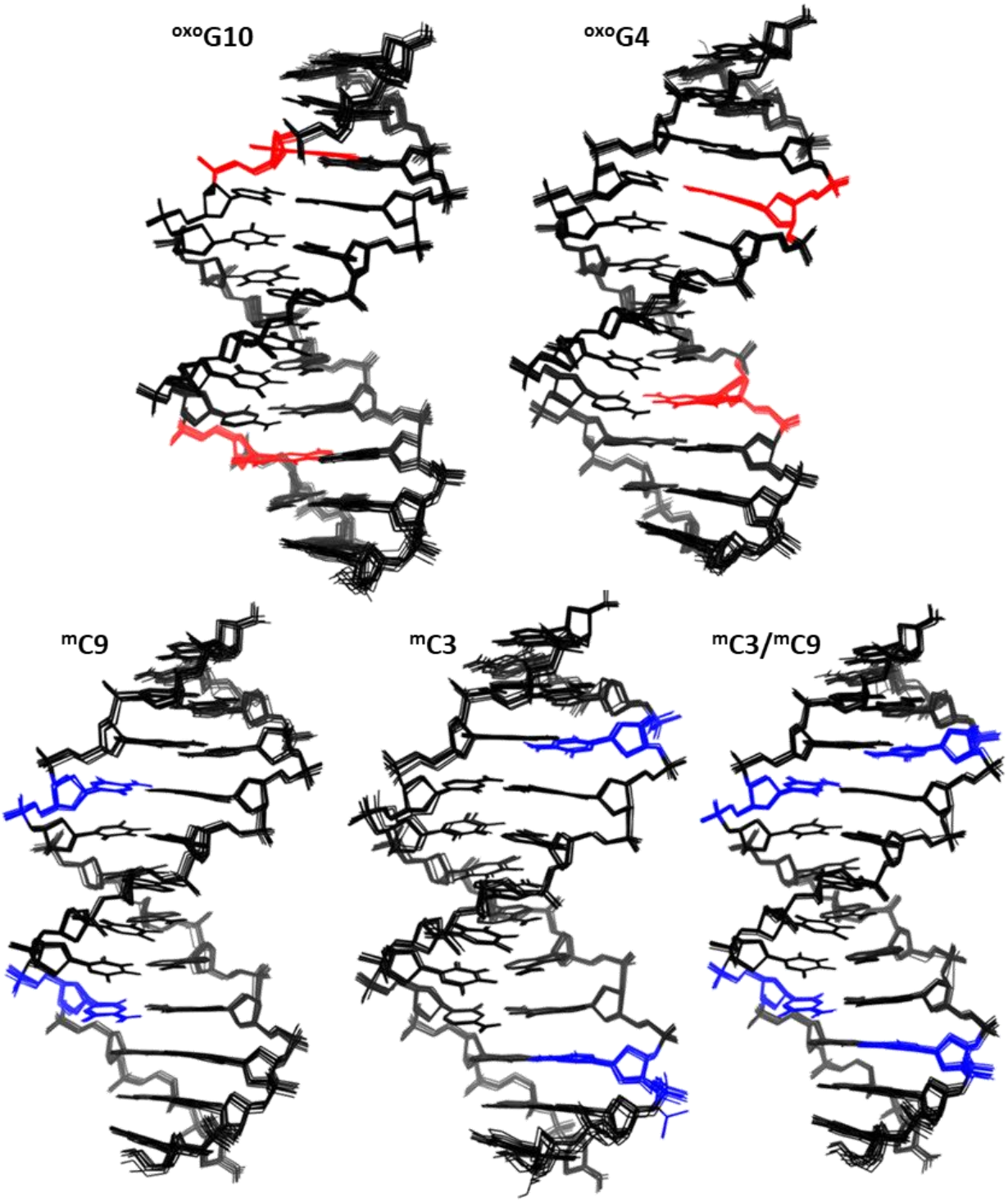



**Figure 3.4a. Solution NMR ensembles of $^{oxo}$G10, $^{oxo}$G4, $^{m}$C9, $^{m}$C3, $^{m}$C3/$^{m}$C9. 5MC nucleotides are colored blue and 8OG nucleotides are colored red. The ensembles display high regularity between samples.**

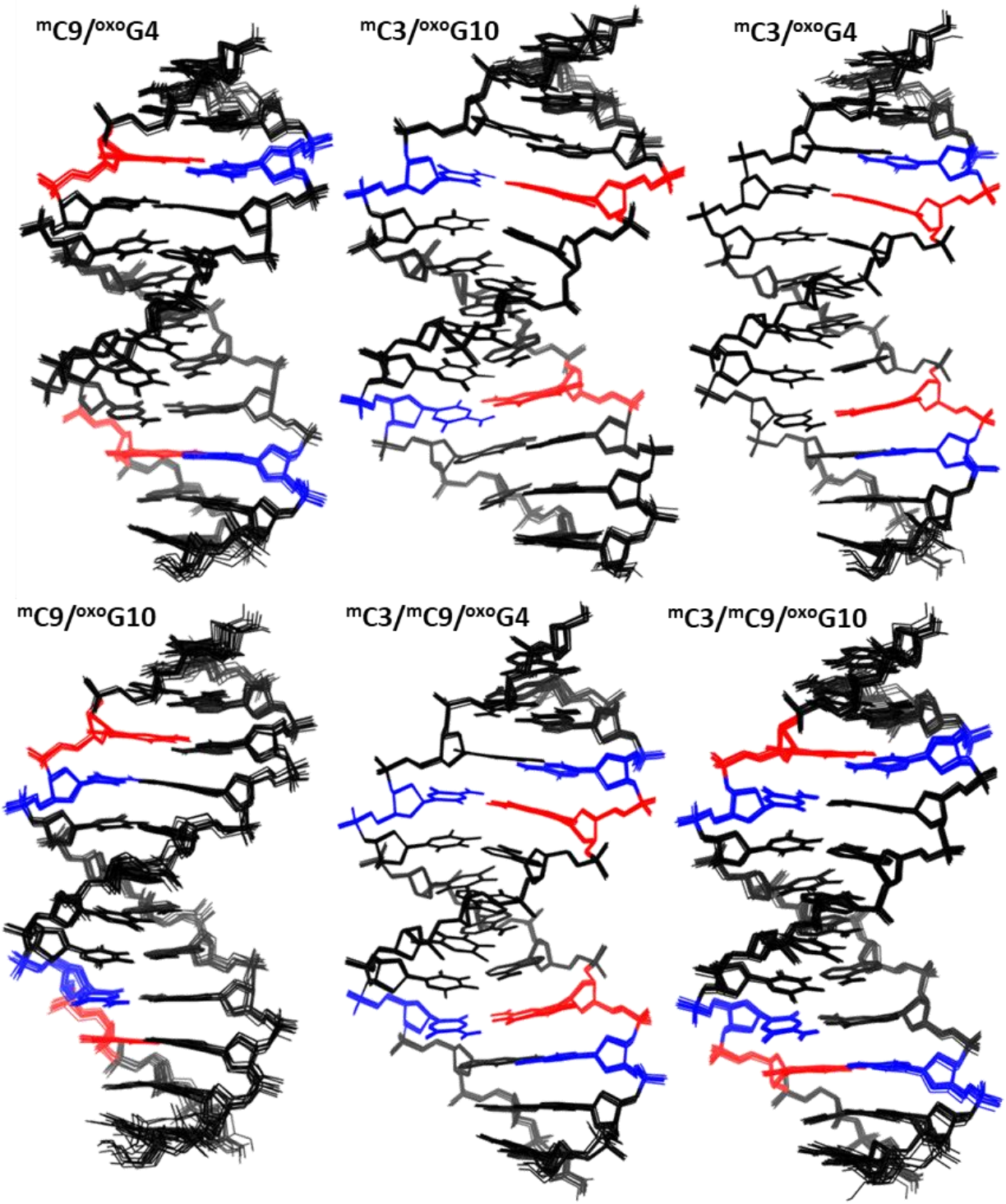



**Figure 3.4b. Solution NMR ensembles of $^{m}$C9/$^{oxo}$G4, $^{m}$C3/$^{oxo}$G10, $^{m}$C3/$^{oxo}$G4, $^{m}$C9/$^{oxo}$G10, $^{m}$C3/$^{m}$C9/$^{oxo}$G4 and $^{m}$C3/$^{m}$C9/$^{oxo}$G10. 5MC nucleotides are colored blue and 8OG nucleotides are colored red. The ensembles display high regularity between samples.**

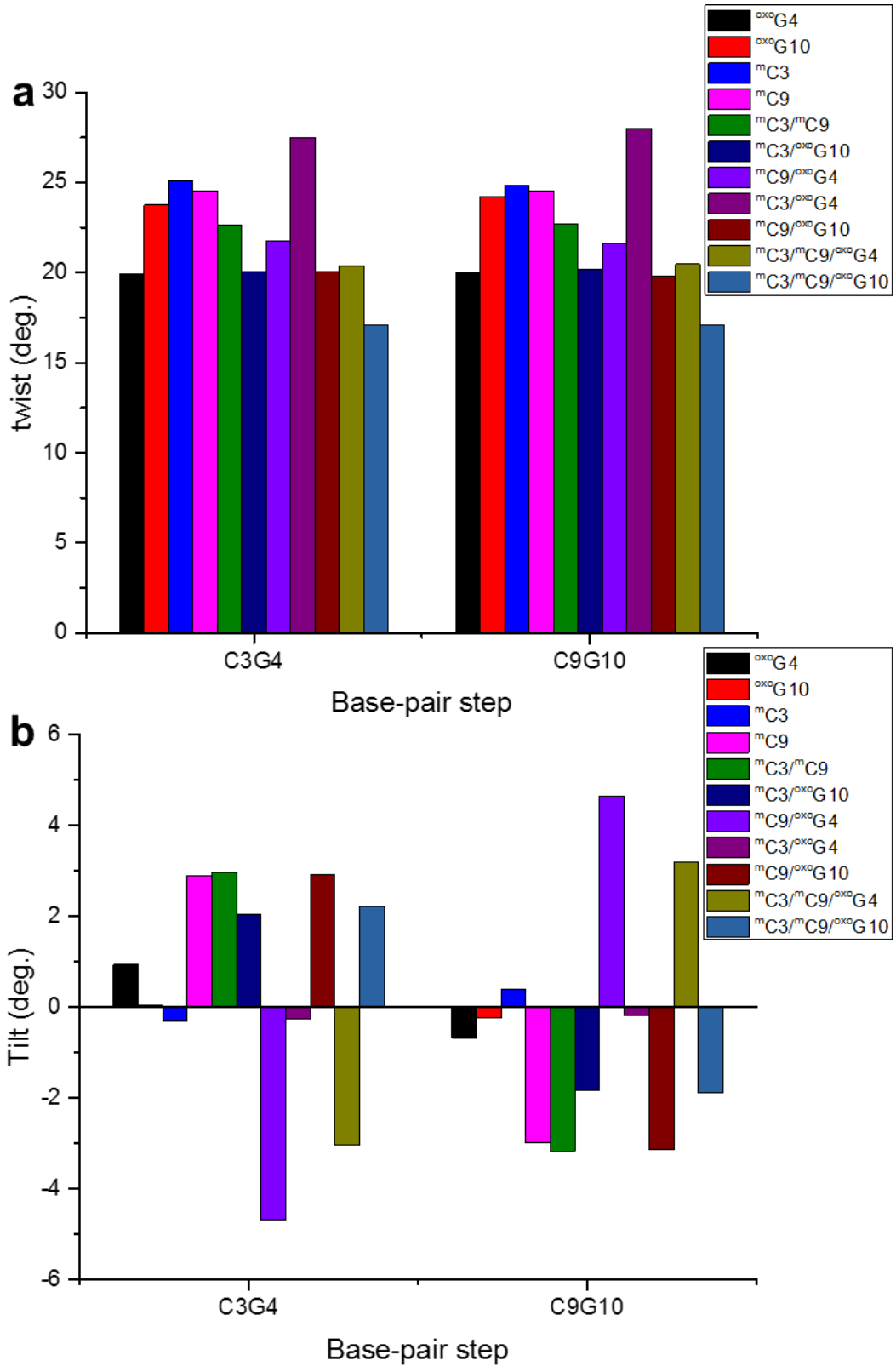



**Figure 3.5. Twist (a) and tilt (b) base-pair step parameter measured for all solution NMR derived structures. The base-pair steps C3pG4 and C9pG10 are shown.**

***Thermodynamic stability of the DNA samples by UV absorbance***. The thermal denaturation of all samples was monitored optically through UV spectrophotometry at 260 nm. The melting temperatures ($T_m$) for all samples are shown in Table 3.4. The $T_m$ values generally follow an expected trend: incorporation of 8OG drives $T_m$ down while addition of 5MC increases the thermal stability of the DNA duplexes. However, in some instances, clustering of 8OG and 5MC results in novel effects. For example, single-methylation of oxidized CpG (*trans*- for G4 and cis- for G10) lowers the overall thermodynamic stability versus the control samples containing lone 8OG.

| Sample | $T_m$ (°C) |
|---|---|
| DDD* | 66.6 ± 0.6 |
| $^{oxo}$G4 | **49.6 ± 1.1** |
| $^{oxo}$G10 | <u>51.3 ± 1.1</u> |
| $^{m}$C3 | 68.0 ± 0.4 |
| $^{m}$C9 | 68.1 ± 0.2 |
| $^{m}$C3/$^{m}$C9 | 66.3 ± 1.1 |
| $^{m}$C3/$^{oxo}$G10 | 51.6 ± 0.5 |
| $^{m}$C9/$^{oxo}$G4 | **46.6 ± 0.3** |
| $^{m}$C3/$^{oxo}$G4 | 50.4 ± 0.7 |
| $^{m}$C9/$^{oxo}$G10 | <u>48.3 ± 0.4</u> |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | 56.9 ± 3.0 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | 56.0 ± 2.5 |

**Table 3.4. Thermodynamic stability of the duplex DNA samples determined by UV absorbance. Addition of single methylation in *trans*- (bold) and *cis*- (underlined) configurations lowers $T_m$ versus respective controls containing lone 8OG modifications.**

***Imino proton NMR line widths and base dynamics***. Imino proton resonance assignment for each oligonucleotide was verified using imino-imino cross peaks from 90% $H_2O$/10% $D_2O$ $^1$H-$^1$H NOESY (Figure 3.1). A $^1$H-NMR spectrum highlighting the imino NH resonances and the 8OG NH7

resonance is shown in Figure 3.6. Five resonances were detected between 12.5 and 13.7 ppm for all samples. The two most downfield shifted peaks belong to thymines T7 and T8 while the remaining three belong to guanines G2, G4 and G10. Guanine G4 and G10 imino proton NMR linewidths (Hz) were analyzed as a function of temperature (5-60 °C) for all target and control samples. The imino proton resonance temperature series of $^{oxo}$G10 (all other sample series omitted) is presented in figure 3.7. The line-broadening effect is pronounced for guanine G2 due to its proximity with the duplex termini. The oxidized guanine displays a similar profile to G2, suggesting that oxidation increases the dynamics (frequency of opening) of the base. Imino lines for internal bases (T7 and T8) are stable to ~55-60 °C, a trend that is consistent for all samples.

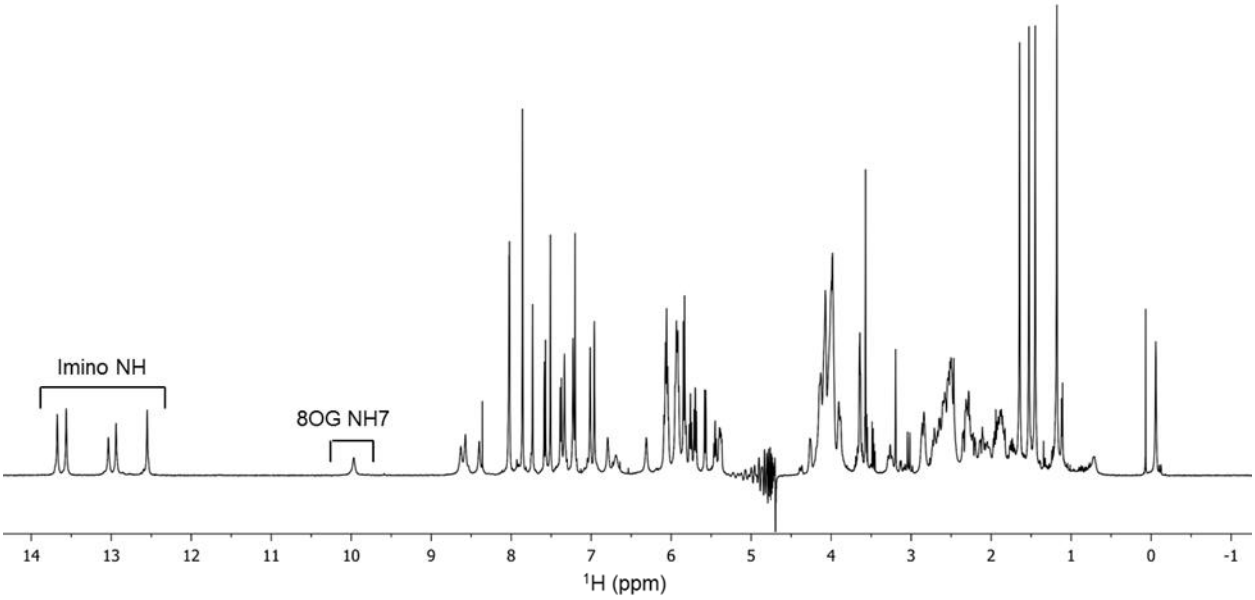



**Figure 3.6. 1D-$^{1}$H NMR spectrum at 298 K on sample $^{m}$C3/$^{m}$C9/$^{oxo}$G10 recorded at 500 MHz. NH7 of 8OG is seen shifted upfield from the group of imino NH resonances.**

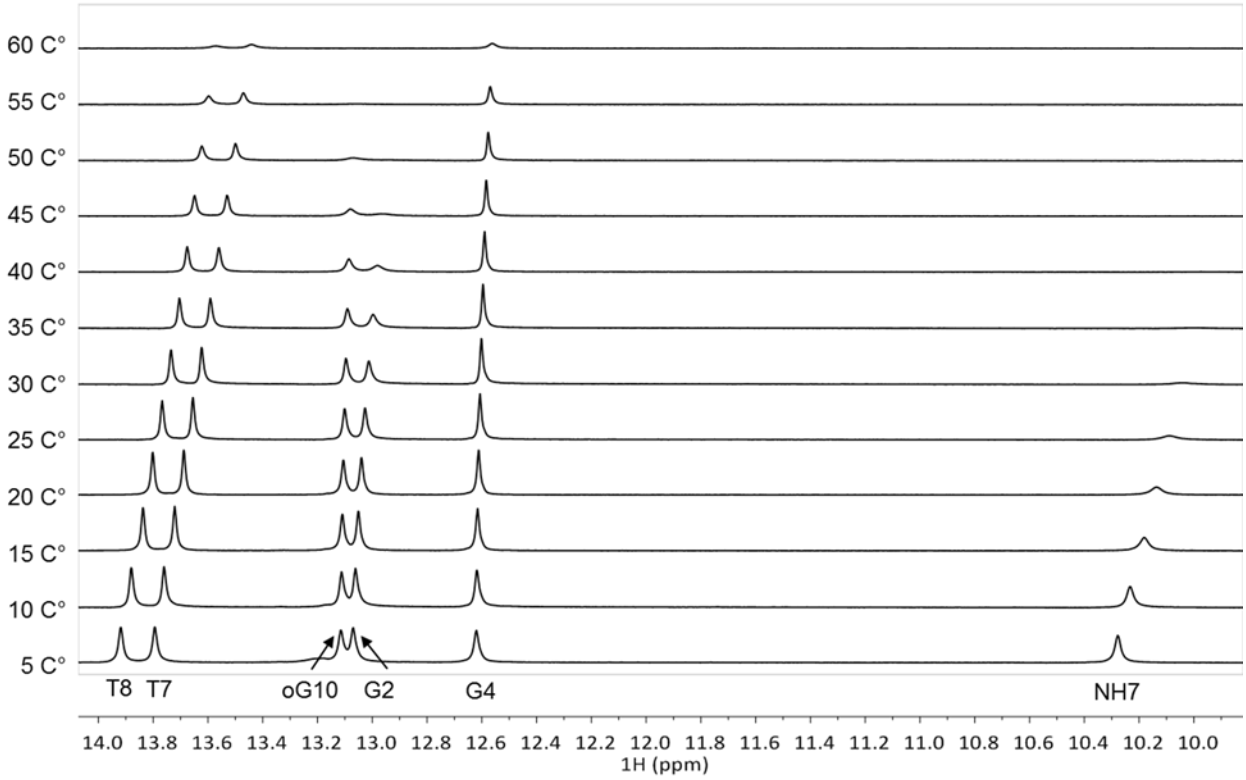



**Figure 3.7. $^{1}$H NMR imino proton resonances for $^{oxo}$G10 as a function of temperature. All imino resonance lines are labelled for their respective bases. The upfield resonance of NH7 belonging to 8OG appears at ~ 10.3 ppm.**

The linewidth (Hz) as a function of temperature for guanines G4 and G10 in all control and target samples is shown in Figures 3.8a/c and Figure 3.9a/c. Consistent with prior published work, guanine oxidation leads to the imino proton line broadening for the oxidized base (Singh et al., 2011). As expected, oxidation of guanine within the fully-methylated CpG sites lead to the imino proton line broadening for the oxidized guanine base for $^{m}$C3/$^{m}$C9/$^{oxo}$G4 and $^{m}$C3/$^{m}$C9/$^{oxo}$G10 (Figures 3.8d and 3.9d, respectively). Surprisingly, the imino proton line width for the non-oxidized G4 or G10 within $^{m}$C3/$^{m}$C9/$^{oxo}$G4 and $^{m}$C3/$^{m}$C9/$^{oxo}$G10 were narrower than those for the same base within their respective oxidized controls ($^{oxo}$G4 and $^{oxo}$G10) (Figures 3.9b and 3.8b).

Further, the non-oxidized guanine within $^{m}$C3/$^{m}$C9/$^{oxo}$G10 displayed narrower lines than the fully-methylated control $^{m}$C3/$^{m}$C9, a result not as distinguishable with $^{m}$C3/$^{m}$C9/$^{oxo}$G4.

For all the hemi-methylated samples ($^{m}$C3/$^{oxo}$G10, $^{m}$C9/$^{oxo}$G4, $^{m}$C3/$^{oxo}$G4 and $^{m}$C9/$^{oxo}$G10) oxidation of guanine causes the expected line-broadening for the oxidized G base (Figures 3.8c and 3.9c, respectively). In addition, we observed line-narrowing for the oxidized guanine caused by the addition methylation of a single cytosine within the CpG site (for both *cis*- and *trans*-methylation vs lone oxidation). The effects of oxidation on the non-oxidized (G4 or G10) are mixed: methylation of C3 causes line broadening for $^{m}$C3/$^{oxo}$G10 (Figure 3.8a) for the non-oxidized guanine bases whereas methylation of C9 causes line-narrowing for the non-oxidized guanine in $^{m}$C9/$^{oxo}$G4 (Figure 3.9a)

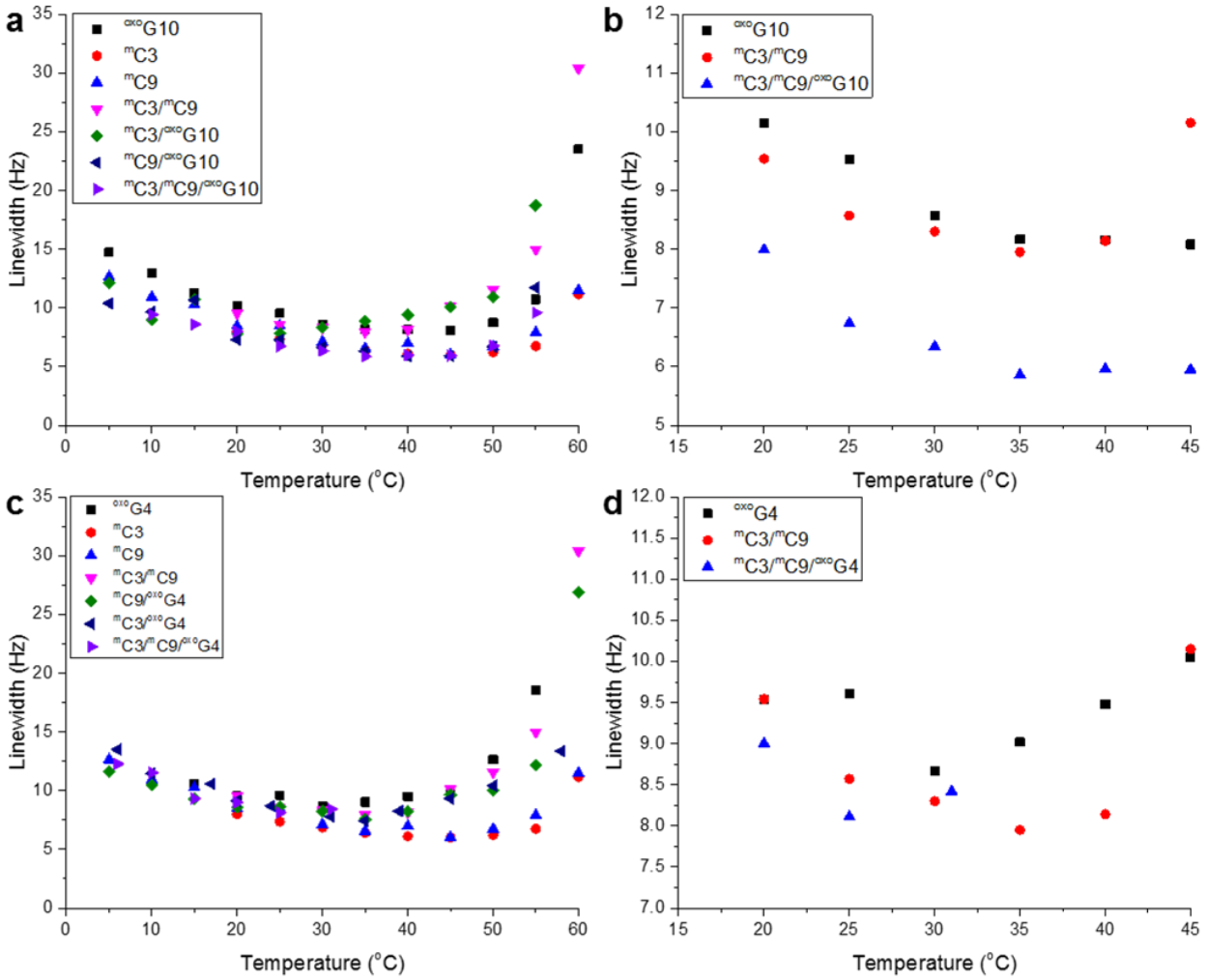



**Figure 3.8. Imino proton linewidths as a function of temperature for guanine G4. The G4 linewidths for samples with incorporated 8OG at position G10 are shown in (a) and (b). The G4 linewidths for samples with incorporated 8OG at position G4 are shown in (c) and (d).**

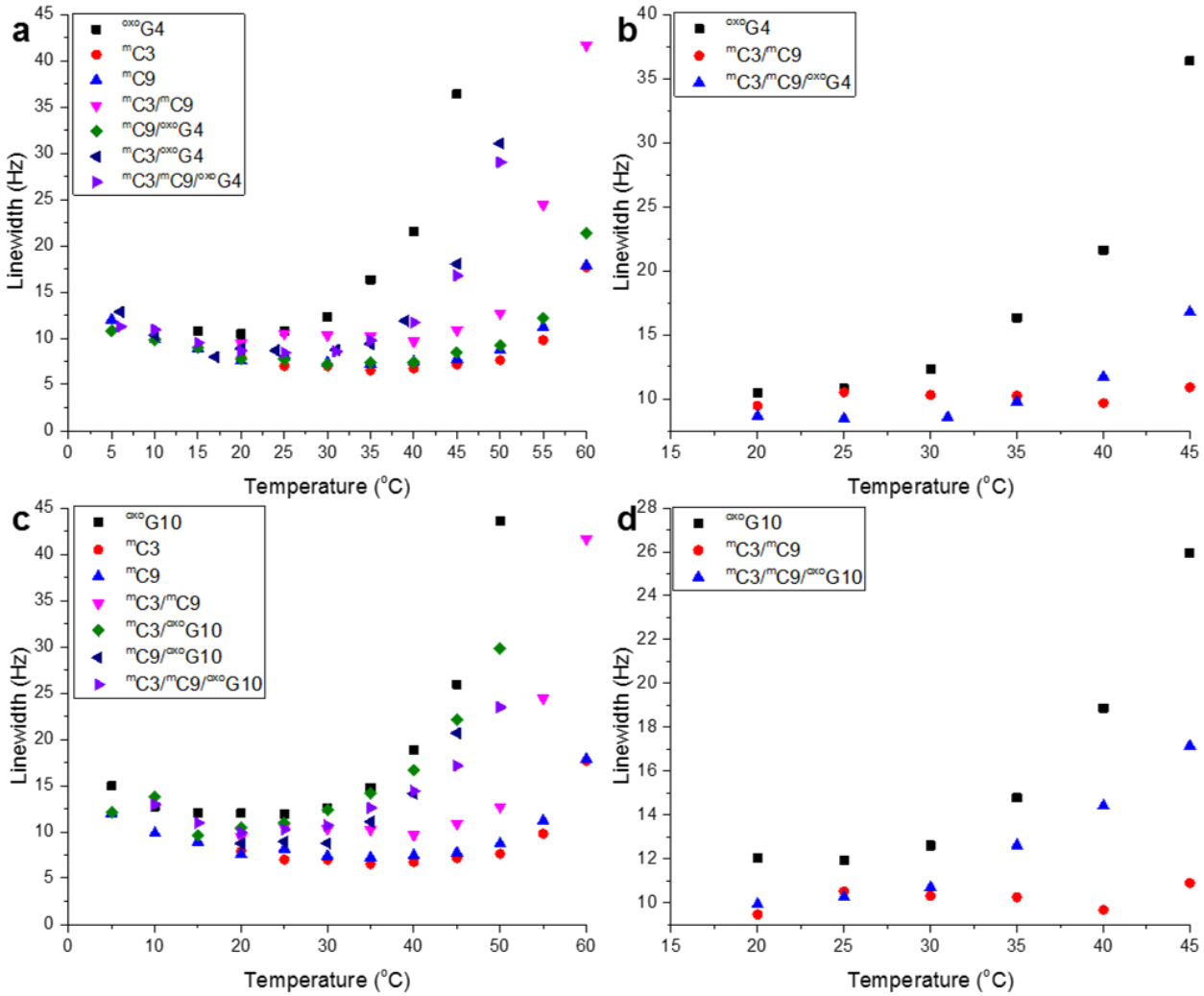



**Figure 3.9. Imino proton linewidths as a function of temperature for guanine G10. The G10 linewidths for samples with incorporated 8OG at position G4 are shown in (a) and (b). The G10 linewidths for samples with incorporated 8OG at position G10 are shown in (c) and (d).**

# *Chapter IV – Molecular dynamics simulations*

***Introduction to molecular dynamics simulations***. Molecular dynamics (MD) is a method of simulating the physical motions of an N-particle system by monitoring atom coordinates in the course of time. This is accomplished by numerically solving Newton's equations of motion. Specifically, Newton's second law is solved, which for a simple atomic system is described in Equations 1a and 1b (Karplus and Petsko, 1990).

$$F_i = m_i \times a_i \qquad (1a)$$

$$F_i = -\frac{\partial V}{\partial r_i} \qquad (1b)$$

Here, all forces, $F_i$, acting upon the atoms are derived from the potential energy function of the system $V(r_i)$, where $r_i$ represents the complete set of Cartesian coordinates for each atom.

Traditionally, the time steps for MD simulations are on the order of 1 femtosecond (fs). Combining each of the time steps generates an overall time-sequenced trajectory of dynamic events which can be used to observe important biological events at the atomic level. At each time step, the starting coordinates and velocities as well as the instantaneous forces acting upon each atom are provided. Integration of Newton's equation of motion is completed to determine the position, velocity and acceleration of each particle over the course of time. For $r_i$ atoms in a system, this calculation quickly becomes complex. In large systems, this commonly gives rise to a many-body problem. Therefore, there is no analytical solution to the function, meaning it must

be solved numerically. The two most commonly used numerical integrators are the Verlet and Leap-frog algorithm, both using finite difference methods to solve the differential equation. The leap-frog algorithm is implemented in AMBER while many other popular MD codes such as NAMD use Verlet.

$$x_{n+1} = x_n + hv_{n+1/2} \qquad (2a)$$

$$V_{n+3/2} = v_{n+1/2} + hF(x_{n+1}) \qquad (2b)$$

Numerical integration using the leapfrog algorithm consists of two functions calculating velocity and position separately. Firstly, position (*x)* is sampled in a sequence of discrete time points at where *h* is the interval between time steps (Equations 2a, 2b). Secondly, velocity (*v*) is calculated at staggered time points relative to position. This scheme has advantages over other numerical integration methods as it is less error prone, time reversible and less demanding of system memory (does not store previous time step information).

***AMBER Force Field*.** The assisted model building with energy refinement (AMBER) force field in its purest definition is a combination of several parameters defining the total potential energy of the system (Cornell et al., 1995). The derivative of the sum of multiple functions relative to position is taken, giving a total potential energy function (Equation 3).

$$E_{\text{total}} = \sum_{\text{bonds}} K_r(r - req)^2 + \sum_{\text{angles}} K_\theta\,(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + cos(n\varphi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{{R_{ij}}^{12}} - \frac{B_{ij}}{{R_{ij}}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] \quad (3)$$

Each of these terms is defined as follows. The first term is the sum of the energy associated with all covalently bonded atoms; with all covalent bonds being described as ideal harmonic springs. Here, $K_r$ is the bond force constant which alters the stiffness of the bond, $r$ is the bond length and $r_{eq}$ is the equilibrium bond length. The second term describes the summation energy associated with all angles of covalently bonded atoms, effectively modeling the interaction of electronic orbitals. All angles are treated harmonically. In this term, $K_\vartheta$ is the valence angle force constant, $\vartheta$ is the valence angle and $\vartheta_{eq}$ is the equilibrium angle. The third function is a calculation of the sinusoidal torsion a bond experiences, whether that be from bond order or the electrostatics of neighboring groups such as lone pairs. Similarly, to the previous terms, $V_n$ is the force constant for the dihedral angle, $n$ is the multiplicity, $\varphi$ is the dihedral angle value and $\gamma$ is the phase angle. The final function represents Van der Waals and electrostatic interactions. $R_{ij}$ represent the distance between two atoms; $A_{ij}$ *and* $B_{ij}$ are constants describing the Van der Waals interactions; ε is the dielectric constant and $q_i$ and $q_j$ are the point charges associated with atoms $i$ and $j$, respectively. The force constants used in each potential energy function were derived from a compilation of structural, vibrational frequency and quantum mechanical calculations on small fragments of proteins and nucleic acids. Further, all bonds and atoms are assumed to be represented as springs and individual point charges, respectively. If any experimental data is available (e.g. NOESY distances) they can be added as components to the basic force field.

MD simulations are often used to investigate the structure, thermodynamics and kinetics of a biological system. MD simulations are employed extensively to complement structural biology, targeting of efficacious new drugs to combat diseases and obtaining specific particle motions as a function of time. Therefore, MD provides a convenient avenue for comparing time-averaged results from traditional structural biology experiments as well as providing insights for details of biomolecular function. The initial 3D coordinates for MD simulations are most commonly derived from traditional structural biology techniques such as NMR or X-Ray crystallography. However, if there is no high-resolution structure available, software such as nucleic acid builder (NAB) and Phyre2 for proteins can provide a suitable predicted starting structure.

Analyzing the raw trajectories to extract information that is relevant to the investigation is always the most time-consuming aspect of any MD experiment. As such, many statistical methods exist to help guide analysis towards dynamics events more important than random diffusion within the system. Once such approach is principal component analysis (PCA), a statistical method used to determine patterns in a series of potentially coordinated observables. In order to extract these motions, the dimensionality of the simulation data is reduced through a linear transformation of a coordinate covariance matrix (3 N x 3 N where N is number of atoms analyzed). The PCs are obtained by diagonalizing, or decomposing, the coordinate covariance matrix into a set of orthogonal modes (eigenvectors). The PCs correspond to concerted atomic motions and major conformational changes of the molecule of interest (Amadei et al., 1993; Yamaguchi et al., 1998; Amadei et al., 1996; Haider et al., 2008) and reveal the most dominant motions in the system of interest. Each resulting eigenvector indicates the direction of motion of the atom and the

corresponding eigenvalue represents the energetic contribution of the component to the motion. The eigenvector with the largest eigenvalue would therefore account for the highest proportion of variance.

## *Results*

***Molecular dynamics simulation stability and convergence***. The backbone root mean square deviations (RMSD) of each 600 ns simulation were monitored as a function of simulation time to assess stability and divergence from the initial starting structure. As shown in Figure 4.1a and 4.1b, the backbone RMSD from the initial starting structure for all twelve simulations was stable, fluctuating between 1.0 and 2.5 Å when monitoring the 8 internal residues of the duplex. The upper limit of this RMSD measurement extends to 3.5 Å when all 12 residues are included in the calculation. Spikes in the running average of the full-length duplex attributed to loss of base pairing at one or both termini. This effect is caused by direct interaction with the water solvent at the termini. Although loss of base pairing occurs fairly regularly, the bases renature on the ns timescale. The average backbone RMSD for each sample over 600ns is shown in Table 4.1. Removing the termini from the calculation reduced the average backbone RMSD 0.19-0.40 (Å) with respect to the full-length duplex RMSD.

| Sample | Ave. backbone RMSD (Å) | Ave. backbone RMSD (Å) excluding termini |
|---|---|---|
| unmodified | 1.91 ± 0.31 | 1.57 ± 0.31 |
| $^{oxo}$G4 | 1.72 ± 0.32 | 1.46 ± 0.32 |
| $^{oxo}$G10 | 1.94 ± 0.37 | 1.73 ± 0.37 |
| $^{m}$C3 | 1.75 ± 0.31 | 1.59 ± 0.31 |
| $^{m}$C9 | 2.00 ± 0.37 | 1.78 ± 0.37 |
| $^{m}$C3/$^{m}$C9 | 1.93 ± 0.35 | 1.74 ± 0.35 |
| $^{m}$C3/$^{oxo}$G10 | 1.89 ± 0.37 | 1.70 ± 0.37 |
| $^{m}$C9/$^{oxo}$G4 | 2.02 ± 0.35 | 1.81 ± 0.35 |
| $^{m}$C3/$^{oxo}$G4 | 1.72 ± 0.33 | 1.50 ± 0.33 |
| $^{m}$C9/$^{oxo}$G10 | 2.04 ± 0.34 | 1.64 ± 0.34 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | 2.04 ± 0.36 | 1.76 ± 0.36 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | 2.02 ± 0.35 | 1.76 ± 0.35 |

**Table 4.1. Average backbone RMSD for the full-length duplex and excluding termini over the course of 600ns simulations time for each sample.**

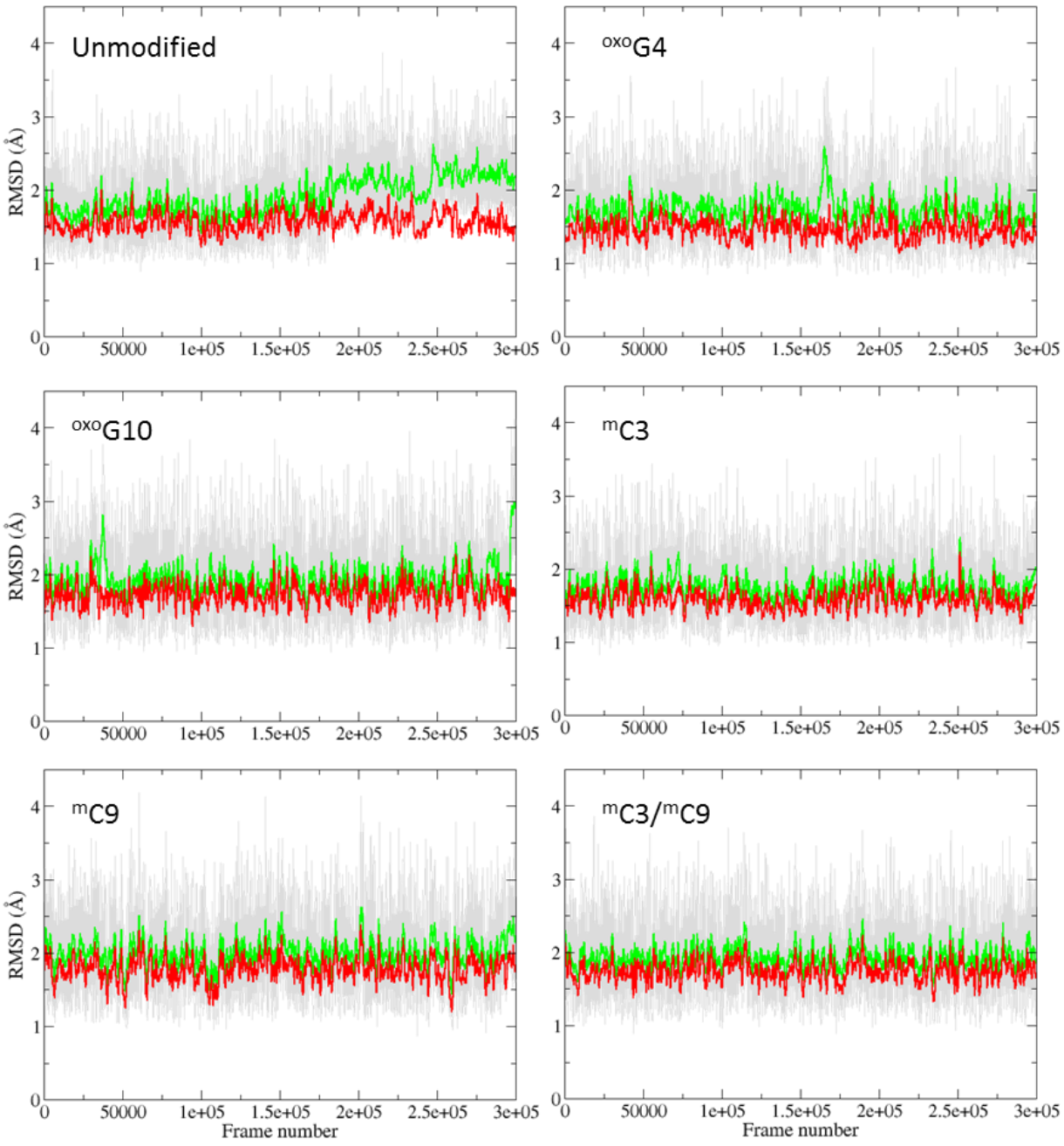



**Figure 4.1a. Backbone RMSD (Å) as a function of frame number (each frame represents 2 ps of simulation time) for all control samples (grey). The plots show two running averages calculated in 2 ns intervals for all backbone atoms (green) and excluding terminal residues (red).**

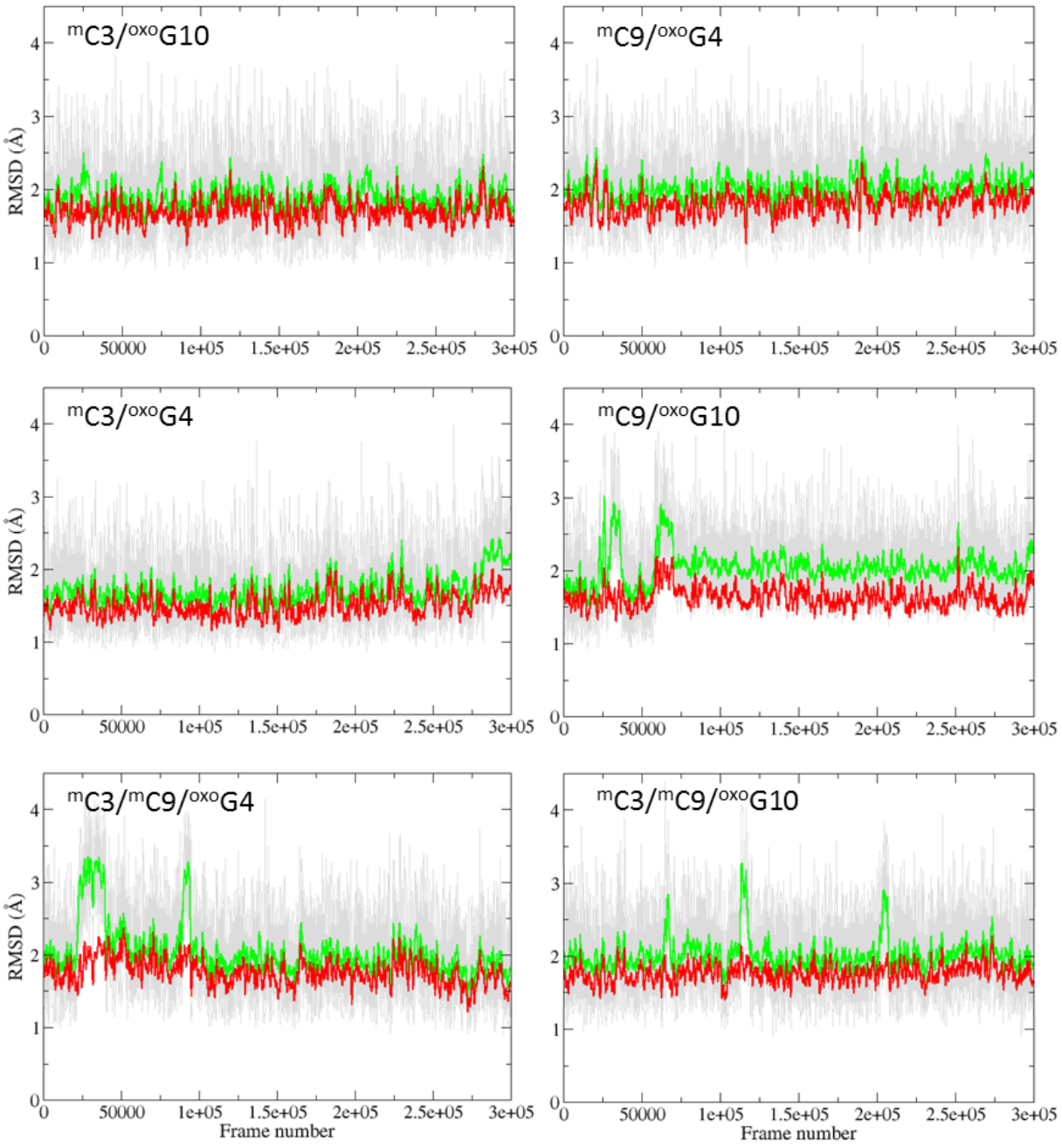



**Figure 4.1b. Backbone RMSD (Å) as a function of frame number (each frame represents 2 ps of simulation time) for all target samples (grey). The plots show two running averages calculated in 2 ns intervals for all backbone atoms (green) and excluding terminal residues (red).**

The average atomic root mean square fluctuations (RMSF) over 555 ns simulation time for each nucleotide are shown in Figure 4.2. The reference structure provided the initial coordinates. Base

pairs at each terminus exhibit the greatest range in fluctuations from 1.75-3.25 Å. As previously stated, the source of the variability is due to base pair opening events. The internal base pairs display greater stability, with nucleotide fluctuations in the range of 0.2-0.4 Å. There were no discernible differences in average fluctuations within CpG sites or neighboring nucleotides of the control and target duplexes.

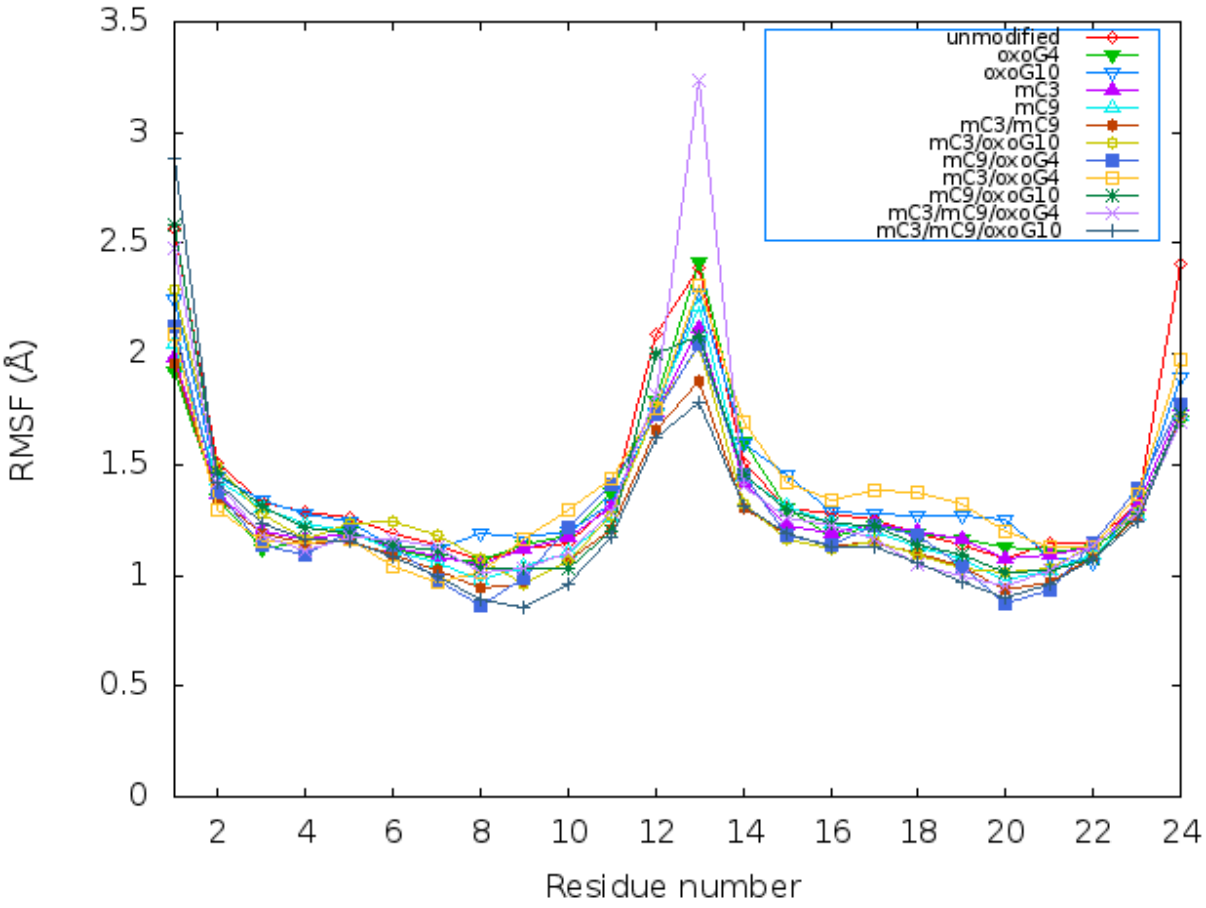



**Figure 4.2. RMSF (Å) as a function of nucleotide for each of the twelve 555ns MD simulations. Each sample is labelled by individual colors and symbols.**

Principal component analysis (PCA) was used to assess the convergence of dynamic properties between each of the twelve simulations. Shown in Figure 4.3 is the projection of the predominant (greatest eigenvalue) principal component (PC). All heavy atoms were used in the calculation,

including the termini. Although PCA cannot quantify the convergence of the independent simulations with respect to each other, it allowed us to determine whether the simulations sampled the same overall motions in 3D Cartesian space. From figure 4.3, we notice that the first PC projections for all twelve systems overlap, indicating the motions for PC 1 are highly similar. The PC 2 and PC 3 were also examined (not shown) and yielded overlapping projections. Visual inspection of principal component trajectories (pseudo-trajectories) generated from the first PC indicate that the most dominant motions are global bending and twisting of the duplex.

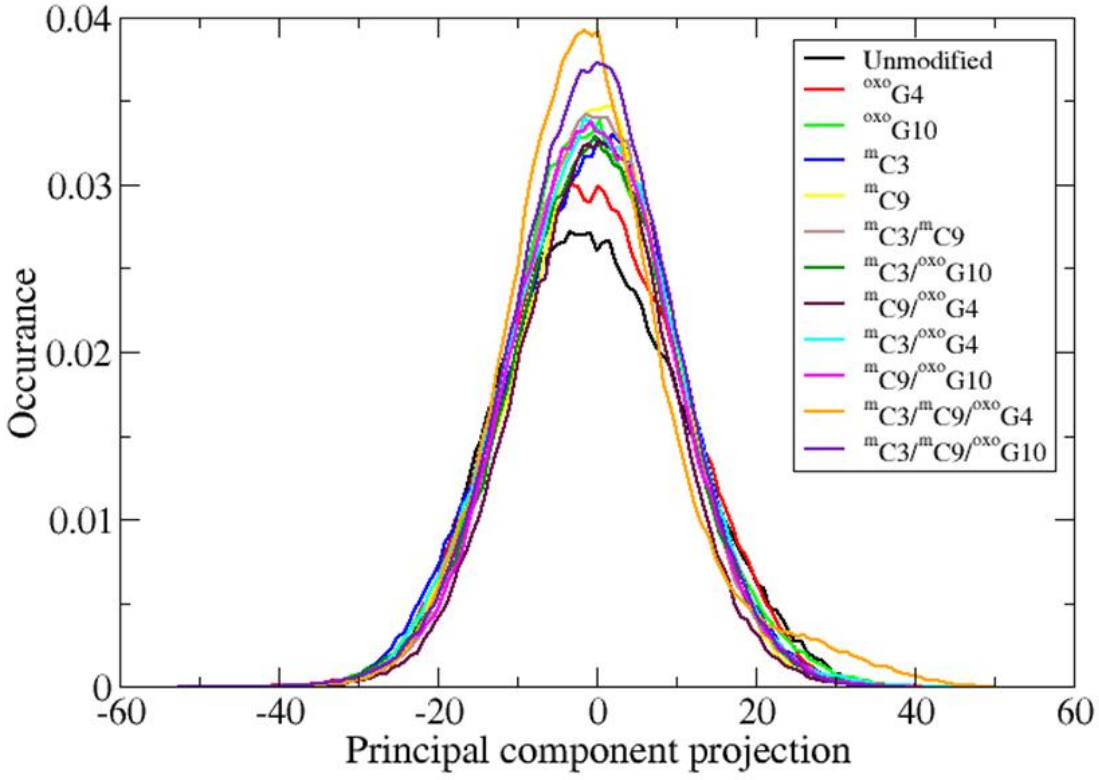



**Figure 4.3. Principal component projection of PC 1 for all samples. All heavy atoms were used in the calculation. Overlap of PC 1 indicates that dynamics properties samples within all twelves simulations were highly similar.**

***Comparison DNA inter- and intra-base helical structural geometries.*** Average values for all inter- and intra-base pair helicoidal parameters for each system were calculated from each 555ns

simulation and are shown in Appendix Figures A1-A4. The terminal base pairs were not included in the analysis as loss of base pairing events led to increased variability between all samples. Average values for the unmodified duplex yielded similar results to those recently reported on μs-timescale dynamics for the same sequence (Dans et al., 2016). The averaged base-pair geometries displayed were highly similar for all base-pairs, including those within the target CpG sites. However, when monitoring base-pair step geometries, dissimilarities between systems were detected. These conformational perturbations were located explicitly on base-pair steps incorporated within the target CpG site or immediately flanking it. Appendix Tables A1-A4 display average and standard deviations for the base-pair steps C3G4, G4A5, C9G10 and G10C11. Specifically, we observed a local unwinding effect caused by the incorporation of 8OG for base-pair step parameters twist and tilt within one standard deviation of the values observed in our solution NMR structures.

***Dynamics of the DNA backbone***. The MD simulations revealed differences in the bimodal distributions of the backbone phosphate conformation (BI or BII). The phosphate step conformation for each system is displayed in Figure 4.4. Phosphate conformations outside of either target CpG site were highly similar. The introduction of 5MC to the 3$^{rd}$ or 9$^{th}$ nucleotide stabilizes the trailing 3'-phosphate (C3pG4 or C9pG10 respectively) in the BI conformation increasing the observed fraction BI to > 0.90. Incorporating 8OG to the 4$^{th}$ nucleotide on the duplex destabilizes BI conformation for the trailing 3'-phosphate (G4pA5), driving the observed BI fraction from ~0.50 without 8OG to 0.10-0.29. Similarly, when 8OG is introduced to the 10$^{th}$ position, the BI conformation of the 3'-phosphate (G10pC11) is disrupted, reducing the BI fraction from ~0.75 to 0.1-0.4. For instances of high BII conformation the following phosphate is

always followed by a high BI population. Furthermore, one notices that the most profound increases in observed BII fraction originate from systems containing clustered 8OG and 5MC. Specifically, duplexes that contain 8OG in a fully-methylated CpG have the greatest increase in BII phosphate conformation followed closely by *trans*-methylated CpGs and then *cis*-methylated CpGs. Additionally, the effect of clustering on phosphate conformation is apparent when monitoring both the transition frequencies (Table 4.2) and average residence times (Table 4.3).

The transitions between BI and BII conformation on target CpG phosphates without 5MC or 8OG modification were high frequency, contributing to similar average residence times for BI and BII. Upon addition of 8OG at the $4^{th}$ or $10^{th}$ position, the frequency of transitions between BI and BII sub-states of G4pA5 and G10pC11 decreases from 10584 to 3686 and 6772 to 2424, respectively. Furthermore, the average time spent in the BII conformation increased from 51 ± 101 ps to 240 ± 417 ps and 43 ± 102 ps to 316 ± 644 ps. When 8OG is contained within a fully-methylated CpG site, the time spent in the BII conformation for G4pA5 was 323 ± 581 ps and G10pC11 was 407 ± 769 ps. The average time spent in BII increases in a step-wise manner from *cis*-, *trans*- and finally fully-methylated CpG for both phosphates that are 3' to 8OG. Methylation at C3 or C9 yielded dramatically increased stiffness to the downstream phosphate conformation. This effect is in agreement with previously published simulation data on samples containing a single methylation (Temiz et al., 2012). BI conformation residence times for C3pG4 increased from 42 ± 152 ps to 377 ± 891 ps upon methylation at C3. A similar effect is observed with methylation of C9 as BI conformation for C9pG10 was increased from 31 ± 125 ps to 260 ± 746 ps. These effects are compounded when the number of modifications to the CpG site is increased. $^{m}$C3/$^{m}$C9/$^{oxo}$G4 and

$^{m}$C3/$^{m}$C9/$^{oxo}$G10 systems observed the greatest impact on residence times and transitions frequencies versus the unmodified duplex.

| Sample | C3pG4 | G4pA5 | C9pG10 | G10pC11 |
|---|---|---|---|---|
| unmodified | 16374 | 10584 | 17016 | 6772 |
| $^{oxo}$G4 | 2558 | 3686 | 21138 | 5388 |
| $^{oxo}$G10 | 17246 | 9004 | 3892 | 2424 |
| $^{m}$C3 | 2734 | 10902 | 19280 | 6192 |
| $^{m}$C9 | 14764 | 11420 | 3696 | 9744 |
| $^{m}$C3/$^{m}$C9 | 2802 | 10772 | 3360 | 9534 |
| $^{m}$C3/$^{oxo}$G10 | 2396 | 9914 | 11472 | 2604 |
| $^{m}$C9/$^{oxo}$G4 | 700 | 3376 | 3694 | 8490 |
| $^{m}$C3/$^{oxo}$G4 | 1066 | 3286 | 22398 | 6534 |
| $^{m}$C9/$^{oxo}$G10 | 17196 | 9940 | 776 | 4566 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | 460 | 3074 | 4358 | 9814 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | 3346 | 10324 | 230 | 2354 |

**Table 4.2. Transition frequency between the BI and BII phosphate conformation over the course of 555ns for each system.**

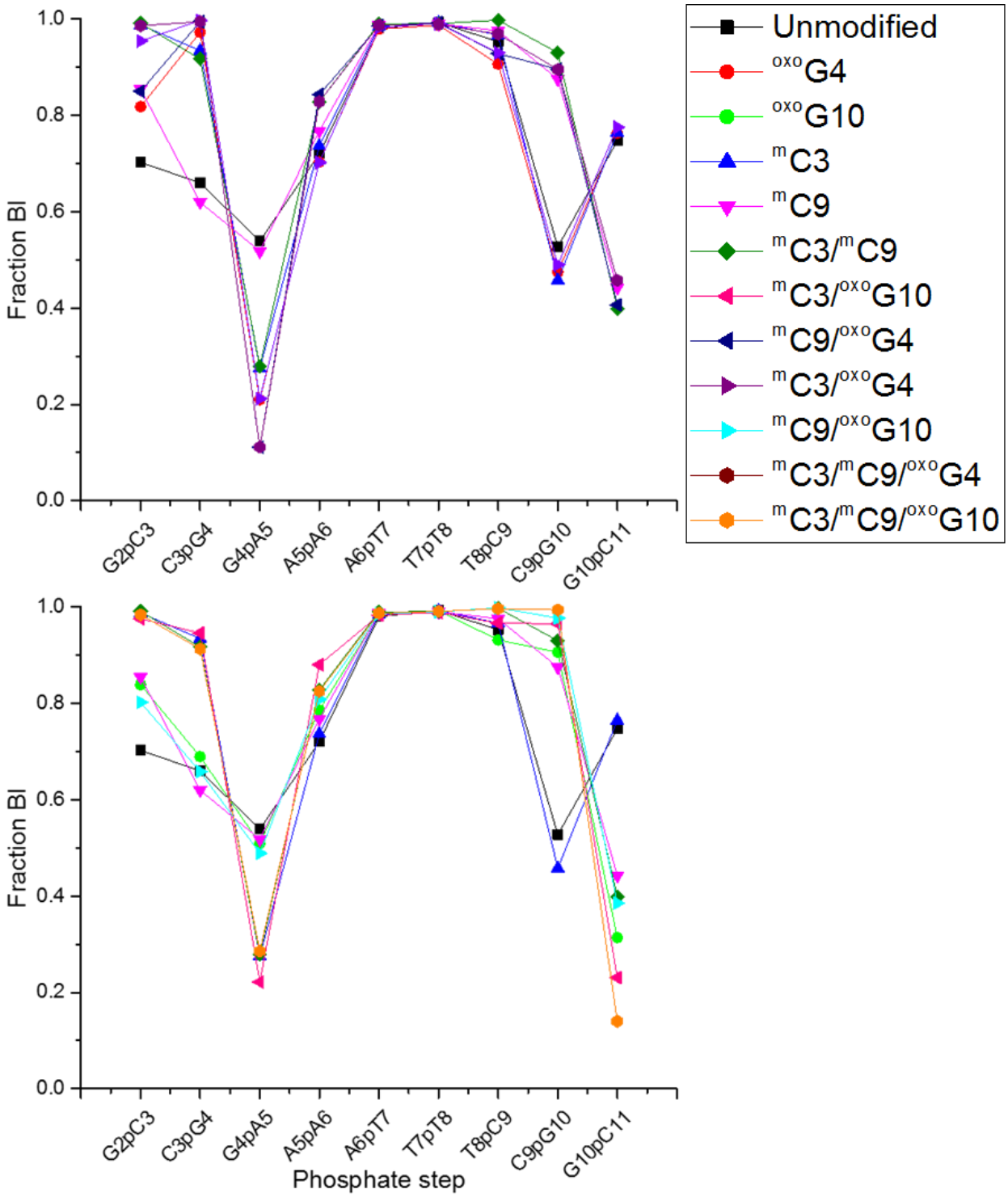



**Figure 4.4. Fraction of each phosphate (excluding termini) occupying the BI conformation. Compares a selection of controls with target samples that include an oxidized guanine on G4 (top) and G10 (bottom)**

| Sample | Conformation | C3pG4 | G4pA5 | C9pG10 | G10pC11 |
|---|---|---|---|---|---|
| unmodified | BI | 42 ± 152 | 53 ± 192 | 31 ± 125 | 120 ± 307 |
| | BII | 25 ± 48 | 51 ± 101 | 34 ± 67 | 43 ± 102 |
| $^{oxo}$G4 | BI | 420 ± 1564 | 60 ± 243 | 21 ± 97 | 154 ± 381 |
| | BII | 13 ± 34 | 240 ± 417 | 31 ± 65 | 51 ± 119 |
| $^{oxo}$G10 | BI | 41 ± 161 | 59 ± 207 | 256 ± 1333 | 141 ± 476 |
| | BII | 22 ± 42 | 63 ± 129 | 29 ± 77 | 316 ± 644 |
| $^{m}$C3 | BI | 377 ± 891 | 25 ± 87 | 22 ± 108 | 134 ± 329 |
| | BII | 28 ± 68 | 76 ± 137 | 34 ± 79 | 44 ± 109 |
| $^{m}$C9 | BI | 43 ± 154 | 47 ± 156 | 260 ± 746 | 47 ± 159 |
| | BII | 31 ± 59 | 49 ± 108 | 39 ± 117 | 65 ± 134 |
| $^{m}$C3/$^{m}$C9 | BI | 361 ± 804 | 25 ± 90 | 305 ± 943 | 43 ± 119 |
| | BII | 34 ± 74 | 77 ± 142 | 25 ± 59 | 72 ± 147 |
| $^{m}$C3/$^{oxo}$G10 | BI | 436 ± 934 | 21 ± 74 | 90 ± 782 | 96 ± 288 |
| | BII | 26 ± 64 | 90 ± 167 | 5 ± 20 | 329 ± 675 |
| $^{m}$C9/$^{oxo}$G4 | BI | 1576 ± 2924 | 33 ± 116 | 267 ± 720 | 50 ± 157 |
| | BII | 9 ± 24 | 295 ± 496 | 33 ± 135 | 80 ± 154 |
| $^{m}$C3/$^{oxo}$G4 | BI | 1036 ± 4384 | 68 ± 280 | 20 ± 92 | 129 ± 355 |
| | BII | 4 ± 14 | 268 ± 515 | 28 ± 52 | 40 ± 100 |
| $^{m}$C9/$^{oxo}$G10 | BI | 39 ± 138 | 51 ± 179 | 1396 ± 7020 | 91 ± 283 |
| | BII | 25 ± 50 | 59 ± 127 | 33 ± 131 | 151 ± 331 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | BI | 2399 ± 6471 | 37 ± 127 | 226 ± 638 | 49 ± 142 |
| | BII | 13 ± 40 | 323 ± 581 | 28 ± 73 | 63 ± 140 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | BI | 301 ± 718 | 27 ± 93 | 4799 ± 11359 | 63 ± 195 |
| | BII | 30 ± 83 | 79 ± 154 | 27 ± 96 | 407 ± 769 |

**Table 4.3. Average residence time (ps) spent in the BI or BII conformation over 555ns simulation time for target CpG site phosphates.**

***Effects of 8OG and 5MC on sugar puckering***. Minor differences on sugar puckering were observed with the incorporation of 8OG and 5MC on target CpG sugars. The cytosine 3 sugar pucker observed a minor increase in the frequency of C1'-exo configuration in systems that have been methylated on C3 (Figure 4.5a). Introducing either 5MC to C3 and/or 8OG to G4 narrowed the sugar pucker distribution of guanine 4 in the C2' endo configuration while samples without these modifications on C3 or G4 observed greater sugar flexibility, with a minor increase in population towards C3'-exo configuration (Figure 4.5a). The sugar puckering on cytosine 9 was shifted towards C1'-exo configuration in samples with a methylated C9 and non-oxidized G10. Finally, the sugar pucker conformation for G10 was dependent on whether G10 was oxidized, C9 was methylated or both. Incorporation of either of these modifications increased the stiffness of the sugar, leading to a narrow distribution around C2-endo configuration. When neither of these modifications was present, the G10 sugar pucker displayed a split population between C2'-endo and C3'-exo configurations (Figure 4.5b).

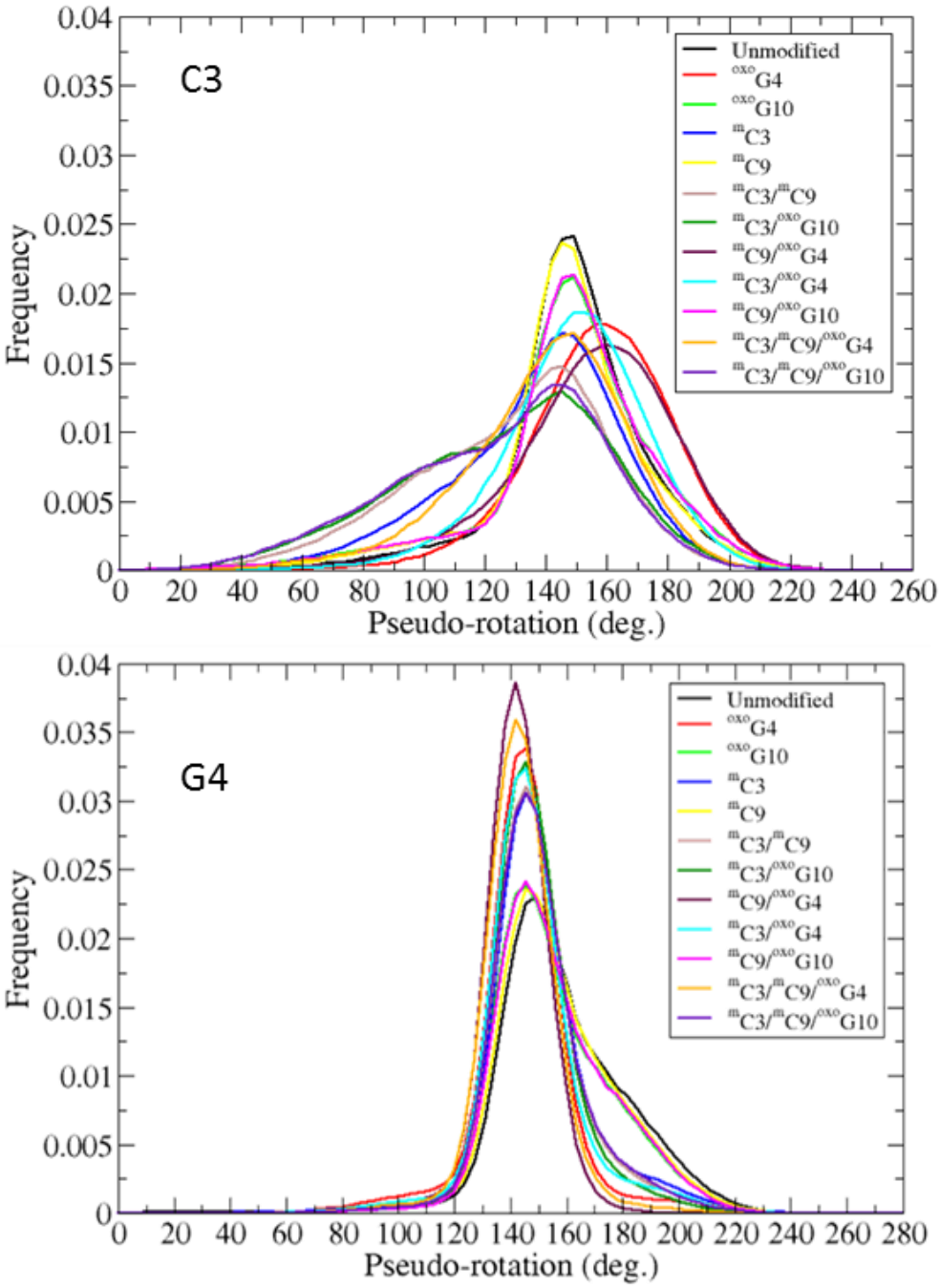



**Figure 4.5a. Sugar pucker distributions for cytosine 3 (top) and guanine 4 (bottom) for all twelve simulations.**

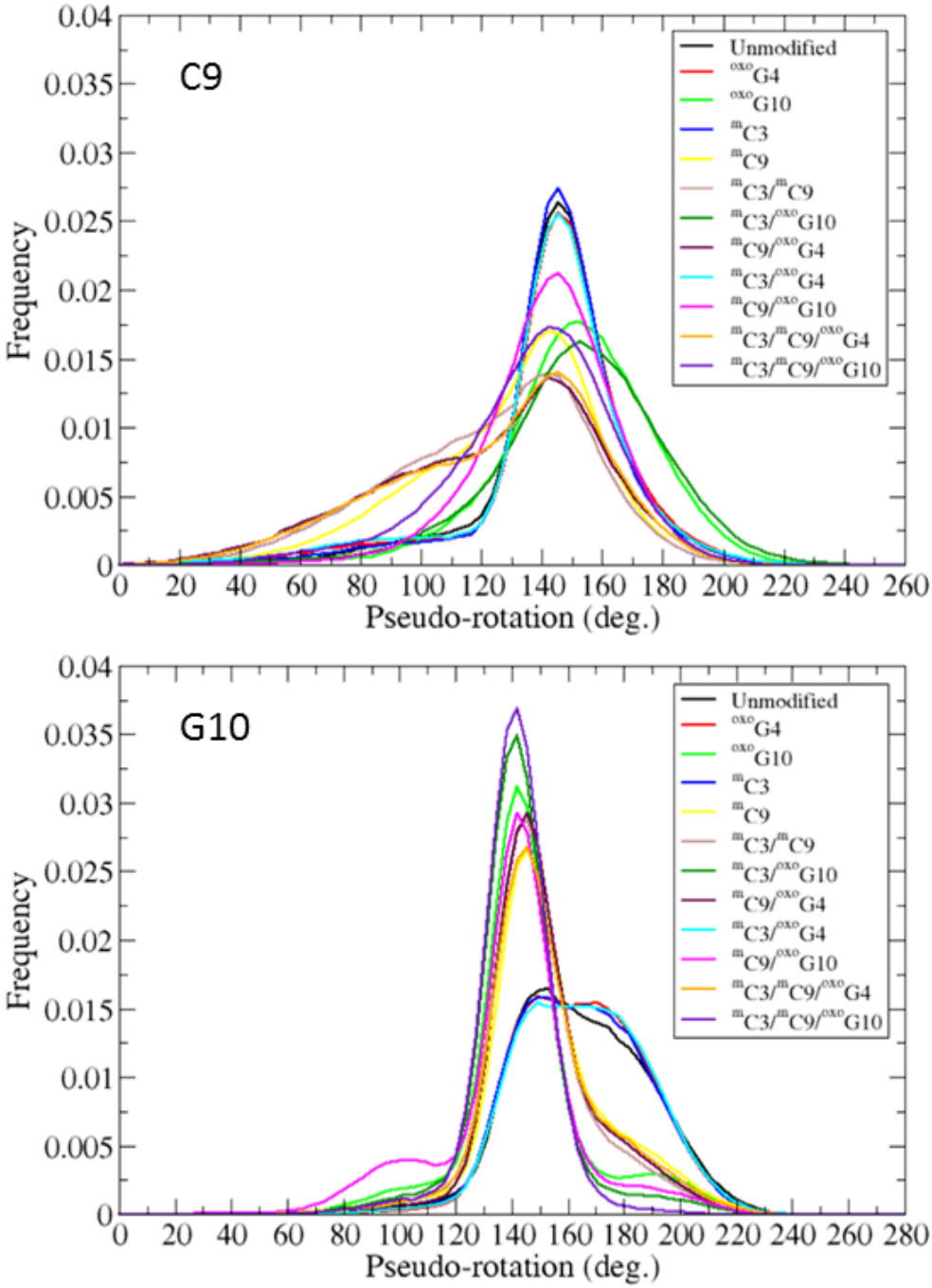



**Figure 4.5b. Sugar pucker distributions for cytosine 9 (top) and guanine 4 (bottom) for all twelve simulations.**

***Effects of 8OG and 5MC clustering on the internal dynamics of CpG sites***. Principal component analysis was applied to all heavy atoms of the CpG site C3pG4-C21pG22. The resulting low-frequency eigenvectors were sorted by magnitude of their associated eigenvalues to determine the most dominant modes of motion of the system (Teeter and Case, 1990). The proportion of variance that each principal component (PC) contributes to the overall dynamic movement in Cartesian space is presented as scree plots ranked by the magnitude of eigenvalues in Figure 4.6. The first PC dominates the overall variance for the CpG sites in all samples analyzed, contributing 26-38 % of the overall variance. Together, the first five modes (PC1 – PC5) contribute >69 % of the overall variance.

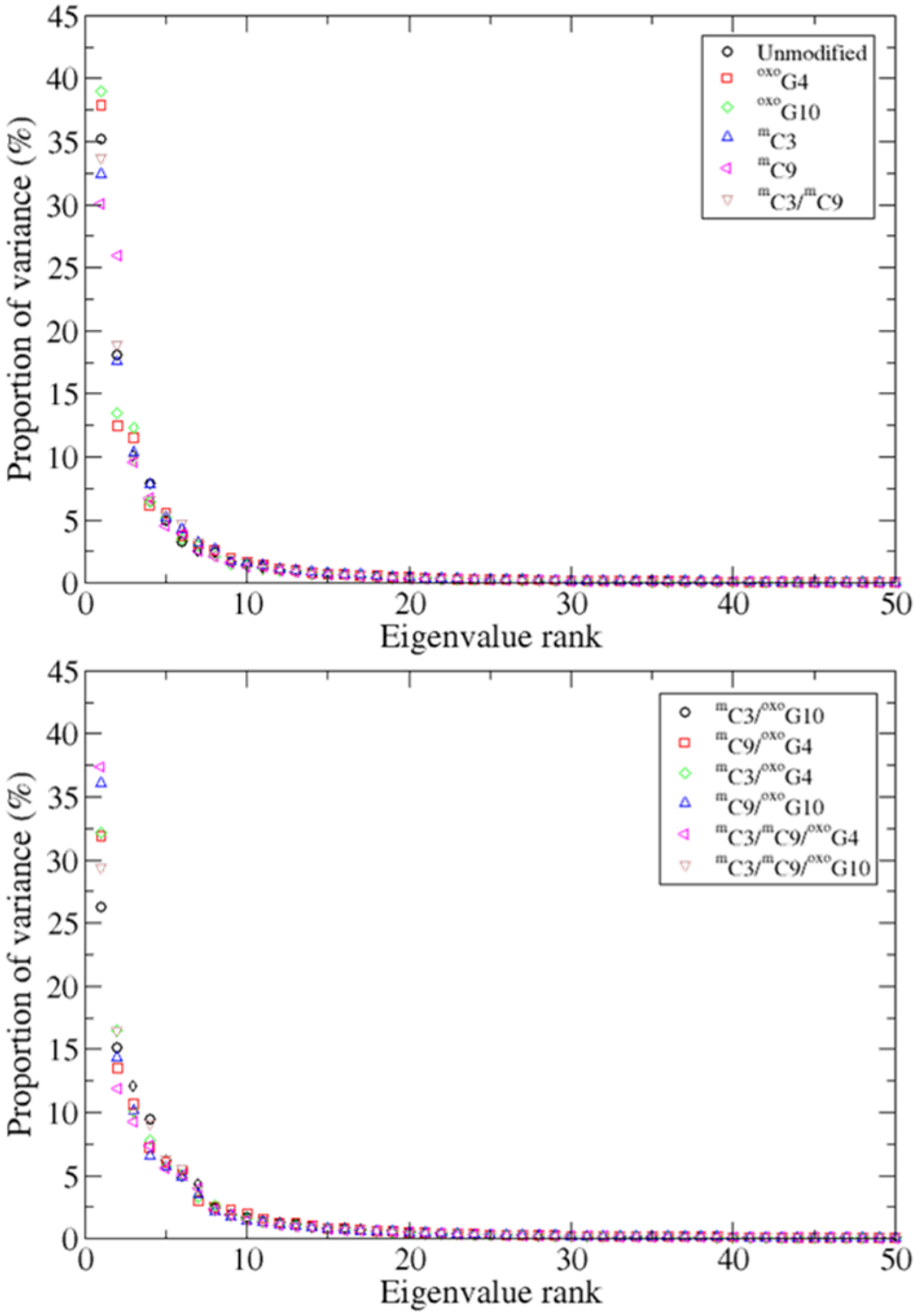



**Figure 4.6. Scree plots indicating the proportion of variance that each principal component contributes to the overall dynamics movement of the CpG site. (top) all control samples, (bottom) all target samples.**

Due to the contribution to the dynamic movements being dominated by the first five PCs, we plotted the principal component projections as histograms to determine if the dynamics had been altered with the addition of 8OG, 5MC or both to the CpG site. The PC 1 projections for the CpG sites in all samples are displayed in Figure 4.7. As depicted, the projection distributions exhibit minimal overlap, indicating the most dominant dynamic movements have been altered with the incorporation and clustering of modifications in distinct sequence contexts. We also analyzed the projections of PC2-5 for all twelve samples. Shown in Figure 4.8 is the PC 2 projection for all samples. As one can see, the histograms display reasonable overlap, indicating the motions are highly similar. Histograms for PC3-5 are not shown, however, the also displayed overlap similar to PC 2. To elucidate the nature of change of these dynamic movements, we relied on two separate visualization techniques.

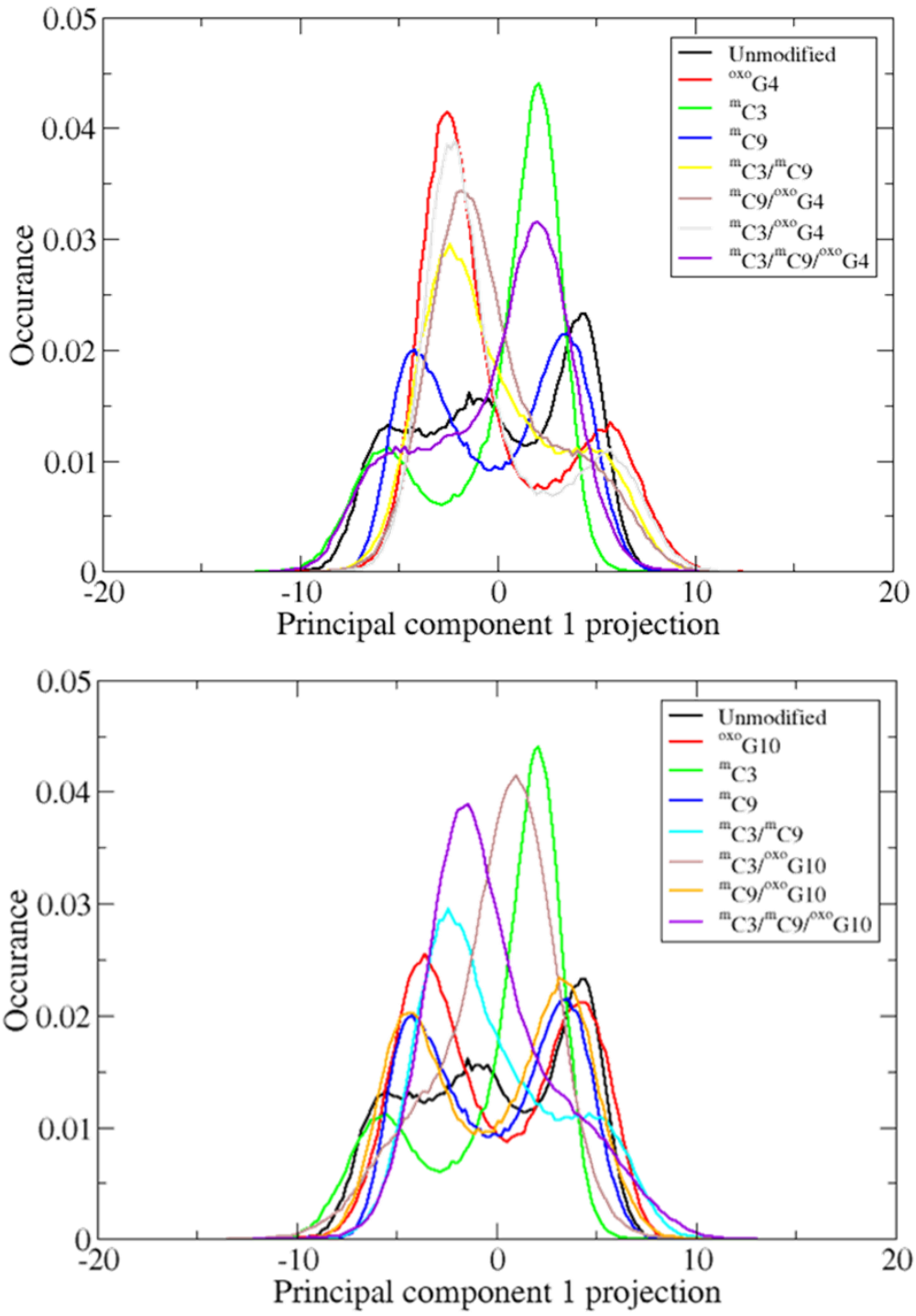



**Figure 4.7. Principal component projections for PC 1 for all samples. (top) comparison of DNA duplexes containing oxidized G4 with controls. (bottom) comparison of DNA duplexes containing oxidized G10 with controls.**

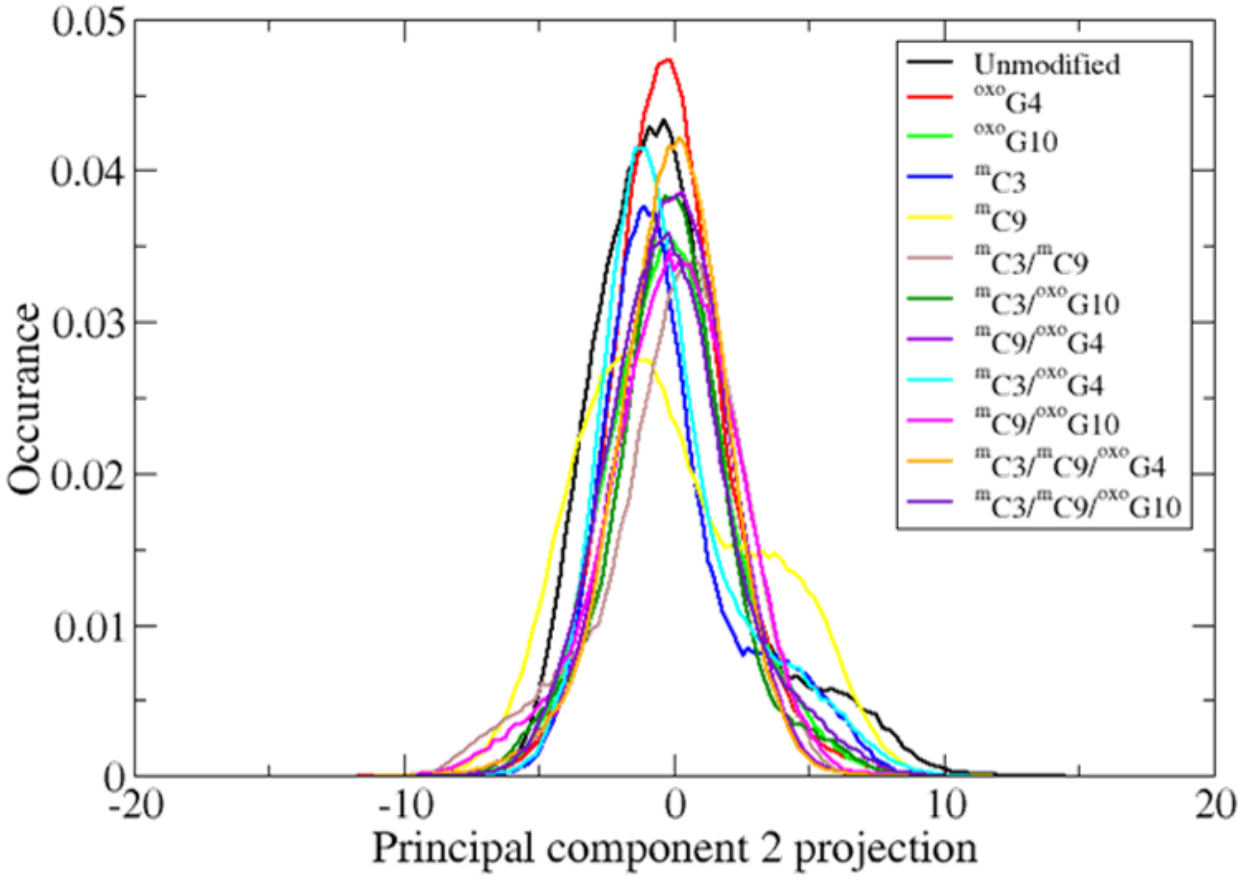



**Figure 4.8. Principal component projection for PC 2 for all samples. Overlap of projections indicate PC2 displays highly similar motions between samples.**

Porcupine plots have been used as a convenient method to visually describe concerted motions of principal components in various MD studies (Haider et al., 2008; Tai et al., 2001; Tai et al., 2002). In Figures 4.9-4.11, we present porcupine plots on one set of *trans*- *cis*- and fully modified samples to highlight perturbations in dynamic movements for clustering of 5MC and 8OG. In addition, we have generated porcupine plots for each CpG (Appendix Figure A5) site to visualize the motions of the most dominant PC. The eigenvectors (direction of motion) and associated eigenvalues (magnitude of motion) for are presented for each of the ~44 (dependent on modification) heavy atoms present in the CpG site.

When 8OG and 5MC are incorporated into the CpG site in the *trans*-methylated configuration ($^{m}$C9/$^{oxo}$G4 and $^{m}$C3/$^{oxo}$G10), the strand containing 5MC showed altered amplitudes and directionality of motion both in the base and backbone (Figure 4.9 and Figure A5). Introducing 5MC and 8OG together so that the CpG site is *cis*-methylated ($^{m}$C3/$^{oxo}$G4 and $^{m}$C9/$^{oxo}$G10) gives rise to dynamic movements that are similar to those of the oxidized CpG sites and very different from those of the methylated-only sample (Figure 4.10 and Figure A5). In the fully-methylated and oxidized samples, the directions of atomic motions are altered with respect to the methylated–only control and amplitudes of motions are increased noticeably with respect to the oxidized control (Figure 4.11).

To further probe the differences in dynamics of PC 1, we visually inspected the resulting pseudo-trajectories and analyzed them for differences in helicoidal geometries and backbone angle transitions, for each CpG site. As a quality control, visual inspection of PC 2 pseudo-trajectories resulted in indistinguishable motions. The range of motion, minimum and maximum values for selected base-pair geometries present in the PC 1 motions are shown in Tables 4.4 and 4.5. Certain combinations of 5MC/8OG clustering display strong effects on the ranges of change of helical parameters. For example, the range of variation of slide for the $^{m}$C3/$^{oxo}$G10 sample (0.56 Å) is much smaller than the corresponding variations for the sample with single oxidation $^{oxo}$G10 (2.31 Å) or only methylated $^{m}$C3 (3.28 Å) samples. Similar reduction of slide variability in PC1 is observed for $^{m}$C9/$^{oxo}$G4 (1.61 Å) when compared to single oxidation oxoG4 (2.19 Å) and single methylation $^{m}$C9 (0.74 Å). Likewise, the $^{m}$C3/$^{m}$C9/$^{oxo}$G10 CpG site leads to a reduction of the range

of variation of slide from 1.60 Å to 1.25 Å observed in the lone oxidation control $^{oxo}$G10. This effect is sequence-specific as the effect is not observed with $^{m}$C3/$^{m}$C9/$^{oxo}$G4. Another helical parameter showing considerable sensitivity to 5MC/8OG clustering is tilt, as its range is reduced dramatically from ~10° to ~ 2° (depending on sequence context) upon clustering of 8OG and *trans*-methylated 5MC with respect to only oxidized and only methylated control samples. This trend is continued to full-methylated and oxidized samples but is not present on *cis*-methylated duplexes.

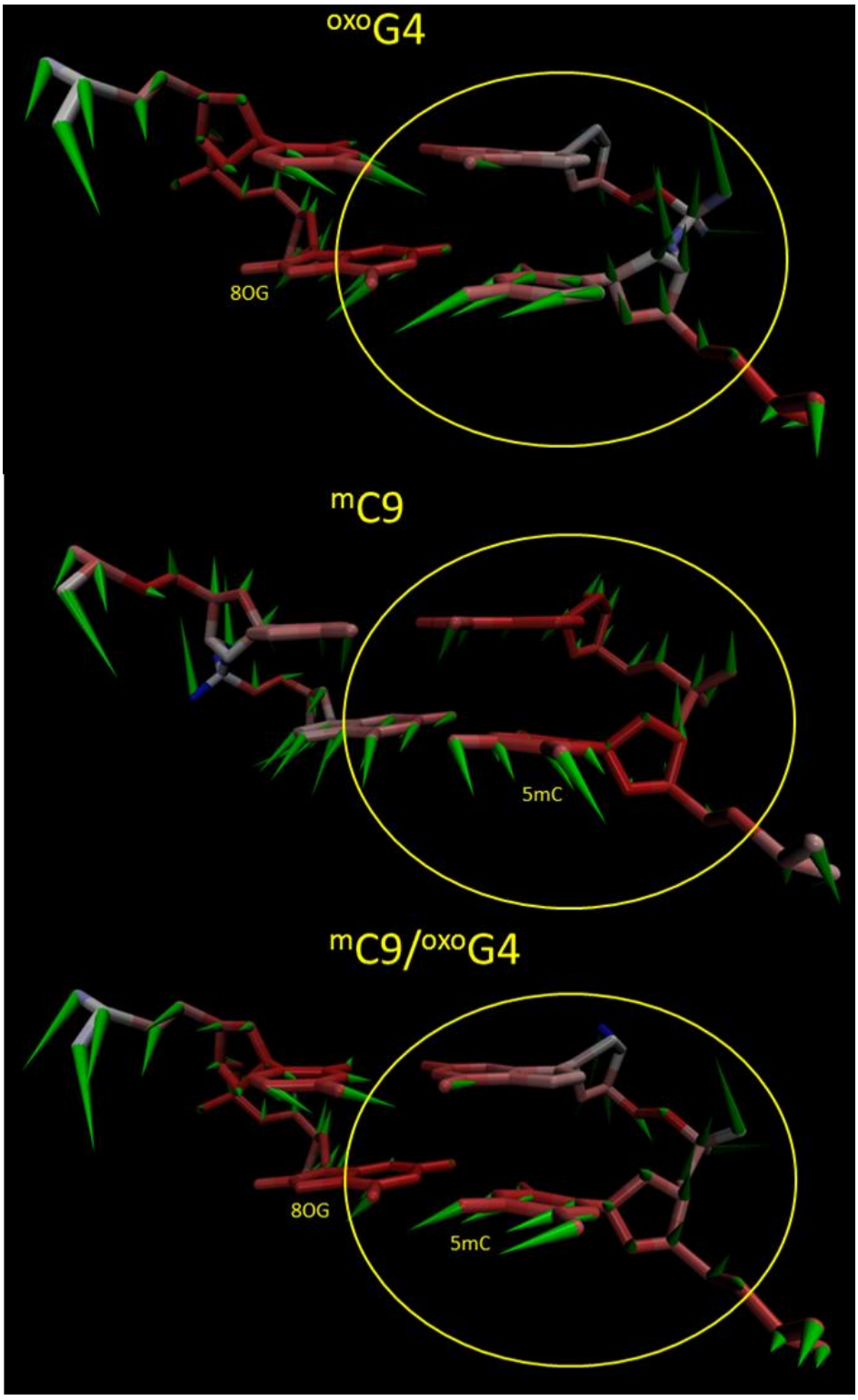



**Figure 4.9. Porcupine plots highlighting the dynamic effect of the 8OG:5MC base-pair on the CpG site. Each arrow projected from each heavy atom indicates the direction of the eigenvector while the length of the arrow indicates the magnitude of the associated eigenvalue. Atom coloring is by mobility, with red being lowest mobility atoms and blue being highest mobility.**

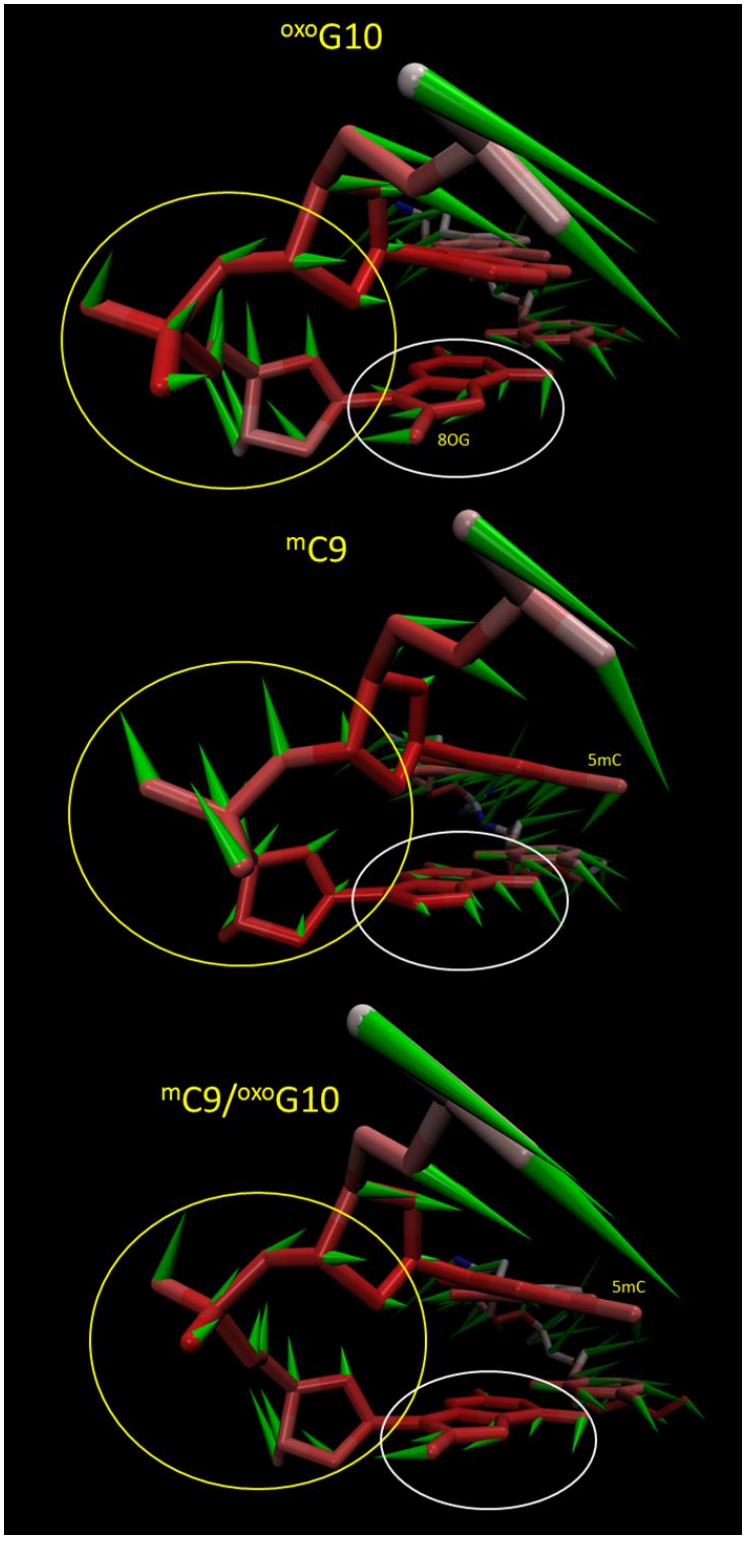



**Figure 4.10. Porcupine plots highlighting the dynamic effect of adjacent 8OG and 5MC on the CpG site. Each arrow projected from each heavy atom indicates the direction of the eigenvector while the length of the arrow indicates the magnitude of the associated eigenvalue. Atom coloring is by mobility, with red being lowest mobility atoms and blue being highest mobility.**

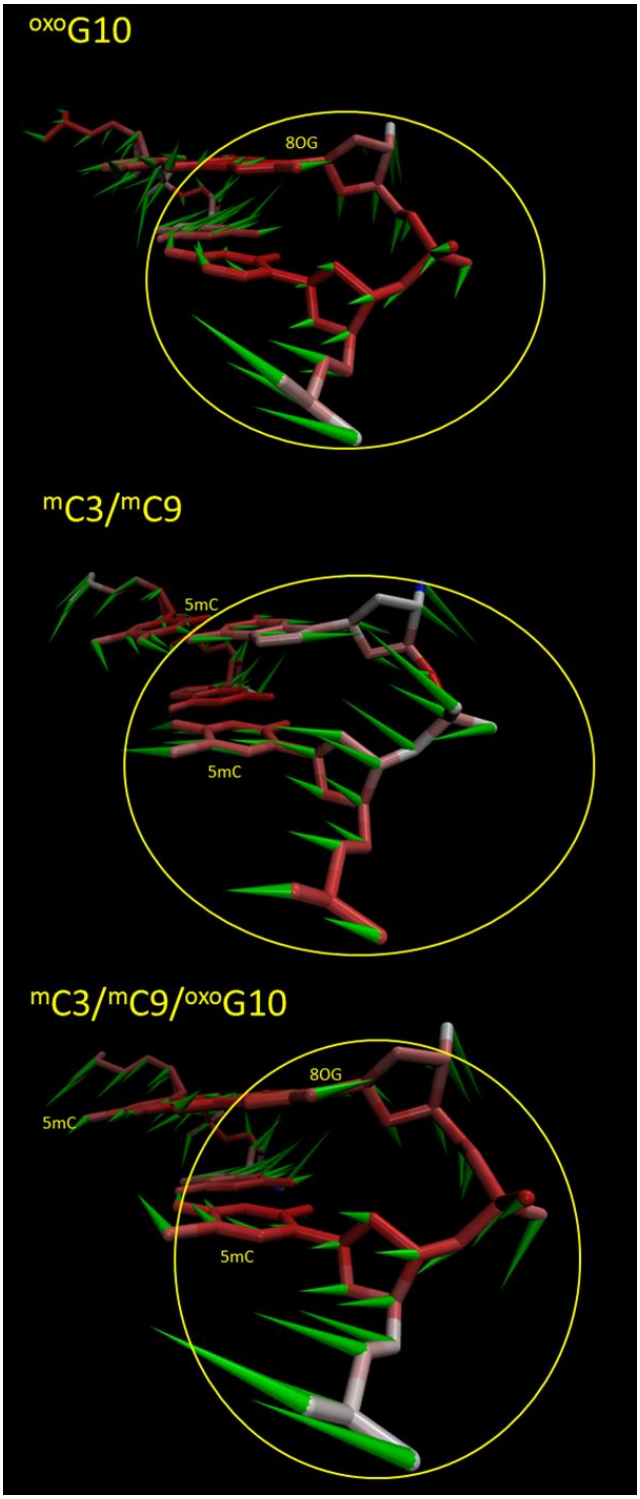



**Figure 4.10. Porcupine plots highlighting the dynamic effect of fully-methylated and oxidized CpG. Each arrow projected from each heavy atom indicates the direction of the eigenvector while the length of the arrow indicates the magnitude of the associated eigenvalue. Atom coloring is by mobility, with red being lowest mobility atoms and blue being highest mobility.**

| Sample | Twist (deg.) | | | Roll (deg.) | | | Tilt (deg.) | | |
|---|---|---|---|---|---|---|---|---|---|
| | range | min | max | range | min | max | range | min | max |
| unmodified | 1.2 | 34.0 | 35.2 | 0.6 | 5.0 | 5.6 | 17.4 | -6.3 | 11.1 |
| $^{oxo}$G4 | 33.5 | 12.2 | 45.6 | 6.7 | 4.5 | 11.2 | 12.9 | -1.1 | 11.9 |
| $^{oxo}$G10 | 36.5 | 7.3 | 43.7 | 3.3 | 6.1 | 9.4 | 10.4 | -6.7 | 3.7 |
| $^{m}$C3 | 30.3 | 16.1 | 46.4 | 6.4 | 4.5 | 11.0 | 12.4 | -0.9 | 11.5 |
| $^{m}$C9 | 22.3 | 18.3 | 40.6 | 5.0 | 4.3 | 9.3 | 12.5 | -7.3 | 5.2 |
| $^{m}$C3/$^{m}$C9 | 37.9 | 5.8 | 43.7 | 8.0 | 6.1 | 14.2 | 3.9 | 0.7 | 4.6 |
| $^{m}$C3/$^{oxo}$G10 | 36.7 | 2.3 | 39.0 | 4.6 | 7.7 | 12.4 | 1.9 | 0.5 | 2.4 |
| $^{m}$C9/$^{oxo}$G4 | 34.0 | 4.3 | 38.3 | 6.4 | 7.2 | 13.6 | 4.4 | 0.6 | 5.0 |
| $^{m}$C3/$^{oxo}$G4 | 33.4 | 12.8 | 46.2 | 5.8 | 5.4 | 11.3 | 12.0 | 0.0 | 11.9 |
| $^{m}$C9/$^{oxo}$G10 | 36.3 | 7.3 | 43.6 | 5.8 | 4.6 | 10.3 | 9.5 | -6.5 | 3.0 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | 33.9 | 4.7 | 38.6 | 6.5 | 7.5 | 14.0 | 5.1 | 0.8 | 5.9 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | 37.4 | 2.1 | 39.5 | 4.7 | 7.8 | 12.5 | 3.8 | -2.0 | 1.8 |

**Table 4.4. Base-pair helicoidal geometry ranges for twist, roll and tilt. Values were calculated from the pseudo-trajectory of PC 1 for each sample.**

| Sample | Rise (Å) | | | Slide (Å) | | | Shift (Å) | | |
|---|---|---|---|---|---|---|---|---|---|
| | range | min | max | range | min | max | range | min | max |
| unmodified | 0.31 | 3.05 | 3.36 | 0.54 | 0.10 | 0.64 | 4.62 | -2.02 | 2.60 |
| $^{oxo}$G4 | 0.59 | 2.90 | 3.49 | 2.19 | -0.89 | 1.3 | 2.62 | -0.35 | 2.27 |
| $^{oxo}$G10 | 0.85 | 2.76 | 3.61 | 1.25 | -0.78 | 0.47 | 2.31 | -1.74 | 0.57 |
| $^{m}$C3 | 0.53 | 2.90 | 3.43 | 2.05 | -0.69 | 1.36 | 3.28 | -0.91 | 2.37 |
| $^{m}$C9 | 0.52 | 2.82 | 3.34 | 0.74 | -0.20 | 0.55 | 4.11 | -2.19 | 1.92 |
| $^{m}$C3/$^{m}$C9 | 0.76 | 2.75 | 3.51 | 2.04 | -1.04 | 1.00 | 1.84 | -0.79 | 1.06 |
| $^{m}$C3/$^{oxo}$G10 | 0.82 | 2.69 | 3.51 | 2.11 | -1.33 | 0.78 | 0.56 | -0.55 | 0.01 |
| $^{m}$C9/$^{oxo}$G4 | 0.79 | 2.69 | 3.48 | 1.61 | -1.05 | 0.56 | 2.07 | -0.54 | 1.53 |
| $^{m}$C3/$^{oxo}$G4 | 0.53 | 2.91 | 3.45 | 2.36 | -0.93 | 1.42 | 2.75 | -0.45 | 2.31 |
| $^{m}$C9/$^{oxo}$G10 | 0.76 | 2.79 | 3.55 | 1.70 | -0.9 | 0.80 | 2.33 | -1.72 | 0.61 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | 0.72 | 2.71 | 3.43 | 1.93 | -1.09 | 0.84 | 2.35 | -0.64 | 1.71 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | 0.83 | 2.71 | 3.54 | 1.60 | -1.05 | 0.55 | 1.01 | -0.84 | 0.18 |

**Table 4.5. Base-pair helicoidal geometry ranges for rise, slide and shift. Values were calculated from the pseudo-trajectory of PC 1 for each sample.**

# *Chapter V—Discussion*

In this study, we examined DNA duplexes containing clustered patterns of two modifications with the goal of unveiling effects in structure, thermodynamics, base dynamics and conformational flexibility upon clustering of these modifications. We characterized the effects with a combination of experimental approaches including 1D-$^{1}$H NMR, 2D homo- and hetero- nuclear NMR, UV-vis spectrophotometry and molecular dynamics.

***Overview of experimental results***

In most regards, the conformational properties of the DNA duplexes determined through solution NMR were analogous to one another. All duplexes were right-handed B-form structures with absent or minor structural deviations located at the site of modification, depending on the modification. This result is consistent with numerous reports on DNA duplexes containing 8OG (Thiviyanathan et al., 2003), 5MC (Theruvathu et al., 2013) as well as many other modifications such as 5-chlororcytosine (Theruvathu et al., 2013), 5-formylcytosine (Hardwick et al., 2017) and phosphorothioate oligonucleotides (Lan et al., 2016). Structural perturbations were found to be induced by 8OG, with and without the presence of 5MC, and were localized around the lesion in all 8OG containing structures. Specifically, downfield $^{31}$P chemical shifts for the phosphodiester 3' to 8OG indicated an increased population of BII conformation relative to undamaged bases. This altered phosphodiester conformation 3' to 8OG caused related unwinding of the helicoidal geometries twist and tilt. The most plausible explanation for this result is that the addition of an electronegative and bulky O8 atom causes intra-nucleotide steric repulsion with the O4' sugar atom. These results are corroborated by previous MD simulations and solution NMR structures

of duplexes containing a lone 8OG modification, but never for these modifications clustered with 5MC (Thiviyanathan et al., 2003; Fujimoto et al., 2005; Naômé et al., 2010).

The thermal denaturation experiments using UV-vis spectrophotometry confirmed that lone 5MC modification increased the $T_m$ of the DNA while lone 8OG decreases the overall stability versus previously reported $T_m$ of DDD. However, clustering of 8OG and 5MC for samples $^{m}$C9/$^{oxo}$G4 and $^{m}$C9/$^{oxo}$G10 was shown to lower the $T_m$ of the duplex by up to 3 °C versus lone 8OG modifications. Therefore, the addition of 5MC at the 9$^{th}$ position to either $^{oxo}$G4 or $^{oxo}$G10 samples induces a destabilizing effect. Furthermore, the effect seems to be sequence specific as the flanking bases for each of these respective samples differ.

1D-$^{1}$H NMR experiments confirmed previously reported data that oxidation of guanine bases leads to imino proton line broadening of the oxidized base (Singh et al., 2011). This was consistent throughout all oxidized samples, as such, the clustering of 5MC with 8OG does not perturb the localized destabilization of the 8OG:C base pair. However, we observed the stabilization of non-oxidized guanines within samples $^{m}$C3/$^{m}$C9/$^{oxo}$G4 and $^{m}$C3/$^{m}$C9/$^{oxo}$G10 vs their respective controls, although the effect observed with $^{m}$C3/$^{m}$C9/$^{oxo}$G4 was less dramatic than that in $^{m}$C3/$^{m}$C9/$^{oxo}$G10. In addition, methylation at C3 destabilized the non-oxidized base pair in $^{m}$C3/$^{oxo}$G10 and stabilization for the non-oxidized base pair in $^{m}$C9/$^{oxo}$G4.

Statistical investigation of free MD simulations on each DNA sample revealed that our simulations were stable and reproduced the conformational aspects of our solution NMR structures.

Specifically, the helicoidal geometries were all within one standard deviation of our reported averaged structure for each DNA sample. Local structural perturbations observed from 8OG in our NMR structures were reproduced by MD including BII phosphodiester conformations for oxidized bases. Principal component analysis was used here to illuminate how clustering of 5MC and 8OG in a CpG site effects motions along eigenvectors with large eigenvalues. We analyzed the most prominent motions as the first five eigenvectors for each simulation accounted for no less than 69% of the overall variance. It was found that significant differences in motion were provided in the first principal component, therefore, the others were excluded from the analysis. For samples that were *trans*-methylated, ($^{m}$C9/$^{oxo}$G4 and $^{m}$C3/$^{oxo}$G10), the strand containing 5MC displayed altered amplitudes of motion in the 5MC:8OG base pair as well as the backbone. Analysis of internal dynamics of *cis*-methylated samples, $^{m}$C3/$^{oxo}$G4 and $^{m}$C9/$^{oxo}$G10, showed dynamic movements similar to that of oxidized only and significantly different to that of methylated only CpG sites. In samples containing a fully-methylated and oxidized CpG ($^{m}$C3/$^{m}$C9/$^{oxo}$G4 and $^{m}$C3/$^{m}$C9/$^{oxo}$G10), directionality and magnitude of atomic motions on the bases and backbone were altered with respect to methylated only control ($^{m}$C3/$^{m}$C9).

***Possible implications on enzymatic removal of 8OG***

Although hOGG1 is able to excise 8OG from the CpG sites containing all methylated contexts investigated in this work, the efficiency of enzymatic repair at two steps ($k_2$, corresponding to cleavage of the *N*-glycosidic bond and $k_3$, associated with product release) is reduced depending on orientation of 5MC to 8OG. A twofold hypothesis has been proposed to explain the reduction in $k_3$ when the CpG is *trans*-methylated with respect to 8OG (Kasymov et al., 2013). First, that the

bulky methyl group opposite 8OG may cause steric clashes with the highly conserved Asn149 residue (Figure 1.9b) that stabilizes the cytosine opposite of 8OG. Second, that a water molecule observed in crystal structures of OGG1 near C5 of the opposing cytosine is displaced (Bruner et al., 2000; Fromme et al., 2003; Chung and Verdine 2004). Crystal structures containing the OGG1—DNA substrate do not contain cytosine adjacent to 8OG so the reduction in $k_2$ cannot be explained in the same manner. However, one plausible explanation for this is conformational restraint caused by close proximity between O8 on 8OG and the bulky methyl group while rotating the 8OG base out of the helix.

Our results derived from PCA on *trans*-methylated CpGs indicate that the addition of 5MC alters the dominant mode of motion for the CpG cross strand from the 8OG. The reduced structural flexibility may make it more difficult to stabilize the unpaired cytosine when the OGG1 everts 8OG from the duplex. Results from our NMR and MD simulations were unable to detect major differences in structures or dominant motions for *cis*-methylated samples, so the reduction in $k_2$ remains unanswered.

***Effects of clustering 5MC and 8OG on DNA methylation***

The previously reported consequences of the 8OG on DNA methylation show the lesion severely interferes with maintenance methylation but not *de novo* methylation (Maltseva et al., 2008; Turk et al., 1995). The report of Maltseva, et al, concluded that the introduction of 8OG adjacent to the target cytosine (analogous to $^{m}$C3/$^{oxo}$G10 and $^{m}$C9/$^{oxo}$G4 samples presented in this thesis) by the catalytic domain of murine DNMT3a caused a 25-fold decrease in methylation rates. In

contrast, their data disclosed that 8OG placed opposite of the target cytosine (corresponding to $^{m}$C3/$^{oxo}$G4 and $^{m}$C9/$^{oxo}$G10 in this manuscript) promoted methylation, leading to a 1.8-fold increase.

Here we report three tenable explanations to the decrease in methylation rates for DNMT substrates $^{m}$C3/$^{oxo}$G10 and $^{m}$C9/$^{oxo}$G4. One such explanation relies on protein—DNA recognition. Our PCA results suggest that for these samples containing *trans*-methylated CpG sites, we observed increased amplitudes and directionality of motions for both the backbone and bases relative to their respective methylated controls. This increase in dynamic flexibility may reduce the favorability of initial protein—DNA recognition. The second explanation is that the enzyme can simply recognize its DNA substrate containing guanine better than its DNA substrate containing 8OG. A third explanation pertains to results from solution NMR derived structures and MD trajectories which indicate that conformational perturbations exist on the 3'-phosphodiester relative to 8OG as well as related unwinding (lower twist and tilt) of base pairs directly 5' of 8OG. These conformational alterations could hinder the recognition of the DNA substrate by DNMTs or the subsequent DNA kinking necessary for enzymatic activity. Specifically, in the case of murine DNMT1, analysis of the crystal structure (Song et al., 2012) reveals a wealth of hydrogen bonding and electrostatic interactions stabilizing the enzyme-DNA interface along the sugar—phosphate backbone as well as intra-helically (Figure 1). Of specific interest, are the interactions between Y1243 and the phosphodiester 3' of 8OG, and amino acid side chains which probe the base pairs on the at the 5' end of 8OG. The 3'-phosphodiester relative to 8OG favoring the non-canonical BII conformation could reduce the capacity for Y1243 to stabilize the backbone in preparation for

the target cytosine to be everted extra-helically. Further, the untwisted nature of the base pairs 5' to 8OG could also interrupt a number of interactions that stabilize the guanine opposite to the target cytosine.

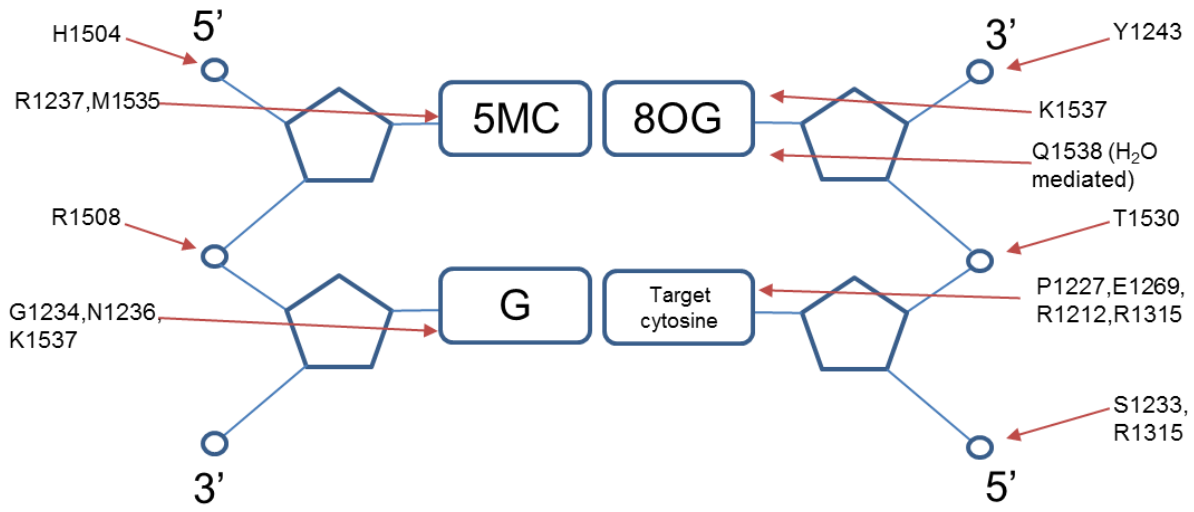



**Figure 5.1. Schematic representation of the murine DNMT1-DNA complex. This minimalistic representation highlights interactions that could be interrupted in the case where 8OG is incorporated adjacent to the target cytosine to become methylated by maintenance DNMTs.**

# *References*


Amadei, A., Linssen, A., & Berendsen, H. J. (1993). Essential dynamics of proteins. *Proteins: Structure, Function, and Bioinformatics*, *17*(4), 412-425.

Amadei, A., Linssen, A. B. M., De Groot, B. L., Van Aalten, D. M. F., & Berendsen, H. J. C. (1996). An efficient method for sampling the essential subspace of proteins. *Journal of Biomolecular Structure and Dynamics*, *13*(4), 615-625.

Ames, B. N., Shigenaga, M. K., & Hagen, T. M. (1993). Oxidants, antioxidants, and the degenerative diseases of aging. *Proceedings of the National Academy of Sciences*, *90*(17), 7915-7922.

Apel, K., & Hirt, H. (2004). Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.*, *55*, 373-399.

Arnott, S., & Hukins, D. W. L. (1972). Optimised parameters for A-DNA and B-DNA. *Biochemical and biophysical research communications*, *47*(6), 1504-1509.

Azzam, E. I., Jay-Gerin, J. P., & Pain, D. (2012). Ionizing radiation-induced metabolic oxidative stress and prolonged cell injury. *Cancer letters*, *327*(1), 48-60.

Balada, E., Ordi-Ros, J., Serrano-Acedo, S., Martinez-Lostao, L., Rosa-Leyva, M., & Vilardell-Tarrés, M. (2008). Transcript levels of DNA methyltransferases DNMT1, DNMT3A and DNMT3B in CD4+ T cells from patients with systemic lupus erythematosus. *Immunology*, *124*(3), 339-347.

Bax, A. (1994). Multidimensional nuclear magnetic resonance methods for protein studies. *Current Opinion in Structural Biology*, *4*(5), 738-744.

Berendsen, H. J. C., Grigera, J. R., & Straatsma, T. P. (1987). The missing term in effective pair potentials. *Journal of Physical Chemistry*, *91*(24), 6269-6271.

Berendsen, H. J., Postma, J. V., van Gunsteren, W. F., DiNola, A. R. H. J., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, *81*(8), 3684-3690.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Brice, M. D., Rodgers, J. R., ... & Tasumi, M. (1977). The protein data bank. *European Journal of Biochemistry*, *80*(2), 319-324.

Bhattacharya, P. K., Cha, J., & Barton, J. K. (2002). 1H NMR determination of base-pair lifetimes in oligonucleotides containing single base mismatches. *Nucleic acids research*, *30*(21), 4740-4750.

Boveris, A., & Chance, B. (1973). The mitochondrial generation of hydrogen peroxide. General properties and effect of hyperbaric oxygen. *Biochemical Journal*, *134*(3), 707-716.

Bruner, S. D., Norman, D. P., & Verdine, G. L. (2000). Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *nature*, *403*(6772), 859-866.

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., ... & Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of computational chemistry*, *26*(16), 1668-1688.

Castagné, C., Murphy, E. C., Gronenborn, A. M., & Delepierre, M. (2000). 31P NMR analysis of the DNA conformation induced by protein binding. *European journal of biochemistry*, *267*(4), 1223-1229.

Cavanagh, J., Fairbrother, W. J., Palmer III, A. G., & Skelton, N. J. (1995). *Protein NMR spectroscopy: principles and practice*. Academic Press.

Cheng, X., Kelso, C., Hornak, V., de los Santos, C., Grollman, A. P., & Simmerling, C. (2005). Dynamic behavior of DNA base pairs containing 8-oxoguanine. *Journal of the American Chemical Society*, *127*(40), 13906-13918.

Chung, S. J., & Verdine, G. L. (2004). Structures of end products resulting from lesion processing by a DNA glycosylase/lyase. *Chemistry & biology*, *11*(12), 1643-1649.

Coman, D., & Russu, I. M. (2005). A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in DNA. *Biophysical journal*, *89*(5), 3285-3292.

Cooke, M. S., Evans, M. D., Dizdaroglu, M., & Lunec, J. (2003). Oxidative DNA damage: mechanisms, mutation, and disease. *The FASEB Journal*, *17*(10), 1195-1214.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., ... & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, *117*(19), 5179-5197.

Crenshaw, C. M., Wade, J. E., Arthanari, H., Frueh, D., Lane, B. F., & Núñez, M. E. (2011). Hidden in plain sight: subtle effects of the 8-oxoguanine lesion on the structure, dynamics, and thermodynamics of a 15-base pair oligodeoxynucleotide duplex. *Biochemistry*, *50*(39), 8463-8477.

Dans, P. D., Danilāne, L., Ivani, I., Dršata, T., Lankaš, F., Walther, J., ... & Orozco, M. (2016). Long-timescale dynamics of the Drew–Dickerson dodecamer. *Nucleic acids research*, gkw264.

Dans, P. D., Faustino, I., Battistini, F., Zakrzewska, K., Lavery, R., & Orozco, M. (2014). Unraveling the sequence-dependent polymorphic behavior of d (CpG) steps in B-DNA. *Nucleic acids research*, *42*(18), 11304-11320.

Dans, P. D., Ivani, I., Portella, G., González, C., & Orozco, M. (2017). How accurate are accurate force-fields for B-DNA?. *Nucleic acids research*, *45*(7), 4217-4230.

David, S. S., O'shea, V. L., & Kundu, S. (2007). Base-excision repair of oxidative DNA damage. *Nature*, *447*(7147), 941-950.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. D. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR*, *6*(3), 277-293.

Dickerson, R. E., & Ng, H. L. (2001). DNA structure from A to B. *Proceedings of the National Academy of Sciences*, *98*(13), 6986-6988.

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1981). Structure of a B-DNA dodecamer: conformation and dynamics. *Proceedings of the National Academy of Sciences*, *78*(4), 2179-2183.

Ehrlich, M., Gama-Sosa, M. A., Huang, L. H., Midgett, R. M., Kuo, K. C., McCune, R. A., & Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic acids research*, *10*(8), 2709-2721.

ESCODD (European Standards Committee on Oxidative DNA Damage. (2002). Comparative analysis of baseline 8-oxo-7, 8-dihydroguanine in mammalian cell DNA, by different methods in different laboratories: an approach to consensus. *Carcinogenesis*, *23*(12), 2129-2133.

Esteller, M. (2005). Aberrant DNA methylation as a cancer-inducing mechanism. *Annu. Rev. Pharmacol. Toxicol.*, *45*, 629-656.

Floyd, R. A. (1990). The role of 8-hydroxyguanine in carcinogenesis. *Carcinogenesis,* 11(9), 1447-1450.

Fromme, J. C., Bruner, S. D., Yang, W., Karplus, M., & Verdine, G. L. (2003). Product-assisted catalysis in base-excision DNA repair. *Nature Structural & Molecular Biology*, *10*(3), 204-211.

Fujimoto, H., Pinak, M., Nemoto, T., O'Neill, P., Kume, E., Saito, K., & Maekawa, H. (2005). Molecular dynamics simulation of clustered DNA damage sites containing 8-oxoguanine and abasic site. *Journal of computational chemistry*, *26*(8), 788-798.

Galindo-Murillo, R., Roe, D. R., & Cheatham III, T. E. (2014). On the absence of intrahelical DNA dynamics on the μs to ms timescale. *Nature communications*, *5*.

Geahigan, K. B., Meints, G. A., Hatcher, M. E., Orban, J., & Drobny, G. P. (2000). The dynamic impact of CpG methylation in DNA. *Biochemistry*, *39*(16), 4939-4946.

Gottlieb, H. E., Kotlyar, V., & Nudelman, A. (1997). NMR chemical shifts of common laboratory solvents as trace impurities. *The Journal of organic chemistry*, *62*(21), 7512-7515.

Güntert, P. (2004). Automated NMR structure calculation with CYANA. *Protein NMR Techniques*, 353-378.

Haider, S., Parkinson, G. N., & Neidle, S. (2008). Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophysical journal*, *95*(1), 296-311.

Hardwick, J. S., Ptchelkine, D., El-Sagheer, A. H., Tear, I., Singleton, D., Phillips, S. E., ... & Brown, T. (2017). 5-Formylcytosine does not change the global structure of DNA. *Nature Structural & Molecular Biology*.

Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., & Reid, B. R. (1983). Assignment of the non-exchangeable proton resonances of d (CGCGAATTCGCG) using two-dimensional nuclear magnetic resonance methods. *Journal of molecular biology*, *171*(3), 319-336.

Hedglin, M., Zhang, Y., & O'Brien, P. J. (2014). Probing the DNA structural requirements for facilitated diffusion. *Biochemistry*, *54*(2), 557-566.

Hermann, A., Goyal, R., & Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *Journal of Biological Chemistry*, *279*(46), 48350-48359.

Hollstein, M., Shomer, B., Greenblatt, M., Soussi, T., Hovig, E., Montesano, R., & Harris, C. C. (1996). Somatic point mutations in the p53 gene of human tumors and cell lines: updated compilation. *Nucleic acids research*, *24*(1), 141-146.

Hoppins, J. J., Gruber, D. R., Miears, H. L., Kiryutin, A. S., Kasymov, R. D., Petrova, D. V., ... & Brockerman, J. A. (2016). 8-Oxoguanine Affects DNA Backbone Conformation in the Eco RI Recognition Site and Inhibits Its Cleavage by the Enzyme. *PloS one*, *11*(10), e0164424.

Hsu, G. W., Ober, M., Carell, T., & Beese, L. S. (2004). Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature*, *431*(7005), 217-221.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics*, *14*(1), 33-38.

Isaacs, R. J., & Spielmann, H. P. (2001). NMR evidence for mechanical coupling of phosphate B I-B II transitions with deoxyribose conformational exchange in DNA. *Journal of molecular biology*, *311*(1), 149-160.

Jeltsch, A. (2006). On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme.

Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Protein NMR Techniques*, 313-352.

Jones, P. A., & Laird, P. W. (1999). Cancer-epigenetics comes of age. *Nature genetics*, *21*(2), 163-167.

Josse, J., Kaiser, A. D., & Kornberg, A. (1961). Enzymatic synthesis of deoxyribonucleic acid VIII. Frequencies of nearest neighbor base sequences in deoxyribonucleic acid. *Journal of biological chemistry*, *236*(3), 864-875.

Joung, I. S., & Cheatham III, T. E. (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *The journal of physical chemistry B*, *112*(30), 9020-9041.

Jovanovic, S. V., & Simic, M. G. (1986). One-electron redox potentials of purines and pyrimidines. *The Journal of Physical Chemistry*, *90*(5), 974-978.

Karplus, M., & Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature*, *347*(6294), 631.

Kasymov, R. D., Grin, I. R., Endutkin, A. V., Smirnov, S. L., Ishchenko, A. A., Saparbaev, M. K., & Zharkov, D. O. (2013). Excision of 8-oxoguanine from methylated CpG dinucleotides by human 8-oxoguanine DNA glycosylase. *FEBS letters*, *587*(18), 3129-3134.

Kim, Y. J., & M Wilson III, D. (2012). Overview of base excision repair biochemistry. *Current molecular pharmacology*, *5*(1), 3-13.

Kozmin, S., Slezak, G., Reynaud-Angelin, A., Elie, C., De Rycke, Y., Boiteux, S., & Sage, E. (2005). UVA radiation is highly mutagenic in cells that are unable to repair 7, 8-dihydro-8-oxoguanine in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(38), 13538-13543.

Kuznetsov, N. A., Koval, V. V., Zharkov, D. O., Nevinsky, G. A., Douglas, K. T., & Fedorova, O. S. (2005). Kinetics of substrate recognition and cleavage by human 8-oxoguanine-DNA glycosylase. *Nucleic acids research*, *33*(12), 3919-3931.

Lafer, E. M., Valle, R. P., Möller, A., Nordheim, A., Schur, P. H., Rich, A., & Stollar, B. D. (1983). Z-DNA-specific antibodies in human systemic lupus erythematosus. *Journal of Clinical Investigation*, *71*(2), 314.

Lan, W., Hu, Z., Shen, J., Wang, C., Jiang, F., Liu, H., ... & Cao, C. (2016). Structural investigation into physiological DNA phosphorothioate modification. *Scientific reports*, *6*.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., ... & Funke, R. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860-921.

Lankaš, F., Cheatham, T. E., Hobza, P., Langowski, J., & Šponer, J. (2002). Critical effect of the N2 amino group on structure, dynamics, and elasticity of DNA polypurine tracts. *Biophysical journal*, *82*(5), 2592-2609.

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews Genetics*, *3*(9), 662-673.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., ... & Edsall, L. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *nature*, *462*(7271), 315-322.

Lu, X. J., & Olson, W. K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic acids research*, *31*(17), 5108-5121.

Lu, X. J., Shakked, Z., & Olson, W. K. (2000). A-form conformational motifs in ligand-bound DNA structures. *Journal of molecular biology*, *300*(4), 819-840.

Maltseva, D. V., Baykov, A. A., Jeltsch, A., & Gromova, E. S. (2009). Impact of 7, 8-dihydro-8-oxoguanine on methylation of the CpG site by Dnmt3a. *Biochemistry*, *48*(6), 1361-1368.

Matta, C. F., Castillo, N., & Boyd, R. J. (2006). Extended weak bonding interactions in DNA: π-stacking (base− base), base− backbone, and backbone− backbone interactions. *The Journal of Physical Chemistry B*, *110*(1), 563-578.

Michaels, M. L., & Miller, J. H. (1992). The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7, 8-dihydro-8-oxoguanine). *Journal of bacteriology*, *174*(20), 6321.

Naômé, A., Schyman, P., Laaksonen, A., & Vercauteren, D. P. (2010). Molecular dynamics simulation of 8-oxoguanine containing DNA fragments reveals altered hydration and ion binding patterns. *The Journal of Physical Chemistry B*, *114*(14), 4789-4801.

Nakabeppu, Y., Kajitani, K., Sakamoto, K., Yamaguchi, H., & Tsuchimoto, D. (2006). MTH1, an oxidized purine nucleoside triphosphatase, prevents the cytotoxicity and neurotoxicity of oxidized purine nucleotides. *DNA repair*, *5*(7), 761-772.

Nikonowicz, E., Roongta, V., Jones, C., & Gorenstein, D. (1989). Two-dimensional 1H and 31P NMR spectra and restrained molecular dynamics structure of an extrahelical adenosine tridecamer oligodeoxyribonucleotide duplex. *Biochemistry*, *28*, 87x4-8725.

Okano, M., Xie, S., & Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature genetics*, *19*(3), 219-220.

Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M., & Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proceedings of the National Academy of Sciences*, *95*(19), 11163-11168.

Ott, J., & Eckstein, F. (1985). Phosphorus-31 NMR spectral analysis of the dodecamer d (CGCGAATTCGCG). *Biochemistry*, *24*(10), 2530-2535.

Park, E. M., Shigenaga, M. K., Degan, P., Korn, T. S., Kitzler, J. W., Wehr, C. M., ... & Ames, B. N. (1992). Assay of excised oxidative DNA lesions: isolation of 8-oxoguanine and its nucleoside derivatives from biological fluids with a monoclonal antibody column. *Proceedings of the National Academy of Sciences*, *89*(8), 3375-3379.

Parker, J. B., Bianchet, M. A., Krosky, D. J., Friedman, J. I., Amzel, L. M., & Stivers, J. T. (2007). Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature*, *449*(7161), 433-437.

Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., ... & Kollman, P. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, *91*(1-3), 1-41.

Pérez, A., Luque, F. J., & Orozco, M. (2007). Dynamics of B-DNA on the microsecond time scale. *Journal of the American Chemical Society*, *129*(47), 14739-14745.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, *25*(13), 1605-1612.

Rhee, I., Bachman, K. E., Park, B. H., Jair, K. W., Yen, R. W. C., Schuebel, K. E., ... & Baylin, S. B. (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature*, *416*(6880), 552-556.

Roberts, M. F., Cui, Q., Turner, C. J., Case, D. A., & Redfield, A. G. (2004). High-resolution field-cycling NMR studies of a DNA octamer as a probe of phosphodiester dynamics and comparison with computer simulation. *Biochemistry*, *43*(12), 3637-3650.

Robertson, K. D., & Jones, P. A. (2000). DNA methylation: past, present and future directions. *Carcinogenesis*, *21*(3), 461-467.

Rodríguez López, C. M., Guzmán Asenjo, B., Lloyd, A. J., & Wilkinson, M. J. (2010). Direct detection and quantification of methylation in nucleic acid sequences using high-resolution melting analysis. *Analytical chemistry*, *82*(21), 9100-9108.

Roe, D. R., & Cheatham III, T. E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation*, *9*(7), 3084-3095.

Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., & Honig, B. (2009). The role of DNA shape in protein–DNA recognition. *Nature*, *461*(7268), 1248-1253.

Rüdisser, S., Hallbrucker, A., & Mayer, E. (1997). B-DNA's conformational substates revealed by Fourier transform infrared difference spectroscopy. *Journal of the American Chemical Society*, *119*(50), 12251-12256.

Ryckaert, J. P., Ciccotti, G., & Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, *23*(3), 327-341.

Saenger, W. (1984). Springer Advanced Texts in Chemistry: Principles of Nucleic Acid Structure (Cantor, CR, ed.).

Schneider, B., Neidle, S., & Berman, H. M. (1997). Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers*, *42*(1), 113-124.

Seeman, N. C., Rosenberg, J. M., & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proceedings of the National Academy of Sciences*, *73*(3), 804-808.

Setoyama, D., Ito, R., Takagi, Y., & Sekiguchi, M. (2011). Molecular actions of Escherichia coli MutT for control of spontaneous mutagenesis. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *707*(1), 9-14.

Shah, K., Kumar, R. G., Verma, S., & Dubey, R. S. (2001). Effect of cadmium on lipid peroxidation, superoxide anion generation and activities of antioxidant enzymes in growing rice seedlings. *Plant Science*, *161*(6), 1135-1144.

Shen, J. C., Rideout, W. M., & Jones, P. A. (1994). The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic acids research*, *22*(6), 972-976.

Shibutani, S., Takeshita, M., & Grollman, A. P. (1991). Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *nature*, *349*(6308), 431.

Singh, S. K., Szulik, M. W., Ganguly, M., Khutsishvili, I., Stone, M. P., Marky, L. A., & Gold, B. (2011). Characterization of DNA with an 8-oxoguanine modification. *Nucleic acids research*, *39*(15), 6789-6801.

Song, J., Teplova, M., Ishibe-Murakami, S., & Patel, D. J. (2012). Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science*, *335*(6069), 709-712.

Still, W. C., Tempczyk, A., Hawley, R. C., & Hendrickson, T. (1990). Semianalytical treatment of solvation for molecular mechanics and dynamics. *Journal of the American Chemical Society*, *112*(16), 6127-6129.

Swartz, M. N., Trautner, T. A., & Kornberg, A. (1962). Enzymatic synthesis of deoxyribonucleic acid XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. *Journal of biological chemistry*, *237*(6), 1961-1967.

Tai, K., Shen, T., Börjesson, U., Philippopoulos, M., & McCammon, J. A. (2001). Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase. *Biophysical journal*, *81*(2), 715-724.

Tai, K., Shen, T., Henchman, R. H., Bourne, Y., Marchot, P., & McCammon, J. A. (2002). Mechanism of acetylcholinesterase inhibition by fasciculin: a 5-ns molecular dynamics simulation. *Journal of the American Chemical Society*, *124*(21), 6153-6161.

Teeter, M. M., & Case, D. A. (1990). Harmonic and quasiharmonic descriptions of crambin. *Journal of Physical Chemistry*, *94*(21), 8091-8097.

Temiz, N. A., Donohue, D. E., Bacolla, A., Luke, B. T., & Collins, J. R. (2012). The role of methylation in the intrinsic dynamics of B-and Z-DNA. *PLoS One*, *7*(4), e35558.

Theruvathu, J. A., Yin, Y. W., Pettitt, B. M., & Sowers, L. C. (2013). Comparison of the structural and dynamic effects of 5-methylcytosine and 5-chlorocytosine in a CpG dinucleotide sequence. *Biochemistry*, *52*(47), 8590-8598.

Thiviyanathan, V., Somasunderam, A., Hazra, T. K., Mitra, S., & Gorenstein, D. G. (2003). Solution structure of a DNA duplex containing 8-hydroxy-2'-deoxyguanosine opposite deoxyguanosine. *Journal of molecular biology*, *325*(3), 433-442.

Turk, P. W., Laayoun, A., Smith, S. S., & Weitzman, S. A. (1995). DNA adduct 8-hydroxyl-2'-deoxyguanosine (8-hydroxyguanine) affects function of human DNA methyltransferase. *Carcinogenesis*, *16*(5), 1253-1255.

Trieb, M., Rauch, C., Wellenzohn, B., Wibowo, F., Loerting, T., & Liedl, K. R. (2004). Dynamics of DNA: BI and BII phosphate backbone transitions. *The Journal of Physical Chemistry B*, *108*(7), 2470-2476.

Turrens, J. F. (1997). Superoxide production by the mitochondrial respiratory chain. *Bioscience reports*, *17*(1), 3-8.

Várnai, P., Djuranovic, D., Lavery, R., & Hartmann, B. (2002). α/γ Transitions in the B-DNA backbone. *Nucleic acids research*, *30*(24), 5398-5406.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., ... & Gocayne, J. D. (2001). The sequence of the human genome. *science*, *291*(5507), 1304-1351.

Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, *171*(737-738), 3-12.

Weitzman, S. A., Turk, P. W., Milkowski, D. H., & Kozlowski, K. (1994). Free radical adducts induce alterations in DNA cytosine methylation. *Proceedings of the National Academy of Sciences*, *91*(4), 1261-1264.

Williams, T., Kelley, C., Bröker, H. B., Campbell, J., Cunningham, R., Denholm, D., ... & Hart, L. Gnuplot 4.6: An interactive plotting program, 2012. *URL http://www. gnuplot. info*.

Winger, R. H., Liedl, K. R., Rüdisser, S., Pichler, A., Hallbrucker, A., & Mayer, E. (1998). B-DNA's BI→ BII conformer substate dynamics is coupled with water migration. *The Journal of Physical Chemistry B*, *102*(44), 8934-8940.

Yamaguchi, H., Van Aalten, D. M., Pinak, M., Furukawa, A., & Osman, R. (1998). Essential dynamics of DNA containing a cis. syn cyclobutane thymine dimer lesion. *Nucleic acids research*, *26*(8), 1939-1946.

Zacharias, W., Jaworski, A., & Wells, R. D. (1990). Cytosine methylation enhances Z-DNA formation in vivo. *Journal of bacteriology*, *172*(6), 3278-3283.

Zhu, J., He, F., Hu, S., & Yu, J. (2008). On the nature of human housekeeping genes. *Trends in genetics*, *24*(10), 481-484.

Zgarbova, M.; Sponer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurecka, P., Refinement of the Sugar-Phosphate Backbone Torsion Beta for the AMBER Force Fields

Improves the Description of Z-DNA and B-DNA. J. Chem. Theory Comput. 2015, 11 (12), 5723-5736.

# *Appendix*

```
/*      Description/Usage:                                          */
/*      AU program which set the temperature before the             */
/*      acquisition is started. The temperature is taken from       */
/*      the parameter TE.                                           */
/************************************************************/
/*      Authors:        Hla Win-Piazza and David R. Gruber          */
/************************************************************/
*/
f1=333.15;
TIMES(12)
STOREPAR("TE", f1);
TESET
TEREADY(1200, 0.1)
IEXPNO
LOCK
ATMA;
XCMD("topshim 3dfast");
XCMD("topshim 1d");
XAU("pulsecal", "quiet");
RGA;
ZG;
F1=f1-5;
END
QUIT
```

**Script 1. AU program used for automated collection of 1D-$^{1}$H NMR experiments.**

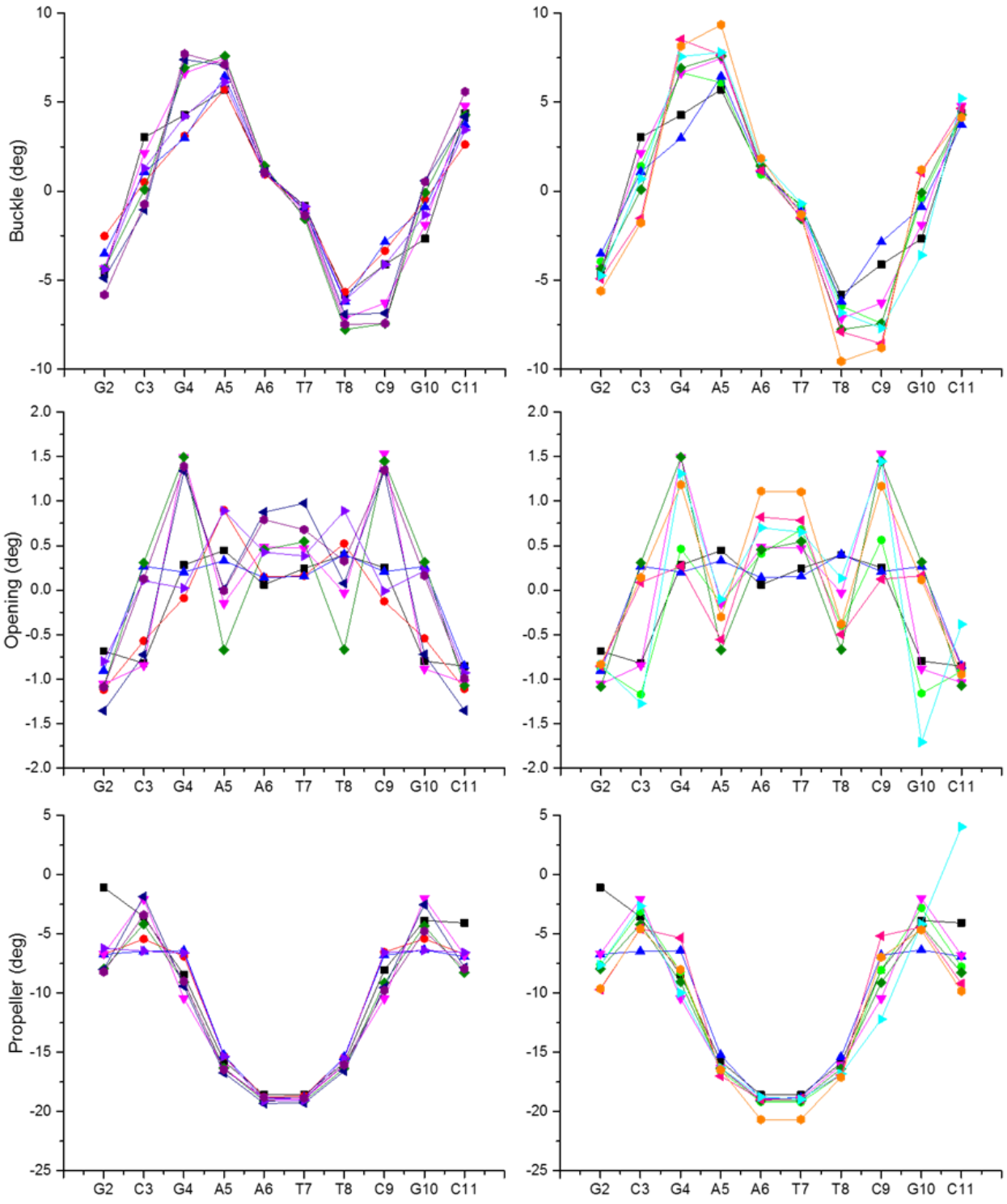



**Figure A1. Average helical base-pair geometries for propeller, opening and buckle obtained from 555ns MD simulation on each DNA sample.**

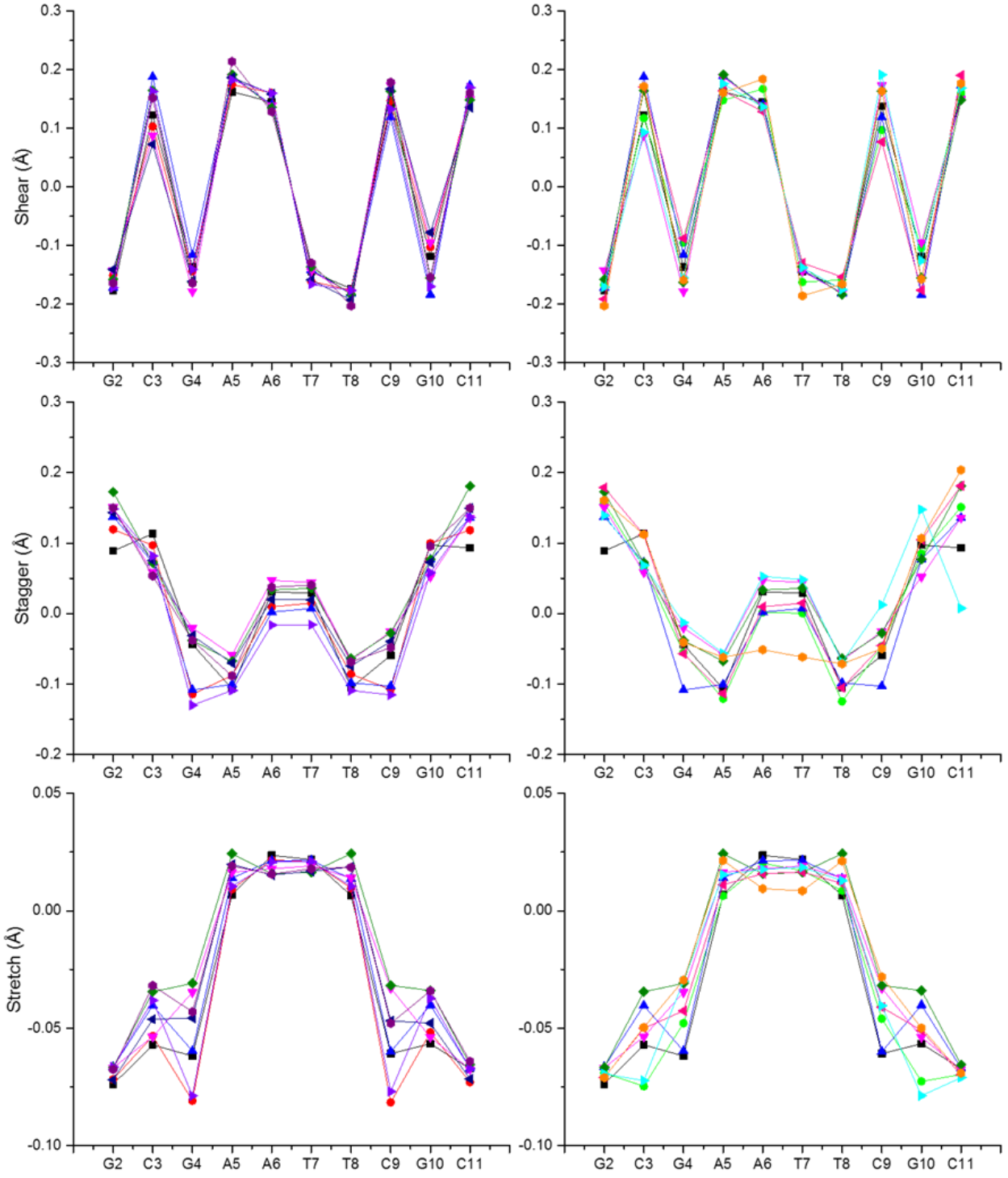



**Figure A2. Average helical base-pair geometries for shear, stagger and stretch obtained from 555ns MD simulation on each DNA sample.**

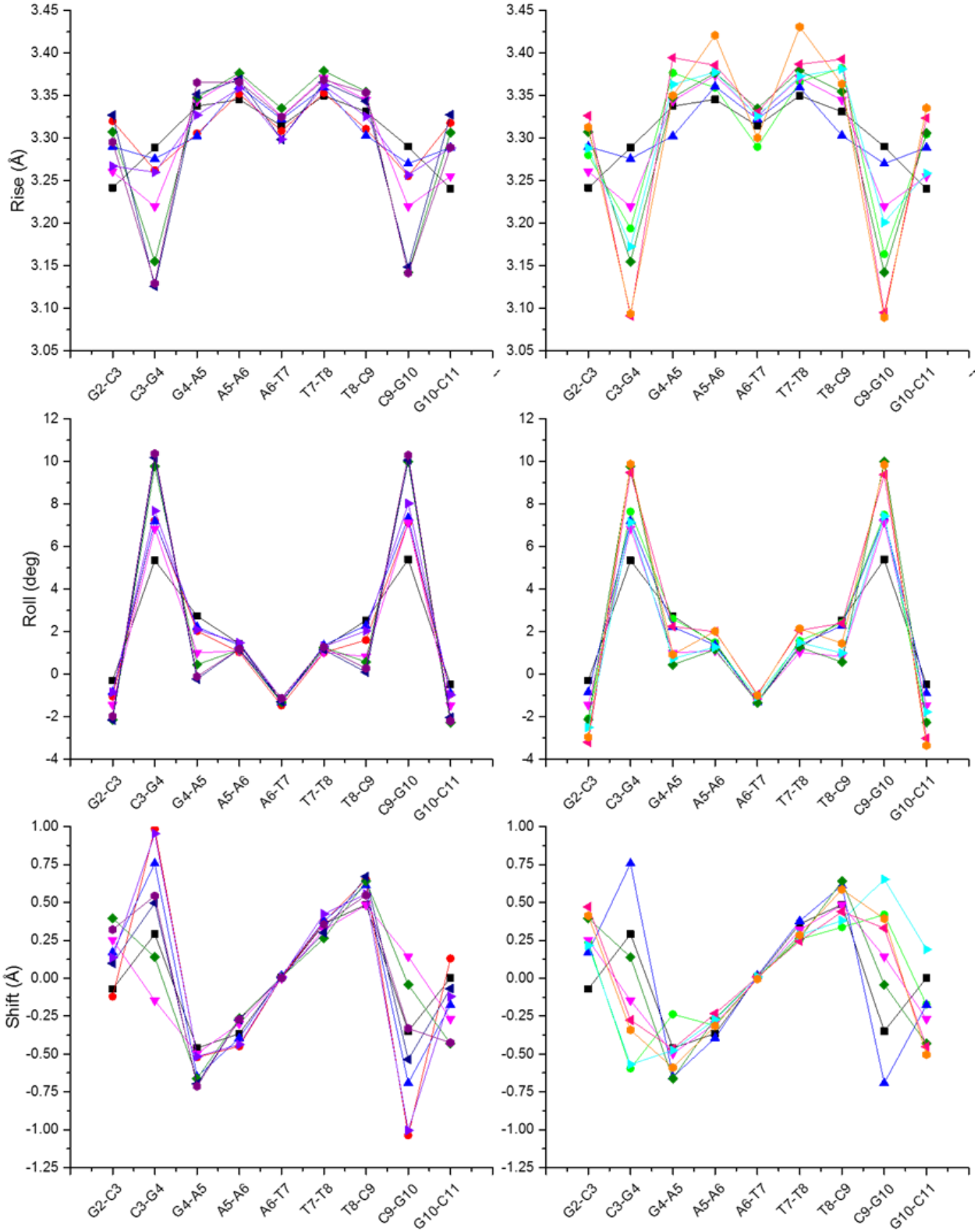



**Figure A3. Average helical base-pair step geometries for rise, roll and shift obtained from 555ns MD simulation on each DNA sample.**

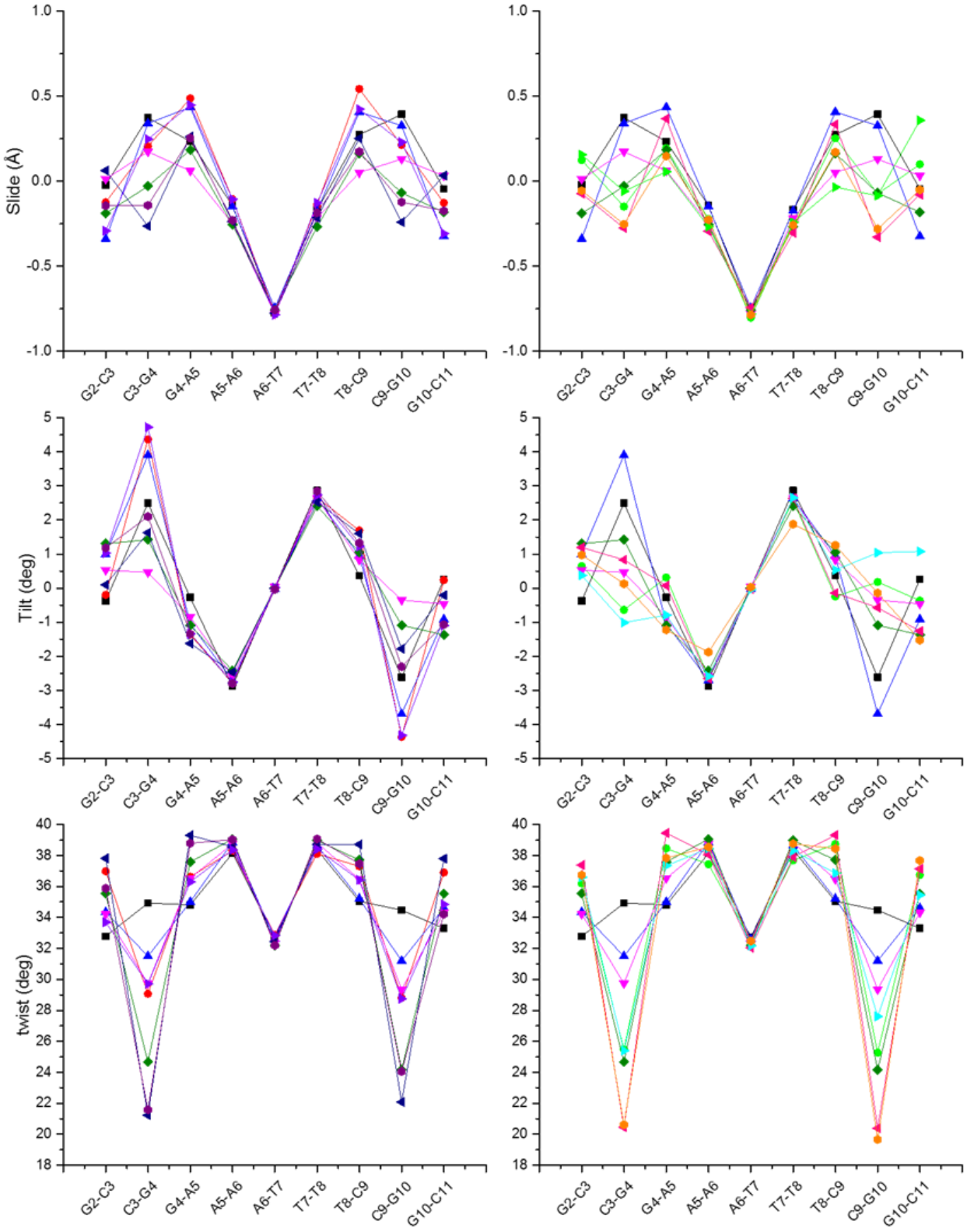
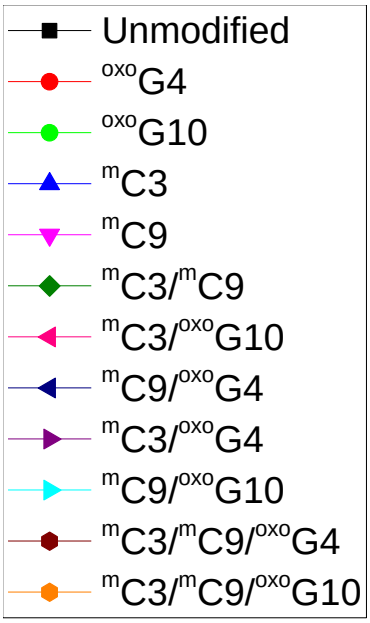



**Figure A4. Average helical base-pair geometries for slide, tilt and twist obtained from 555ns MD simulation on each DNA sample.**




**Legend for Figures A1-A5**

| C3-G4 | Twist (deg) | roll (deg) | tilt (deg) | rise (Å) | slide (Å) | shift (Å) |
|---|---|---|---|---|---|---|
| unmodified | 34.9 ± 6.8 | 5.4 ± 6.3 | 2.5 ± 5.8 | 3.29 ± 0.30 | 0.37 ± 0.55 | 0.29 ± 1.00 |
| $^{oxo}$G4 | 29.1 ± 7.2 | 7.2 ± 6.3 | 4.4 ± 5.1 | 3.26 ± 0.32 | 0.20 ± 0.63 | 0.99 ± 0.72 |
| $^{oxo}$G10 | 25.5 ± 8.0 | 7.6 ± 5.8 | -0.6 ± 5.0 | 3.19 ± 0.34 | -0.15 ± 0.59 | -0.60 ± 0.72 |
| $^{m}$C3 | 31.5 ± 7.1 | 7.2 ± 6.4 | 3.9 ± 5.1 | 3.28 ± 0.30 | 0.34 ± 0.58 | 0.76 ± 0.83 |
| $^{m}$C9 | 29.7 ± 8.0 | 6.8 ± 6.2 | 0.5 ± 5.2 | 3.22 ± 0.31 | 0.17 ± 0.57 | -0.15 ± 0.93 |
| $^{m}$C3/$^{m}$C9 | 24.7 ± 7.6 | 9.8 ± 5.9 | 1.4 ± 4.6 | 3.16 ± 0.29 | -0.03 ± 0.54 | 0.14 ± 0.76 |
| $^{m}$C3/$^{oxo}$G10 | 20.5 ± 6.5 | 9.5 ± 5.5 | 0.8 ± 4.5 | 3.09 ± 0.30 | -0.28 ± 0.56 | -0.28 ± 0.65 |
| $^{m}$C9/$^{oxo}$G4 | 21.2 ± 6.3 | 10.2 ± 5.4 | 1.6 ± 4.4 | 3.13 ± 0.31 | -0.26 ± 0.51 | 0.50 ± 0.59 |
| $^{m}$C3/$^{oxo}$G4 | 29.7 ± 7.3 | 7.7 ± 6.2 | 4.7 ± 5.0 | 3.26 ± 0.31 | 0.25 ± 0.62 | 0.95 ± 0.77 |
| $^{m}$C9/$^{oxo}$G10 | 25.4 ± 8.0 | 7.1 ± 5.7 | -1.0 ± 4.9 | 3.17 ± 0.32 | -0.06 ± 0.60 | -0.57 ± 0.77 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | 21.6 ± 6.9 | 10.4 ± 5.4 | 2.1 ± 4.5 | 3.13 ± 0.30 | -0.14 ± 0.54 | 0.54 ± 0.68 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | 20.6 ± 6.8 | 9.9 ± 5.4 | 0.1 ± 4.4 | 3.09 ± 0.30 | -0.25 ± 0.52 | -0.34 ± 0.65 |

**Table A.1 Average values and standard deviations for C3G4 base pair step geometries obtained from 555ns MD simulation.**

| G4-A5 | twist (deg) | roll (deg) | tilt (deg) | rise (Å) | slide (Å) | shift (Å) |
|---|---|---|---|---|---|---|
| unmodified | 34.8 ± 6.3 | 2.7 ± 5.3 | -0.3 ± 4.6 | 3.34 ± 0.3 | 0.23 ± 0.66 | -0.46 ± 0.71 |
| $^{oxo}$G4 | 36.6 ± 6.5 | 2.0 ± 5.4 | -1.4 ± 4.5 | 3.31 ± 0.3 | 0.49 ± 0.60 | -0.52 ± 0.70 |
| $^{oxo}$G10 | 38.5 ± 5.5 | 2.6 ± 5.6 | 0.3 ± 4.4 | 3.38 ± 0.29 | 0.19 ± 0.61 | -0.24 ± 0.65 |
| $^{m}$C3 | 35.0 ± 6 | 2.2 ± 5.3 | -1.0 ± 4.5 | 3.30 ± 0.30 | 0.43 ± 0.58 | -0.65 ± 0.62 |
| $^{m}$C9 | 36.5 ± 5.4 | 1.0 ± 4.7 | -0.8 ± 3.9 | 3.34 ± 0.28 | 0.06 ± 0.58 | -0.50 ± 0.64 |
| $^{m}$C3/$^{m}$C9 | 37.6 ± 4.9 | 0.5 ± 4.7 | -1.1 ± 3.8 | 3.35 ± 0.27 | 0.18 ± 0.52 | -0.66 ± 0.56 |
| $^{m}$C3/$^{oxo}$G10 | 39.4 ± 4.8 | 2.2 ± 5.3 | 0.1 ± 4.3 | 3.39 ± 0.28 | 0.37 ± 0.54 | -0.47 ± 0.56 |
| $^{m}$C9/$^{oxo}$G4 | 39.3 ± 4.7 | -0.2 ± 4.7 | -1.6 ± 3.8 | 3.35 ± 0.28 | 0.26 ± 0.48 | -0.70 ± 0.57 |
| $^{m}$C3/$^{oxo}$G4 | 36.3 ± 6.2 | 2.1 ± 5.4 | -1.3 ± 4.4 | 3.33 ± 0.30 | 0.45 ± 0.57 | -0.52 ± 0.71 |
| $^{m}$C9/$^{oxo}$G10 | 37.3 ± 5.2 | 0.7 ± 4.7 | -0.8 ± 3.9 | 3.36 ± 0.28 | 0.06 ± 0.56 | -0.48 ± 0.60 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | 38.8 ± 4.8 | -0.1 ± 4.7 | -1.3 ± 3.8 | 3.37 ± 0.27 | 0.25 ± 0.46 | -0.71 ± 0.56 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | 37.8 ± 4.8 | 0.9 ± 4.9 | -1.2 ± 3.8 | 3.35 ± 0.27 | 0.15 ± 0.52 | -0.59 ± 0.57 |

**Table A.2 Average values and standard deviations for G4A5 base pair step geometries obtained from 555ns MD simulation.**

| C9-G10 | twist (deg) | roll (deg) | tilt (deg) | rise (Å) | slide (Å) | shift (Å) |
|---|---|---|---|---|---|---|
| unmodified | 34.5 ± 6.9 | 2.5 ± 6.3 | -2.6 ± 5.8 | 3.29 ± 0.30 | 0.39 ± 0.56 | -0.35 ± 0.99 |
| $^{oxo}$G4 | 28.8 ± 7 | 1.6 ± 6.3 | -4.4 ± 5 | 3.26 ± 0.31 | 0.21 ± 0.62 | -1.04 ± 0.69 |
| $^{oxo}$G10 | 25.3 ± 8.1 | 2.4 ± 5.8 | 0.2 ± 5 | 3.16 ± 0.33 | -0.07 ± 0.64 | 0.42 ± 0.80 |
| $^{m}$C3 | 31.2 ± 7.3 | 2.3 ± 6.4 | -3.7 ± 5.1 | 3.27 ± 0.30 | 0.33 ± 0.58 | -0.69 ± 0.85 |
| $^{m}$C9 | 29.4 ± 8.2 | 0.8 ± 6.1 | -0.4 ± 5.2 | 3.22 ± 0.31 | 0.13 ± 0.57 | 0.14 ± 0.91 |
| $^{m}$C3/$^{m}$C9 | 24.2 ± 7.2 | 0.6 ± 5.8 | -1.1 ± 4.5 | 3.14 ± 0.29 | -0.07 ± 0.52 | -0.04 ± 0.74 |
| $^{m}$C3/$^{oxo}$G10 | 20.4 ± 6.4 | 2.4 ± 5.4 | -0.6 ± 4.5 | 3.10 ± 0.30 | -0.33 ± 0.52 | 0.33 ± 0.61 |
| $^{m}$C9/$^{oxo}$G4 | 22.1 ± 6.8 | 0.1 ± 5.6 | -1.8 ± 4.6 | 3.15 ± 0.31 | -0.24 ± 0.54 | -0.54 ± 0.64 |
| $^{m}$C3/$^{oxo}$G4 | 28.8 ± 7.2 | 2.0 ± 6.2 | -4.3 ± 4.9 | 3.26 ± 0.31 | 0.23 ± 0.60 | -1.00 ± 0.76 |
| $^{m}$C9/$^{oxo}$G10 | 27.6 ± 7.8 | 1.0 ± 5.9 | 1.0 ± 4.9 | 3.20 ± 0.31 | -0.09 ± 0.61 | 0.65 ± 0.80 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | 24.1 ± 8.1 | 0.3 ± 5.5 | -2.3 ± 4.5 | 3.14 ± 0.30 | -0.12 ± 0.61 | -0.33 ± 0.74 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | 19.7 ± 6.1 | 1.4 ± 5.3 | -0.1 ± 4.3 | 3.09 ± 0.29 | -0.28 ± 0.49 | 0.39 ± 0.61 |

**Table A.3 Average values and standard deviations for C9G10 base pair step geometries obtained from 555ns MD simulation.**

| G10-C11 | twist (deg) | roll (deg) | tilt (deg) | rise (Å) | slide (Å) | shift (Å) |
|---|---|---|---|---|---|---|
| unmodified | 33.3 ± 6.3 | 5.4 ± 5.4 | 0.3 ± 4.5 | 3.24 ± 0.29 | -0.05 ± 0.52 | 0.00 ± 0.70 |
| $^{oxo}$G4 | 36.9 ± 5.1 | 7.1 ± 5.4 | 0.2 ± 4.2 | 3.32 ± 0.28 | -0.13 ± 0.55 | 0.13 ± 0.61 |
| $^{oxo}$G10 | 36.7 ± 6.4 | 7.5 ± 5.8 | -0.4 ± 4.5 | 3.30 ± 0.31 | 0.10 ± 0.52 | -0.17 ± 0.69 |
| $^{m}$C3 | 34.6 ± 5.1 | 7.3 ± 4.8 | -0.9 ± 4 | 3.29 ± 0.28 | -0.32 ± 0.48 | -0.18 ± 0.6 |
| $^{m}$C9 | 34.3 ± 6.1 | 7.1 ± 5.5 | -0.5 ± 4.5 | 3.26 ± 0.29 | 0.03 ± 0.51 | -0.27 ± 0.66 |
| $^{m}$C3/$^{m}$C9 | 35.5 ± 4.7 | 10.0 ± 5.1 | -1.4 ± 3.9 | 3.31 ± 0.27 | -0.18 ± 0.47 | -0.43 ± 0.59 |
| $^{m}$C3/$^{oxo}$G10 | 37.1 ± 5 | 9.4 ± 5.2 | -1.3 ± 4 | 3.32 ± 0.27 | -0.08 ± 0.45 | -0.45 ± 0.61 |
| $^{m}$C9/$^{oxo}$G4 | 37.8 ± 4.6 | 10.0 ± 5.5 | -0.2 ± 4.2 | 3.33 ± 0.27 | 0.03 ± 0.55 | -0.07 ± 0.60 |
| $^{m}$C3/$^{oxo}$G4 | 34.8 ± 4.9 | 8.0 ± 4.8 | -1.1 ± 3.9 | 3.29 ± 0.27 | -0.31 ± 0.47 | -0.12 ± 0.58 |
| $^{m}$C9/$^{oxo}$G10 | 35.4 ± 5.9 | 7.4 ± 5.8 | 1.1 ± 4.4 | 3.26 ± 0.29 | 0.36 ± 0.49 | 0.19 ± 0.63 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G4 | 34.2 ± 5.3 | 10.3 ± 5.1 | -1.1 ± 4 | 3.29 ± 0.27 | -0.17 ± 0.48 | -0.42 ± 0.62 |
| $^{m}$C3/$^{m}$C9/$^{oxo}$G10 | 37.7 ± 4.7 | 9.9 ± 5.1 | -1.5 ± 3.9 | 3.34 ± 0.27 | -0.05 ± 0.44 | -0.5 ± 0.58 |

**Table A.1 Average values and standard deviations for G10C11 base pair step geometries obtained from 555ns MD simulation.**

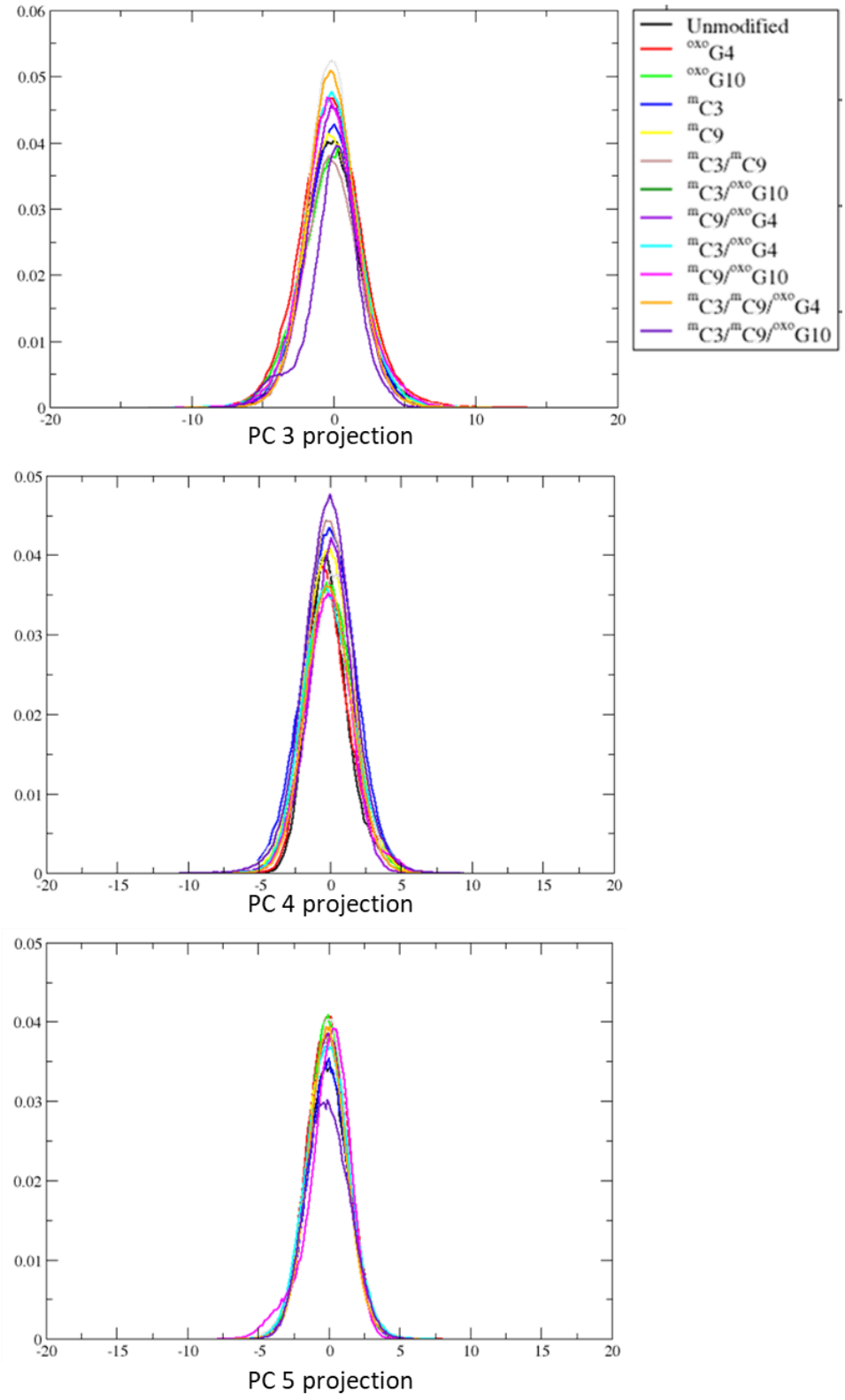



**Figure A5. Principal component projections 3-5 of target CpG sites from MD simulations.**

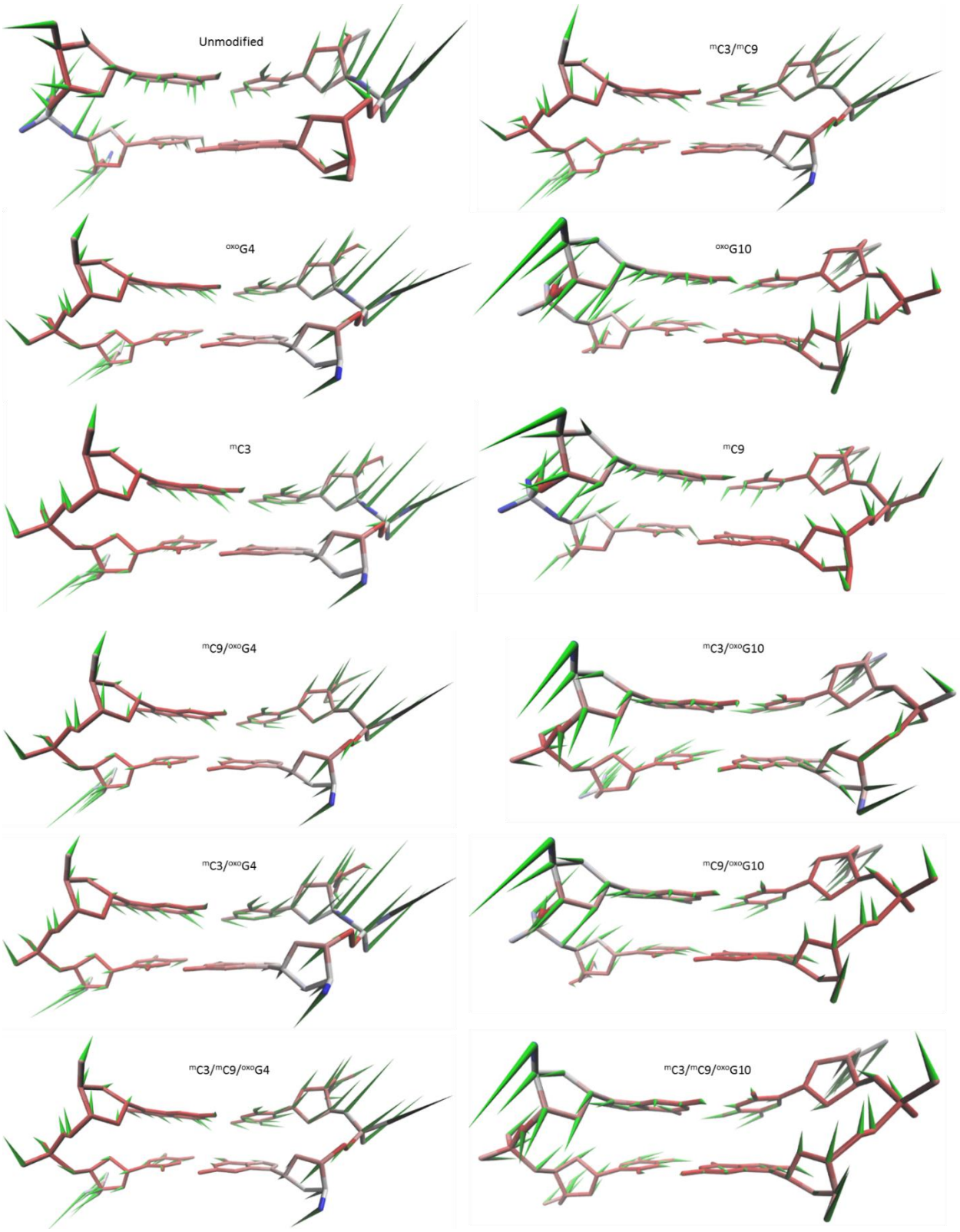



**Figure A6. Porcupine plots for CpG sites of the first eigenvectors for the MD simulation on the all twelve samples. The arrows attached to each heavy atom indicates the direction of the**

**eigenvector and the arrow length depicts the magnitude of the associated eigenvalue. The models are colored according to mobility where red is least mobile and blue is highest mobility.**