Summer 2016

Comprehensive Solution Structure Analysis of DNA Duplexes Containing Chemical Base Alterations

Joanna Hoppins
Western Washington University, hoppinjj@gmail.com

COMPREHENSIVE SOLUTION STRUCTURE ANALYSIS OF DNA DUPLEXES CONTAINING CHEMICAL BASE ALTERATIONS

By

Joanna Hoppins

Accepted in Partial Completion
of the Requirements for the Degree
Master of Science

Kathleen L. Kitto, Dean of the Graduate School

ADVISORY COMMITTEE

Chair, Dr. Sergey L. Smirnov

Dr. P. Clint Spiegel

Dr. John Antos

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.

Joanna Hoppins
July 19th, 2016

COMPREHENSIVE SOLUTION STRUCTURE ANALYSIS OF DNA DUPLEXES CONTAINING CHEMICAL BASE ALTERATIONS

A Thesis
Presented to
The Faculty of Western Washington University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

by
Joanna Hoppins
July 2016

Abstract

Most types of chemical modifications of DNA bases are endogenous processes which are sensitive to the intracellular conditions. For example, the enzymatically catalyzed methylation of canonical cytosine to 5-methylcytosine (5metC) is a key form of epigenetic regulation of gene expression patterns. This modification is catalyzed and controlled by DNA methyltransferases (DNMTs). Perturbed rates of enzymatic DNA methylation leads to hyper- or hypo-methylation, both of which are a common initiating step in several forms of cancer. In addition to epigenetics, DNA bases can be chemically altered, or damaged, in response to reactive oxygen species (ROS). The oxidation of DNA by ROS can lead to formation of various types of adducts, with 8-oxoguanine (oxoG) being one of the most prolific and toxic. oxoG is a deleterious modification which has been linked to cancer and neurological disorders. The first step in oxoG damage repair, oxoG glycosylase (hOGG1) recognizes and excises the oxidized base. Oxidation of guanine and methylation of cytosine can occur simultaneously within CpG sites. Moreover, in these sites the both enzymology of oxoG repair is compromised by the adjacent 5metC, and the enzymatic methylation of cytosine is altered by oxoG. This manuscript describes the structural study of the DNA substrates where oxoG and 5metC modifications are clustered in a CpG site to aid in the understanding of the enzymatic effects of such clustering. The NMR solution structure is shown of six of these proposed duplex DNA samples, two with a single oxidation, two with a single methylation, one sample with a fully methylated CpG site and a modified sample with both oxoG and 5metC occurring on opposing strands in one single base pair. No global structural changes are reported amongst these structures, with all of these structures featuring elements of right-handed A/B DNA. One local structural change was observed in all samples with oxoG, namely that oxoG causes the BII backbone conformation 3' of the modification site. This BII backbone conformation may be the link between these modifications and lower enzymatic activity, as both hOGG1 and DNMT1 make direct contact with the 3' backbone following an oxoG.

Acknowledgments

I'd like to thank Dr. Serge Smirnov for taking me into his lab the Fall of 2012, and being available and engaged in all my work since day one. This lab has become my home, and I find it very difficult to leave. I'd also like to thank my committee members, Dr. Clint Spiegel and Dr. John Antos. Dr. Spiegel very generously allowed the shared use of both space and equipment, without which this thesis would not have been possible, as well as being available to answer any panicked questions. Dr. Antos has my thanks for allowing the shared use of the lyophilizer, as well as my thanks for being an excellent Graduate School advisor this past year. I'd like to thank all my committee members for reading and editing this thesis.

In addition, I'd like to thank my past and current lab members. The structures reported in this thesis were a team effort, and without Heather Miears and David Gruber we would not have so many of them to report today. The past lab members, Stas Fedechkin and Jacob Brockerman paved the way by determining appropriate NMR experiments and pulse sequences to apply and analyze for the NMR solution structures.

Table of Contents

Abstract iv

Acknowledgments iv

List of Figures and Tables x

List of Abbreviations xiii

Chapter 1 - Introduction 1

DNA structure 1

Nucleobase structure 1

Base pairing 2

Tertiary structure 2

Backbone and sugar pucker 4

DNA in biological systems 5

Replication 5

Transcription and translation 6

Epigenetics 6

DNA methyltransferases 7

DNMT mechanisms 9

DNA damage and repair 10

8-oxoguanine glycosylases 11

DNA modifications and CpG sites 13

Presence of oxoG affects CpG methylation by DNMTs 13

5metC slows 8oxoG excision by hOGG1 14

Possible consequences of oxoG and 5metC in a CpG site 15

Identifying a target DNA substrate for 8oxoG-5metC investigations 15

Published DNA oligomers of relevance 17

References 20

Chapter 2 - Drew-Dickerson DNA solution structure determination and analysis 24

Sample Preparation 24

Sample origins 24

In-house pre-NMR preparations 24

NMR recording 26

On buffer conditions 26

1D experiments 28

Homonuclear ^{1}H 2D NMR experiments 28

Heteronuclear 2D experiments: ^{31}P – ^{1}H HETCOR 36

Restrained Molecular Dynamics 37

NOESY-derived distance restraints 38

Hydrogen bonding restraints 39

Torsion angle restraints 39

Starting structures 40

Simulated annealing protocol 40

Determination of the representative ensemble 41

Structure Analysis 41

References 42

Chapter 3 – Results Part I: Reported structures 44

Control samples 45

Sample 2 - CGC(oxoG)AATTCGCG 45

Sample 3 - CGCGAATTC(oxoG)CG 48

Sample 4 - CG(5metC)GAATTCGCG 51

Sample 5 - CGCGAATT(5metC)GCG 53

Sample 6 - CG(5metC)AATT(5metC)GCG 55

Target samples 57

Sample 8 – CGC(oxoG)AATT(5metC)GCG 57

References 60

Chapter 4 – Results Part II: Comparisons 61

Comparisons between reported structures 61

Oxidized control samples 61

Methylated control samples 63

Target sample 64

Comparison to published structures 66

1BNA – 1.9 Å resolution crystal structure of the Drew-Dickerson sequence 68
2DAU – Solution NMR structure of the Drew-Dickerson sequence 70
1NAJ – Liquid crystalline NMR Drew-Dickerson sequence 72
183D – CCA(8oxoG)CGCTGG containing crystal structure with 1.6 Å resolution 73
4MKW – Crystal structure at 1.22 Å resolution of CG(5metC)GAATTCGCG 74
4C63 – Crystal structure at 1.32 Å resolution of CGCGAATT(5metC)GCG 76
Internal Quality Control Comparisons 77
Chemical shift comparison 77
Restraint comparison 78
References 79
Chapter 5 – Conclusions and Future Work 80
oxoG and 5metC modifications do not dramatically alter the structure of the DDD 80
BII may play a role in OGG1 recognition of 8oxoG 80
BII may slow the addition of 5metC by mDNMT1 82
Future work 83
Target sample structures 83
Imino proton dynamics 83
Enzymatic studies 84
References 85
Appendix 86

List of Figures and Tables

Figure 1.1. Two nucleotides, cytosine and guanine, connected through the 3' of cytosine and the 5' phosphate of guanine.

Figure 1.2. Hydrogen bonding between canonical base pairs.

Figure 1.3. Representations of B-, A- and Z-DNA.

Figure 1.4. Base pair parameters in their respective categories.

Figure 1.5. The major DNA torsion angles.

Figure 1.6. 5-methylcytosine.

Figure 1.7. Murine DNMT1 with bound hemi-methylated DNA demonstrating the target cytosine in the catalytic cleft near the AdoMet cofactor.

Figure 1.9. oxoG base pairing.

Figure 1.10. hOGG1 flips both oxoG and guanine into the binding site extra-helically.

Figure 1.11. The CpG target sample scheme.

Figure 2.1. The 1D ^{1}H NMR spectrum of DD03 in H_2O.

Figure 2.2. The full H_2O NOESY for DD03.

Figure 2.3. A portion of an H_2O NOESY for DD03.

Figure 2.4. The full 260 ms D_2O NOESY for DD06.

Figure 2.5. The connections of base protons to sugar protons termed the “walk”.

Figure 2.6. The “walk region” of the 260 ms D_2O NOESY for DD06.

Figure 2.7. The full 120 ms TOCSY for DD06.

Figure 2.8. The DQFCOSY for DD06

Figure 2.9. The HETCOR for DD03.

Figure 3.1. The imino region of the H_2O NOESY for DD02

Figure 3.2. The H3' to P region of the ^{1}H-^{31}P HETCOR for DD02.

Appendix Table 1. Chemical shifts from DD02 D_2O spectra taken at 25°C.

Appendix Table 2. Helical parameters for DD02.

Appendix Table 3. Torsion angles for DD02.

Figure 3.3. The representative ensemble for DD02.

Figure 3.4. Three base pairs of the DD02 structure.

Figure 3.5. The DD02 BII backbone conformation.

Figure 3.6. The imino region of the H_2O NOESY for DD03.

Figure 3.7. The H3' to P region of the ^{1}H-^{31}P HETCOR for DD03.

Appendix Table 4. Chemical shifts from DD03 D_2O spectra taken at 25°C.

Appendix Table 5. Helical parameters for DD03.

Appendix Table 6. Torsion angles for DD03.

Figure 3.8. The representative ensemble for DD03.

Figure 3.9. Three base pairs of the DD03 structure.

Figure 3.10. The DD03 BII backbone conformation.

Figure 3.11. The imino region of the H_2O NOESY for DD04.

Appendix Table 7. Chemical shifts from DD04 D_2O spectra taken at 25°C.

Appendix Table 8. Helical parameters for DD04.

Appendix Table 9. Torsion angles for DD04.

Figure 3.12. The representative ensemble for DD04.

Figure 3.13. Three base pairs of the DD04 structure.

Figure 3.14. The imino region of the H_2O NOESY for DD05.

Appendix Table 10. Chemical shifts from DD05 D_2O spectra taken at 25°C.

Appendix Table 11. Helical parameters for DD05.

Appendix Table 12. Torsion angles for DD05.

Figure 3.15. The representative ensemble for DD05.

Figure 3.16. Three base pairs of the DD05 structure.

Figure 3.17. The imino region of the H_2O NOESY for DD06.

Appendix Table 13. Chemical shifts from DD06 D_2O spectra taken at 25°C.

Appendix Table 14. Helical parameters for DD06.

Appendix Table 15. Torsion angles for DD06.

Figure 3.18. The representative ensemble for DD06.

Figure 3.19. Three base pairs of the DD06 structure.

Figure 3.20. The H3' to P region of the ^{1}H-^{31}P HETCOR for DD08.

Appendix Table 16. Chemical shifts from DD08 D_2O spectra taken at 25°C.

Appendix Table 17. Helical parameters for DD08.

Appendix Table 18. Torsion angles for DD08.

Figure 3.21. The representative ensemble for DD08.

Figure 3.22. Three base pairs of the DD08 structure.

Figure 3.23. The DD08 BII backbone conformation.

Figure 4.1. The ε-ζ, slide, shift and twist values for DD02 and DD03.

Figure 4.2. The ε-ζ, buckle, and twist values for DD04, DD05 and DD06.

Figure 4.3. The ε-ζ, buckle, roll, twist, shift, opening, slide, and propeller values for DD02, DD05 and DD08.

Figure 4.4. The difference in buckle of the base pair underneath the modification sites in DD08, DD02 and DD05.

Figure 4.5. The propeller and twist values for structures of the Drew-Dickerson sequence.

Figure 4.6. The stretch values for structures with the Drew-Dickerson sequence.

Figure 4.7. The hydrogen bonding distances of the G_2:C_{23} base pair in the Drew-Dickerson sequence for 1BNA, 1NAJ and DD02.

Figure 4.8. The ε–ζ values for structures with the Drew-Dickerson sequence.

Figure 4.9. ^{31}P 1D spectra for DD03, DD04 and DD06.

Figure 4.10. The ε–ζ values for 183D, DD02 and DD03.

Figure 4.11. The opening, buckle and slide values for 4MKW and DD04.

Figure 4.12. The ε–ζ values for 4MKW, 4C63, DD04, DD05, and DD06.

Figure 4.13. The opening, buckle and slide values for 4C63 and DD05.

Figure 4.14. The "walk" region of DD02, DD05 and DD08 changes only near the modification sites.

Figure 4.15. A comparison of NOE-derived distance restraints used in the solution structure calculations.

Figure 5.1. The direct contact between the 3' backbones of canonical guanine to the catalytic residue K249 of hOGG1 from the PDB ID structure of IYQK and the lack of contact between K249 of hOGG1 and the backbone 3' of oxoG in PDB ID 2NOZ.

Figure 5.2. The contacts between murine DNMT1 and four base pairs of the target DNA substrate.

Figure 5.3. The direct contact of murine DNMT1 to the phosphate 3' of the *trans* guanine that base pairs to the 5metC present in hemi-methylated DNA from PDB ID 4DA4.

List of Abbreviations

DNA – Deoxyribonucleic acid

RNA – ribonucleic acid

C – Cytosine

G- Guanine

T – Thymine

A – Adenine

mRNA – messenger RNA

5metC – 5-methylcytosine

DNMT – DNA methyltransferase

CpG site – cytosine phosphodiester linked to guanine

ROS – reactive oxygen species

oxoG- 8-oxoguanine

OGG1 – 8-oxoguanine glycosylase 1

hOGG1 – the human 8-oxoguanine glycosylase 1

DDD- Drew-Dickerson dodecamer

DD## - shorthand notion of Drew-Dickerson sequence modification number (##) according to the sample scheme provided in this thesis

NMR – Nuclear magnetic resonance spectroscopy

EDTA – ethylenediaminetetraacetic acid

FPLC – fast pace liquid chromatography

1D – one dimensional

2D- two dimensional

NOE – nuclear Overhauser effect

NOESY – nuclear Overhauser effect spectroscopy

TOCSY – total correlation spectroscopy

DQFCOSY – double quantum filtered correlation spectroscopy

HETCOR – heteronuclear correlation

rMD – restrained molecular dynamics

RMSD – root mean square deviation

1BNA - 1.9 Å resolution crystal structure of the Drew-Dickerson sequence

2DAU – Solution NMR structure of the Drew-Dickerson sequence

1NAJ – Liquid crystalline NMR Drew-Dickerson sequence

183D – CCA(8oxoG)CGCTGG containing crystal structure with 1.6 Å resolution

4MKW – Crystal structure at 1.22 Å resolution of CG(5metC)GAATTCGCG

4C63 – Crystal structure at 1.32 Å resolution of CGCGAATT(5metC)GCG

Chapter 1 - Introduction

Deoxyribonucleic acid (DNA), along with closely related ribonucleic acid (RNA), is the foundation for the current understanding of life at the molecular level. These two macromolecules fall under the broader category of nucleic acids. At the most basic level, DNA has a major role in carrying genetic information, reproduction, cell differentiation and evolution.

DNA structure

Nucleobase structure

DNA is a biopolymer with four main repeating units, which connect through phosphodiester linkages into strands. Each nucleic acid unit consists of a 2-deoxyribose pentose sugar that is attached to the interchangeable base at C1. DNA differs from RNA in that DNA is deoxygenated at C2 of the sugar, while RNA has the full ribose.

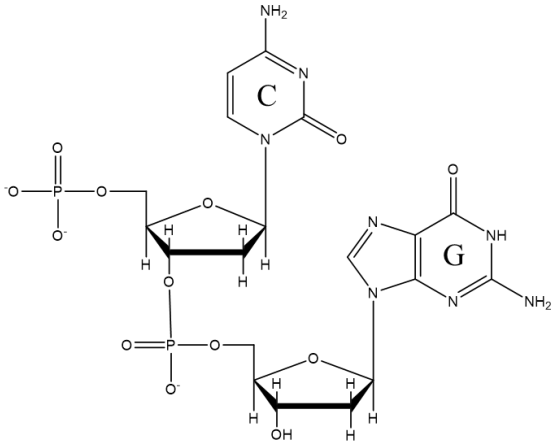


Figure 1.1. Two nucleotides, cytosine (C) and guanine (G), connected through the 3' of cytosine and the 5' phosphate of guanine. The bases are shown with their representative one letter code.

Through a dehydration reaction where the 3' hydroxyl group of a nucleotide performs a nucleophilic attack upon the 5' phosphate of the next unit, the nucleotides polymerize to form a single strand consisting of a combination of linear units with a directional backbone **(Figure 1.1)**. Cytosine (C), adenine (A), thymine (T) and guanine (G) bases each individually form a glycosidic bond with the unchanging sugar to complete the nucleobase unit. In solution and living organisms, DNA forms an antiparallel double stranded structure. The strands are originally associated due to the highly favorable hydrogen bonding of the bases.

Base pairing

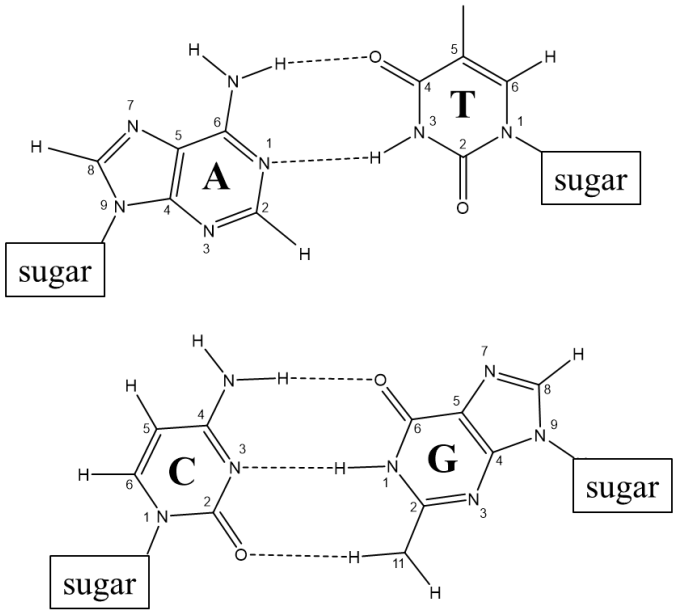


Figure 1.2. Hydrogen bonding between canonical base pairs shown with corresponding one letter codes.

The four canonical bases fall into two distinct categories, purines and pyrimidines. The purines, so named because of the fused 5- and 6-membered heterocyclic rings in the base, are adenine and guanine. The pyrimidine bases, cytosine and thymine, contain one single heterocyclic 6-membered ring. Typical Watson-Crick base pairing occurs between one purine and one pyrimidine: cytosine and guanine come together to form a CG pair, while adenine and thymine form an AT pair **(Figure 1.2)** [1]. AT pairs contain two hydrogen bonds, between adenine H6[1] to thymine O4 and adenine N1 to thymine H3. GC base pairs consist of three hydrogen bonds: cytosine H4[1] to guanine O6, cytosine N3 to guanine H1 and cytosine O2 to guanine H2[1]. Due to keto forms being much more likely to exist in solution than the enol form of the molecule, one DNA base has only one possible base pairing partner, resulting in high fidelity base pair matching.

Tertiary structure

Once hydrogen bonding has occurred, the anti-parallel strands twist into helices to create pi-stacking interactions between bases and stabilize the macromolecular structure. The three dimensional structure contains a set of distinct features found in all double stranded DNA **(Figure 1.3a)**. Twisting of strands can happen in either direction, leading to a structure with distinct

handedness. The two strands form two distinct faces, called grooves, which track down the helix. These two grooves are referred to as the "major" and "minor" grooves because of their appearance in B-DNA [2]. The structure of B-DNA, first discovered by Watson and Crick [1], is the canonical form of DNA in solution and is thought to be the main form of DNA *in vivo* **(Figure 1.3b)**. The other commonly found form of DNA, A-DNA, differs from B-DNA mainly by how the backbone twists around the central axis [3]. The more compact A-DNA has a higher number of base pairs per full turn (pitch), less contrasting groove widths and a greater diameter [4]. Both A- and B-DNA have right-handed helices. Left-handed Z-DNA also exists [5], but it is far less common than either the A- or B-form.

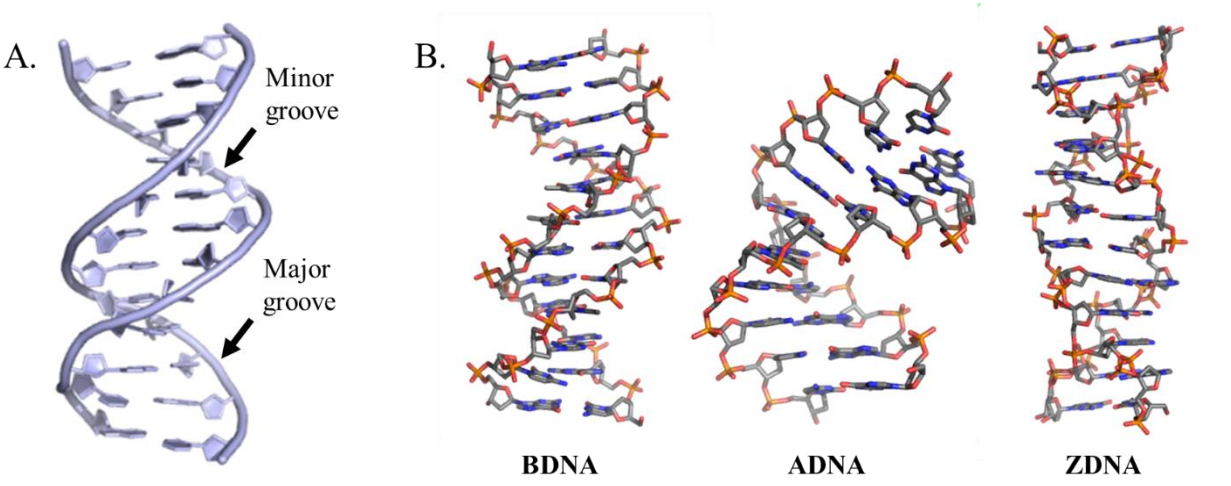


Figure 1.3. A. A ribbon representation of the crystal structure of B-DNA with PDB ID 1BNA [2]. Arrows highlight the locations of the major and minor grooves. **B.** Stick models of short DNA segments with PDB IDs 1BNA [2], 4IZQ [4] and 4OCB [5], showing the differences within B-, A- and Z-form DNA.

Besides the more global descriptors such as pitch, handedness and groove width, double stranded DNA can be characterized by the position and interaction of the bases **(Figure 1.4)**. The base interactions fall into three categories: base pair parameters, step parameters and helical parameters [6]. Base pair parameters describe the interactions within a single complementary base pair, e.g. between the C and G on opposing strands in a CG base pair. Step parameters treat each individual

complementary base pair as one unit (calling the entire base pair one single step), and describe its interaction with the next base pair step on the 5′ end of strand I and the 3′ end of strand II. Similar to step parameters are helical parameters, which consider a more local (typically two step) environment to calculate the central axis instead of a more global axis definition [6]. These parameters provide an entirely new set of structural descriptions that is unique to the DNA duplex, allowing for more in depth comparison and analysis of DNA structures.

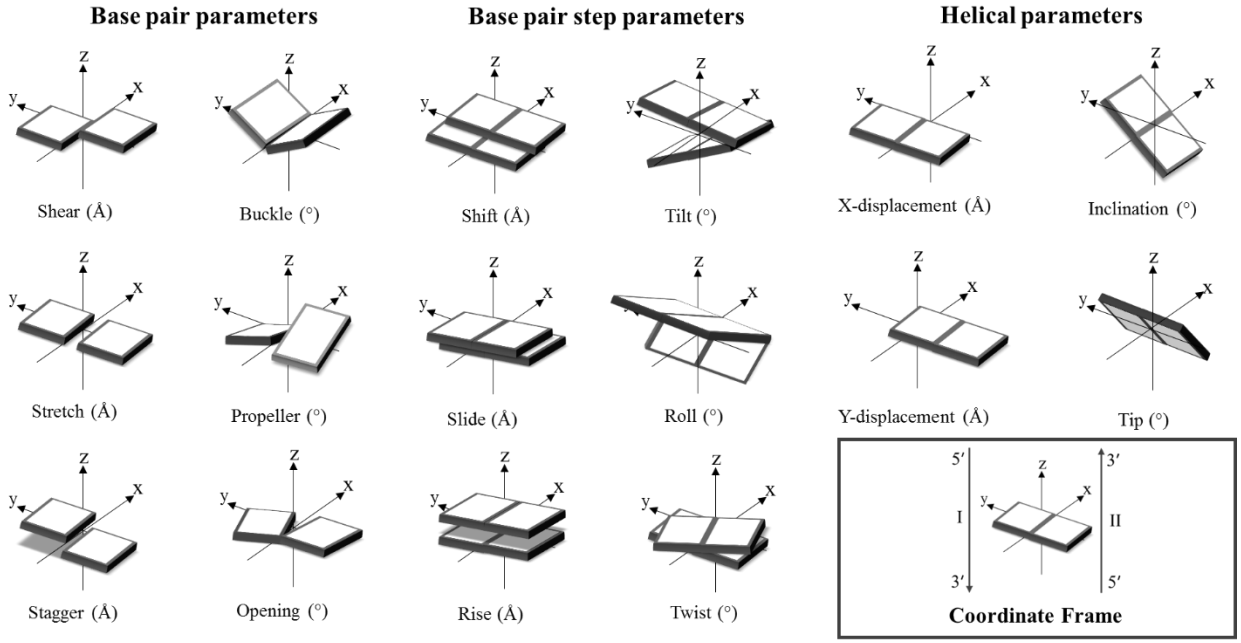


Figure 1.4. Base pair parameters in their respective categories: base pair, base pair step and helical parameters. Shown with view into the major groove, with the 3′ to 5′ strand on the right and the 5′ to 3′ strand on the left of each individual axis. [6]

Backbone and sugar pucker

Like any complex biomolecule, DNA is composed of a set of dihedral angles which describe the local geometry of the four atoms [7], [8]. In total, there are 6 main chain angles which describe the backbone, α, β, γ, δ, ε, and ζ **(Figure 1.5)**. The glycosidic bond, the bond between the sugar and

the base, has a corresponding angle known as χ. The sugar contains many small angles known as τ_0, τ_1, τ_2, τ_3, τ_4 which culminate in an overall sugar ring pucker. For the sugar ring, overall pucker is favored as a descriptor and torsion angles are not typically reported. In a similar vein, certain backbone angles, though still often used, correspond to a larger set of backbone conformations. Notably, two distinct populations of differing ε, and ζ values exist within B-DNA, denoting two main B-form backbone conformations: the BI (ε - ζ < 20°) and the less common BII (ε - ζ > 20°) [3].

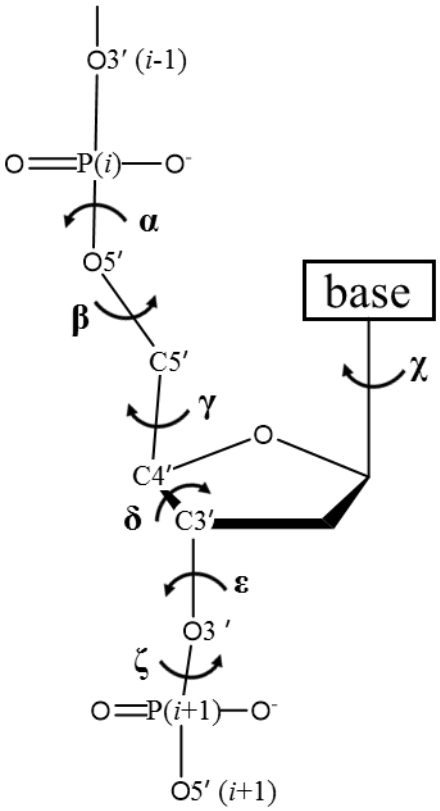


Figure 1.5. The major DNA torsion angles.

DNA in biological systems

Replication

After Watson and Crick's discovery of the DNA duplex [1], it was clear that intrinsic to DNA structure is a mechanism for replication, a mandatory step in cell division. The high-fidelity base pairing ensures that if one strand is present, the complementary strand can be replicated in a process called semiconservative replication. One duplex can be dissociated into two strands, each of which is used as a template to replicate the complementary strand. One complete duplex can thus be replicated to two identical duplexes, with one original strand and one daughter strand per duplex. Hydrogen bonding, as a method of associating the complimentary strands, is flexible: strong enough for duplex formation under most conditions, but weak enough to enable enzyme-mediated strand dissociation. Thus, strand dissociation can be accomplished for the purpose of DNA replication.

Transcription and translation

The oft-repeated central dogma of biology highlights the main components of cellular life: DNA is transcribed to RNA that is then translated to proteins. In this central dogma, DNA as a genetic code is of importance for the information it carries for protein production. The intermediate step between the DNA code and protein production, transcription, involves RNA polymerases which read the DNA 3′ to 5′ and create a complementary messenger RNA (mRNA) strand 5′ to 3′ [9]. The nascent mRNA undergoes additional processing before it is ready for translation. Translation, done by the ribosome, produces an amino acid chain based on a three letter mRNA code [10]. This amino acid chain folds, and may experience additional modification before producing the final protein.

Epigenetics

The pure nucleotide sequence alone does not solely influence gene expression [11]. In response to environmental factors, key changes can be created in the DNA independent of sequence. The field of epigenetics strives to explain how the environment effects gene expression, with particular emphasis on the heritability of such responses [12].

Recent developments in epigenetics have shown that there are many ways to achieve epigenetic gene regulation. Two of the main mechanisms are histone modification and the methylation of the cytosine base at C5. Histone modifications effect chromatin packing [13]. Chemical modifications of a histone, often methylation or acetylation of key lysines, change histone interactions. The resulting modified histones are more or less likely to associate, depending on the type and placement of the modification. When the histones are packed tightly in higher order chromatin packing, the DNA wrapped around them is incapable of performing interactions with necessary

transcriptional proteins. When the opposite phenomenon is observed, a modified histone in a tightly packed region can cause unwinding of DNA fibers, making the DNA more readily available for transcription. 5-methylcytosine or 5metC **(Figure 1.6)** works in an analogous way to histone packing. The modification of the cytosine can block the interaction of DNA and transcriptional regulators [14]. In addition, cytosine methylation has been linked to histone acetylation, and thus tighter histone packing [15]. Tight histone packing is associated with silencing genes, while histone dissociation generally enhances gene expression.

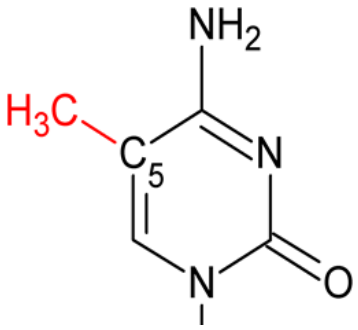


Figure 1.6. 5-methylcytosine. The modification for the cytosine base is shown in red. The attached sugar is not shown.

The most impactful features of epigenetics lie in the reversibility, heritability and speed at which it controls gene expression. Unlike changes in the base sequence, epigenetic marks are entirely reversible. The methods with which they are achieved are non-permanent and often require upkeep to maintain. Though reversible, many epigenetic modifications are also heritable; cytosine methylation has been shown to have stable heritability from eukaryotic parent to progeny [16]. This heritable modification can occur at rates much faster than traditional evolution, and the consequences of such a subtle method of gene regulation are still being investigated.

DNA methyltransferases

Cytosine methylation patterns are established, maintained and passed down through generations by several different DNA methyltransferases (DNMTs). While not all DNMTs specifically methylate cytosine at C5, two of the most important classes of mammalian DNMTs perform exactly that function. Most mammalian cytosine methylations occur in CpG sites, where cytosine is attached through the phosphodiester bond to a guanine on the same strand [17] **(Figure 1.1)**.

DNMT1 primarily maintains an already existing methylation pattern [18]. The optimum substrate for DNMT1 is on a newly synthesized or repaired strand of DNA which already contains a methylation in the CpG on the opposing strand, referred to as a hemi-methylated CpG site. In addition to the maintenance DNMT1, DNMT3s perform *de novo* methylation [19]. This is particularly important in the heritability of methylation patterns and parent origin allele-specific gene imprinting. A zygotic cell needs only to maintain currently existing methylation patterns that were passed from the parent cell and persist throughout the following somatic cell lines. However, all methylation patterns are completely removed in primordial germ cells. Primordial germ cells undergo meiosis and eventually form gametes. The methylation patterns are completely redeveloped in gametogenesis of both sexes and include a concert of complicated functions performed by DNMT3a, DNMT3b and DNMT3L. Improper function of DNMT1 can lead to hyper- or hypo-methylation which have both been identified as common cancer triggers [20]. DNMT3 malfunction typically leads to failed embryonic development, but has also been linked to developmental diseases [21], [22]. Correctly established and maintained cytosine methylation patterns through DNMTs are extremely important for gene expression, tumor suppression, differentiation, overall cell health and proper epigenetic response.

DNMT mechanisms

The enzymatic mechanism of human DNMTs has been under much recent scrutiny due to the role of DNMTs in both epigenetics and cancer. Enzymatic studies have focused on understanding the mechanism of DNMTs as a possible target for therapeutics and to gain further knowledge of enzyme-DNA interactions. The maintenance DNMT1 has been shown to be a processive enzyme, which rarely switches strands and does not often skip targets, proceeding from one target site to the next downstream target in 98% of cases [23]. From these observations, it has been proposed that DNMT1 slides along one strand of DNA until matched with the correct substrate for methylation. DNMT1 shows a 15 fold preference for hemimethylated CpG sites over blank CpG sites, but also has been shown to have a lesser *de novo* methylation preference for the CCGG sequence [24]. DNMT1 truncation studies have shown that it contains three separate DNA binding domains on its N-terminus [25]. Differing substrate preferences exist due to autoinhibition of DNMT1 by other portions of the N-terminal fragment. Upon meeting the preferred substrates, the

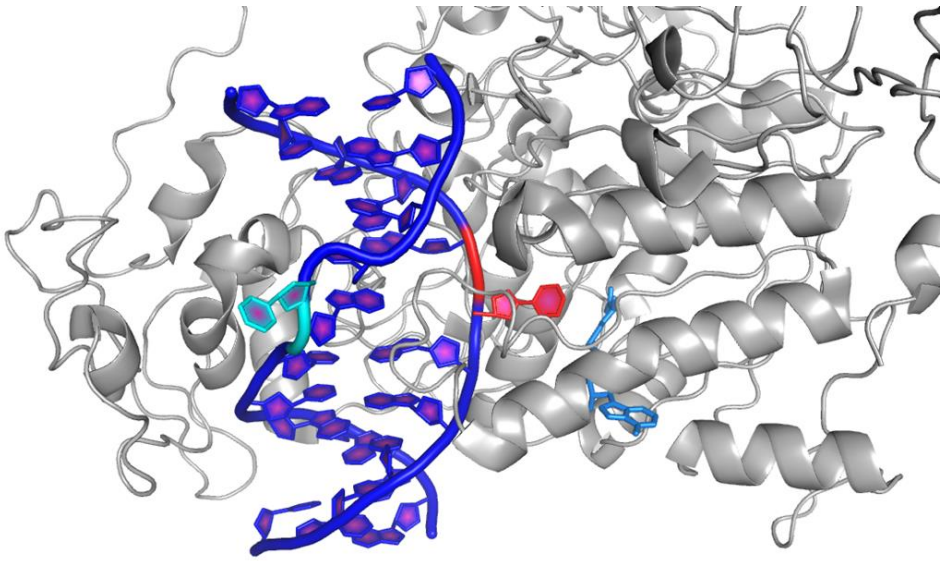


Figure 1.7. Murine DNMT1 with bound hemi-methylated DNA demonstrating the target cytosine in the catalytic cleft near the AdoMet cofactor [25]. DNMT1 is in grey, the AdoMet cofactor in marine blue, the DNA substrate is shown in blue with the target cytosine in red and the hemi-methylated 5metC in cyan. Both the target cytosine and pre-existing 5metC are extra helical, suggesting that this is the mechanism of the DNMT1 preference for hemi-methylated DNA.

C-terminal methyltransferase domain in DNMT1 performs a nucleophilic attack at C6 [18]. The double bond between C5 and C6 is saturated by this nucleophilic attack and leaves the position at C5 open for the addition of a methyl group. The methyl is provided by the common cofactor and methyl group source S-adenosyl methionine (known as AdoMet or SAM), and is added at C5. DNMT3a has been proposed to have a similar mechanism to DNMT1, though the N-terminal domain is not necessary for enzymatic activity [26]. While processive on its own, DNMT3a relies upon DNMT3L for processivity enhancement [27]. Evidence suggests that DNMT1 is localized near the replication fork, and the DNMT3 family is near the heterochromatin, which further solidifies their proposed roles in epigenetic regulation [28], [29].

DNA damage and repair

DNA damage, and subsequent repair, are another set of processes by which DNA is chemically modified. DNA damage is typically deleterious and requires regular removal to prevent mutations that could negatively impact cell or organism survival. Unlike DNMTs, where the chemical modification to the DNA was added on by the enzyme, enzymes involved in DNA damage repair are specialized towards the removal of a pre-existing modification. Exposure to mutagens, radiation, or just general metabolic byproducts can cause damaging chemical modifications. Common lesions created by those sources are double stranded breaks, pyrimidine dimers, depurination, methylation and oxidation. DNA damage repair mechanisms include: removal of the nucleotide (nucleotide excision repair), removal of the

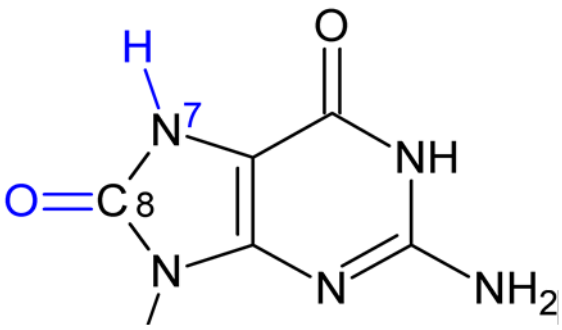


Figure 1.8. 8-oxoguanine. The modifications to the guanine base are shown in blue. The attached sugar is not shown.

affected base only (base excision repair or BER), non-homologous end joining and homologous recombination.

Without additional external stress such as radiation, the oxidative metabolic byproducts alone cause significant endogenous damage that is maintained at least in low levels at all times [30]. Though there are many differing ways for oxidative damage to affect DNA, the oxidation of guanine at the C8 with an additional hydrogen at the N7, known as 8-oxoguanine or oxoG, is among the most common and toxic **(Figure 1.8)** [31]. Left without repair, oxoG can cause GC → TA transversions [32]. Transversions occur because oxoG does not have the base pairing fidelity that canonical bases inherently contain. oxoG can rotate around the glycosidic bond into the *syn* conformation to form a Hoogsten base pair with adenine **(Figure 1.9)**. This oxoG:A mispairing can lead to replacement of guanine with thymine on the replicated strand.

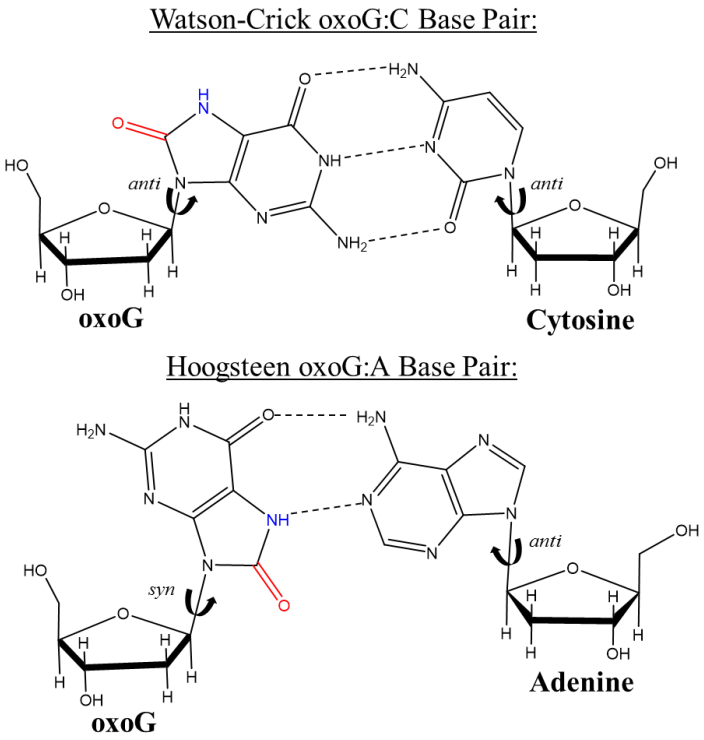


Figure 1.9. oxoG base pairing. oxoG can form a canonical Watson-Crick base pair with cytosine, or a Hoogsteen base pair with adenine. The oxoG chemical modifications are shown in red and blue, and the arrows indicate the glycosidic bond. In Watson-Crick base pairing, both nucleotide bases are *anti* to the sugar. In Hoogsteen base pairing the oxoG base is *syn* with respect to the sugar ring.

8-oxoguanine glycosylases

8-oxoguanine is repaired through BER, which is initiated by dedicated glycosylases that identify the oxidized guanine base and sever the glycosidic bond between the base lesion and the

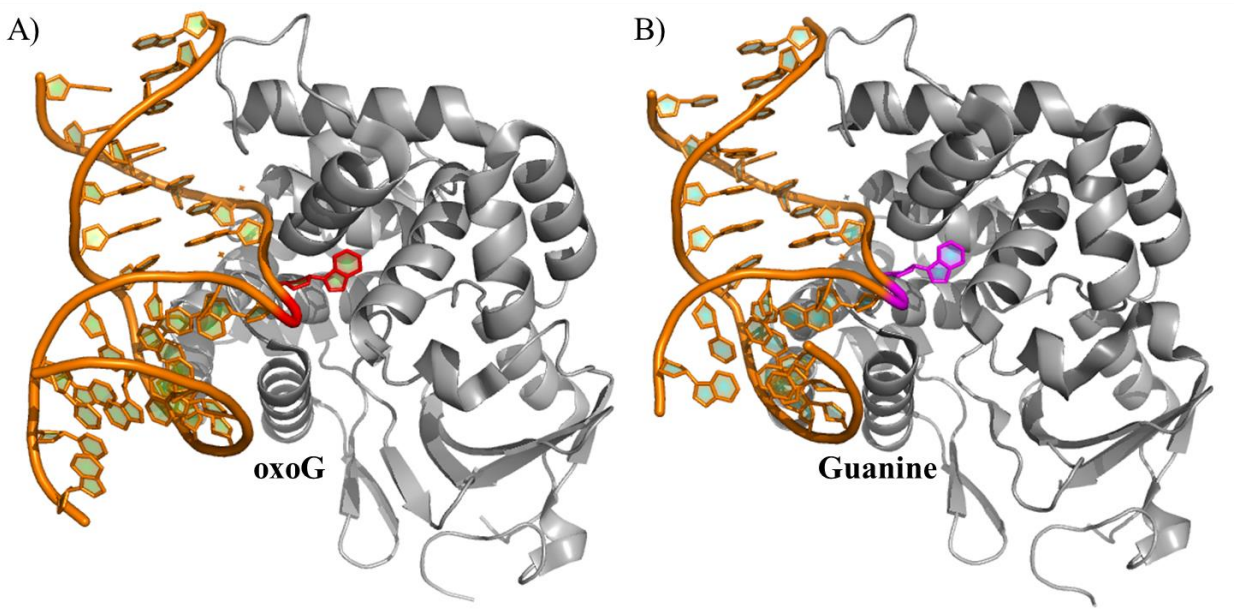


Figure 1.10. hOGG1 flips both oxoG and guanine into the binding site extra-helically. **A)** Crystal structure of PDB ID 2NOL, a catalytically inactive hOGG1 with oxoG in the binding site. [33] **B)** The crystal structure with PDB ID 1YQL, showing catalytically inactive hOGG1 with guanine in the binding site [36]. hOGG1 is shown in grey, the DNA substrate in orange, the oxoG is feature in red and canonical guanine is in magenta.

undamaged sugar. The human isoform of this enzyme is known as human 8-oxoguanine glycosylase 1, or hOGG1. Though the complete mechanism of oxoG recognition by hOGG1 is not known, much is known about 8oxoG binding and activity. hOGG1 binds to the backbone of the oxoG-containing strand, appears to flip out the oxoG base into the recognition pocket and bends the backbone around 8oxoG to insert it into the active site [33] [34] **(Figure 1.10)**. Once in the active site, the key catalytic residue lysine 249 performs a nucleophilic attack upon the C1' of oxoG and initiates a Schiff base reaction, leading to excision of the base [35]. In addition to the binding of oxoG, guanine can also be extra-helically inserted into the binding pocket but does not proceed to the active site because it does not interact with lysine 249 in the correct manner [36]. It is not known if this is the main step in distinguishing oxoG from G, or simply a singular form of extra-helical discrimination. The opposing cytosine also plays a role in the recognition of guanine analogues over other bases. In order to flip the oxoG into the recognition site, the base stacking of

the opposing cytosine is disrupted and the interaction is replaced with residues on the glycosylase. With other opposing bases, this interaction is not favored enough to allow base flipping into the enzyme recognition site.

DNA modifications and CpG sites

The chemical modifications 5metC and oxoG can occur simultaneously in a CpG site. A single strand CpG site can contain both oxoG and 5metC modifications adjacent to one another. However, due to the canonical base pair partners, one CpG site is base paired to another on the opposite strand. This leads to more possible additions of either modification in the duplex CpG site, *syn* (on the same strand) or *trans* (on the opposite strand) **(Figure 1.11)**. The previously discussed enzymes DNMTs and hOGG1 rely upon mechanisms which are not completely elucidated for recognition of their target substrate. The addition of either of the modifications in a single CpG, 5metC and oxoG, can decrease overall activity and thus, efficacy of both hOGG1 and the DNMTs. Both base excision by hOGG1 and cytosine methylation by DNMT-mediated epigenetics are extremely important cellular processes which can lead to improper development, diseases or cancers if their function is disrupted.

Presence of oxoG affects CpG methylation by DNMTs

The presence of oxoG in a CpG site has been found to affect the enzymatic rates of both the maintenance DNMTs and the *de novo* DNMTs. In murine DNMT3a, introducing an oxoG adjacent to the target cytosine in a hemimethylated CpG site slows the methylation of the target site [37]. Adding the oxoG directly across from the target cytosine, adjacent to the methylated cytosine on the opposite strand, accelerated target methylation. In human DNMT1, a similar effect was observed [38]. When oxoG was added next to the target cytosine in a hemimethylated site, the enzymatic activity was only slightly reduced, but when added adjacent to the target site the

enzymatic activity was reduced by an order of magnitude. From both studies, it was clear that the presence of oxoG in a CpG site alone can alter the rates and activities of the enzymatic reaction, but the location of the oxoG was more important to the overall impact on the DNMTs.

5metC slows 8oxoG excision by hOGG1

hOGG1 does not follow classical Michaelis-Menten enzyme kinetic reaction. Rather, the reaction distinguishes between the substrate and the product as separate observable states, where the hOGG1 is "E", the DNA substrate is "S", the enzyme substrate complex is "ES", the enzyme product complex is "EP" and "P" is the unbound product (nucleotide excised DNA).

$$E + S \underset{}{\overset{k_1}{\rightleftharpoons}} ES \overset{k_2}{\rightleftharpoons} EP \overset{k_3}{\rightleftharpoons} E + P$$

Equation 1.1: $E + S \rightleftharpoons ES \rightleftharpoons EP \rightleftharpoons E + P$

The rate constant k_1 characterizes the first step, the binding of the enzyme to the substrate. k_2 describes the rate of conversion from the enzyme-bound substrate to the enzyme-bound product and k_3 is the rate of release of the product. Similar to the DNMT experiments, it was observed that the addition of 5metC to the CpG of an oxoG affected hOGG1 rates [39]. When a single 5metC was added adjacent to the oxoG in a CpG site, or *syn* methylation, the k_2 was halved, but the k_3 did not differ significantly. Adjacent 5metC therefore only impacts the conversion of substrate to product, but not the subsequent product release. The 5metC modification on the opposing strand from the oxoG in the CpG site, *trans* methylation, caused k_2 to remain unchanged, but roughly halved k_3. Full methylation, methylation of both the adjacent and cross-strand cytosine in the CpG site, shared the same results as the *trans* CpG methylation. In these cases, the conversion from substrate to product was not slowed, but the product was released at a decreased rate.

Possible consequences of oxoG and 5metC in a CpG site

Despite clear changes in enzymatic activity associated with the co-occurrence of oxoG and 5metC in a CpG site and the devastating consequences for the disruption of hOGG1 and DNMT function, there is no known method for the cell to deal with this problem. Under normal cellular conditions, when the cell is not exposed to outside stressors, oxoG occurs once every 10^6 guanines [40]. This large number of oxoGs can interfere with DNMT function, and possibly lead to loss of methylation patterns. Approximately 1% of the human genome consists of CpG sites, and 70-80% of those sites contain a 5metC [41]. With a genome size of roughly 6 billion base pairs in humans, each and every human cell should contain at least 120 CpG sites that contain both 5metC and oxoG. Of these sites, all 120 of these would significantly slow down the function of hOGG1. This could lead to a build-up of oxoG, and possibly be linked to a greater rate of GC → TA transversions.

Identifying a target DNA substrate for 8oxoG-5metC investigations

Determining why the DNMTs and hOGG1 have decreased rates at oxidized/methylated CpG sites is a complex problem. Before investigations with an enzyme can be performed, the target substrate must be identified. Prior to investigating sources of the decreased enzymatic activity in the enzyme itself, it is necessary to identify the DNA substrate that will yield the largest difference in enzymatic rates. Therefore, a more precise investigation of the DNA substrates that were used in the DNMT and hOGG1 studies was undertaken.

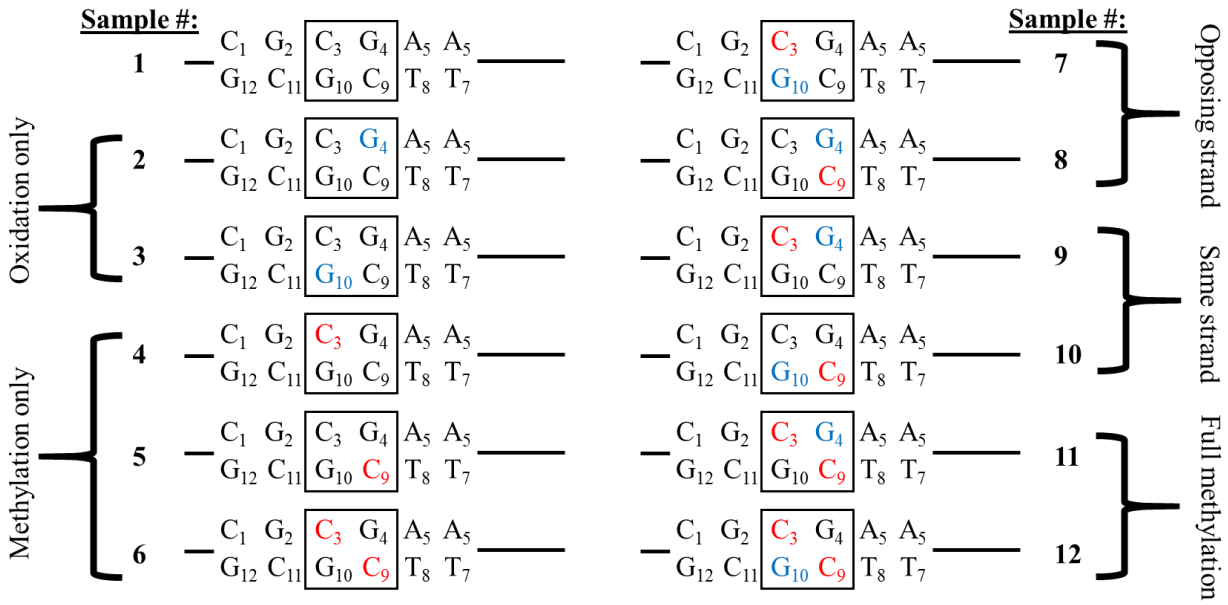


Figure 1.11. The CpG target sample scheme. Shown in boxes are the target CpG sites, with red denoting methylation at that site, and blue representing oxidation. Target samples are grouped by their type, and the grouping is labeled as shown.

A set of possible target DNA substrates was generated off the highly studied CpG-containing dodecamer first studied by Drew and Dickerson [2]. This sequence, CGCGAATTCGCG, contains four total CpG sites. The Drew-Dickerson dodecamer is a well characterized, short, symmetric sequence that is the standard sequence for DNA investigation by NMR [42]. The terminal CpG site of this sequence, and any sequence, is likely to be unstable. Therefore, the target CpG sites must be those which are not terminal. Both symmetric CpG sites are flanked by two differing environments, both an AT core and a CG terminus at either the 5' or 3' end. Due to the symmetry of this sequence, it is ideal for studying the hemi- and full-methylation patterns that are observed *in vivo*. The full scheme for all biologically relevant variations of 5metC and oxoG are shown in **Figure 1.11**. Throughout the rest of the text, the samples will be referred to by their sample number, or as "DD##" signifying that they are modified Drew-Dickerson sequences. *In this thesis, the solution structures of many of these samples will be compared and conclusions will be drawn as to the substrate's role in non-canonical enzyme-substrate interactions.*

Published DNA oligomers of relevance

When producing the structures of the modified Drew-Dickerson sequence shown in **Figure 1.11**, it is necessary to compare the structures which are reported in this thesis to those which have already been published by others to validate our results. The unmodified Drew-Dickerson sequence has several available structures which have been solved through various methods, as it represents an established model DNA sequence for structure determination. Of the modified structures we propose to study, two of the methylation-only samples (DD04 and DD05) have already been studied by other groups and their crystal structures have been reported. In addition, there has been a crystal structure published with the oxoG modification, oxoG:C base pair in a non-Drew-Dickerson sequence context. An overview of these structures is given below, and they will be used in later chapters to compare to our reported structures.

Unmodified Drew-Dickerson sequence structures

The structure of the unmodified Drew-Dickerson sequence has been extensively studied. The very first single-molecule crystal structure was of this sequence, reported by Drew and Dickerson, and resolved to 1.9 Å resolution [2]. From this crystal structure, new methods of solving DNA structures could be tested and compared to the standard B-DNA structure reported by Drew and Dickerson. So this sequence has been used as the basis for attempting the new methods for DNA structure solution: both solution NMR [43] and liquid crystalline NMR [42]. These differing methods of structure determination have their strengths and weakness, but provided a larger picture of this sequence in the different buffer conditions it was studied under and can be considered together. This unmodified sequence forms a right-handed helix, with Watson-Crick hydrogen bonding and overall has traits which most closely resemble B-DNA. The PDB IDs for these different structures are: 1BNA for the crystal structure, 2DAU for the NMR solution structure and

1NAJ for the liquid crystalline NMR structure solved with the residual dipolar coupling (RDC) restraints.

Methylated Drew-Dickerson sequence structures

The studies of the unmodified Drew-Dickerson sequence make this an ideal basis for studying DNA modifications, as we have seen in this thesis. Thus, other groups have looked into methylation of the Drew-Dickerson sequence at the exact positions that we have proposed to in DD04 and DD05. The dynamics of the backbone of these samples have been studied by solution NMR [44], and crystal structures have been solved for both of these methylated samples [45], [46]. However, their crystallization conditions differ, making the structures difficult to compare exactly to determine the effect of methylation upon the structure of the Drew-Dickerson duplex. Like the unmodified structures, these methylated Drew-Dickerson sequence structures are still right-handed helices, with Watson-Crick hydrogen bonding and have the general features of B-DNA. From this, we do not expect our methylated modified samples to produce any large global changes, such as mispairing, kinking or change in handedness. These structures serve as an excellent basis to assess the general quality of our own reported methylated structures, and could validate, or invalidate, any of the findings therein based upon the comparison of those reported in this thesis and the previously published structures.

oxoG-containing structure

Of DNA duplex structures which contain oxoG as a base modification, there is a sole crystal structure which has oxoG as the only modification in the sequence [47]. Other DNA-only structures featuring this modification also include further lesions such as abasic sites [48], rendering the effect of oxoG on these structures something that cannot be uncoupled through the clustering of these lesions. Though this decamer sequence, CCA(8oxoG)CGCTGG, differs from

the Drew-Dickerson sequence, the DNA duplex structure is overall similar. The oxoG-containing structure is a right-handed helix, with Watson-Crick base pairing and features general B-DNA parameters (rise, twist, groove width, pitch etc.). Based upon this structure, it is again expected that this modification should not cause large global changes in the DNA structure overall and thus it serves a basis for comparison on the oxoG-only modified structures to measure the overall reliability of our structural calculations.

References

[1] J. D. Watson and F. H. Crick, "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid," *Nature*, vol. 171, no. 4356, pp. 737–738, Apr. 1953.

[2] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson, "Structure of a B-DNA dodecamer: conformation and dynamics," *Proc. Natl. Acad. Sci.*, vol. 78, no. 4, pp. 2179–2183, Apr. 1981.

[3] Z. Shakked and D. Rabinovich, "The effect of the base sequence on the fine structure of the DNA double helix," *Prog. Biophys. Mol. Biol.*, vol. 47, no. 3, pp. 159–195, 1986.

[4] P. M. Keane, F. E. Poynton, J. P. Hall, I. V. Sazanovich, M. Towrie, T. Gunnlaugsson, S. J. Quinn, C. J. Cardin, and J. M. Kelly, "Reversal of a Single Base-Pair Step Controls Guanine Photo-Oxidation by an Intercalating Ruthenium(II) Dipyridophenazine Complex," *Angew. Chem. Int. Ed.*, vol. 54, no. 29, pp. 8364–8368, Jul. 2015.

[5] Z. Luo, M. Dauter, and Z. Dauter, "Phosphates in the Z-DNA dodecamer are flexible, but their P-SAD signal is sufficient for structure solution," *Acta Crystallogr. D Biol. Crystallogr.*, vol. 70, no. Pt 7, pp. 1790–1800, Jul. 2014.

[6] X.-J. Lu and W. K. Olson, "3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures," *Nucleic Acids Res.*, vol. 31, no. 17, pp. 5108–5121, Sep. 2003.

[7] M. L. M. Beckers and L. M. C. Buydens, "Multivariate analysis of a data matrix containing A-DNA and B-DNA dinucleoside monophosphate steps: Multidimensional Ramachandran plots for nucleic acids," *J. Comput. Chem.*, vol. 19, no. 7, pp. 695–715, May 1998.

[8] M. Sundaralingam, "Stereochemistry of nucleic acids and their constituents. IV. Allowed and preferred conformations of nucleosides, nucleoside mono-, di-, tri-, tetraphosphates, nucleic acids and polynucleotides," *Biopolymers*, vol. 7, no. 6, pp. 821–860, Jun. 1969.

[9] M. C. Thomas and C.-M. Chiang, "The General Transcription Machinery and General Cofactors," *Crit. Rev. Biochem. Mol. Biol.*, vol. 41, no. 3, pp. 105–178, Jan. 2006.

[10] C. E. Aitken and J. R. Lorsch, "A mechanistic overview of translation initiation in eukaryotes," *Nat. Struct. Mol. Biol.*, vol. 19, no. 6, pp. 568–576, Jun. 2012.

[11] R. Holliday, "Epigenetics: A Historical Overview," *Epigenetics*, vol. 1, no. 2, pp. 76–80, Apr. 2006.

[12] S. L. Berger, T. Kouzarides, R. Shiekhattar, and A. Shilatifard, "An operational definition of epigenetics," *Genes Dev.*, vol. 23, no. 7, pp. 781–783, Apr. 2009.

[13] N. McCarthy, "Epigenetics: Histone modification," *Nat. Rev. Cancer*, vol. 13, no. 6, pp. 379–379, Jun. 2013.

[14] P. A. Jones and D. Takai, "The Role of DNA Methylation in Mammalian Epigenetics," *Science*, vol. 293, no. 5532, pp. 1068–1070, Aug. 2001.

[15] T. H. Bestor, "Gene silencing: Methylation meets acetylation," *Nature*, vol. 393, no. 6683, pp. 311–312, May 1998.

[16] M. G. Goll and T. H. Bestor, "Eukaryotic cytosine methyltransferases," *Annu. Rev. Biochem.*, vol. 74, pp. 481–514, 2005.

[17] J. Sandoval, H. Heyn, S. Moran, J. Serra-Musach, M. A. Pujana, M. Bibikova, and M. Esteller, "Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome," *Epigenetics*, vol. 6, no. 6, pp. 692–702, Jun. 2011.

[18] S. S. Smith, B. E. Kaplan, L. C. Sowers, and E. M. Newman, “Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 89, no. 10, pp. 4744–4748, May 1992.
[19] M. Okano, S. Xie, and E. Li, “Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases,” *Nat. Genet.*, vol. 19, no. 3, pp. 219–220, Jul. 1998.
[20] T. Chen, S. Hevi, F. Gay, N. Tsujimoto, T. He, B. Zhang, Y. Ueda, and E. Li, “Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells,” *Nat. Genet.*, vol. 39, no. 3, pp. 391–396, Mar. 2007.
[21] Y. Kato, M. Kaneda, K. Hata, K. Kumaki, M. Hisano, Y. Kohara, M. Okano, E. Li, M. Nozaki, and H. Sasaki, “Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse,” *Hum. Mol. Genet.*, vol. 16, no. 19, pp. 2272–2280, Oct. 2007.
[22] M. Okano, D. W. Bell, D. A. Haber, and E. Li, “DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development,” *Cell*, vol. 99, no. 3, pp. 247–257, Oct. 1999.
[23] A. Hermann, R. Goyal, and A. Jeltsch, “The Dnmt1 DNA-(cytosine-C5)-methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites,” *J. Biol. Chem.*, vol. 279, no. 46, pp. 48350–48359, Nov. 2004.
[24] R. Goyal, R. Reinhardt, and A. Jeltsch, “Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase,” *Nucleic Acids Res.*, vol. 34, no. 4, pp. 1182–1188, Jan. 2006.
[25] J. Song, O. Rechkoblit, T. H. Bestor, and D. J. Patel, “Structure of DNMT1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA Methylation,” *Science*, vol. 331, no. 6020, pp. 1036–1040, Feb. 2011.
[26] T. Yokochi and K. D. Robertson, “Preferential Methylation of Unmethylated DNA by Mammaliande Novo DNA Methyltransferase Dnmt3a,” *J. Biol. Chem.*, vol. 277, no. 14, pp. 11735–11745, Apr. 2002.
[27] C. Holz-Schietinger and N. O. Reich, “The Inherent Processivity of the Human de Novo Methyltransferase 3A (DNMT3A) Is Enhanced by DNMT3L,” *J. Biol. Chem.*, vol. 285, no. 38, pp. 29091–29100, Sep. 2010.
[28] P. M. Vertino, J. A. Sekowski, J. M. Coll, N. Applegreen, S. Han, R. J. Hickey, and L. H. Malkas, “DNMT1 is a Component of a Multiprotein DNA Replication Complex,” *Cell Cycle*, vol. 1, no. 6, pp. 416–423, Nov. 2002.
[29] K. E. Bachman, M. R. Rountree, and S. B. Baylin, “Dnmt3a and Dnmt3b Are Transcriptional Repressors That Exhibit Unique Localization Properties to Heterochromatin,” *J. Biol. Chem.*, vol. 276, no. 34, pp. 32282–32287, Aug. 2001.
[30] R. D. Bont and N. van Larebeke, “Endogenous DNA damage in humans: a review of quantitative data,” *Mutagenesis*, vol. 19, no. 3, pp. 169–185, May 2004.
[31] A. P. Grollman and M. Moriya, “Mutagenesis by 8-oxoguanine: an enemy within,” *Trends Genet.*, vol. 9, no. 7, pp. 246–249, Jul. 1993.
[32] M. L. Wood, M. Dizdaroglu, E. Gajewski, and J. M. Essigmann, “Mechanistic studies of ionizing radiation and oxidative mutagenesis: genetic effects of a single 8-hydroxyguanine (7-hydro-8-oxoguanine) residue inserted at a unique site in a viral genome,” *Biochemistry (Mosc.)*, vol. 29, no. 30, pp. 7024–7032, Jul. 1990.

[33] C. T. Radom, A. Banerjee, and G. L. Verdine, "Structural characterization of human 8-oxoguanine DNA glycosylase variants bearing active site mutations," *J. Biol. Chem.*, vol. 282, no. 12, pp. 9182–9194, Mar. 2007.
[34] S. D. Bruner, D. P. Norman, and G. L. Verdine, "Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA," *Nature*, vol. 403, no. 6772, pp. 859–866, Feb. 2000.
[35] H. M. Nash, R. Lu, W. S. Lane, and G. L. Verdine, "The critical active-site amine of the human 8-oxoguanine DNA glycosylase, hOgg1: direct identification, ablation and chemical reconstitution," *Chem. Biol.*, vol. 4, no. 9, pp. 693–702, Sep. 1997.
[36] A. Banerjee, W. Yang, M. Karplus, and G. L. Verdine, "Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA," *Nature*, vol. 434, no. 7033, pp. 612–618, Mar. 2005.
[37] D. V. Maltseva, A. A. Baykov, A. Jeltsch, and E. S. Gromova, "Impact of 7,8-Dihydro-8-oxoguanine on Methylation of the CpG Site by Dnmt3a†," *Biochemistry (Mosc.)*, vol. 48, no. 6, pp. 1361–1368, Feb. 2009.
[38] P. W. Turk, A. Laayoun, S. S. Smith, and S. A. Weitzman, "DNA adduct 8-hydroxyl-2′-deoxyguanosine (8-hydroxyguanine) affects function of human DNA methyltransferase," *Carcinogenesis*, vol. 16, no. 5, pp. 1253–1255, May 1995.
[39] R. D. Kasymov, I. R. Grin, A. V. Endutkin, S. L. Smirnov, A. A. Ishchenko, M. K. Saparbaev, and D. O. Zharkov, "Excision of 8-oxoguanine from methylated CpG dinucleotides by human 8-oxoguanine DNA glycosylase," *FEBS Lett.*, vol. 587, no. 18, pp. 3129–3134, Sep. 2013.
[40] ESCODD (European Standards Committee on Oxidative DNA Damage), "Comparative analysis of baseline 8-oxo-7,8-dihydroguanine in mammalian cell DNA, by different methods in different laboratories: an approach to consensus," *Carcinogenesis*, vol. 23, no. 12, pp. 2129–2133, Dec. 2002.
[41] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.-F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu,

L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H.-C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. R. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. A. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S.-P. Yang, R.-F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, A. Patrinos, and M. J. Morgan, “Initial sequencing and analysis of the human genome,” *Nature*, vol. 409, no. 6822, pp. 860–921, Feb. 2001.

[42] Z. Wu, F. Delaglio, N. Tjandra, V. B. Zhurkin, and A. Bax, “Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and 31P chemical shift anisotropy,” *J. Biomol. NMR*, vol. 26, no. 4, pp. 297–315, Aug. 2003.

[43] A. Y. Denisov, E. V. Zamaratski, T. V. Maltseva, A. Sandström, S. Bekiroglu, K. H. Altmann, M. Egli, and J. Chattopadhyaya, “The solution conformation of a carbocyclic analog of the Dickerson-Drew dodecamer: comparison with its own X-ray structure and that of the NMR structure of the native counterpart,” *J. Biomol. Struct. Dyn.*, vol. 16, no. 3, pp. 547–568, Dec. 1998.

[44] Y. Tian, M. Kayatta, K. Shultis, A. Gonzalez, L. J. Mueller, and M. E. Hatcher, “31P NMR Investigation of Backbone Dynamics in DNA Binding Sites,” *J. Phys. Chem. B*, vol. 113, no. 9, pp. 2596–2603, Mar. 2009.

[45] J. A. Theruvathu, W. Yin, B. M. Pettitt, and L. C. Sowers, “Comparison of the structural and dynamic effects of 5-methylcytosine and 5-chlorocytosine in a CpG dinucleotide sequence,” *Biochemistry (Mosc.)*, vol. 52, no. 47, pp. 8590–8598, Nov. 2013.

[46] L. Lercher, M. A. McDonough, A. H. El-Sagheer, A. Thalhammer, S. Kriaucionis, T. Brown, and C. J. Schofield, “Structural insights into how 5-hydroxymethylation influences transcription factor binding,” *Chem. Commun. Camb. Engl.*, vol. 50, no. 15, pp. 1794–1796, Feb. 2014.

[47] L. A. Lipscomb, M. E. Peek, M. L. Morningstar, S. M. Verghis, E. M. Miller, A. Rich, J. M. Essigmann, and L. D. Williams, “X-ray structure of a DNA decamer containing 7,8-dihydro-8-oxoguanine,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 92, no. 3, pp. 719–723, Jan. 1995.

[48] J. Zálešák, M. Lourdin, L. Krejčí, J.-F. Constant, and M. Jourdan, “Structure and dynamics of DNA duplexes containing a cluster of mutagenic 8-oxoguanine and abasic site lesions,” *J. Mol. Biol.*, vol. 426, no. 7, pp. 1524–1538, Apr. 2014.

Chapter 2 - Drew-Dickerson DNA solution structure determination and analysis

The twelve DNA sample scheme given at the end of the previous chapter was treated as twelve individual samples for structure determination, albeit with high similarity. The elucidation of the NMR solution structures and subsequent analysis followed the same general protocol. In this chapter, the protocol from sample preparation to structure analysis is given; the following chapters will highlight the individual differences between the completed samples and compare the samples to one another.

Sample Preparation

Samples were prepared for NMR in house, but the duplex DNA dodecamers were synthesized elsewhere.

Sample origins

The synthesis of the individual samples of the twelve-sample scheme were dependent upon the modification. The methylation-only samples (samples 4, 5 and 6) were reliably obtained commercially through Midland Certified Reagent Co (Midland, TX). The oxidized samples were synthesized from commercially available phosphoramidites (Glen Research, Sterling, VA) under supervision of our collaborator Professor Dmitry Zharkov (Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia).

In-house pre-NMR preparations

Upon reception of a lyophilized sample, the oligonucleotide was resuspended in buffer consisting of 50 mM NaCl, 10 mM sodium phosphate and 1 mM EDTA at pH 6.8. After suspension, the samples were re-annealed by raising the temperature in the NMR bore or a water bath to approximately 70°C, allowing the DNA sample to remain there for 20 minutes, then were slowly

cooled to room temperature for an additional 20 minutes. This was done to remove any possible mismatched strand conformation induced through the shipping and storage processes. Initial 1D ^{1}H NMR spectra were acquired to check for contaminants present in solution, and this required the addition of 10% (v/v) D_2O to obtain a sufficient lock. Contaminants were defined as any present peaks on the 1D NMR spectrum which were higher than the general peak populations during water suppression, excluding the water peak itself which appears at ~4.8 ppm.

If NMR revealed a contaminant was present, and a non-DNA species was present in every sample received, fast-paced liquid chromatography (FPLC) was performed. The sample was injected onto a Hi-Prep 16/60 Sephacryl S-100 HR column with an AKTA Prime Plus FPLC system as per the standard protocol with the initial buffer conditions used throughout the process. After injection of the sample onto the column, the buffer flow rate was kept at 0.8 ml/min. Fractions collected from the FPLC were of too low a concentration for NMR detection, so the fractions were further concentrated using spin concentration with 15 mL Amicon Ultra centrifugation filtration units with a 3000 Da molecular weight cut off. The centrifugation filtration units were prepped before addition of DNA to strip the membrane of contaminants that interfere with NMR signal. The washing step performed to the units consisted of 15 mL of 0.1 mM NaOH, 30 mL of water and 15 mL of buffer passed through the membrane through centrifugation at 4000 x g until cleared. Prior to cleaning, all centrifugation filtration units were checked by NMR, first by running a small amount of 99% D_2O through the membrane, then analyzing 500 uL of the filtrate by 1D ^{1}H NMR inspection. In the absence of any contaminants, the units were stored with fresh 15 mL of buffer overnight to equilibrate with the membrane before addition of a DNA sample. DNA samples were concentrated using the centrifugation units at 4000 x g until the desired concentration of at least 0.1 mM DNA was achieved. DNA concentration was determined by spectroscopically on a Take3

place using a BioTek Epoch Microplate Spectrophotometer set at a wavelength of 260 nm. Preparations for a sample in H_2O were completed at this stage, and 500 uL of sample with 10% (v/v) D_2O were moved to an eight inch in length Norrell 508-UP NMR tube optimized for 500 MHz NMR recordings.

The bulk of the NMR data was taken in 99.9% D_2O conditions, as will be explained in the NMR section below; thus, some fraction of the samples underwent further processing. To remove the 100% water solvent, lyophilization was performed overnight with an Air Products CSW-202 Displex Cryogenic Refrigerator after flash freezing 500 uL of the purified, concentrated DNA sample. The sample was resuspended in 500 uL of 99.9% D_2O and placed in a clean eight inch 508-UP Norrell NMR tube.

NMR recording

In order to solve the solution structure, a variety of NMR spectra were acquired. The majority of data for structure solution was obtained through two dimensional (2D) NMR spectra. In this section, each type of NMR experiment is described and its importance to the solution structure is demonstrated.

On buffer conditions

Only spin ½ nuclei can be detected through conventional solution NMR spectroscopy utilized in the Smirnov lab. In order to properly match the observed frequency for water, some amount of D_2O is always necessary in an NMR sample. Water is in high concentration in any aqueous solution and contains two hydrogens for every one water molecule. The presence of water accounts for the majority of the signal in ^{1}H-NMR spectra. Therefore, a suppression of the water signal was applied to each type of acquired ^{1}H (or proton) NMR spectra. Once sufficient water suppression is achieved, in a sample that is 90% H_2O and 10% D_2O, any exchangeable protons are likely to be

populated with hydrogens and can be observed. Exchangeable protons are so called because they can undergo an acid/base exchange with water and can exchange with the protons from the water in solution. In DNA, this largely gives insight into the hydrogens attached to nitrogen atoms. The non-exchangeable protons, those attached to carbon, can be viewed in all the spectra. This leads to many peaks in the spectra, which can be difficult to resolve. While the exchangeable protons are important to observe, the exact environment of each individual non-exchangeable proton yields the majority of the information about the DNA structure. Thus, the solvent is changed from being mostly water with its spin ½ nuclei hydrogens, to the isotopically labeled ^{2}H or deuterium atom (spin 1, shortened as D) heavy water D_2O. This allows the exchangeable protons to equilibrate with the solution and be replaced by deuterium, and thus are not visible on a proton NMR spectrum. This allows for fewer observed peaks, with more similar environments and thus a narrow

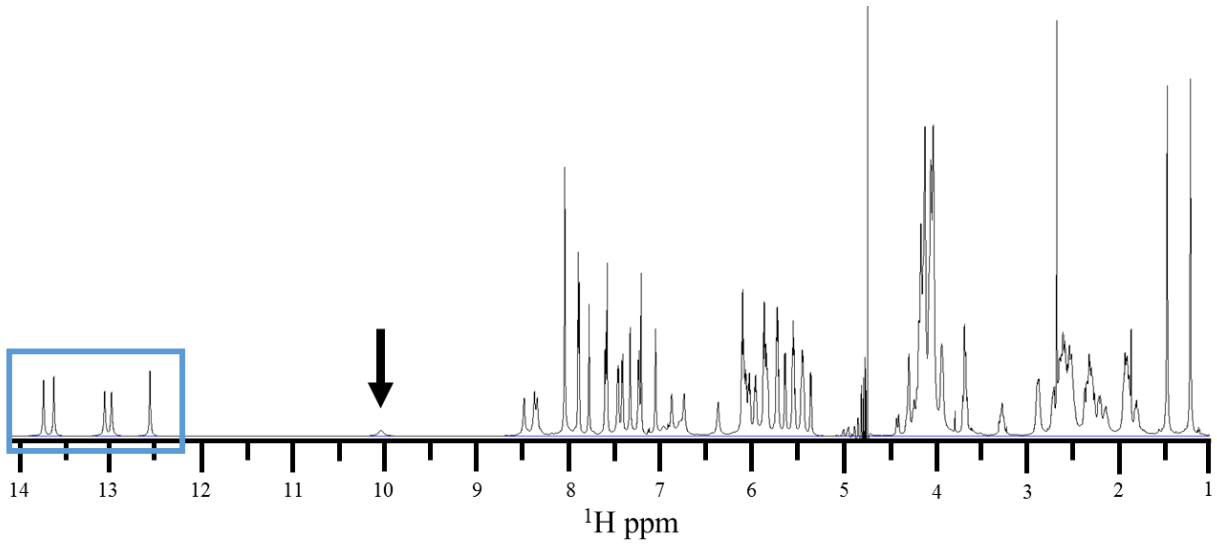


Figure 2.1. The 1D ^{1}H NMR spectrum of DD03 in H_2O taken on a 700 MHz Bruker Avance NMR with a cryoprobe at 25°C. The blue box shows the imino peaks between 12 and 14 ppm. The arrow highlights the H7 proton that should be present in an oxidized sample.

range of chemical shifts. The chemical shift range can be sampled with more points over less time, and increases the indirect resolution of the 2D NMR spectra.

1D experiments

Single dimension (1D) NMR experiments were not used directly for the solution of the DNA duplex structures, but were used as a quality control measure. Prior to any 2D experiments, the 1D of each nuclei was taken, whether it was ^{1}H or ^{31}P. The water suppression in ^{1}H NMR spectra was achieved through first collecting a 1D spectrum, finding the location of the water peak in that particular sample, and suppressing the signal with the power and duration which is appropriate for the buffer (90% or 0.01% H_2O) [1]. No solvent suppression was needed for ^{31}P experiments. The presence and number of the imino protons **(Figure 2.1)** were checked in the H_2O samples before proceeding to the H_2O NOESY [2].

The quality of each modified DNA sample (DD02 through DD12) was checked first through 1D NMR to confirm if the NMR signatures specific to the modification(s) for that sample could be observed. For methylated samples, the low ppm methyl peak should be observed. In the Drew-Dickerson sequence, there are already two methyl groups present, located on each canonical thymine. In the methylated samples, the additional methyl should join the pre-existing two methyl peaks (either 1 or 2 additional peaks depending on the number of additional methyls). For the oxidized samples, the presence of the proton on N7 was checked for at around 10 ppm **(Figure 2.1)**. The absence of the base proton (H8) was the other signature of the oxidized guanine bases (oxoG) **(Figure 1.2)**.

Homonuclear ^{1}H 2D NMR experiments

Homonuclear proton NMR experiments provided the bulk of the data needed to solve the solution structure of the duplex DNA samples [3]. The relative environments of each proton, including the distance and covalent bonds separating individual protons, yields much information about the structure and was the main component that drove restrained molecular dynamics protocols. The

following NMR experiments in this section all contain strictly proton-proton interactions. These homonuclear NMR experiments were referenced to one another using the chemical shift ever-present water peak (featured around 4.8 ppm as the largest peak in **Figure 2.1** and the large central black lines in **Figure 2.4**). The chemical shift of the reference water peak was adjusted for the temperature of each experiment using a pre-established relationship between chemical shift and temperature [4].

2D NOESY recorded in 90% H_2O

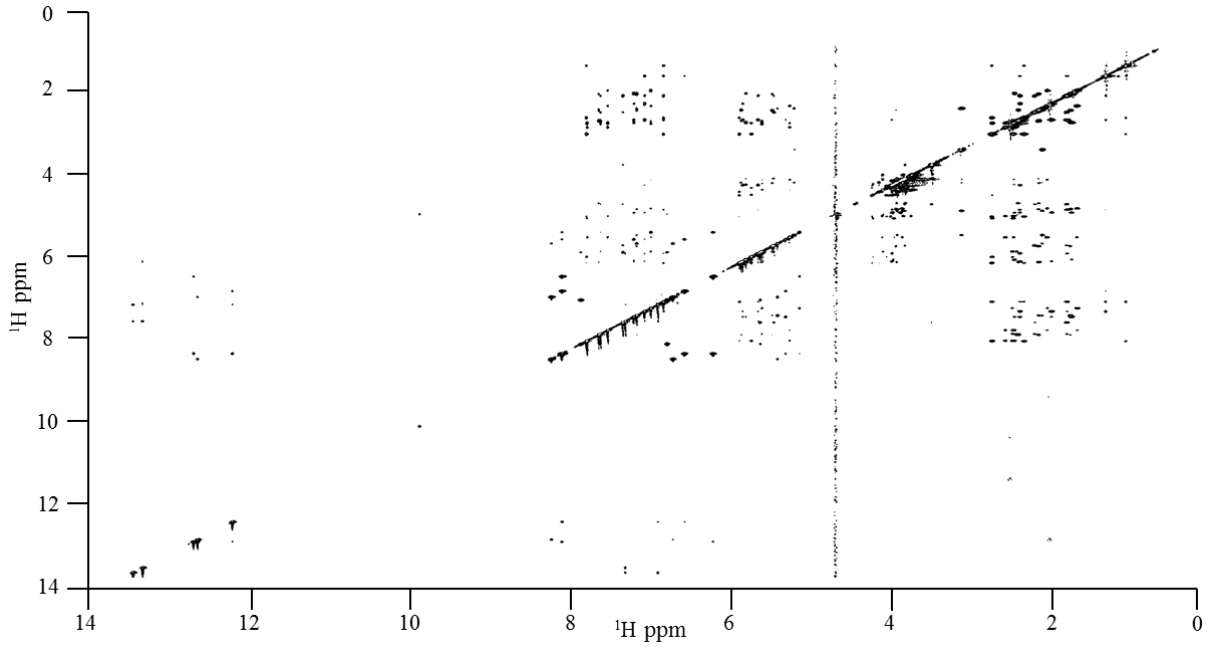


Figure 2.2. The full H_2O NOESY for DD03 taken on a 700 MHz Bruker Avance NMR with a cryoprobe at 5 °C.

Nuclear Overhauser effect spectroscopy (or NOESY) are experiments that are performed to elucidate through-space interactions of spin ½ nuclei. In the nuclear Overhauser effect (NOE), spin excitation is transferred between polarized nuclei through cross relaxation. By extending the relaxation time delay in the NMR pulse sequence, greater inter-nuclei distances can be sampled. For the duplex DNA samples in H_2O, a relaxation time of 200 ms was used, which shows proton-

proton interactions through space within around 4.0 Å. As mentioned before, the NOESY spectra recorded in H_2O contained many peaks for both the non-exchangeable and exchangeable protons. The majority of the peaks observed lay within the range of 1.0 to 9.0 ppm **(Figure 2.2)**. The highly exchangeable nucleotide base imino protons which are sensitive to base pairing interactions are within the ppm range of 13 to 14 ppm. The amino protons that they form the NOESY cross-peaks with lie between 6 to 9 ppm **(Figure 2.3)**. In order to capture a slower dynamic and exchange rate for these base pairing protons, the recordings were performed at lower temperature (typically 5 °C).

The pattern of peaks observed in this region is the target information of the H_2O NOESY, as they give insight into the fundamental nature of duplex DNA base pairing [2]. If regular Watson-Crick base pairing has occurred, a few expected NOESY interactions can be observed. For the GC base pair, the proton attached to N1 should be within distance range of both of the protons attached to the cytosine N4. This would appear as two separate peaks for each GC base pair. For the AT base pair, a Watson-Crick hydrogen bonding pair would show the proton attached to the thymine N3 within the NOE distance of the proton bonded to the adenine C2. This would result in two peaks for each AT

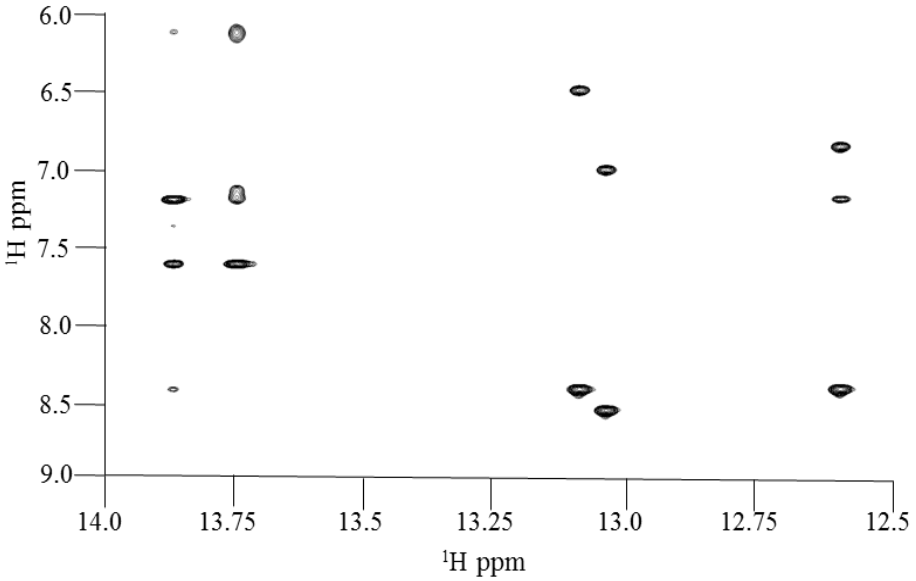


Figure 2.3. A portion of an H_2O NOESY for DD03 taken on a 700 MHz Bruker Avance NMR with a cryoprobe at 5 °C. This zoomed in region shows the interactions between imino and amino protons between hydrogen bonded bases.

base pair. Assignment of these peaks can be completed by tracing several interactions back to the bulk of the non-exchangeable proton assignments completed for the D_2O NOESY. If patterns such as these are not observed, then non-Watson-Crick hydrogen bonding or single-strand DNA state had occurred.

2D NOESY recorded in 100% D_2O

To unclutter the 2D NOESY spectrum from the signals involving exchangeable proteins and dramatically increase the resolving power, the NOESY spectra recorded in 99.9% (referred to as 100% for ease of reference) D_2O were used to obtain the majority of the information used in restrained molecular dynamics. In D_2O, there are no peaks observed downfield of 9 ppm. The non-exchangeable hydrogen interactions observed are much fewer and thus have better separation than the H_2O NOESY **(Figure 2.4)**.

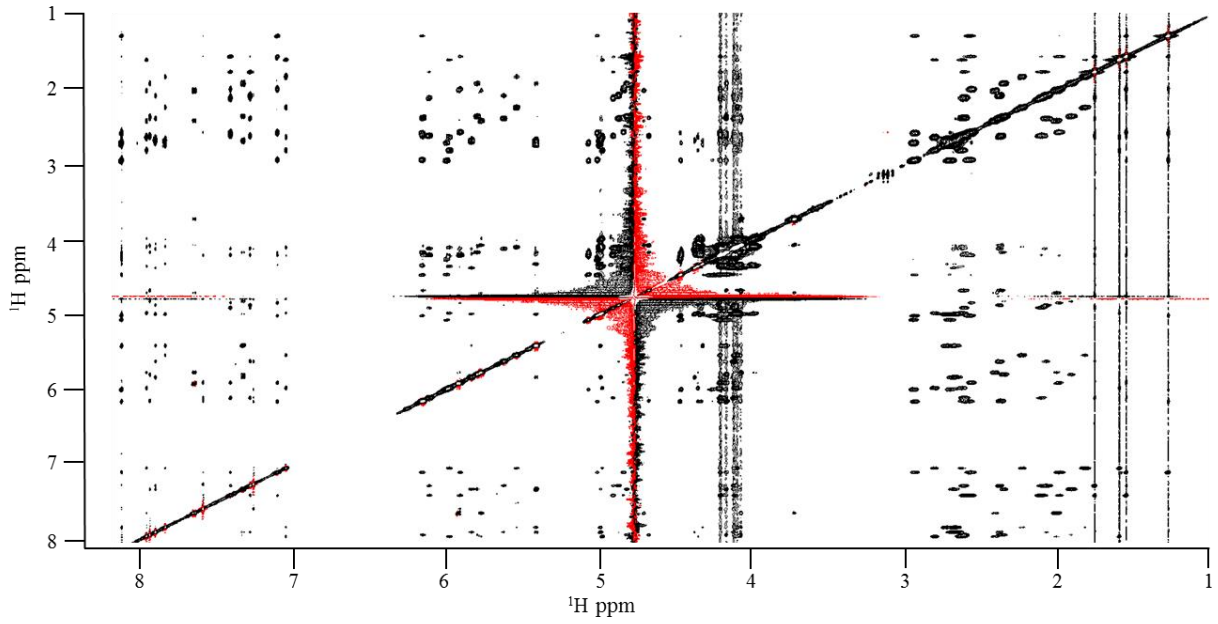


Figure 2.4. The full 260 ms D_2O NOESY for DD06 taken on a 500 MHz Varian Inova NMR at 25°C.

With D_2O NOESYs, the quality of synthesis of each sample was addressed in-depth. The cytosine methylation should lead to strong upfield methyl peaks which interact (in the NOESY sense) with

the protons of the surrounding bases. The oxidation of guanine causes the loss of an observed base proton (H8) for the modified guanine. The placement of the modification and the handedness of the duplex DNA were assessed by beginning peak assignments in the "walk region." In this region, the peaks corresponding to the interacting protons usually show the interactions between the sugar (H1') and base protons (H6/H8) of a residue, and the residue that directly follows it in the 3' direction **(Figure 2.5)**. From the knowledge of the sequence, the interactions from base proton (H8/H6) *i* to sugar proton (H1') *i* to base proton (H8/H6) *i+1* can be traced in a right-handed DNA helix which allows the extension of the NOESY peak and NMR chemical shift assignment through the oligonucleotide **(Figure 2.6)**. Further assignments were completed for the rest of the peaks using the walk region peaks as a starting point [3]. Each resolved peak was assigned to its corresponding proton-proton interaction and the oval-approximated volume of the peak was measured. All peaks were measured at the same noise cutoff threshold to assure the volumes were not artificially inflated or deflated relative to one another. Any peaks that were not well resolved were included if at all possible, but marked for possibly inflated volumes for possible removal from the structure determination at a later stage.

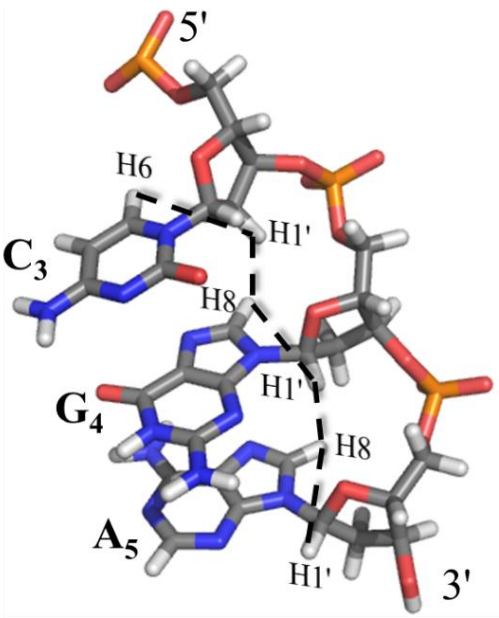


Figure 2.5. The connections of base protons to sugar protons termed the "walk". Coloration is by element type with labels for the 5' and 3' ends, the bases and the protons whose interactions are shown. The black dashed line shows both the inter- and intra-base connections between the base and sugar protons.

Unlike the other types of spectra, the D_2O NOESY was sampled at various mixing times (e.g., 90, 140, 200 and 260 ms) allowing for the detection of various through space proton-proton resonance transfers through distances up to 5.5 Å. The use of different mixing times mitigated the adverse relaxation effects associated with each individual mixing time NOESY. As relaxation occurs, the resonance sampled can approach the *J*-coupling frequency experienced for an individual set of resonance transfers between specific protons [5]. This effect can be directly viewed from the unusual peak pattern shown for the two cytosine H5-H6 peaks in **Figure 2.6**. A set of mixing time values allows each peak set produced through spectral assignment to be added to a pool of volume-based distances, and thus is not as likely to have a systematic problem due to relaxation effects. The specific mixing times differed between samples, but at least three differing mixing times were used in the production of each solution structure.

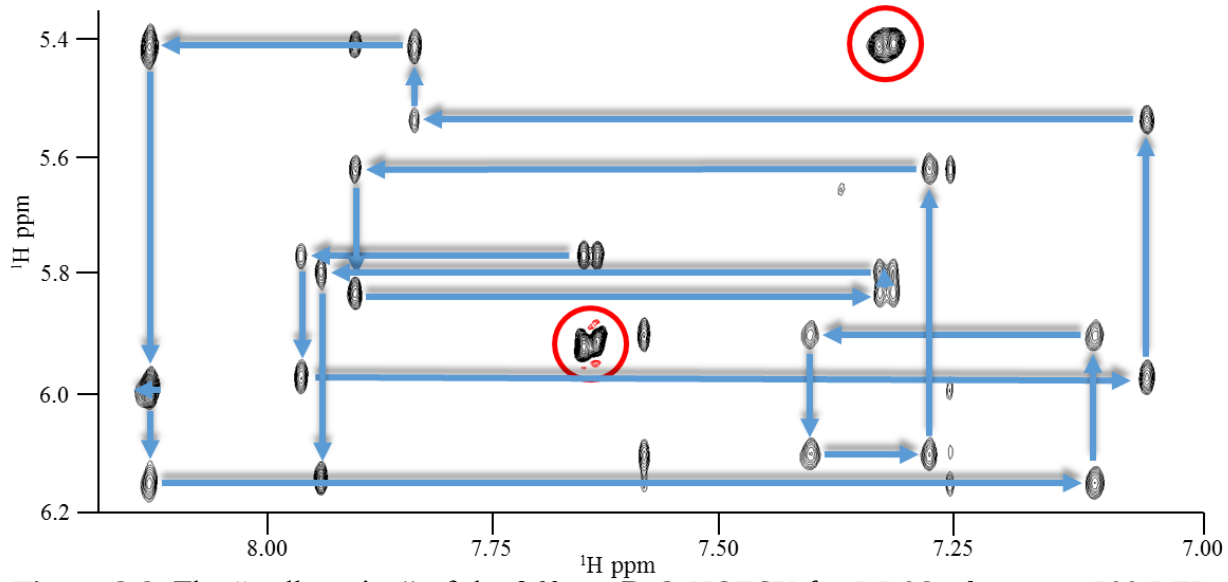


Figure 2.6. The "walk region" of the 260 ms D_2O NOESY for DD06 taken on a 500 MHz Varian Inova NMR at 25 °C. The blue arrows show the walk, starting at the first cytosine base proton to its own H1' and follows the connections from base to base. This walk confirms the sample is a right handed helix and precedes all other peak assignments. Red circles show the peak corresponding to a cytosine H5-H6 connection.

2D TOCSY recorded in 100% D_2O

The main role of the total correlation spectroscopy (TOCSY) data was to aid in peak assignment, but the spectra were also instrumental in sample quality control process. Due to features in the TOCSY pulse sequence, TOCSY's transfer of magnetization between protons through short sequences of covalent bonds, not through space as a NOESY does. However, NOESY and TOCSY are similar in that the extent (distance or number of bonds) of the resonance transfer observed is dependent on the mixing time. A TOCSY mixing time of 120 ms was used to observe the interactions between protons that are separated by 5 bonds or fewer. This allows the corroboration of NOESY data. All of the peaks observed in a 120 ms TOCSY are also observable in the NOESY, but all of the peaks in a NOESY are not observable in TOCSY **(Figures 2.4 and 2.7**. This narrows down the possibilities for NOESY peak assignment. In addition, if the process of synthesis of the oligonucleotide was compromised, it was usually readily observable by the unusual pattern of TOCSY peaks. In this way, the TOCSY was used as a general quality control measure to detect

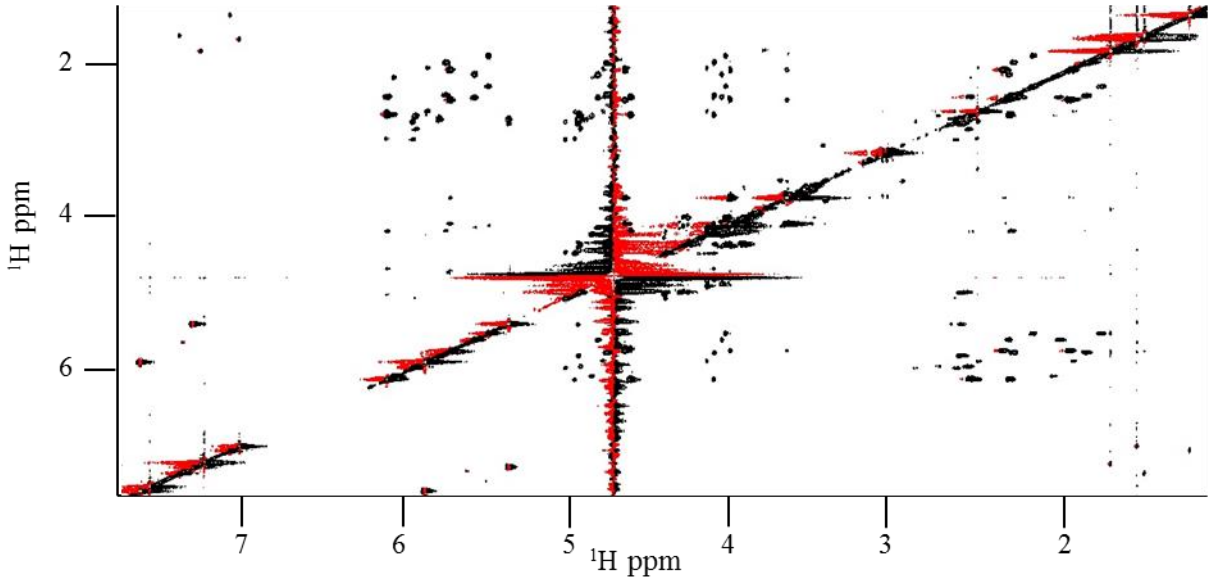


Figure 2.7. The full 120 ms TOCSY for DD06 taken on a 500 MHz Varian Inova NMR at 25 °C.

unexpected bonds or through bond interactions before the DNA structure determination process began.

DQFCOSY recorded in 100% D_2O

Double quantum filtered correlation spectroscopy (DQFCOSY) has similar elements to the TOCSY, but provides further insight into sugar pucker. DQFCOSY is still correlation spectroscopy, as TOCSY is, so the magnetization transfer through bonds is observed. The advantage of a using DQFCOSY lies in the double quantum filtering. This filtering removes the response from uncoupled spins, like ^{1}H in water, and increases spectral resolution. With the increased resolution, the multiplet structure of the peaks is observed **(Figure 2.8)** and from those multiplets the *J*-coupling constants can be found. The proton-proton couplings can elucidate much about the local structure of the molecule [6]. The majority of the peaks seen in the DQFCOSY are those which are representative of intra-sugar interactions. Thus, with the knowledge of the peak assignment from the TOCSY and NOESY, the multiplet patterns from the DQFCOSY yield information about sugar pucker. This type of spectra can be very useful for quantitative assessment of *J*-coupling constants and qualitative assessment of general patterns to ensure that the sugar

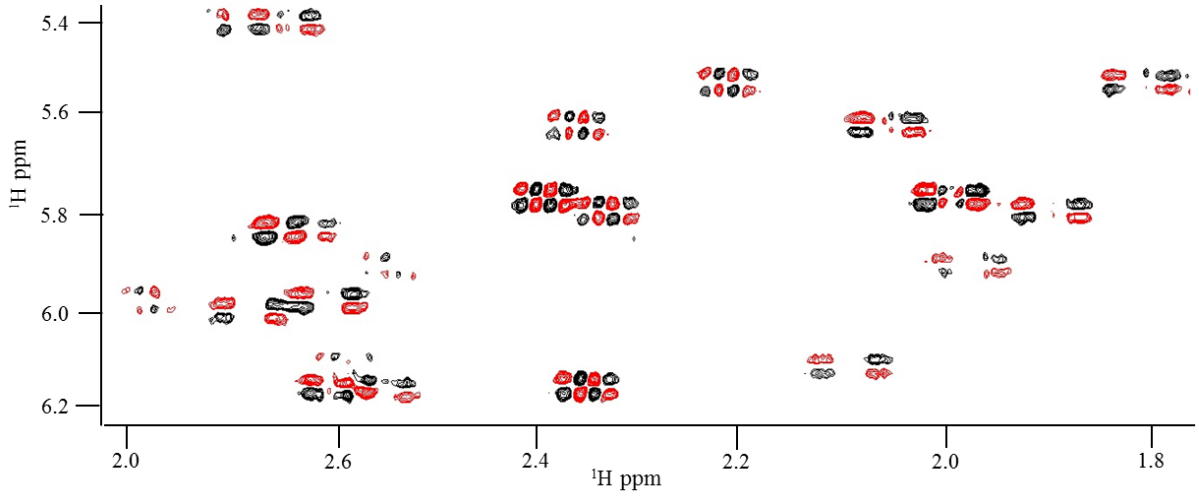


Figure 2.8. The DQFCOSY for DD06 taken on a 500 MHz Varian Inova NMR at 25 °C. The region shown highlights the H1' to H2'/H2" interactions and their multiplet peak patterns.

puckers contained within the sample are generally of A/B DNA type and do not have outstanding outliers which need to be addressed.

Heteronuclear 2D experiments: ^{31}P – ^{1}H HETCOR

An important part of the DNA structure that is not sampled sufficiently by the homonuclear 2D NMR experiments is the backbone conformation. The phosphates on the backbone contain protons reporting highly degenerate chemical shift values hindering their reliable assignment in NMR spectra. To mitigate this, a 2D NMR experiment probing *J*-correlation between the highly abundant spin ½ nuclei ^{31}P and the nearby protons were taken. The pulse sequence used, HETCOR or heteronuclear correlation, works much the same way as the previous correlation pulse sequences or heteronuclear analogs like HSQC **(Figure 2.9)**. HETCOR samples the magnetization transfer through covalent bonds between the ^{1}H (H3' and H4') and the ^{31}P. Having already obtained the assignments for those protons that are near the phosphate through the prior proton experiments, the ^{31}P

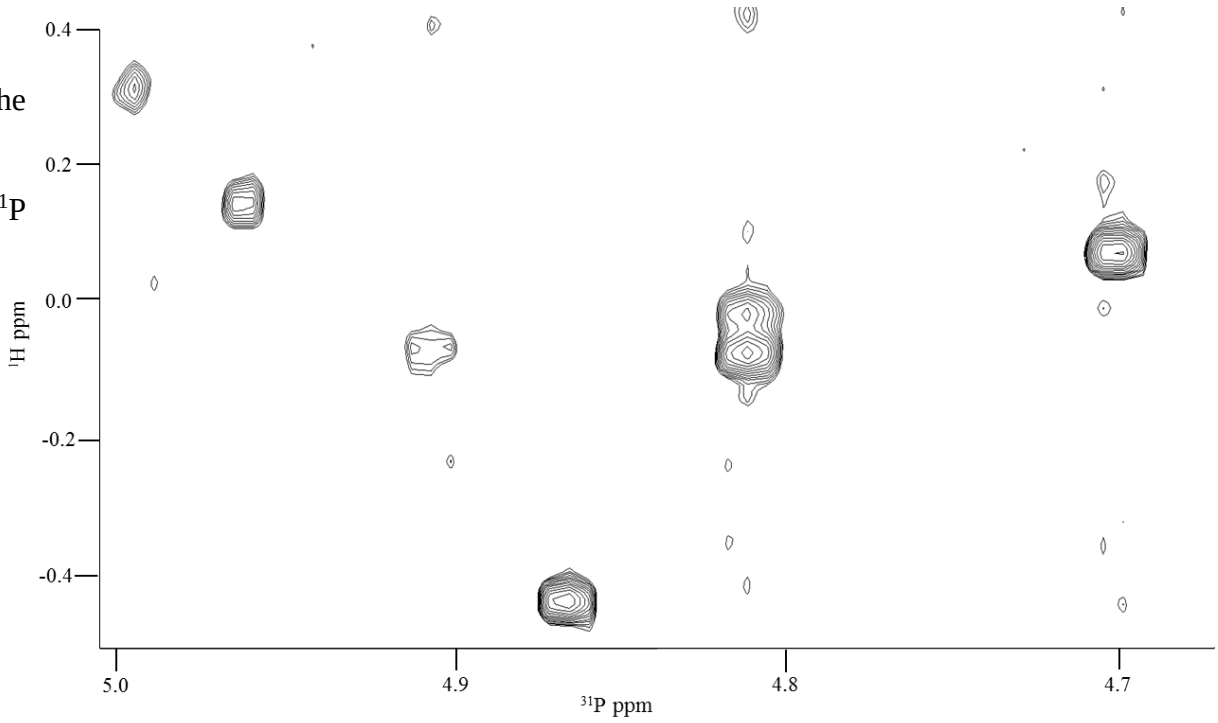


Figure 2.9. The HETCOR for DD03 taken on a 700 MHz Bruker Avance NMR with a cryoprobe at 25°C. The region shown highlights the P to H3' interactions.

dimension assignments could be made [7]. In the Drew-Dickerson sequence the majority of ^{31}P chemical shifts cluster tightly within ~0.5 ppm due to the highly similar BI conformation of every backbone step [8]. However, in some of the modified samples from our sample scheme, there is a clear outlier that does not overlap with the rest of the ^{31}P chemical shifts. Prior experiments have determined that the ^{31}P chemical shift is sensitive to the backbone conformation, and that a strong shift downfield correlates with a larger population of the backbone in the BII conformation [9]. Presence or absence of an assigned ^{31}P peak shifted downfield informed the allowed backbone conformation ranges for restrained molecular dynamics. The chemical shifts of phosphorus experiments were indirectly referenced using tri-methyl sulfate as the reference standard [10].

Restrained Molecular Dynamics

While the data for solution DNA structures is taken from NMR spectra, it needs to be translated into the structural features itself. In order to apply the NMR information acquired, all the data requires compiling into a usable form for structure determination. These are referred to as structural restraints. Restraints are applied to a structure calculation as an additive component of the overall energy of the system, for which violating the restraints increases the overall energy of the molecule [11]. As per the laws of thermodynamics, the lowest energy state is the most likely conformation. While restraints are important for the solution structure, there is much general knowledge about a molecule that can be applied to the calculation of a final structure that can also be used and is typically included in a force field [12], a highly-elaborate and carefully tuned set of parameters defining the energy of all possible types of atom-atom interactions (covalent, Van der Waals, electrostatic etc). In order to best follow the given set of restraints in a way that has the lowest overall systemic energy following the force field and the restraints, a protocol must be utilized to allow the structure to find the global energy minimum. Simulating annealing is one of

modern molecular dynamics processes which aims to find the global energy minimum with a reasonable amount of computational power [13].

NOESY-derived distance restraints

NOESY peak volumes are inversely correlated to inter-proton distance by R^6, where R is the inter-proton radius. Using this relationship the volumes of all assigned NOESY peaks were taken and used to calculate into distance restraints. Calibration of distance restraints was done using the program CYANA 2.1 [14]. The calibrated distances were compared to H5-H6 distances (cytosine bases), which are independent of the conformation. Each sample had a set of NOESY volumes recorded at various mixing times. These differing NOESYs each had their own peak set whose volumes were converted to distances. The sets of distance restraints for each proton-proton pair were averaged over the available NOESY data and combined into final NOESY-derived restraints. Due to symmetrical nature of the sequence, and absence of any evidence to the contrary, the conformation of the two strands were assumed to be identical and the NOE-distances were replicated for the opposite strand.

The peaks from which the distances were derived have varying amount of uncertainty due to relaxation effects, overlaps and other reasons. Peaks with smaller volumes have a lower signal-to-noise ratio, and thus more uncertainty. Since the relationship between peak volume and distance is inversely proportional, longer NOE-derived distances have a greater uncertainty than shorter distances Therefore, when calculating the structure, the NOE-derived distances (called simply NOEs) were not expected to strictly adhere to the exact calculated NOE distance. Instead, the distance between two atoms was allowed to deviate away from the calculated NOE distances in a distance-dependent manner to fall within a range of values without penalty. Three groups were created: those NOEs which have short distances of less than 3 Å, those which have medium

distances between 3 and 5 Å, and long distance NOEs which are greater than 5 Å. The shorter distance NOEs were given the least amount of no penalty tolerance (typically 0.5 Å) to deviate away from the assigned distances and the tolerance given was increased as the distances were increased (typically 0.7 Å for the medium distance interactions and 0.9 Å for the long distance interactions). Methyl-group containing distances were sectioned off into their own category, as each of the three individual protons attached to the methyl carbon contribute to the NOE peak volume, markedly increasing the peaks' uncertainty. Methyl-group containing NOEs were given a wide no penalty tolerance across all samples of 1.8 Å.

Hydrogen bonding restraints

Hydrogen bonding restraints were generated using AMBER 12 based off the Watson-Crick hydrogen bonding pattern observed in the 2D NOESY recorded in 90% H_2O. The peak pattern observed allowed the justified use of enforcement of canonical hydrogen bonds. The values used for hydrogen bond distances were generated automatically from AMBER using the oligonucleotide sequence and the FF14 force field [15].

Torsion angle restraints

Torsion angle restraints were confined to a general range (BI for the backbone, anti- for the glycosidic Chi angle) unless the modification was found to disturb the backbone at any point. The general torsion angle restraints were applied from a statistical analysis of a compilation of previously solved structures [16]. If a downfield phosphorus chemical shift was observed in the HETCOR, that phosphate was considered to be in BII. BII is defined by the ε-$\zeta > 20°$. To establish this, the ε was confined to 215-295° and ζ within 155-205° [17].

Starting structures

To begin the structure calculations, AMBER package requires a starting structure which is used to apply restraints and perform simulated annealing. From the Drew-Dickerson sequence and force field modifications for the modified residues, a starting structure of perfect A- and B-DNA was generated for each modified sequence using the AMBER package. This was accomplished by first making perfect canonical Drew-Dickerson B- and A-DNA, loading the modified residues into the sequence and performing energy minimization upon the created structures.

Simulated annealing protocol

Restrained molecular dynamics simulations (rMD) were conducted via a simulated annealing protocol with AMBER 12 using the parm99 force field [18]. The generalized Born implicit solvation model was used [19]. The starting structures used were of A- and B-DNA type models. The starting structures for each sample were subjected to an initial energy minimization of 100 cycles starting with 50 steps using the steepest descent algorithm. A force constant of 32 kcal mol^{-1} A^{-2} was applied to all NOE and H-bonding restraints. For backbone torsion restraints, a force constant of 32 kcal mol^{-1} deg^{-2} was applied for all angles with the exception of epsilon and zeta of each oxoG nucleotide, which had a force constant of 512 kcal mol^{-1} deg^{-2}. The temperature of the system was increased from 0 K to the high ("target") temperature during the first 5 ps with a coupling periodicity of 0.4 ps. The weight of all restraints was gradually increased from 0.1 to 1.0 over the first 3 ps. The simulated annealing protocols were performed using varying high "target" temperature values (580 K, 600 K and 620 K) and held at that temperature for varying times (90 ps, 100 ps and 110 ps). Cooling the system from the target temperature to 100 K was accomplished over 100 ps with a coupling of 4 ps. The system was cooled from 100 K to 0 K over 10 ps with a coupling of 1.0 to 0.05 ps. The resulting structures were subjected to energy minimization until

convergence. This protocol produced a total of 20 structures, 10 from A-DNA and 10 from B-DNA starting structures for each modified sample.

Determination of the representative ensemble

Of the total 20 structures generated through the rMD protocol, the best were chosen as the representative ensemble. The best structures were chosen by first creating an averaged minimized structure of all 20 rMD produced structures using the CPPTRAJ software tool [20]. The RMSD between each individual structure and the averaged minimized structure was taken. Then, the overall energy of each produced structure was graphed against the against-average RMSD. Those structures which were located in the low-RMSD low-energy population were chosen to be a part of the representative ensemble. After completion of this step, the restraint violations for the members of the representative ensemble were scrutinized for unacceptable violations (more than 0.1 Å or 15° for distance and torsion angle restraint, respectively). From the representative ensemble, a representative structure was created by averaging the representative ensemble structures with the CPPTRAJ module in AMBER.

Structure Analysis

Detailed structural analysis was performed using a variety of programs. For structure visualization and RMSD calculation the program Visual Molecular Dynamics was used [21]. RMSDs were considered with all possible residues fit; both the fitting of structures with all atoms excluding hydrogens and backbone only RMSDs were considered. For analysis of helical parameters, 3DNA was used [22]. Helical parameters were compared between individual samples, and previously solved structures of the same sequence when available, and were considered against defined standards [23].

References

[1] B. A. Messerle, G. Wider, G. Otting, C. Weber, and K. Wüthrich, "Solvent suppression using a spin lock in 2D and 3D NMR spectroscopy with H 2O solutions," *J. Magn. Reson.*, vol. 85, pp. 608–613, 1989.

[2] J. Kemmink, R. Boelens, T. Koning, G. A. van der Marel, J. H. van Boom, and R. Kaptein, "1H NMR study of the exchangeable protons of the duplex d(GCGTTGCG).d(CGCAACGC) containing a thymine photodimer containing a thymine photodimer," *Nucleic Acids Res.*, vol. 15, no. 11, pp. 4645–4653, Jun. 1987.

[3] D. R. Hare, D. E. Wemmer, S.-H. Chou, G. Drobny, and B. R. Reid, "Assignment of the non-exchangeable proton resonances of d(C-G-C-G-A-A-T-T-C-G-C-G) using two-dimensional nuclear magnetic resonance methods," *J. Mol. Biol.*, vol. 171, no. 3, pp. 319–336, Dec. 1983.

[4] H. E. Gottlieb, V. Kotlyar, and A. Nudelman, "NMR chemical shifts of common laboratory solvents as trace impurities," *J. Org. Chem.*, vol. 62, no. 21, pp. 7512–7515, 1997.

[5] T. L. James, "Relaxation matrix analysis of two-dimensional nuclear Overhauser effect spectra," *Curr. Opin. Struct. Biol.*, vol. 1, no. 6, pp. 1042–1053, Dec. 1991.

[6] M. R. Conte, C. J. Bauer, and A. N. Lane, "Determination of sugar conformations by NMR in larger DNA duplexes using both dipolar and scalar data: Application to d(CATGTGACGTCACATG)2," *J. Biomol. NMR*, vol. 7, no. 3, pp. 190–206, May 1996.

[7] S. Schroeder, J. M. Fu, C. R. Jones, and D. G. Gorenstein, "Assignment of phosphorus-31 and nonexchangeable proton resonances in a symmetrical 14 base pair lac pseudooperator DNA fragment," *Biochemistry (Mosc.)*, vol. 26, no. 13, pp. 3812–3821, Jun. 1987.

[8] J. Ott and F. Eckstein, "Phosphorus-31 NMR spectral analysis of the dodecamer d(CGCGAATTCGCG)," *Biochemistry (Mosc.)*, vol. 24, no. 10, pp. 2530–2535, May 1985.

[9] B. Heddi, N. Foloppe, N. Bouchemal, E. Hantz, and B. Hartmann, "Quantification of DNA BI/BII Backbone States in Solution. Implications for DNA Overall Structure and Recognition," *J. Am. Chem. Soc.*, vol. 128, no. 28, pp. 9170–9177, Jul. 2006.

[10] T. Maurer and H. R. Kalbitzer, "Indirect Referencing of31P and19F NMR Spectra," *J. Magn. Reson. B*, vol. 113, no. 2, pp. 177–178, Nov. 1996.

[11] G. M. Clore, A. M. Gronenborn, A. T. Brünger, and M. Karplus, "Solution conformation of a heptadecapeptide comprising the DNA binding helix F of the cyclic AMP receptor protein of Escherichia coli," *J. Mol. Biol.*, vol. 186, no. 2, pp. 435–455, Nov. 1985.

[12] S. Pérez, A. Imberty, S. B. Engelsen, J. Gruza, K. Mazeau, J. Jimenez-Barbero, A. Poveda, J.-F. Espinosa, B. P. van Eyck, G. Johnson, A. D. French, M. L. C. E. Kouwijzer, P. D. J. Grootenuis, A. Bernardi, L. Raimondi, H. Senderowitz, V. Durier, G. Vergoten, and K. Rasmussen, "A comparison and chemometric analysis of several molecular mechanics force fields and parameter sets applied to carbohydrates," *Carbohydr. Res.*, vol. 314, no. 3–4, pp. 141–155, Dec. 1998.

[13] S. Kirkpatrick, "Optimization by simulated annealing: Quantitative studies," *J. Stat. Phys.*, vol. 34, no. 5–6, pp. 975–986, Mar. 1984.

[14] P. Güntert, "Automated NMR Structure Calculation With CYANA," in *Protein NMR Techniques*, A. K. Downing, Ed. Humana Press, 2004, pp. 353–378.

[15] D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, and others, "AMBER 12; University of California, San Francisco: San Francisco, CA, 2012," *There No Corresp. Rec. This Ref.*

[16] S. Arnott and D. W. L. Hukins, “Optimised parameters for A-DNA and B-DNA,” *Biochem. Biophys. Res. Commun.*, vol. 47, no. 6, pp. 1504–1509, Jun. 1972.
[17] M. Trieb, C. Rauch, B. Wellenzohn, F. Wibowo, T. Loerting, and K. R. Liedl, “Dynamics of DNA: BI and BII Phosphate Backbone Transitions,” *J. Phys. Chem. B*, vol. 108, no. 7, pp. 2470–2476, Feb. 2004.
[18] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham III, C. A. Laughton, and M. Orozco, “Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers,” *Biophys. J.*, vol. 92, no. 11, pp. 3817–3829, Jun. 2007.
[19] V. Tsui and D. A. Case, “Theory and applications of the generalized born solvation model in macromolecular simulations,” *Biopolymers*, vol. 56, no. 4, pp. 275–291, Jan. 2000.
[20] D. R. Roe and T. E. Cheatham, “PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data,” *J. Chem. Theory Comput.*, vol. 9, no. 7, pp. 3084–3095, Jul. 2013.
[21] W. Humphrey, A. Dalke, and K. Schulten, “VMD: Visual molecular dynamics,” *J. Mol. Graph.*, vol. 14, no. 1, pp. 33–38, Feb. 1996.
[22] X.-J. Lu and W. K. Olson, “3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures,” *Nucleic Acids Res.*, vol. 31, no. 17, pp. 5108–5121, Sep. 2003.
[23] W. K. Olson, M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X. J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C. S. Tung, E. Westhof, C. Wolberger, and H. M. Berman, “A standard reference frame for the description of nucleic acid base-pair geometry,” *J. Mol. Biol.*, vol. 313, no. 1, pp. 229–237, Oct. 2001.

Chapter 3 – Results Part I: Reported structures

All of the structures of modified DNA sequences reported in this thesis are consistent with canonical A/B-DNA. The modified samples are referred to by their short-hand name assigned in the original sample scheme; please refer back to the introduction section for inquiry about the sample numbers. Samples will be referred to as either “sample #”, or “DD##” to designate that they originate from the Drew-Dickerson sequence, for ease of reference. Reported here are the structures of all five control samples (DD02-DD06), and one target sample (DD08). Of the modified samples, all completed thus far are right-handed double-stranded helices with Watson-Crick base pairing. All information reported here reflects the mostly canonical nature of these helices. Due to the remarkable number of spectral and structural similarities between structures, in this thesis only the differences will be highlighted. The largest structural variations observed among the samples include a change in backbone conformation 3' of 8oxoG. All three of the oxidized samples reported, DD02, DD03 and DD08, contain evidence in the NMR spectra to indicate the backbone conformation 3' of the 8oxoG is in the BII range of the epsilon and zeta dihedral angles. Overall, the observed differences in structure are either small, or adjacent to the modification site, highlighting the stability of the reported duplex DNA solution NMR structures.

Control samples

Sample 2 - CGC(oxoG)AATTCGCG

NMR Data

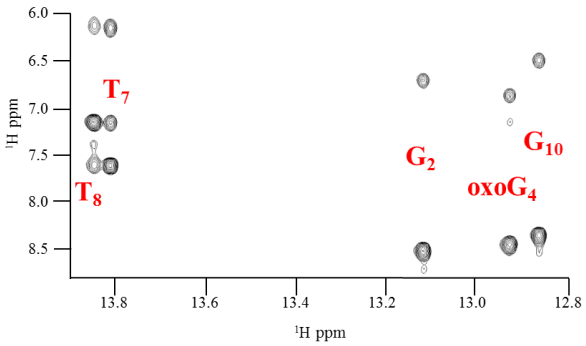


Figure 3.1. The imino region of the H_2O NOESY for DD02 taken on a 700 MHz Bruker Avance with a cryoprobe at 5 °C. This region demonstrates the Watson-Crick base pairing and the strand association within the sample. The labels in red demonstrate the base assignment for the direct dimension.

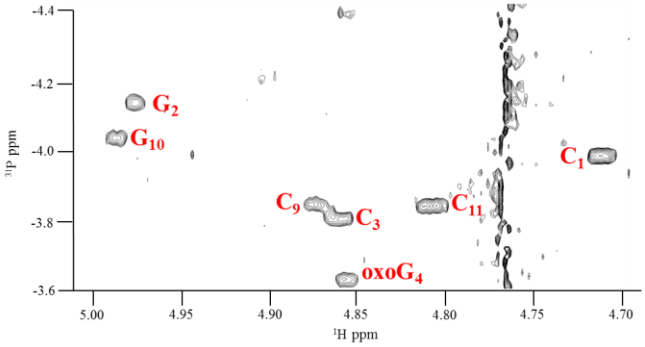


Figure 3.2. The H3' to P region of the ^{1}H-^{31}P HETCOR taken on a 700 MHz Bruker Avance with a cryoprobe at 25 °C for DD02. Each peak is labelled in red with the residue number to which to the H3' chemical shift aligns.

All NMR data for DD02 was obtained using a 700 MHz Bruker Avance spectrometer with a cryoprobe at the facilities of the Siberian branch of the Russian Academy of Sciences in coordination with our collaborator, Dr. Dmitry Zharkov (Institute of Chemical Biology and Fundamental Medicine). The acquired spectra include a 200 ms H_2O NOESY **(Figure 3.1)** taken at 5 °C. All other spectra were taken at 25 °C in D_2O and include: a 120 ms TOCSY, a 120 ms DQF-COSY, a 200 ms NOESY, a 120 ms NOESY, a 70 ms NOESY, and a ^{1}H/^{31}P HETCOR **(Appendix Table #1)**. Of note, the HETCOR showed one ^{31}P chemical shift that was farther downfield than the other peaks by approximately 0.3 ppm **(Figure 3.2)**. Through spectral assignment, it was determined that this peak corresponds to the phosphate between oxoG$_4$ and A$_5$. In the NOESY, there were no corresponding oxoG$_4$ H2' to A$_5$

H2' peak to corroborate that the oxoG_4-A_5 step was in the BII. The lack of this peak may be due to the remarkably low sample concentration, as typically these long-distance peaks do not have a high signal to noise ratio. Other than the HETCOR, all spectra were similar to the published NMR data for the unmodified sequence [1]. The observed differences can be directly attributed to the modification of G_4 to oxoG_4, such as the loss of a guanine base proton at that position and the addition of the HN7 in the H_2O NOESY.

Representative Ensemble

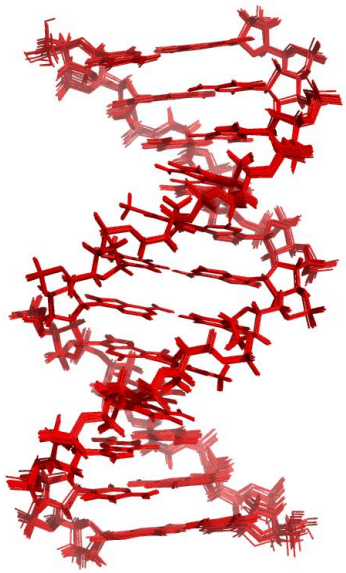


Figure 3.3. The representative ensemble for DD02 shown with all members of the ensemble in red. All members of the ensemble are fit to the averaged minimized structure.

The NMR solution structure representative ensemble for DD02 was completed using restrained molecular dynamics (rMD) with AMBER 12. There were 189 unique NOE-derived distance restraints, with 50 of those in the short-distance category, 138 in the mid-distance category and 1 NOE in the long-distance category. The no penalty tolerance allowed to each NOE-derived distance was dependent upon the category, with tolerances of 0.45, 0.65 and 0.85 Å used for each category ascending in distance. The NOE-derived restraints were duplicated for the other strand due to the symmetry of the structure observed in the NOESY spectra. The other applied experimental restraints included 64 hydrogen bonding distances derived from the NOESY data recorded in 90% H_2O and 120 dihedral restraints based on the observed right-handed A/B-nature of the duplex (NOESY, COSY NMR data). Of the 20 structures generated from the rMD, 13 were determined to be the lowest energy structures and included in the representative ensemble **(Figure 3.3)**. The averaged minimized structure was

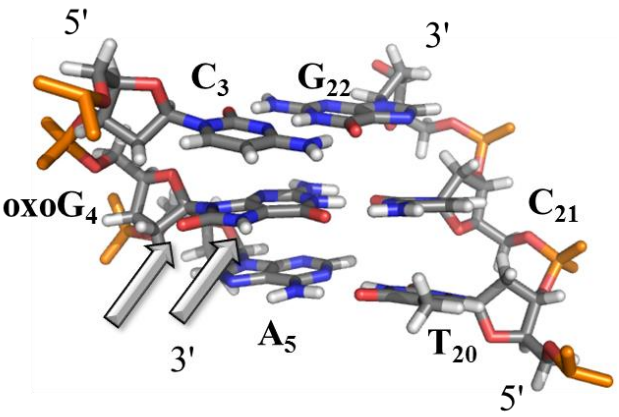


Figure 3.4. Three base pairs of the DD02 structure featuring the target CpG site and the base pair which proceeds it. The modification is highlighted with the grey arrows, both the 8-oxo and HN7 are shown in the major groove. Coloration is in accordance with the elemental composition, while labels are given for the bases as well as the 5' and 3' ends of backbone.

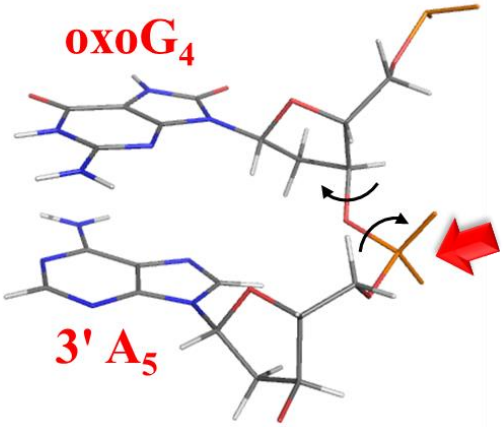


Figure 3.5. The DD02 BII backbone conformation directly 3' of the oxoG modification site. Black arrows show the torsion angles ε and ζ which differ in the BII conformation when compared to the BI conformation. The red arrow highlights the phosphate which is responsible for the signal in the ^{31}P NMR spectra. The bases are labelled in red.

created from these 13 lowest energy structures. The root mean squared deviation (RMSD) among each member of the ensemble compared to the average was less than 0.71 Å. When comparing every member of the ensemble to all other members, all the pair-wise RMSD values were less than 0.78 Å. Of the most non-canonical helical parameters in the representative structure, a single base pair's shear at C_3:G_{22} and two base pair steps' twists, C_3-oxoG_4 and C_9-G_{10}, deviated from the standard helical parameters for A and B-DNA by more than 2σ **(Figure 3.4, Appendix Table 2)** [2]. Local helical parameters incline, X-displacement and helical rise also deviated by 2σ at the base steps C_3-oxoG_4 and C_9-G_{10}. The backbone is in BII between base pairs oxoG_4 and A_5, and in BI for all other steps **(Figure 3.5, Appendix Table 3).** Importantly, all these notably perturbed structural features are adjacent to the oxoG_4 modified residue.

Sample 3 - CGCGAATTC(oxoG)CG

NMR Data

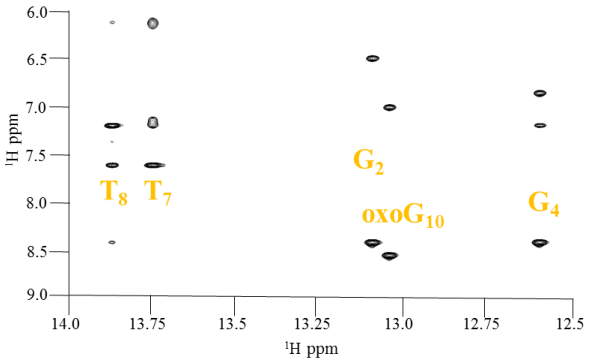


Figure 3.6. The imino region of the H_2O NOESY for DD03 taken on a 700 MHz Bruker Avance with a cryoprobe at 5 °C. This region demonstrates the Watson-Crick base pairing and the strand association within the sample. The labels in orange demonstrate the base assignment for the direct dimension.

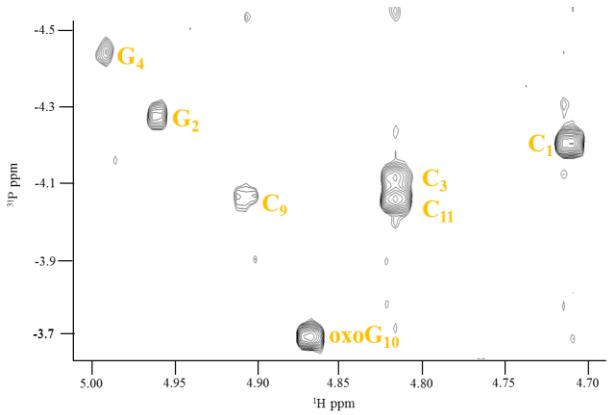


Figure 3.7. The H3' to P region of the ^{1}H-^{31}P HETCOR taken on a 700 MHz Bruker Avance with a cryoprobe at 25 °C for DD03. Each peak is labelled in orange with the residue number to which to the H3' chemical shift aligns.

All NMR data for DD03 was obtained using a 700 MHz Bruker Avance spectrometer with a cryoprobe at the facilities of the Siberian branch of the Russian Academy of Sciences in coordination with our collaborator, Dr. Dmitry Zharkov (Institute of Chemical Biology and Fundamental Medicine). The acquired spectra include a 200 ms H_2O NOESY **(Figure 3.6)** taken at 5 °C. All other spectra were taken at 25 °C in D_2O and include: a 120 ms TOCSY, a 120 ms DQF-COSY, NOESY spectra (with mixing times of 200, 140 and 70 ms), and a ^{1}H/^{31}P HETCOR **(Appendix Table 4)**. Of note, the HETCOR showed the G_{10} ^{31}P chemical shift that was farther downfield than the other peaks by around 0.3 ppm **(Figure 3.7).** This is indicative of the phosphate between G_{10} and C_{11} adopting the BII conformation. Additionally, in the D_2O NOESY, a G_{10} H2' to C_{11} H2' peak was observed in all three mixing times, further corroborating that this step features a BII backbone conformation. Other than the HETCOR, all spectra were similar to the published

NMR data for the unmodified sequence [1]. The observed differences can be directly attributed to the modification of G_{10} to oxoG_{10}, such as the loss of a guanine base proton at that position.

Representative Ensemble

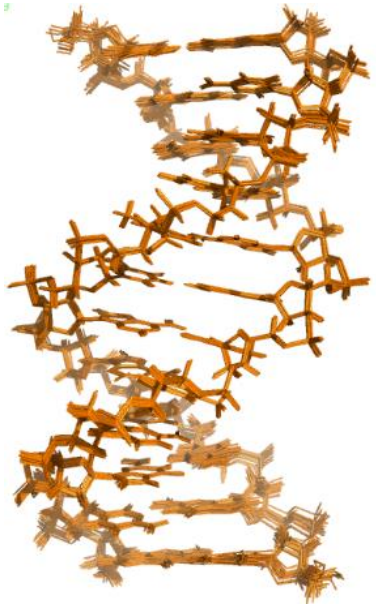


Figure 3.8. The representative ensemble for DD03 shown with all members of the ensemble in orange. All members of the ensemble are fit to the averaged minimized structure.

The NMR solution structure representative ensemble for DD03 was completed using restrained molecular dynamics using AMBER 12. There were 231 unique NOE-derived distance restraints, with 53 of those in the short-distance category, 161 in the mid-distance category and 17 NOEs in the long-distance category. The no penalty tolerance allowed to each NOE-derived distance was dependent upon the category with tolerances of 0.45, 0.65 and 0.85 Å used for each category ascending in distance. The NOE-derived restraints were duplicated for the other strand due to the symmetry of the structure observed in the NOESY spectra. The other applied restraints included 64 hydrogen bonding distance restraints and 120 dihedral restraints. Of the 20 structures generated from the rMD, 14 were determined to be the lowest energy structures and included in the representative ensemble **(Figure 3.8)**. The averaged minimized structure was created from these 14 lowest energy structures. The RMSD amongst each member of the ensemble compared to the average was less than 0.83 Å. When comparing every member of the ensemble to all other members, the pair-wise RMSD values were lower than 0.92 Å. Of the helical parameters for the representative structure, two base pairs' shear, at base pairs G_4:C_{21} and C_9:G_{16}, deviated from the standard helical parameters for A and B-DNA by more than 2σ **(Figure 3.9, Appendix Table 5)** [2]. Stagger also deviated from the standard

parameters by more than 2σ at base pairs A_5:T_{20} and T_8:A_{17}, as well as slide at base pair steps C_3-G_4 and C_9-oxoG_{10}. Local helical parameters inclination, X-displacement and helical rise deviated by 2σ at C_3-G_4 and C_9-oxoG_{10}. The backbone is in BII between base pairs oxoG_{10} and C_{11}, and in BI for all other steps **(Figure 3.10, Appendix Table 6).**

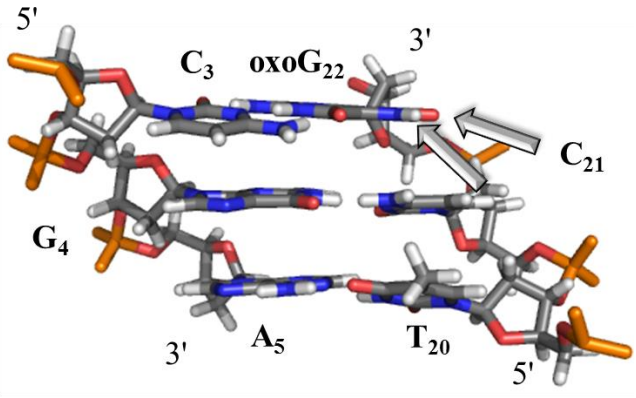


Figure 3.9. Three base pairs of the DD03 structure featuring the target CpG site and the base pair which proceeds it. The modification is highlighted with the grey arrows, both the 8-oxo and HN7 are shown in the major groove. Coloration is in accordance with the elemental composition, while labels are given for the bases as well as the 5' and 3' ends of backbone.

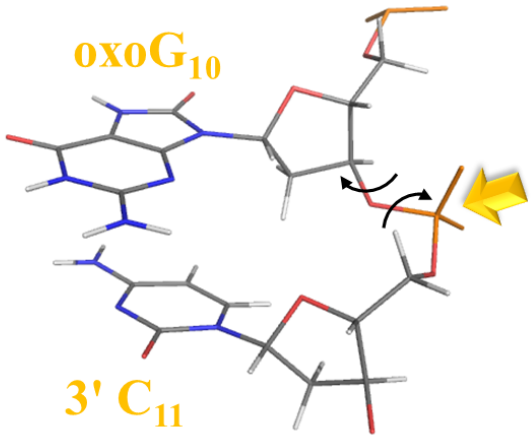


Figure 3.10. The DD03 BII backbone conformation directly 3' of the oxoG modification site. Black arrows show the torsion angles ε and ζ which differ in the BII conformation when compared to the BI conformation. The orange arrow highlights the phosphate which is responsible for the signal in the ^{31}P NMR spectra. The bases are labelled in orange.

Sample 4 - CG(5metC)GAATTCGCG

NMR Data

All NMR data for DD04 was obtained at a 500 MHz on a Varian Inova. The acquired spectra include a 200 ms H_2O NOESY **(Figure 3.11)** taken at 5 °C. All other spectra were taken at 25 °C in D_2O and include: a 120 ms TOCSY, a 120 ms DQF-COSY, NOESY (with mixing times of 260, 200 and 140 ms), **(Appendix Table 7)**. One dimensional ^{31}P NMR was also taken that showed there were no ^{31}P peaks shifted away from the usual range of peaks clustered within 0.55 ppm. All spectra were similar to the published NMR data for the unmodified sequence [1], with observed differences that can be directly attributed to the modification of C_3 to $5metC_3$, such as the addition of interactions incorporating the new methyl.

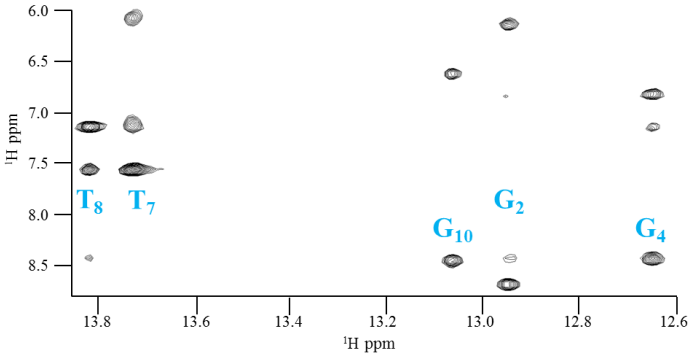


Figure 3.11. The imino region of the H_2O NOESY for DD04 taken on a 500 MHz Varian Inova at 5 °C. This region demonstrates the Watson-Crick base pairing and the strand association within the sample. The labels in cyan demonstrate the base assignment for the direct dimension.

Representative Ensemble

The NMR solution structure representative ensemble for DD04 was completed using restrained molecular dynamics with AMBER 12. There were 159 unique NOE-derived distance restraints, with 35 of those in the short-distance category, 122 in the mid-distance category and 2 NOEs in the long-distance category. The no penalty tolerance allowed to each NOE-derived distance was dependent upon the category with tolerances of 0.5, 0.7 and 0.9 Å used for each category ascending

in distance. The NOE-derived restraints were duplicated for the other strand due to the symmetry of the structure observed in the NOESY spectra. The other applied restraints included 64 hydrogen bonding distance restraints and 120 dihedral restraints. Of the 20 structures generated from the rMD, 8 were determined to be the lowest energy structures and included in the representative ensemble **(Figure 3.12)**. The averaged minimized structure was created from these 8 lowest energy structures. The RMSD amongst each member of the ensemble compared to the average was less than 0.86 Å. When comparing every member of the ensemble to all other members, the pair-wise RMSD values were less than 1.0 Å. Of the helical parameters for the representative structure, shear for base pairs G_2:C_{21}, C_3:G_{22}, G_4:C_{21}, C_9:G_{16}, G_{10}:C_{15}, and C_{11}:G_{14} deviated from the standard helical parameters for A and B-DNA by more than 2σ **(Figure 3.13, Appendix Table 8)** [2]. Roll deviated by more than 2σ as well, at base steps

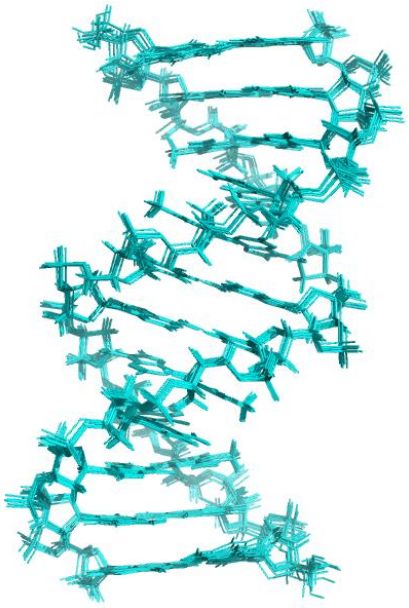


Figure 3.12. The representative ensemble for DD04 shown with all members of the ensemble in cyan. All members of the ensemble are fit to the averaged minimized structure.

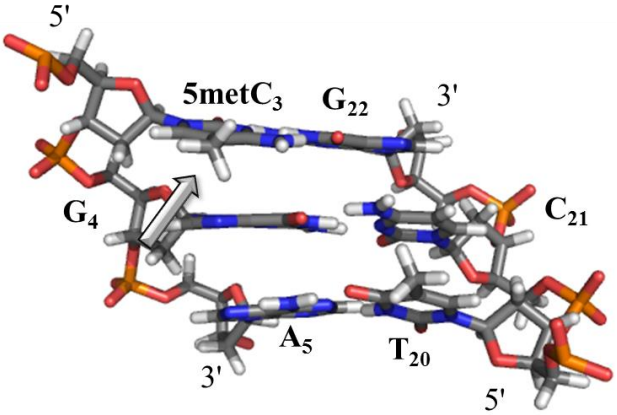


Figure 3.13. Three base pairs of the DD04 structure featuring the target CpG site and the base pair which proceeds it. The modification is highlighted with the grey arrow, with the methyl shown in the major groove. Coloration is in accordance with the elemental composition, while labels are given for the bases as well as the 5' and 3' ends of backbone.

5metC_3-G_4 and C_9-G_{10}. Local helical parameters also deviated by 2σ for inclination, X-displacement and helical rise at steps 5metC_3-G_4 and C_9-G_{10}. The differences in helical parameters all occur within the modified CpG site as expected. The backbone is in BI for all steps **(Appendix Table 9).**

Sample 5 - CGCGAATT(5metC)GCG

NMR Data

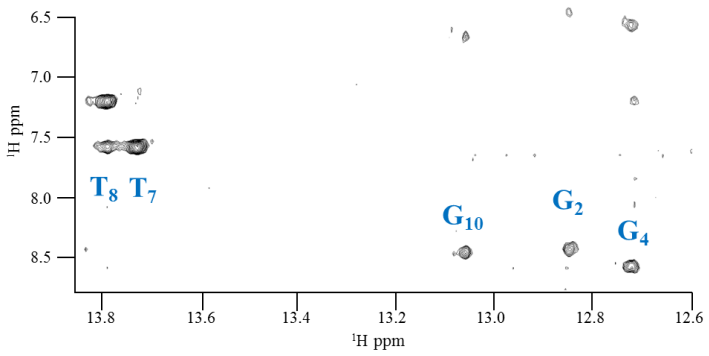


Figure 3.14. The imino region of the H_2O NOESY for DD05 taken on a 500 MHz Varian Inova at 5 °C. This region demonstrates the Watson-Crick base pairing and the strand association within the sample. The labels in blue demonstrate the base assignment for the direct dimension.

NMR data for DD05 was obtained using a 500 MHz Varian Inova and a 500 MHz Bruker Avance III HD with a Prodigy cryoprobe at a Bruker Co. facility (Fremont, CA). The acquired spectra include a 200 ms H_2O NOESY **(Figure 3.14)** taken at 5 °C on the Varian Instrument. Spectra acquired on the 500 MHz Varian Inova instrument also includes a 120 ms TOCSY and a 120 ms DQF-COSY, both in D_2O at 25 °C. The remaining D_2O spectra were taken on the Bruker Avance and include: 260 ms NOESY, a 200 ms NOESY, and a 140 ms NOESY **(Appendix Table 10)**. One dimensional ^{31}P NMR was also taken that showed there were no ^{31}P peaks shifted away from the usual range of clustered peaks. All spectra were similar to the published NMR data for the unmodified sequence [1], with observed differences that can be directly attributed to the modification of C_9 to 5metC_9, such as the addition of interactions incorporating the new methyl.

Representative Ensemble

The NMR solution structure representative ensemble for DD05 was completed using restrained molecular dynamics with AMBER 12. There were 158 unique NOE-derived distance restraints, with 43 of those in the short-distance category, 113 in the mid-distance category and 2 NOEs in the long-distance category. The no penalty tolerance allowed to each NOE-derived distance was dependent upon the category with tolerances of 0.5, 0.7 and 0.9 Å used for each category, ascending in distance. The NOE-derived restraints were doubled for the other strand due to the symmetry of the structure observed in the NOESY spectra. The other applied restraints included 64 hydrogen bonding distance restraints and 120 dihedral restraints. Of the 20 structures generated from the rMD, 8 were determined to be the lowest energy structures and included in the representative ensemble **(Figure 3.15)**. The averaged minimized structure was created from these 8 lowest energy structures. The all atom pair-wise RMSD amongst each member of the ensemble compared to the average was less than 0.49 Å. When

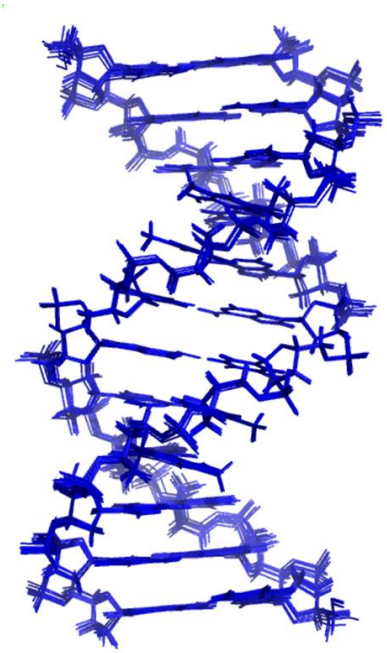


Figure 3.15. The representative ensemble for DD05 shown with all members of the ensemble in blue. All members of the ensemble are fit to the averaged minimized structure.

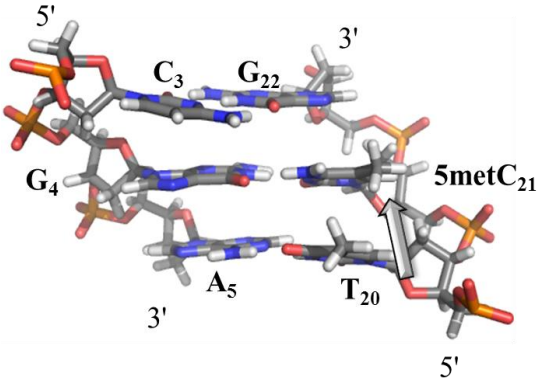


Figure 3.16. Three base pairs of the DD05 structure featuring the target CpG site and the base pair which proceeds it. The modification is highlighted with the grey arrow, with the methyl shown in the major groove. Coloration is in accordance with the elemental composition, while labels are given for the bases as well as the 5' and 3' ends of backbone.

comparing every member of the ensemble to all other members, the pair-wise RMSD was less than 0.89 Å. Of the helical parameters for the representative structure, shear for base pairs G_2:C_{21}, C_3:G_{22}, G_4:C_{21}, C_9:G_{16}, G_{10}:C_{15}, and C_{11}:G_{14} deviated from the standard helical parameters for A and B-DNA by more than 2σ **(Figure 3.16, Appendix Table 11)** [2]. Roll deviated as well, at base steps C_3-G_4 and 5metC_9-G_{10}. Helical parameters also deviated by 2σ for inclination, X-displacement and helical rise at steps C_3-G_4 and 5metC_9-G_{10}. All deviations appear at or adjacent to the modified CpG site, as anticipated. The backbone is in BI for all steps **(Appendix Table 12).**

Sample 6 - CG(5metC)AATT(5metC)GCG

NMR Data

All NMR data for DD06 was obtained using a 500 MHz Varian Inova. The acquired spectra include a 200 ms H_2O NOESY **(Figure 3.17)** taken at 5 °C. All other spectra were taken at 25 °C in D_2O and include: a 120 ms TOCSY, a 120 ms DQF-COSY, a 260 ms NOESY, a 200 ms NOESY, a 140 ms NOESY, and a 90 ms NOESY **(Appendix Table 13)**. One dimensional ^{31}P NMR was also taken that showed there were no ^{31}P peaks shifted away from the usual range of clustered peaks. All spectra were similar to the published NMR data for the unmodified sequence [1], with observed differences that can be directly attributed to the modification of C_3 to 5metC_3 and C_9 to 5metC_9.

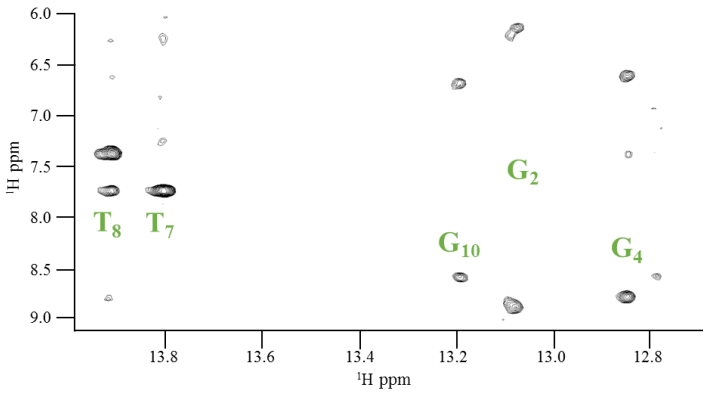


Figure 3.17. The imino region of the H_2O NOESY for DD06 taken on a 500 MHz Varian Inova at 5 °C. This region demonstrates the Watson-Crick base pairing and the strand association within the sample. The labels in green demonstrate the base assignment for the direct dimension.

Representative Ensemble

The NMR solution structure representative ensemble for DD06 was completed using restrained molecular dynamics using AMBER 12. There were 140 unique NOE-derived distance restraints, with 34 of those in the short-distance category, 103 in the mid-distance category and 3 NOEs in the long-distance category. The no penalty tolerance allowed to each NOE-derived distance was dependent upon the category with tolerances of 0.5, 0.7 and 0.9 Å used for each category ascending in distance. The NOE-derived restraints were doubled for the other strand due to the symmetry of the structure observed in the NOESY spectra. The other applied restraints included 64 hydrogen bonding distance restraints and 120 dihedral restraints. Of the 20 structures generated from the rMD, 10 were determined to be the lowest energy structures and included in the representative ensemble **(Figure 3.18)**. The averaged minimized structure was created from these 10 lowest energy structures. The RMSD amongst

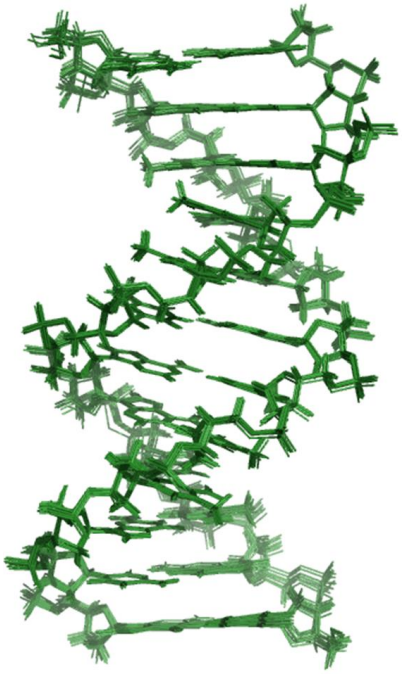


Figure 3.18. The representative ensemble for DD06 shown with all members of the ensemble in green. All members of the ensemble are fit to the averaged minimized structure.

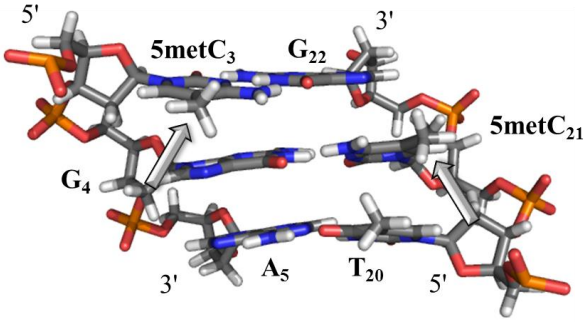


Figure 3.19. Three base pairs of the DD06 structure featuring the target CpG site and the base pair which proceeds it. The modifications are highlighted with the grey arrows, with both the methyl groups shown in the major groove. Coloration is in accordance with the elemental composition, while labels are given for the bases as well as the 5' and 3' ends of backbone.

each member of the ensemble compared to the average was less than 0.83 Å. When comparing every member of the ensemble to all other members, the pair-wise RMSD was less than 0.91 Å. Of the helical parameters for the representative structure, shear, roll, twist and shift deviated from the standard helical parameters for A and B-DNA by more than 2σ **(Appendix Table 14)** [2]. The base pairs where the differences occured were: shear for base pairs G_2:5metC_{21}, 5metC_3:G_{22}, G_{10}:5metC_{15}, C_{11}:G_{14}, shift at G_4-A_5 and T_8-5metC_9, twist and roll for steps 5metC_3-G_4 and 5metC_9-G_{10} **(Figure 3.19)**. Local helical parameters also deviated by 2σ for inclination, X-displacement and helical rise at steps 5metC_3-G_4 and 5metC_9-G_{10}. The backbone is in BI for all steps **(Appendix Table 15)**, though the step G_4-A_5 approaches BII closer than any other BI step**.**

Target samples

Sample 8 – CGC(oxoG)AATT(5metC)GCG

NMR Data

NMR data for DD08 was obtained using 500 MHz Bruker Avance III HD spectrometer with a cryoprobe at a Bruker Co. facility in Fremont, CA. All spectra were taken at 25 °C in D_2O and include: a 120 ms TOCSY, a 120 ms DQF-COSY, a 260 ms NOESY, a 200 ms NOESY, a 140 ms NOESY, a 90 ms NOESY, and a $^1H/^{31}P$ HETCOR **(Appendix Table 16)**. Of note, the HETCOR showed one ^{31}P chemical shift that was farther downfield than the other peaks by around 0.3 ppm **(Figure 3.20)**. Other than the HETCOR, all spectra were similar to the published NMR data for the unmodified sequence [1].

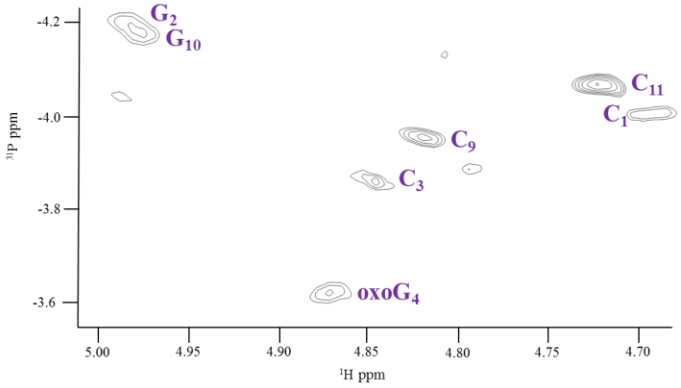


Figure 3.20. The H3' to P region of the ^1H-^{31}P HETCOR taken on a 500 MHz Bruker Avance with a cryoprobe at 25 °C for DD08. Each peak is labelled in purple with the residue number to which to the H3' chemical shift aligns.

The observed spectral differences can be directly attributed to the modification of G_4 to oxoG_4 and C_9 to 5metC_9.

Representative Ensemble

The NMR solution structure representative ensemble for DD08 was completed using restrained molecular dynamics with AMBER 12. There were 172 unique NOE-derived distance restraints, with 45 of those in the short-distance category, 117 in the mid-distance category and 10 NOEs in the long-distance category. The no penalty tolerance allowed to each NOE-derived distance was dependent upon the category with tolerances of 0.5, 0.7 and 0.9 Å used for each category ascending in distance. The NOE-derived restraints were doubled for the other strand due to the symmetry of the structure observed in the NOESY spectra. The other applied restraints included 64 hydrogen bonding distance restraints and 120 dihedral restraints. Of the 20 structures generated from the rMD, 11 were determined to be the lowest energy structures and included in the representative ensemble **(Figure 3.21)**. The averaged minimized structure was created from these 11 lowest energy structures. The RMSD amongst each member of the ensemble compared to the average was less than 0.81 Å. When comparing every member of the ensemble to all other members, the pair-wise RMSD was less than 0.89 Å. Of the helical parameters for the representative structure, shear for base pairs G_2:C_{21}, C_3:G_{22}, G_{10}:5metC_{15}, and C_{11}:G_{14}, twist for steps C_3-G_4 and 5metC_9-G_{10}, tilt for steps A_5-A_6 and T_7 to T_8, and propeller for base pairs A_5:T_{20} and T_8:A_{17} deviated from the standard helical

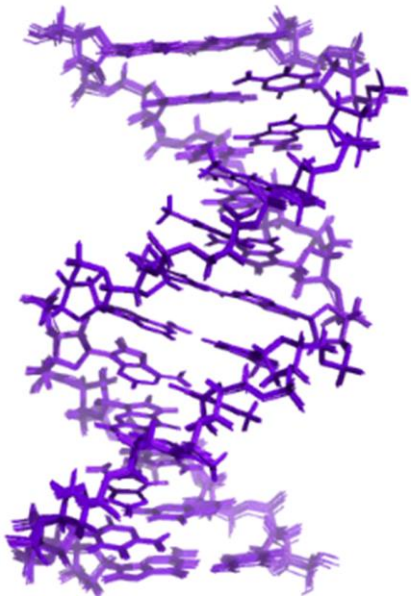


Figure 3.21. The representative ensemble for DD08 shown with all members of the ensemble in purple. All members of the ensemble are fit to the averaged minimized structure.

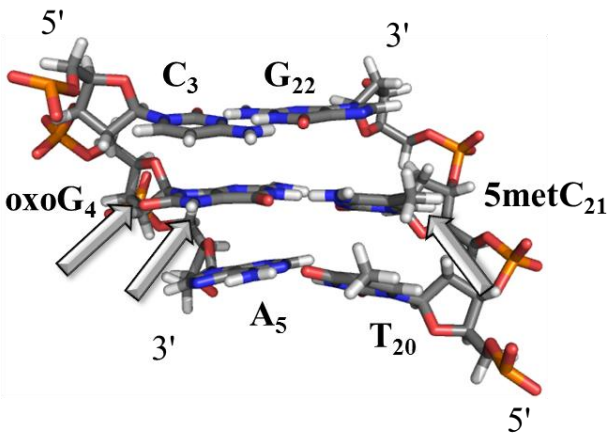


Figure 3.22. Three base pairs of the DD08 structure featuring the target CpG site and the base pair which proceeds it. The modifications are highlighted with the grey arrows, with the methyl group, 8-oxo and HN7 shown in the major groove. Coloration is in accordance with the elemental composition, while labels are given for the bases as well as the 5' and 3' ends of backbone.

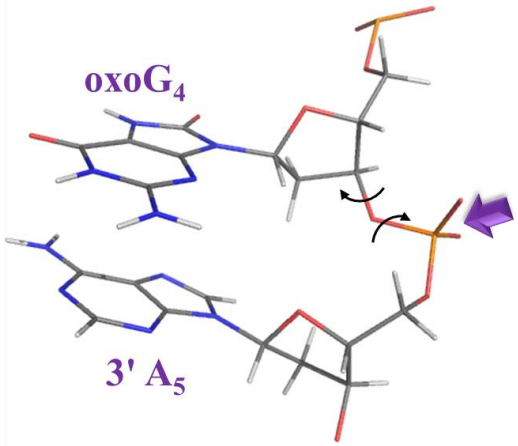


Figure 3.23. The DD08 BII backbone conformation directly 3' of the oxoG modification site. Black arrows show the torsion angles ε and ζ which differ in the BII conformation when compared to the BI conformation. The purple arrow highlights the phosphate which is responsible for the signal in the ^{31}P NMR spectra. The bases are labelled in orange.

parameters for A and B-DNA by more than 2σ **(Appendix Table 17)** [2]. Local helical parameters also deviated by 2σ for inclination, tip, X-displacement, Y-displacement, helical twist and helical rise at steps C_3-oxoG_4 and 5metC_9-G_{10} **(Figure 3.22)**. The backbone is in BII for G_4-A_5, while all other steps are in BI though the A_5-A_6 was closer to BII than any other BI step. **(Figure 3.23, Appendix Table 18)**.

References

[1] W. Nerdal, D. R. Hare, and B. R. Reid, “Solution structure of the EcoRI DNA sequence: refinement of NMR-derived distance geometry structures by NOESY spectrum back-calculations,” *Biochemistry (Mosc.)*, vol. 28, no. 26, pp. 10008–10021, Dec. 1989.

[2] W. K. Olson, M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X. J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C. S. Tung, E. Westhof, C. Wolberger, and H. M. Berman, “A standard reference frame for the description of nucleic acid base-pair geometry,” *J. Mol. Biol.*, vol. 313, no. 1, pp. 229–237, Oct. 2001.

Chapter 4 – Results Part II: Comparisons

Comparisons between reported structures

From the reported structures, it is clear that the modification of the CpG site with oxoG and/or 5metC does not reflect any global structural changes, but does induce local structure perturbations. Chiefly, for the first time in an NMR solution structure we report that oxoG appears to be linked to the non-canonical BII backbone conformation directly 3' of the modified base. This backbone conformation change due to oxoG persists, throughout both change in placement in the CpG site and with the 5metC modification occurring *trans* of the oxoG. 5metC does not seem to induce any further noticeable changes beyond the simple addition of the new methyl group. The target sample with *tran* oxoG and 5metc in the CpG site showed further structural differences in the positions of the base pairs that flank the modified base pair. This comparison of structures outlines a general protocol for analysis of the remaining target structures in the future.

Oxidized control samples

The oxidized control samples, DD02 and DD03, show high agreement when compared to each other and have an all atom RMSD of 0.86 Å. Despite the replacement of guanine with 8oxoG at differing positions, the low RMSD and helical parameters show that oxoG does not perturb the local structure in a dramatic fashion. The largest observed structural difference was the backbone

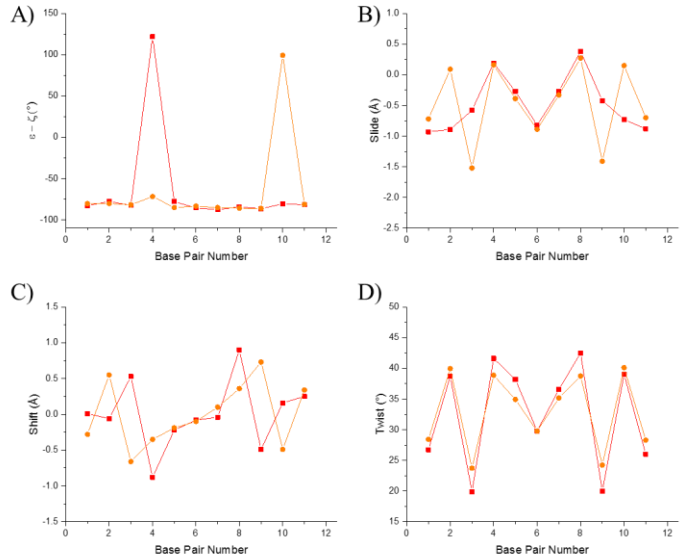


Figure 4.1. The ε-ζ **(A)**, slide **(B)**, shift **(C)** and twist **(D)** values for DD02 and DD03. DD02 is shown is in red closed squares, and DD03 in orange closed circles.

conformation in BII 3' of the modification site between steps oxoG$_4$-A$_5$ or oxoG$_{10}$-C$_{11}$ **(Figure 4.1A)**. The helical parameters were also in good agreement. This is to be expected due to their low RMSD when compared to one another. The largest difference in helical parameters between DD02 and DD03 is in the base pair slide and shift, which may be caused by the differing positions of 8oxoG in the sequence **(Figure 4.1B,C)**. In particular, as the structure of target samples continue to be solved, it would be interesting to see the values for target structures with the same oxidation. Overall, the structures are similar, with the backbone conformation changing locally depending upon the placement of the modification in the sequence.

Methylated control samples

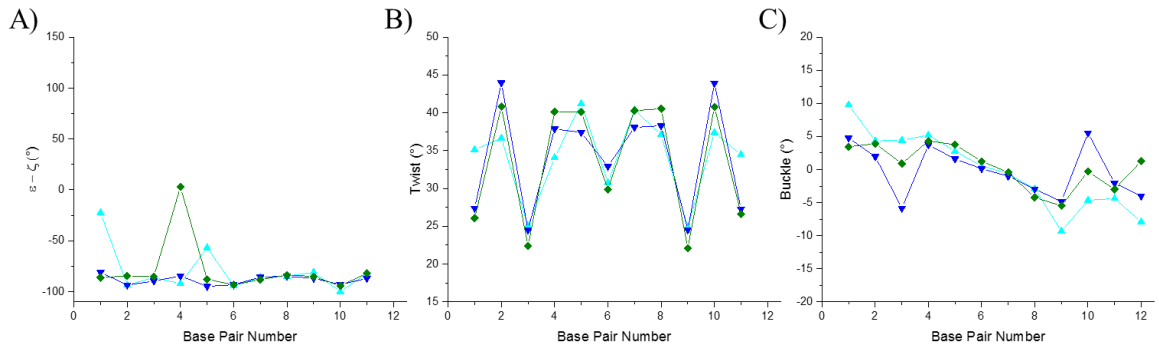


Figure 4.2. The ε-ζ **(A)**, buckle **(B)**, and twist **(C)** values for DD04, DD05 and DD06. DD04 is shown in cyan triangles, DD05 is shown by blue inverted triangles and DD06 in green diamonds.

The methylated control samples, DD04, DD05 and DD06 were in good agreement and showed no remarkable structural changes. The RMSD between the hemi-methylated samples DD04 and DD05, 0.54 Å, was the lowest RMSD observed amongst all the structures considered in this thesis, both published and reported. In accordance with that, the helical parameters changed little throughout the two structures and the backbone conformations were also comparable **(Figure 4.2)**. A notable minor change in buckle is seen at the modification sites of each which is not shared in DD06. For the fully methylated sample DD06, the RMSD against DD04 and DD05 was observed to be 1.0 and 0.89 Å, respectively. For DD06 all helical parameters were comparable to DD04 and DD05. However, there is an increased ε-ζ at the G_4 to A_5 step, which does not quite reach the BII cutoff of 20° or higher. An NMR solution structure is largely an average between rapidly exchanging conformational states [1], so it stands to reason that there is a higher BII propensity in this structure as compared to others, even given the same BI-favoring angle restraints in the rMD. As was discussed previously, the $5metC_3$ and $5metC_9$ substitutions have been shown to increase BII frequency at the modification site [2]. In addition to this increased BII frequency, the structure is slightly untwisted at the site of both methylations **(Figure 4.2C)**. The presence of both methyls in the major groove, along with the pre-existing methyls of the proceeding thymines in the major

groove, could cause this slight untwisting at this step which might contribute to the increased propensity of BII. A similar effect is observed in DD03, where there is a dip in twist for the step that precedes the BII site **(Figure 4.1A, D)**.

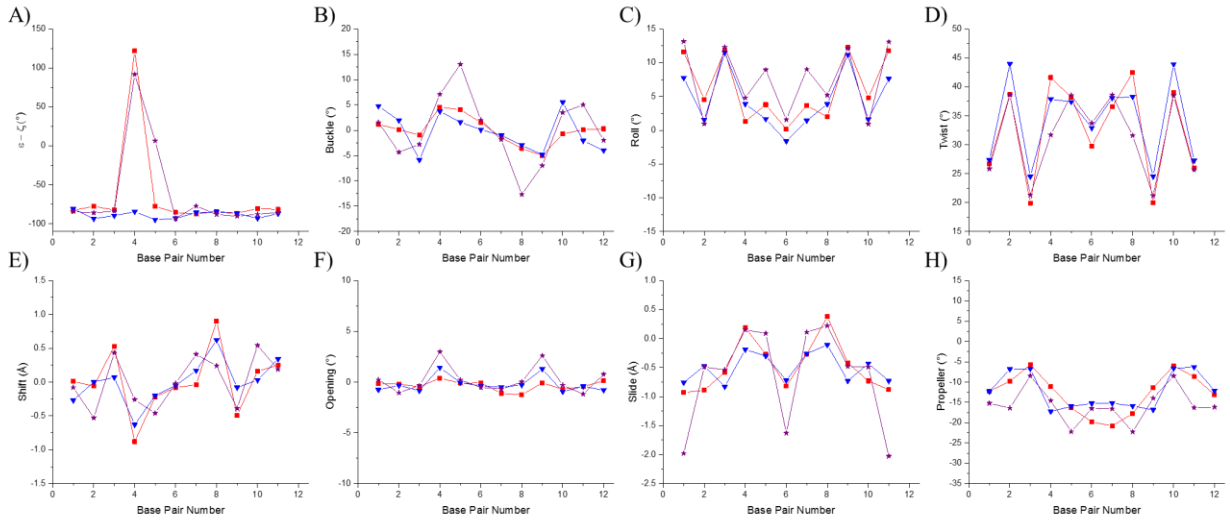


Figure 4.3. The ε-ζ **(A)**, buckle **(B)**, roll **(C)**, twist **(D)**, shift **(E)**, opening **(F)**, slide **(G)**, and propeller **(H)** values for DD02, DD05 and DD08. DD02 is shown in red squares, DD05 in blue inverted triangles and DD08 in purple stars.

Target sample

The target sample DD08 has the modifications featured in DD02 and DD05, oxoG_4 and 5metC_9. Thus, the control samples DD02 and DD05 are by design the best structures to compare DD08 to judge what differences the modifications in tandem produce in the structure. Not surprisingly, DD08 as the most heavily modified structure contains the largest and most numerous differences when compared to the published structures and to our reported structures. DD08 is more similar to its oxidized control sample DD02 than its methylated control sample DD05. The RMSD between DD08 and DD02 is 1.46 Å, while the RMSD against DD03 is just slightly higher at 1.56 Å. DD08 RMSDs against DD04, DD05 and DD08 are 1.80 Å, 1.88 Å, and 1.73 Å, respectively. DD08 more closely matching its oxidized control sample is to be expected, as the largest observed structural change throughout the samples seems to be the backbones states; oxidized samples have

one step outside the CpG site in strict BII conformation, which the methylated samples did not share. In addition to the expected BII conformation after the oxoG modification, DD08 displayed an additional step which approached BII in the center of the structure at the A_5-A_6 step **(Figure 4.3A)**. Like the additional BII-like step observed in DD06, this did not quite reach the ε-ζ value that defines BII, but was dramatically different from all others which shared the same angle restraints. While unusual for an ApA step to approach BII, it is not unprecedented [3].

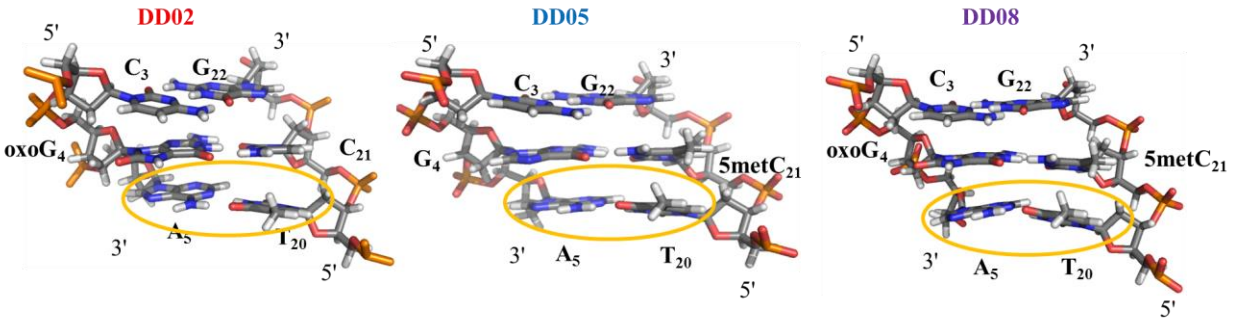


Figure 4.4. The difference in buckle of the base pair underneath the modification sites in DD08 when compared to the control samples DD02 and DD05. The base pair whose buckle value changes in DD08 is shown inside the yellow oval. The labels for each residue are shown in black, with the directionality of the backbone also labelled. DD08 has the A_5:T_{20} base pair buckled upwards towards the modification sites.

The helical parameters for DD08 differ from the control DD02 and DD05 helical parameters in several instances **(Figure 4.3)**. The sources of these differences seem to trace back to either the additional modification in the CpG site, and/or the unusual backbone conformation of the central ApA. The most striking differences are seen in the buckle and slide **(Figure 4.3 B,G)**. The buckle reaches the highest and lowest points seen in any of the reported structures in the step directly after the $oxoG_4$ and the step directly preceding the $5metC_9$. These steps are symmetric steps in the Drew-Dickerson sequence so these extreme high and lows are complementary and make internal sense. These values of the buckle at both steps indicate that the base pairs at each step are shifted along the center axis of the duplex in a way that points the base pair hydrogen bonding contacts closer to the CpG steps and their corresponding end nearer to the glycosidic bond away from the CpG

site **(Figure 4.4)**. As for the parameter slide, it takes a sharp dip directly in the center of the sequence. Slide represents the shifting of base pairs towards either strand's backbone. The extreme negative slide value represents a base pair which is moved away from the backbone of the oxoG-containing strand. Despite these interesting structural differences, the majority of the deviations away from the helical parameters of the control samples suggest that the *trans* substitution of oxoG and 5metC in the same base pair ($oxoG_4$:$5metC_{21}$ and $5metC_9$:$oxoG_{16}$) produces only local structural changes at the CpG site while having some impact on the central base pairs of the structure.

Comparison to published structures

The Drew-Dickerson sequence CGCGAATTCGCG is a widely studied sequence whose structure has been solved using many differing techniques. Due to the well-known nature of this sequence, it is a prime target for modification studies [4]–[6]. Here, all six of the solved structures are compared to three Drew-Dickerson sequence structures solved through three differing techniques. These structures include a crystal structure (PDB ID: 1BNA), an NMR solution structure (PDB ID: 2DAU) and a liquid crystalline NMR structure (PDB ID: 1NAJ). These structures, for ease of reference, will be referred to as their PDB IDs throughout the rest of this chapter. In addition, there have been crystal structures solved for our single methylation modified DD04 and DD05. A structure including oxoG, in a different sequence context, has also been previously reported. This comparison to previously published structures is not the focus of this thesis, but does provide necessary validation of the modified structures by comparing the previously known structural information about the Drew-Dickerson sequence and the subsequent modifications to the reported structures. This external comparison demonstrates the overall stability of the duplex DNA scaffold,

robust nature of our methods and validates the internal comparisons between reported modified structures.

A double-stranded DNA structure is dependent upon salt concentration, pH, temperature, and sequence [7], [8]. None of the previously published structures matched our buffer conditions of 10 mM sodium phosphate buffer (pH 6.8), 50 mM NaCl, and 0.1 mM EDTA exactly. When comparing the reported structures against the published structures, this was taken under consideration. The reported structures have helical parameters that fall within the range of those in the published structures, as well as consistent RMSDs against those structures. The 5metC modified DD04 and DD05 show less similarity to the crystal structures of the same sequence, with an important difference in backbone structure seen at the modification step G_{10} to C_{11}. The only previously published structure that contains oxoG does not have the Drew-Dickerson sequence. However, our oxoG-containing structures feature a BII backbone conformation 3' of the oxoG which is also observed in the oxoG-containing crystal structure. Overall, there were no global differences such as change in handed-ness, helicity or hydrogen bonding to report when comparing the published structures to those reported in this thesis.

1BNA – 1.9 Å resolution crystal structure of the Drew-Dickerson sequence

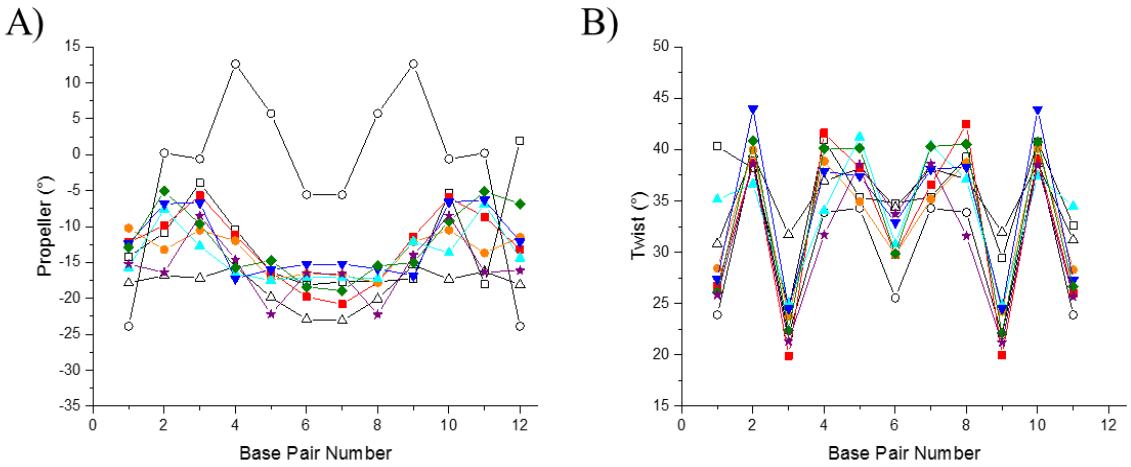


Figure 4.5. The propeller **(A)** and twist **(B)** values for structures of the Drew-Dickerson sequence. The marker and color scheme is as follows: 1BNA (the Drew-Dickerson crystal structure) in black open squares, 2DAU (the Drew-Dickerson NMR solution structure) is in black open circles, 1NAJ (the Drew-Dickerson liquid crystalline NMR structure) in black open triangles, DD02 is in red closed squares, DD03 in orange closed circles, DD04 in cyan triangles, DD05 in blue inverted triangles, DD06 is green diamonds and DD08 in purple stars.

The Drew-Dickerson sequence was the first duplex DNA sequence to have a solved single molecule crystal structure, and the corresponding PDB ID for that structure is 1BNA [9]. 1BNA, at the time, was remarked for its striking difference from perfect B-DNA in the helical parameters twist and propeller. The pattern and values for these parameters are matched in all of the reported structures, as well as for 1NAJ, while 2DAU did not match the other structures in propeller **(Figure 4.5)**. The pattern of twist and propeller appears to be largely sequence-dependent [10]. The largest difference between 1BNA helical parameters and the reported modified structures lies in the parameter stretch **(Figure 1.4, 4.6)**. The longer stretch for our solution structures as compared to the 1BNA crystal structure is due to the hydrogen bonding restraints imposed upon our structures during restrained molecular dynamics and match the ideal hydrogen bonding length. 1BNA has hydrogen bonding lengths which are slightly less than generally accepted, which may have been caused by the salt concentration or general differing conditions that are necessary to crystallize

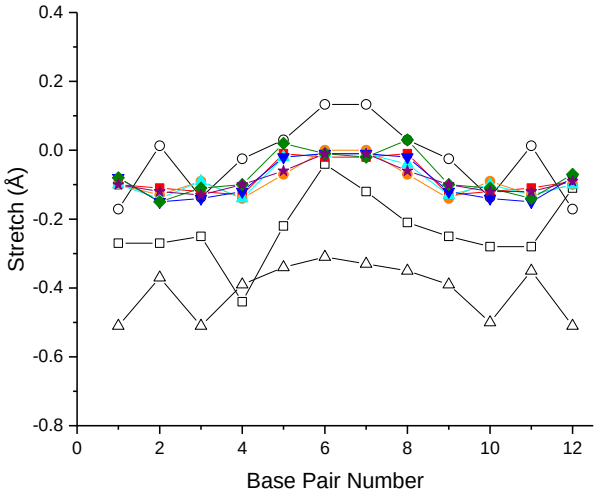


Figure 4.6. The stretch values for structures with the Drew-Dickerson sequence. The marker and color scheme is as follows: 1BNA (the Drew-Dickerson crystal structure) in black open squares, 2DAU (the Drew-Dickerson NMR solution structure) is in black open circles, 1NAJ (the Drew-Dickerson liquid crystalline NMR structure) in black open triangles, DD02 is in red closed squares, DD03 in orange closed circles, DD04 in cyan triangles, DD05 in blue inverted triangles, DD06 is green diamonds and DD08 in purple stars.

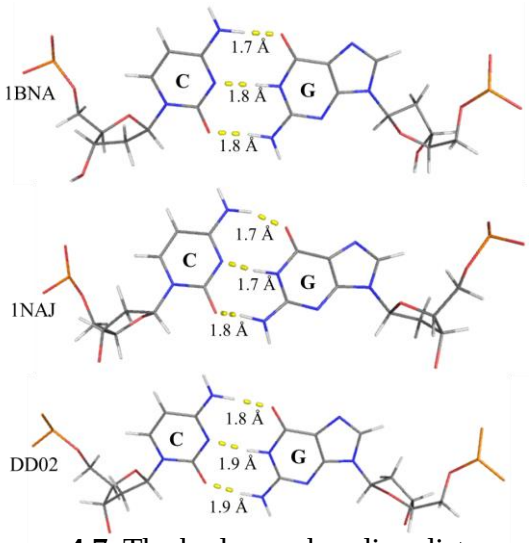


Figure 4.7. The hydrogen bonding distances of the G_2:C_{23} base pair in the Drew-Dickerson sequence for 1BNA, 1NAJ and DD02 showing the differences between the structures. 1BNA (the Drew-Dickerson crystal structure) and 1NAJ (the Drew-Dickerson sequence liquid crystalline NMR structure) have slightly shorter hydrogen bond distances than DD02. In addition, there is some shifting of the bases in relation to one another in the case of 1NAJ.

(Figure 4.7). Alternatively, through X-ray crystallography, there is a more direct observation of the distance between base pairs, as extrapolated from the diffraction pattern. Our observations were largely qualitative based upon H_2O NOESY patterns and lead to the use of force-field-specific hydrogen bonding restraints without directly measuring the distance between the hydrogen bonding donor and acceptor for each base pair. Thus, 1BNA may have hydrogen bonding/stretch values which are closer to DNA *in vivo* than our NMR solution structures contain.

Despite the similarity between 1BNA and our reported structures, the pair-wise RMSDs between them are rather high. The all atom RMSDs between 1BNA and the reported structures range from 3.7 Å for DD05 to 4.6 Å for DD08. The lowest backbone RMSD between 1BNA and the modified samples was 2.2 Å for DD05, with the highest backbone RMSD of 3.9 Å when compared to DD08.

This suggests that the greatest difference between the structures lies in the overall backbone, not the bases themselves. This is in agreement with the observation that the structures have relatively similar base pair helical parameters. The fact that the backbone has different conformations between the two structures is also demonstrated by a major difference between the backbones at the step G_{10}-C_{11}. This step is shown to be in the BII conformation in 8oxoG-containing DNA and this observation is not shared in our sequences which do not contain an 8oxoG at that site **(Figure 4.8, 4.9)**.

2DAU – Solution NMR structure of the Drew-Dickerson sequence

The PDB ID 2DAU is a structure which was solved through traditional NMR solution structure methods [11]. Perhaps surprisingly, the NMR solution structure 2DAU, whose methods more closely match those reported in this thesis, is more dissimilar to our reported structures than the other two unmodified Drew-Dickerson sequence structures solved through different techniques **(Figures 4.5, 4.6)**. However, unlike both 1BNA and 1NAJ, this unmodified Drew-Dickerson structure was not the main focus of the study in which it appeared. While 1BNA and 1NAJ structures were the focus of their papers and used techniques which were relatively novel for

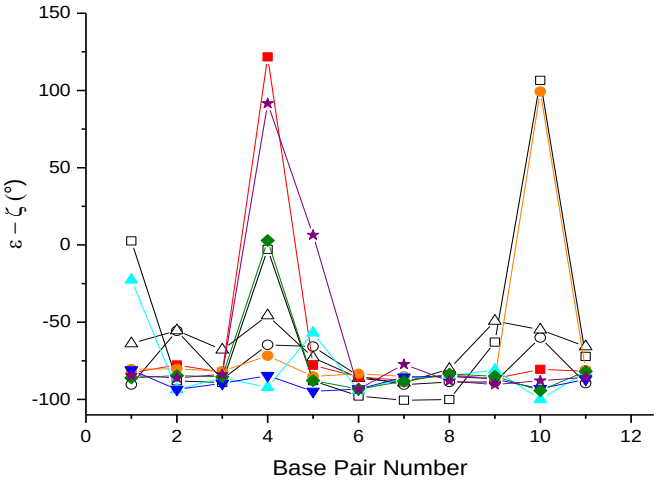


Figure 4.8. The ε–ζ values for structures with the Drew-Dickerson sequence. An ε–ζ > 20° indicates the backbone conformation is in BII The marker and color scheme is as follows: 1BNA (the Drew-Dickerson crystal structure) in black open squares, 2DAU (the Drew-Dickerson NMR solution structure) is in black open circles, 1NAJ (the Drew-Dickerson liquid crystalline NMR structure) in black open triangles, DD02 is in red closed squares, DD03 in orange closed circles, DD04 in cyan triangles, DD05 in blue inverted triangles, DD06 is green diamonds and DD08 in purple stars.

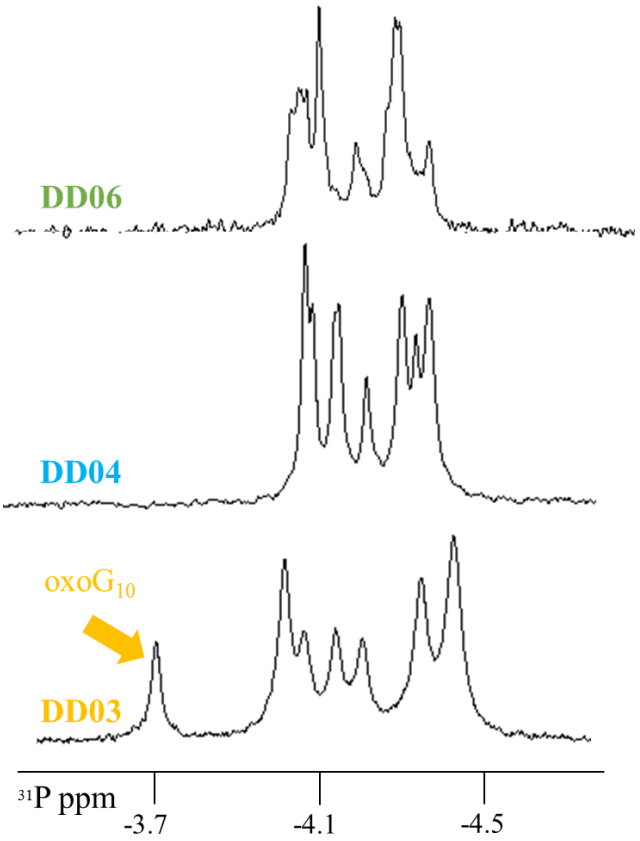


Figure 4.9. ^{31}P 1D spectra for DD03, DD04 and DD06. DD03 represents the oxidation only samples, DD04 represents the single methylation samples and DD06 represents the only fully methylated sample. A downfield peak is clearly seen in DD03, which indicates BII. This downfield peak is not seen in DD04 or DD06, neither is it seen in DD05 (not shown). This demonstrates that we do not have evidence for BII in our methylation-only samples.

their respective publishing dates, 2DAU was simply a control sample for a modified sequence structure study similar to our own. In addition, they do not state which force field they used for restrained molecular dynamics in this study. There have been great improvements in the DNA-specific force fields since 1998, when the structure was solved. The force field has been found to play a role in the backbone conformation [12], which explains the backbone RMSD difference observed between our structures and 2DAU (ranging from 2.2 Å for DD08 to 3.2 Å for DD06). To further add to the noticeable differences, 2DAU reportedly was solved in a buffer solution that contains 10 µM sodium phosphate buffer as opposed to the 10 mM sodium phosphate buffer used as a standard in solution structure studies. The three orders of magnitude difference in buffer concentration and the fact the DNA is more abundant than the buffer means that the pH of this solution is not controlled by the buffer. Additionally, the authors of 2DAU do not report the pH of the sample. For these reasons, 2DAU was not considered for use as our unmodified control sample despite the fact that it is an NMR solution structure.

1NAJ – Liquid crystalline NMR Drew-Dickerson sequence

Solved through NMR with chemical shift anisotropy (CSA) and residual dipolar coupling (RDC) data in liquid crystalline media, PDB ID 1NAJ is the most recent structure of the unmodified Drew-Dickerson sequence [13]. With the addition of the CSA and RDC restraints, the authors hoped to overcome the common shortcomings, namely lack of numerous directly-observed long-range restraints, of solving NMR structures in solution or in liquid crystalline media. This method is much more expensive and time-consuming when compared to solving the crystal or NMR solution structures, but takes some of the most sophisticated and demanding techniques from both. This structure was solved, when it was in solution, in a buffer of 10 mM sodium phosphate, 50mM KCl and 6 mM NaN_3 at pH 6.8, which is similar to our own conditions, though with KCl instead of NaCl. Like the crystal structure 1BNA, 1NAJ has more direct observations, which are reflected in the structure. This can again be seen in the stretch values, which were set to ideal hydrogen bonding distances for our structure based off qualitative observations, and are shorter than ideal in 1NAJ **(Figure 4.6, 4.7)**. The backbone RMSDs between the reported structure are smaller than those of 1BNA, ranging from 1.3 Å for DD05 to 3.4 Å for DD08. The all atom RMSDs, however, are very similar to those seen in 1BNA, ranging from 3.4 Å for DD05 to 4.2 Å for DD08. Correspondingly, there is a greater difference between the helical parameters for 1NAJ and the reported structures than when compared with 1BNA. This difference, however, is generally minor. For these reasons, 1NAJ and 1BNA in combination are a good representation of the unmodified Drew-Dickerson sequence DD01 and are treated as such for future comparisons to limit the already overwhelming number of structures solved by the Smirnov lab.

183D – CCA(8oxoG)CGCTGG containing crystal structure with 1.6 Å resolution

The crystal structure with PDB ID 183D is a duplex DNA decamer with the sequence of CCA(8oxoG)CGCTGG which refracted to 1.6 Å [4]. This is the only currently reported duplex crystal structure available which contains 8oxoG without an abasic site. Due to the difference in sequence and length of this DNA decamer, it is difficult to compare this structure to those reported in this thesis. However, observations that are made in our structures are shared by both this structure and the extended molecular dynamics simulation based upon it [14]. The addition of 8oxoG to both the Drew-Dickerson sequence and 183D's sequence causes a slight untwisting 3' of the modification site as characterized by lower values of the helical parameter "twist". Perhaps a more important and larger structural change caused by the addition of 8oxoG is the increase in BII propensity just 3' of the 8oxoG which is not observed in absence of 8oxoG **(Figure 4.10)**. 183D also features a few other steps in BII which are not tied to the addition of 8oxoG. There is no structure currently available for this sequence without 8oxoG, so there is no way of knowing whether this is due to 8oxoG, the sequence or other experimental conditions without further study. In the molecular dynamics study launched from this sequence, the BII frequency was not observed to be over 50% at those non-

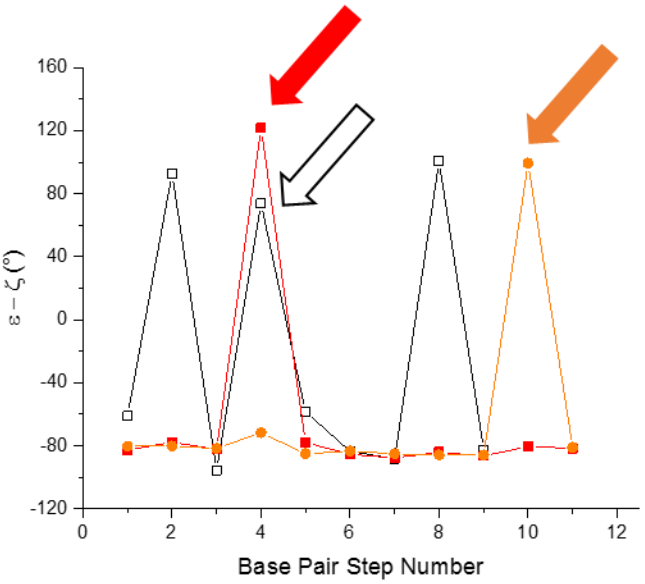


Figure 4.10. The ε–ζ values for 183D, DD02 and DD03. An ε–ζ > 20° indicates the backbone conformation is in BII. The marker and color scheme is as follows: 183D (the oxoG-contain crystal structure of a different sequence) in black open squares, DD02 is in red closed squares, DD03 is shown by orange closed circles. The arrows of the color corresponding to each structure show the position of the 8oxoG in each modified sequence.

8oxoG induced BII sites [14]. This perhaps indicates that the other BII sites are not sequence-dependent and that the disagreement between MD and crystal structure could be due to conditions in either experiment. This slight disagreement aside, our studies, the crystal structure and the MD simulations all demonstrated that the step 3' of 8oxoG has profound BII propensity. This common observation validates our most striking change in structure for the 8oxoG modification of a DNA duplex.

4MKW – Crystal structure at 1.22 Å resolution of CG(5metC)GAATTCGCG

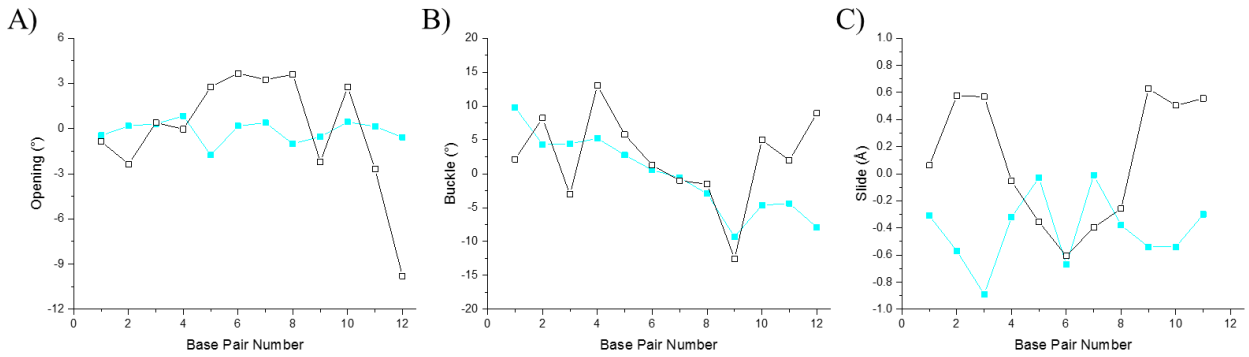


Figure 4.11. The opening **(A)**, buckle **(B)** and slide **(C)** values for 4MKW and DD04. 4MKW (the crystal structure with the same sequence as DD04) is shown in black open squares, and DD04 in cyan triangles.

The crystal structure PDB ID 4MKW was solved to 1.22 Å resolution and has the same sequence as our DD04, with the substitution of 5metC in the third position [5]. Despite the exact match between the sequence of DD04 and 4MKW, the helical parameters between this crystal structure and our NMR solution structure do not match up quite as well as those of 1BNA. The helical parameters that differed the greatest amount were opening, slide and buckle **(Figure 4.11)**. In addition to this, 4MKW showed that the G_{10}-C_{11} step is in BII **(Figure 4.12)**. In our NMR data, we saw no evidence of the markers that indicate BII in any of our methylation-only modified samples. The ^{31}P chemical shifts were grouped together with no significant outliers to indicate that any of the steps in our sequence show a backbone conformation in BII for DD04, DD06 and DD06

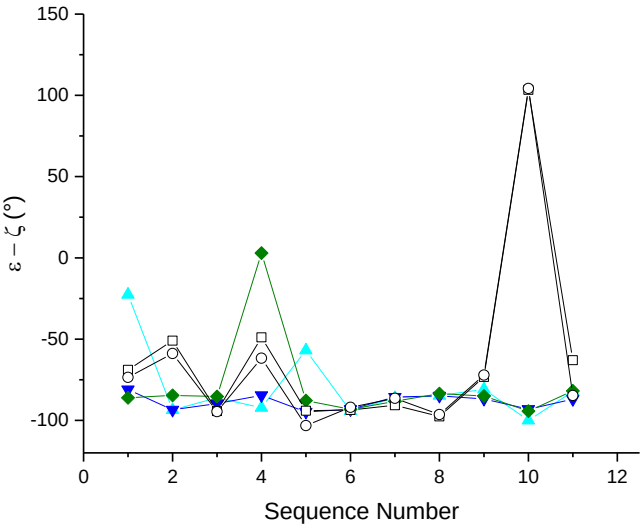


Figure 4.12. The ε–ζ values for 4MKW, 4C63, DD04, DD05, and DD06. An ε–ζ > 20° indicates the backbone conformation is in BII. The marker and color scheme is as follows: 4MKW (the crystal structure with the same sequence as DD04) in black open squares, 4C63 in black open circles, DD04 is in cyan triangles, DD05 is shown by blue inverted triangles and DD06 in green diamonds.

(Figure 4.9). In addition, the other evidence for BII, a NOESY peak indicating an interaction between the H2's of sequential residues was also missing in our NMR data despite clear differences in chemical shifts that would make appearance of this peak non-ambiguous. The RMSD between 4MKW and DD04 is higher than those compared to 1BNA, with each being 4.0 and 3.8 Å, respectively. These observed differences between the crystal and solution structures are not entirely unexpected, due to the difference in buffer conditions. DNA at a higher phosphate buffer concentration of 100 mM PHOS at pH 7.4 has an increased percentage of BII for the DD04 sequence than those shown in our structure, or those shown in other unmodified Drew-Dickerson sequence structures [2]. Based on this observation, it is entirely possible that the dissimilar buffer conditions are responsible for the difference between these structures.

4C63 – Crystal structure at 1.32 Å resolution of CGCGAATT(5metC)GCG

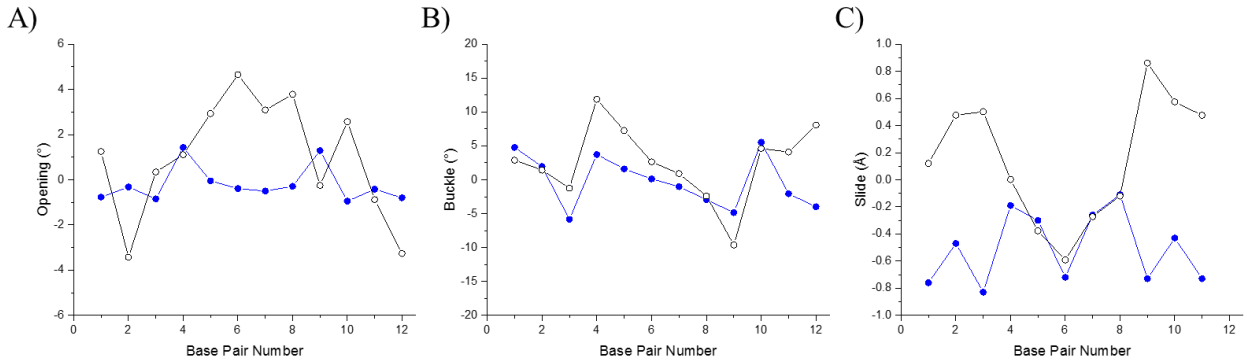


Figure 4.13. The opening **(A)**, buckle **(B)** and slide **(C)** values for 4C63 and DD05. 4C63 (the crystal structure with the same sequence as DD05) is shown in black open circles, and DD05 in blue circles.

A crystal structure matching our DD05 sequence has been solved to 1.32 Å resolution [6]. Much like 4MKW, there are differences between the helical parameters of our structure and 4C63 **(Figure 4.13)**. 4C63 also shows BII in the same position that 4MKW demonstrates, whereas we do not see any steps in BII in our DD05 structure **(Figure 4.12)**. It is not surprising that 4C63 and 4MKW both show BII at the steps G_{10} to C_{11}, as their crystallization conditions are very similar. The addition of this structure done in an independent study shows that buffer may very well be a contributing factor in the structural differences. Unfortunately, the authors of 1BNA do not report their buffer conditions, so those cannot be considered in this comparison of 1BNA to 4MKW and DD05.

Internal Quality Control Comparisons

Chemical shift comparison

Prior to structure calculations, the chemical shifts were analyzed to discover any unexpected differences. Based off spectral digital resolution, chemical shifts in the direct dimension had an error of 0.1 ppm and the indirect dimension has an error of 0.4 ppm. The differences in chemical shifts were taken between samples that are appropriate to compare (i.e. DD04 vs. DD05 vs. DD06, DD02 vs. DD03 and DD02 vs. DD05 vs. DD08). Besides those directly attributed to the chemical modification of the sequence, there were no important base-to-sugar chemical shift differences which reached beyond the possible error. For example, between DD08 and its corresponding control samples DD02 and DD05 there is no noticeable change in the pattern of the "walk" region of the D_2O NOESY, besides the loss of the oxoG base proton and the 5metC changing chemical shift to be nearer to the thymines **(Figure 4.14)**. DD02, DD05 and DD08 are

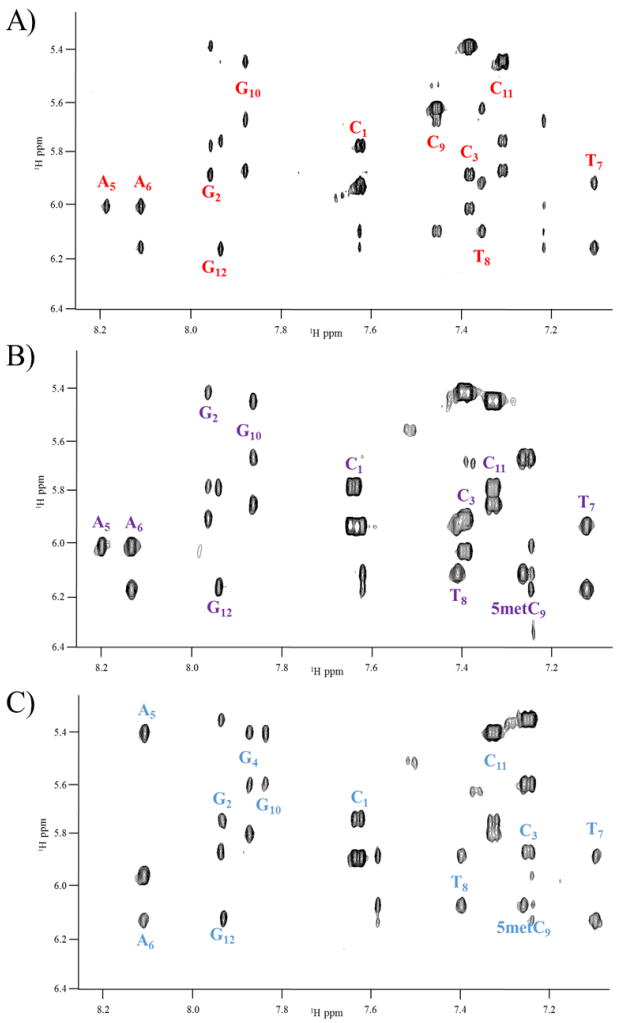


Figure 4.14. The "walk" region of DD02, DD05 and DD08 changes only near the modification sites. **A)** The "walk" region of DD02, with the base assignment given in the direct (x-) dimension in red letters. **B)** The "walk" region of DD08, with the base assignment given in the direct (x-) dimension in purple letters. **C)** The "walk" region of DD05, with the base assignment given in the direct (x) dimension in blue letters.

representative of all of our reported structures, and highlight the high similarity in chemical shifts between samples.

Restraint comparison

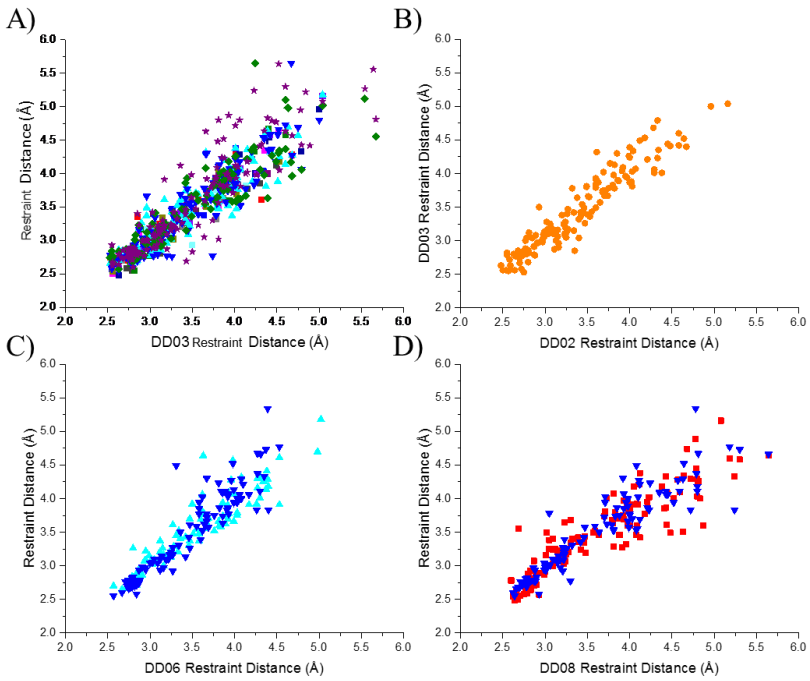


Figure 4.15. A comparison of NOE-derived distance restraints used in the solution structure calculations. The color scheme is as follows: DD02 is in red closed squares, DD03 in orange closed circles, DD04 in cyan triangles, DD05 in blue inverted triangles, DD06 is green diamonds and DD08 in purple stars. **(A)** All restraints are graphed against DD03, the sample which gave the most numerous distance restraints. **(B)** The correlation between DD02 and DD03 restraints. **(C)** The correlation of DD04 and DD05 with DD06 restraints. **(D)** DD02 and DD05 restraints compared to the DD08 restraints.

The NOE-derived distance restraints for the modified samples were in good agreement overall, showing positive correlations when interactions of the same type were compared. **(Figure 4.15A)** This agreement was observed despite the independent generation of the NOEs upon each new sample. As observed in the structures, DD02 and DD03 showed the strongest similarity, with an R^2 value of 0.90 **(Figure 4.15B)**. Of the methylated control samples, the hemi-methylated samples DD04 and DD05 had an R^2 of 0.82 when compared. DD04 against DD06 yielded an R^2 of 0.87, and DD05 against DD06 gave an R^2 of 0.86 **(Figure 4.15C)**. The target sample DD08 had an R^2 of 0.86 and 0.83 when compared to its control samples DD02 and DD05, respectively **(Figure 4.15D)**.

References

[1] O. Jardetzky, "On the nature of molecular conformations inferred from high-resolution NMR," *Biochim. Biophys. Acta BBA - Protein Struct.*, vol. 621, no. 2, pp. 227–232, Feb. 1980.

[2] Y. Tian, M. Kayatta, K. Shultis, A. Gonzalez, L. J. Mueller, and M. E. Hatcher, "31P NMR Investigation of Backbone Dynamics in DNA Binding Sites," *J. Phys. Chem. B*, vol. 113, no. 9, pp. 2596–2603, Mar. 2009.

[3] D. Djuranovic and B. Hartmann, "DNA fine structure and dynamics in crystals and in solution: The impact of BI/BII backbone conformations," *Biopolymers*, vol. 73, no. 3, pp. 356–368, Feb. 2004.

[4] L. A. Lipscomb, M. E. Peek, M. L. Morningstar, S. M. Verghis, E. M. Miller, A. Rich, J. M. Essigmann, and L. D. Williams, "X-ray structure of a DNA decamer containing 7,8-dihydro-8-oxoguanine," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 92, no. 3, pp. 719–723, Jan. 1995.

[5] J. A. Theruvathu, W. Yin, B. M. Pettitt, and L. C. Sowers, "Comparison of the structural and dynamic effects of 5-methylcytosine and 5-chlorocytosine in a CpG dinucleotide sequence," *Biochemistry (Mosc.)*, vol. 52, no. 47, pp. 8590–8598, Nov. 2013.

[6] L. Lercher, M. A. McDonough, A. H. El-Sagheer, A. Thalhammer, S. Kriaucionis, T. Brown, and C. J. Schofield, "Structural insights into how 5-hydroxymethylation influences transcription factor binding," *Chem. Commun. Camb. Engl.*, vol. 50, no. 15, pp. 1794–1796, Feb. 2014.

[7] N. Borochov, H. Eisenberg, and Z. Kam, "Dependence of DNA conformation on the concentration of salt," *Biopolymers*, vol. 20, no. 1, pp. 231–235, Jan. 1981.

[8] G. S. Kumar and M. Maiti, "DNA Polymorphism Under the Influence of Low pH and Low Temperature," *J. Biomol. Struct. Dyn.*, vol. 12, no. 1, pp. 183–201, Aug. 1994.

[9] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson, "Structure of a B-DNA dodecamer: conformation and dynamics," *Proc. Natl. Acad. Sci.*, vol. 78, no. 4, pp. 2179–2183, Apr. 1981.

[10] T. Geršata, A. Pérez, M. Orozco, A. V. Morozov, J. Sponer, and F. Lankaš, "Structure, Stiffness and Substates of the Dickerson-Drew Dodecamer," *J. Chem. Theory Comput.*, vol. 9, no. 1, pp. 707–721, 2013.

[11] A. Y. Denisov, E. V. Zamaratski, T. V. Maltseva, A. Sandström, S. Bekiroglu, K. H. Altmann, M. Egli, and J. Chattopadhyaya, "The solution conformation of a carbocyclic analog of the Dickerson-Drew dodecamer: comparison with its own X-ray structure and that of the NMR structure of the native counterpart," *J. Biomol. Struct. Dyn.*, vol. 16, no. 3, pp. 547–568, Dec. 1998.

[12] S. Y. Reddy, F. Leclerc, and M. Karplus, "DNA Polymorphism: A Comparison of Force Fields for Nucleic Acids," *Biophys. J.*, vol. 84, no. 3, pp. 1421–1449, Mar. 2003.

[13] Z. Wu, F. Delaglio, N. Tjandra, V. B. Zhurkin, and A. Bax, "Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and 31P chemical shift anisotropy," *J. Biomol. NMR*, vol. 26, no. 4, pp. 297–315, Aug. 2003.

[14] T. Geršata, M. Kara, M. Zacharias, and F. Lankaš, "Effect of 8-oxoguanine on DNA structure and deformability," *J. Phys. Chem. B*, vol. 117, no. 39, pp. 11617–11622, Oct. 2013.

Chapter 5 – Conclusions and Future Work

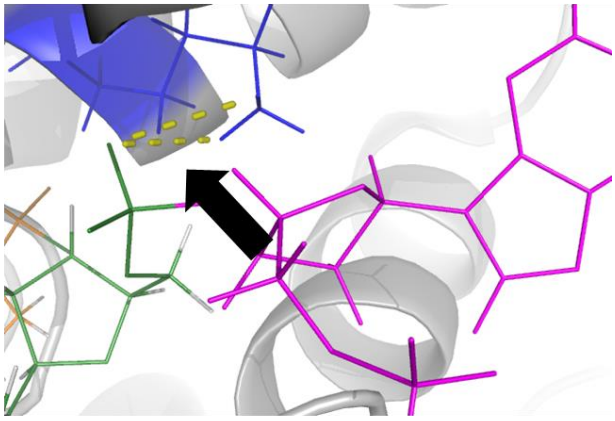
oxoG and 5metC modifications do not dramatically alter the structure of the DDD

The structure of the DNA Drew-Dickerson dodecamer does not display global structural changes upon incorporating the 5metC and oxoG modifications independently or in tandem. The helical parameters remain relatively stable throughout the modified structures, and the RMSDs between each of the reported structures reach a maximum at 1.87 Å. Only local changes, such as alterations in helical parameters or backbone conformation are observed upon modification. These local changes can lead to the enzymatic consequences suggested in this text, but also suggest that structure may not be the only property that is altered by modification. Insight into the motional dynamics and thermodynamic stability of the modified samples is necessary to determine possible sources of DNA substrate-driven differences in enzymatic activity.

BII may play a role in OGG1 recognition of 8oxoG

Recently, the backbone conformation BII has been hypothesized to affect protein-DNA interactions [1], so it is of great interest that we report in these NMR solution structures that the oxoG modification is linked to BII 3' of the modification site in a variety of sequence and modifications contexts. It has been suggested that the human oxoG glycosylase, hOGG1, forms early contact with the backbone 3' of the oxoG site, and that this contact can play a role in the differentiation of canonical guanine from 8oxoG [2]. There is a key catalytic residue, K249, necessary for base excision that directly contacts the phosphate 3' of the extra-helical canonical guanine. This contact is absent in oxoG-containing crystal structures [2], [3] **(Figure 5.1)**. This contact between the 3' phosphate of guanine and K249 has been proposed to halt the further

A)



B)

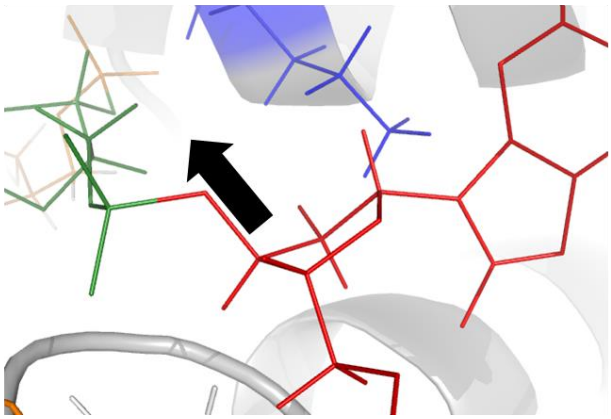


Figure 5.1. (A) The direct contact between the 3' backbones of canonical guanine to the catalytic residue K249 of hOGG1 from the PDB ID structure of IYQK [2]. hOGG1 is shown in grey, with its K249 in blue. The target guanine is shown in magenta and the guanine 3' of the target guanine in green. The dashed lines show the contact between the K249 amino protons, the DNA backbone phosphate and backbone O3'. The P to NH_3^+ distance is 2.7 Å and the O3' to NH_3^+ distance is 2.5 Å. **(B)** The lack of contact between K249 of hOGG1 and the backbone 3' of oxoG in PDB ID 2NOZ [3]. hOGG1 is shown in grey, with its K249 in blue. The flipped out oxoG is shown in red and the guanine 3' of the oxoG in green. The black arrows show the contact/non-contact areas.

movement of guanine into the catalytic site, saving the enzyme from excising the guanine. Both canonical guanine and oxoG are flipped from helix for hOGG1 to determine if the base can be excised, because of this the DNA is highly distorted by the enzyme. Despite this, it is possible that the BII-like backbone conformation induced by oxoG is either slightly preserved throughout the hinging process or that the prior BII conformation leads to a different lowest-energy conformation upon hOGG1 swinging the base into the recognition site. Thus, it is reasonable to propose that the BII conformation may play a role in the recognition of oxoG over canonical guanine by hOGG1.

BII may slow the addition of 5metC by mDNMT1

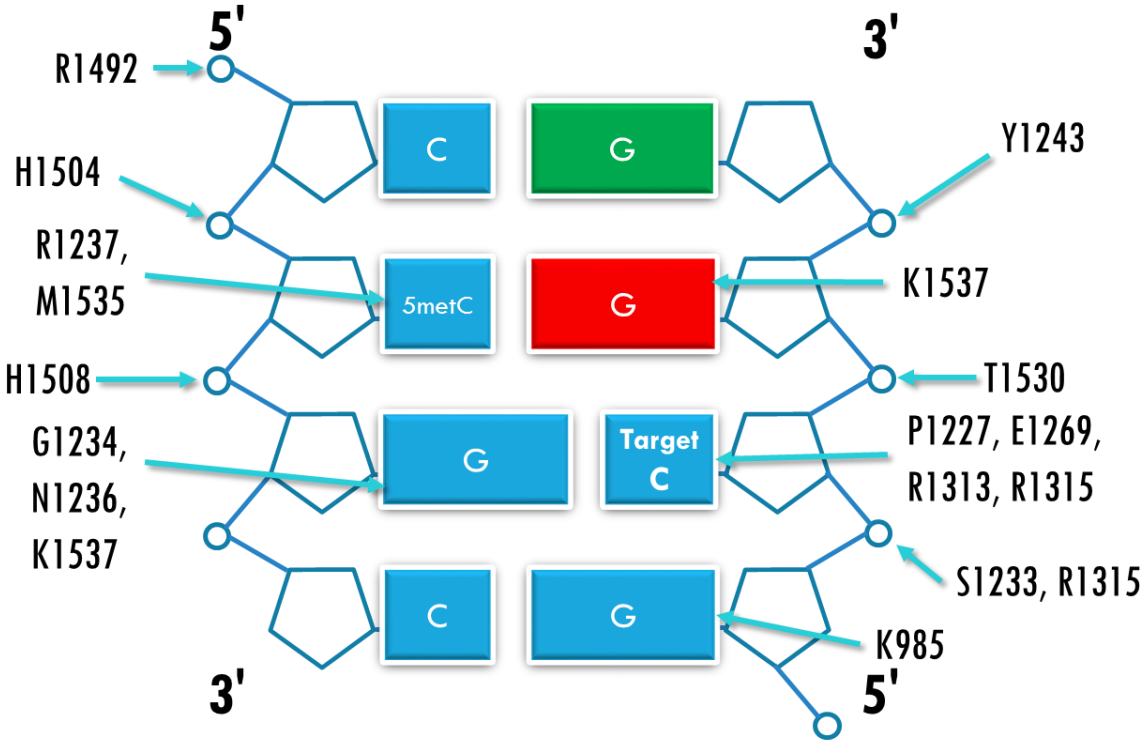


Figure 5.2. The contacts between murine DNMT1 and four base pairs of the target DNA substrate [5]. The bases are shown in boxes, with their corresponding sugars as pentagons attached to the backbone. The open circles between sugars represents the phosphate group. All contacts are shown with arrows, with the tip pointed towards the general location one the DNA where the contact occurs and the base simply listing the DNMT1 residues which interact with the DNA. The this cartoon demonstrates the lack contact directly 3' of the target CpG site on the strand which contains 5metC, which is featured in the target cytosine strand with the contact of the 3' CpG site phosphate with Y1243.

In human DNMT1, the addition of 8oxoG was shown to affect the enzymatic activity in a position-dependent manner [4]. When oxoG was introduced to hemi-methylated DNA *cis*, on the same strand as the already established 5metC, the activity was slowed from 128 fmol CH_3 incorporated/min to 123 fmol CH_3 incorporated/min. When the oxoG was modified *trans* from the 5metC, the activity was slowed from 128 fmol CH_3 incorporated/min to 2 fmol CH_3 incorporated/min. So, the incorporation of oxoG produced a dramatic effect when modified *trans* to the 5metC, but not *cis*. Our DD08 structure is a *trans* oxidized sequence, and shows the backbone 3' of the oxoG site in BII conformation. In light of this, it is interesting to note that the

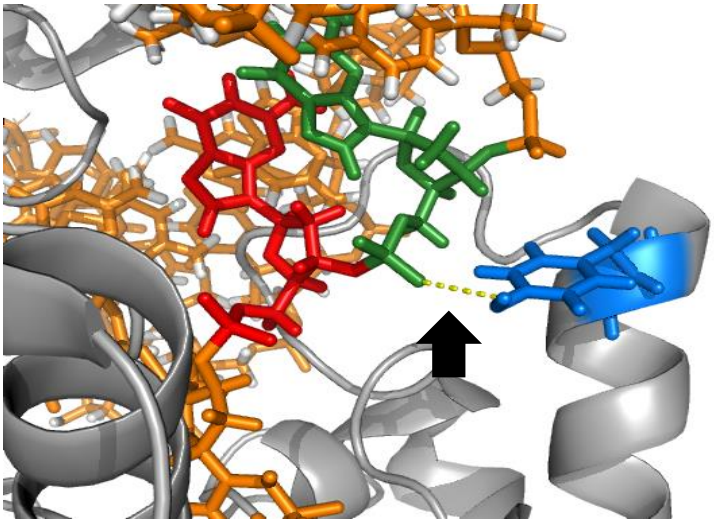


Figure 5.3. The direct contact of murine DNMT1 to the phosphate 3' of the *trans* guanine which base pairs to the 5metC present in hemi-methylated DNA from PDB ID 4DA4 [5]. Shown in grey is the DNMT1, in blue is the contact residue Y1243. The DNA is shown in orange, with the *trans* guanine shown in red and the residue 3' of the guanine shown in green. The distance measurement of 3.8 Å is shown between the oxygen on the phosphate and the hydroxyl hydrogen of Y1243. The black arrow points to the contact represented by a dashed line.

structure of highly similar murine DNMT1, with a sequence homology of 77%, demonstrates a direct hydrogen bonding contact to the *trans* 3' phosphate but not the *cis* 3' phosphate **(Figure 5.2, Figure 5.3)** [5]. *Thus, we propose that the BII backbone conformation that occurs 3' of the oxoG modification interrupts the contact between the phosphate and DNMT1, which slows the enzymatic activity.*

Future work

Target sample structures

Of the twelve samples proposed in the sample scheme to investigate the effects of 5metC and oxoG in the CpG site, six of the samples are reported here, and six more remain. The target samples DD07 and DD09 through DD12 still remain to be solved. During the writing of this thesis, NMR spectra have been collected for DD07, DD10 and DD12, with DD12 currently undergoing structural calculations. Data for these indicate that oxoG induces the 3' BII conformation in these sequence contexts as well (not shown).

Imino proton dynamics

In addition to the structure studies, thermodynamic data may yield insight into enzymatic phenomena. 5metC has been shown to increase DNA melting temperatures, making it a stabilizing

modification [6]. In contrast, oxoG is a de-stabilizing modification [7]. The combination of the stabilizing 5metC and de-stabilizing oxoG could yield interesting thermodynamic results. To investigate this, a UV-melting temperature study of each of our samples is being performed to determine the melting temperature of each sample. Further, to investigate where the melting is occurring in a temperature-dependent manner, the NMR analysis of imino protons is being performed. As discussed before, the imino protons NMR spectral lines are indicative of the Watson-Crick hydrogen bonding. Thus, hydrogen bonding patterns within base pairs can be analyzed through line-width analysis to determine which hydrogen bonding systems are dissociating at each temperature step [8]. A change in DNA stability could have enzymatic consequences, such as increasing the activation energy required to modify the bases, or rendering the site more or less accessible to the enzyme. With the ongoing thermodynamic and imino protons NMR line-width analysis, we can enrich the value of our structural results in relation to the DNMT1 and hOGG1 activity observed in presence of the modifications.

Enzymatic studies

In addition to the DNA-substrate based studies, enzyme focused studies are of even greater interest. The Drew-Dickerson dodecamer contains the *EcoRI* recognition site, as well as having the preferred substrate for DNMT1. Our collaborators are performing enzymatic activity studies of DNMT1, hOGG1, and *EcoRI* of longer DNA sequences that contain our oxidized/methylated CpG sequences. Similar studies have been previously reported for similar sequences, but none have been completed for our full set of modifications. Enzymatic studies will back the structural, thermodynamic, and dynamic findings that will compose the entirety of the investigation into the modification of the Drew-Dickerson sequence with oxoG and 5metC.

References

[1] B. Heddi, N. Foloppe, N. Bouchemal, E. Hantz, and B. Hartmann, "Quantification of DNA BI/BII Backbone States in Solution. Implications for DNA Overall Structure and Recognition," *J. Am. Chem. Soc.*, vol. 128, no. 28, pp. 9170–9177, Jul. 2006.

[2] A. Banerjee, W. Yang, M. Karplus, and G. L. Verdine, "Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA," *Nature*, vol. 434, no. 7033, pp. 612–618, Mar. 2005.

[3] C. T. Radom, A. Banerjee, and G. L. Verdine, "Structural characterization of human 8-oxoguanine DNA glycosylase variants bearing active site mutations," *J. Biol. Chem.*, vol. 282, no. 12, pp. 9182–9194, Mar. 2007.

[4] P. W. Turk, A. Laayoun, S. S. Smith, and S. A. Weitzman, "DNA adduct 8-hydroxyl-2′-deoxyguanosine (8-hydroxyguanine) affects function of human DNA methyltransferase," *Carcinogenesis*, vol. 16, no. 5, pp. 1253–1255, May 1995.

[5] J. Song, M. Teplova, S. Ishibe-Murakami, and D. J. Patel, "Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation," *Science*, vol. 335, no. 6069, pp. 709–712, Feb. 2012.

[6] P. Guldberg, J. Worm, and K. Grønbæk, "Profiling DNA methylation by melting analysis," *Methods*, vol. 27, no. 2, pp. 121–127, Jun. 2002.

[7] G. E. Plum, A. P. Grollman, F. Johnson, and K. J. Breslauer, "Influence of the Oxidatively Damaged Adduct 8-Oxodeoxyguanosine on the Conformation, Energetics, and Thermodynamic Stability of a DNA Duplex," *Biochemistry (Mosc.)*, vol. 34, no. 49, pp. 16148–16160, Dec. 1995.

[8] S. Cheung, K. Arndt, and P. Lu, "Correlation of lac operator DNA imino proton exchange kinetics with its function," *Proc. Natl. Acad. Sci.*, vol. 81, no. 12, pp. 3665–3669, Jun. 1984.

Appendix

Table 1. Chemical shifts from dd02 D_2O spectra taken at 25°C.

Residue Number	H1'	H2	H2'	H2''	H3'	H4'	H5	H6	H71	H72	H73	H8	P
1	5.7636		1.9632	2.4086	4.704	4.0655	5.9171	7.6323					
2	5.8763		2.6702	2.7057	4.9699		---					7.9635	-4.0245
3	6.0087		1.8454	2.5537	4.8559	4.2368	5.3859	7.3901					-4.1737
4	5.1581		3.3526	2.4127	4.8491								-3.8461
5	6.0002	7.2244	2.7115	2.933	5.0589	4.438						8.1953	-3.6786
6	6.1576	7.6335	2.5647	2.9342	4.9989	4.4565						8.119	
7	5.91		1.9803	2.5645	4.8167	4.2343		7.1122	1.2723	1.2723	1.2723		
8	6.0945		2.1641	2.5512	4.8931	4.1413		7.3623	1.5139	1.5139	1.5139		
9	5.6765		2.0406	2.3903	4.8661	4.1275	5.625	7.4624					
10	5.8635		2.6231	2.6973	4.9798	4.3637						7.8867	-3.8902
11	5.7492		1.8788	2.3239	4.8134	4.1352	5.4466	7.3151					-4.0732
12	6.164		2.6126	2.3706	4.6726	4.1726						7.9414	-3.8841

Table 2. Helical parameters for dd02.

Local base pair parameters							
Step Number	**Base Pair**	**Shear**	**Stretch**	**Stagger**	**Buckle**	**Propeller**	**Opening**
1	C-G	0.34	-0.1	0.09	1.15	-12.23	-0.16
2	G-C	-0.38	-0.11	0.1	0.15	-9.84	-0.21
3	C-G	0.43	-0.12	0.02	-0.94	-5.69	-0.46
4	G-C	-0.33	-0.13	-0.07	4.53	-11.17	0.38
5	A-T	0.12	-0.01	-0.12	4.08	-16.31	-0.1
6	A-T	0.08	-0.02	0.01	1.52	-19.82	-0.08
7	T-A	-0.08	-0.02	0.02	-1.44	-20.81	-1.11
8	T-A	-0.13	-0.01	-0.17	-3.58	-17.76	-1.26
9	C-G	0.32	-0.13	-0.09	-4.99	-11.42	-0.08
10	G-C	-0.42	-0.12	-0.02	-0.74	-5.98	-0.71
11	C-G	0.4	-0.11	0.07	0.09	-8.67	-0.44
12	G-C	-0.3	-0.09	0.14	0.28	-13.14	0.12
Local base pair step parameters							
Step Number	**Step**	**Shift**	**Slide**	**Rise**	**Tilt**	**Roll**	**Twist**
1	CG/CG	0.01	-0.93	3.2	-0.07	11.6	26.73
2	GC/GC	-0.06	-0.89	3.31	0.54	4.55	38.7
3	CG/CG	0.53	-0.58	3.12	0.93	11.82	19.91
4	GA/TC	-0.88	0.19	3.24	-2.57	1.28	41.61
5	AA/TT	-0.22	-0.27	3.23	-2.84	3.79	38.19
6	AT/AT	-0.08	-0.82	3.27	-0.27	0.18	29.77
7	TT/AA	-0.04	-0.27	3.22	2.66	3.67	36.57
8	TC/GA	0.9	0.38	3.26	2.67	1.97	42.48
9	CG/CG	-0.49	-0.42	3.15	-0.68	12.3	19.96
10	GC/GC	0.16	-0.73	3.27	-0.42	4.83	39.04
11	CG/CG	0.25	-0.88	3.17	-0.17	11.79	26
Local base pair helical parameters							
Step Number	**Step**	**X-disp**	**Y-disp**	**h-Rise**	**Incl.**	**Tip**	**h-Twist**
1	CG/CG	-4.2	-0.03	2.58	23.72	0.13	29.09
2	GC/GC	-1.88	0.15	3.19	6.83	-0.82	38.96
3	CG/CG	-5.19	-1.02	2.42	30.89	-2.44	23.14
4	GA/TC	0.14	0.97	3.29	1.8	3.61	41.7
5	AA/TT	-0.88	-0.02	3.2	5.77	4.32	38.47
6	AT/AT	-1.63	0.1	3.27	0.34	0.53	29.77
7	TT/AA	-0.92	0.41	3.17	5.83	-4.22	36.84
8	TC/GA	0.32	-0.97	3.32	2.71	-3.68	42.6
9	CG/CG	-4.91	1	2.48	31.85	1.76	23.43
10	GC/GC	-1.65	-0.28	3.16	7.19	0.63	39.33
11	CG/CG	-4.26	-0.54	2.54	24.66	0.35	28.51

Table 3. Torsion angles for dd02.

	Strand	I							
Step Number	**Base**	**Alpha**	**Beta**	**Gamma**	**Delta**	**Epsilon**	**Zeta**	**Chi**	**Epsilon-Zeta**
1	C	---	---	55.6	132.9	-164.7	-81.9	-127.6	-82.8
2	G	-75.7	172.5	50.2	124.8	179.7	-102.6	-111.8	-77.7
3	C	-67.8	-178.1	56.1	125.8	-165.9	-83.5	-114.4	-82.4
4	G	-74.7	168.2	44.9	133.1	-83.3	155	-94.2	121.7
5	A	-87.1	149.9	45.4	134.5	-177.6	-99.7	-111.2	-77.9
6	A	-69.8	179.7	55.9	127.3	-175.9	-90.4	-111.2	-85.5
7	T	-66	166.3	59.3	100.7	-177.5	-89.8	-131.5	-87.7
8	T	-62.2	171.5	60.2	120.6	-171.4	-87.3	-119.9	-84.1
9	C	-69.2	167.8	57.4	105.2	-173.4	-86.9	-120	-86.5
10	G	-66.7	170.9	54.2	124.8	179.8	-99.7	-112.6	-80.5
11	C	-67.8	179.7	55.7	120.6	-168.7	-87	-120.9	-81.7
12	G	-72.4	168.8	52.9	114.3	---	---	-123.3	---

	Strand	II							
Step Number	**Base**	**Alpha**	**Beta**	**Gamma**	**Delta**	**Epsilon**	**Zeta**	**Chi**	**Epsilon-Zeta**
1	G	-73.5	171.7	53.4	117.1	---	---	-123.3	---
2	C	-68.4	-179.7	55.5	120.9	-169.6	-89.4	-122.6	-80.2
3	G	-66.3	171.6	54.4	125.7	-179.8	-100.7	-113.5	-79.1
4	C	-69.7	167.2	57.1	102.7	-174.4	-87.2	-122.2	-87.2
5	T	-62.1	172.2	59.6	121.6	-171.9	-87.7	-118.9	-84.2
6	T	-64.5	166.5	59.2	104	-177.3	-90.3	-130.2	-87
7	A	-68.8	174.4	56.2	120.6	-176.5	-90.2	-114.5	-86.3
8	A	-86.2	149.7	45.6	133	-176.6	-93.9	-110.1	-82.7
9	G	-74.5	167.2	45	132.4	-84.3	154.9	-93.2	120.8
10	C	-67.4	-177.6	56.1	127.1	-165.1	-82.9	-111.7	-82.2
11	G	-74.7	173.9	50.9	126.4	180	-103.6	-111.3	-76.4
12	C	---	---	56.9	126.2	-165.3	-83.8	-128.8	-81.5

Table 4. Chemical shifts from dd03 D_2O spectra taken at 25°C.

Residue Number	**H1'**	**H2**	**H2'**	**H2''**	**H3'**	**H4'**	**H5**	**H6**	**H71**	**H72**	**H73**	**H8**	**P**
1	5.76173		1.974	2.40526	4.69281	4.06335	5.90696	7.63994					
2	5.8809		2.63587	2.70271	4.95521	4.33381						7.94095	-4.19521
3	5.58673		1.84922	2.25328	4.80982	4.09554	5.3978	7.28338					-4.26476
4	5.49947		2.662	2.76677	4.98941	4.32791						7.8213	-4.05228
5	6.00863	7.24834	2.67491	2.93388	5.04725	4.44495						8.08715	-4.42941
6	6.15182	7.61986	2.54415	2.92663	4.99538	4.46798						8.08684	
7	5.90602		1.98229	2.57482	4.80905	4.2061		7.10506	1.26198	1.26198	1.26198		
8	6.11767		2.19035	2.58243	4.89838	4.20072		7.38009	1.51695	1.51695	1.51695		
9	6.07597		1.94879	2.64339	4.90135	4.10512	5.59539	7.51144					
10	5.48947		3.31942	2.32476	4.85921								-4.05504
11	5.7736		1.93588	2.3513	4.81163	4.07769	5.67846	7.4633					-3.69836
12	6.1461		2.60568	2.36221	4.66218	4.15688							-4.10727

Table 5. Helical parameters for dd03.

Local base pair parameters							
Step Number	**bp**	**Shear**	**Stretch**	**Stagger**	**Buckle**	**Propeller**	**Opening**
1	C-G	0.28	-0.1	0.28	-0.14	-10.24	-1.22
2	G-C	-0.36	-0.13	0.04	-2.13	-13.22	-2.2
3	C-G	0.31	-0.09	0.02	7.33	-10.52	0.98
4	G-C	-0.47	-0.14	-0.22	0.09	-11.99	0.32
5	A-T	0.01	-0.07	-0.37	-1.01	-17.39	3.29
6	A-T	0.12	0	-0.05	-2.29	-16.59	-1.28
7	T-A	-0.13	0	-0.06	1.26	-16.74	-1.64
8	T-A	-0.02	-0.07	-0.37	-0.29	-17.8	2.72
9	C-G	0.45	-0.14	-0.21	-1.45	-12.12	-0.07
10	G-C	-0.28	-0.09	0.08	-6.04	-10.51	0.85
11	C-G	0.36	-0.13	0.04	3.11	-13.68	-2.24
12	G-C	-0.28	-0.1	0.29	0.21	-11.49	-1.23
Local base pair step parameters							
Step Number	**step**	**Shift**	**Slide**	**Rise**	**Tilt**	**Roll**	**Twist**
1	CG/CG	-0.28	-0.72	3.21	1.97	10.72	28.42
2	GC/GC	0.55	0.09	3.1	0.65	-1.96	39.95
3	CG/CG	-0.66	-1.52	3.44	0.03	10.58	23.74
4	GA/TC	-0.35	0.16	3.28	-0.26	7.62	38.86
5	AA/TT	-0.19	-0.39	3.22	-3.61	3.29	34.92
6	AT/AT	-0.1	-0.89	3.15	0.02	0.4	29.76
7	TT/AA	0.1	-0.33	3.22	3.57	3.23	35.16
8	TC/GA	0.36	0.27	3.28	0.06	7.63	38.73
9	CG/CG	0.73	-1.41	3.36	-0.24	10.49	24.23
10	GC/GC	-0.49	0.15	3.12	-0.2	-2.67	40.11
11	CG/CG	0.34	-0.7	3.23	-1.87	10.87	28.29
Local base pair helical parameters							
Step Number	**step**	**X-disp**	**Y-disp**	**h-Rise**	**Incl.**	**Tip**	**h-Twist**
1	CG/CG	-3.4	0.91	2.74	20.88	-3.84	30.4
2	GC/GC	0.34	-0.73	3.1	-2.86	-0.95	40
3	CG/CG	-6.18	1.48	2.53	24.24	-0.08	25.96
4	GA/TC	-0.66	0.48	3.25	11.32	0.39	39.57
5	AA/TT	-1.13	-0.21	3.18	5.46	5.98	35.25
6	AT/AT	-1.81	0.2	3.14	0.78	-0.04	29.76
7	TT/AA	-1.01	0.36	3.17	5.31	-5.87	35.48
8	TC/GA	-0.52	-0.52	3.27	11.38	-0.09	39.44
9	CG/CG	-5.72	-1.66	2.53	23.63	0.53	26.38
10	GC/GC	0.51	0.68	3.11	-3.89	0.29	40.2
11	CG/CG	-3.42	-1	2.75	21.25	3.65	30.32

Table 6. Torsion angles for dd03.

	Strand	I							
Step Number	**Base**	**Alpha**	**Beta**	**Gamma**	**Delta**	**Epsilon**	**Zeta**	**Chi**	**Epsilon-Zeta**
1	C	---	---	55	134.7	-168.3	-87.9	-120.1	-80.4
2	G	-75.9	174.3	52.2	118.1	-178.9	-98.7	-114.3	-80.2
3	C	-64.9	170.1	63.6	127.4	-177.1	-95.4	-112.7	-81.7
4	G	-71.5	-179.1	51.4	119.6	-176	-104.3	-118.7	-71.7
5	A	-67.4	-179.1	56.5	132.6	-178	-92.9	-106.6	-85.1
6	A	-67.7	172.3	56.8	116.2	-175.6	-92.2	-120.1	-83.4
7	T	-65.8	169	58.8	106.7	-177.1	-92	-127.5	-85.1
8	T	-63.4	169	61.9	116	-171.2	-85.4	-124.3	-85.8
9	C	-68	172.3	53.5	136.7	-179.9	-93.9	-99.5	-86
10	G	-68.8	172.7	58	122.6	-77.6	-176.9	-121.8	99.3
11	C	-67.8	121.7	45.5	108.4	-176.8	-95.7	-109.7	-81.1
12	G	-71.3	172.9	54.2	122.7	---	---	-116.2	---
	Strand	**II**							
Step Number	**Base**	**Alpha**	**Beta**	**Gamma**	**Delta**	**Epsilon**	**Zeta**	**Chi**	**Epsilon-Zeta**
1	G	-71.9	174	54	124.1	---	---	-115.7	---
2	C	-68.7	123.2	45.2	109.8	-176.9	-97	-110.1	-79.9
3	G	-69.4	173.3	57.8	123.8	-78.6	-174.9	-122.8	96.3
4	C	-68.2	172.6	53.4	137.2	-179.9	-94.8	-99.7	-85.1
5	T	-63.1	168.3	61.8	115.2	-171.1	-85.4	-125.5	-85.7
6	T	-65.2	167.5	58.8	104.1	-176.8	-90.7	-128.8	-86.1
7	A	-67.3	171.6	56.9	115.9	-175.5	-90.7	-119.1	-84.8
8	A	-67	-179.1	56.4	133	-178.2	-92.5	-104.5	-85.7
9	G	-71.1	-179.4	52.1	118.9	-176.4	-103.1	-117.6	-73.3
10	C	-64.8	170.6	63.6	127.6	-177.2	-95.9	-112.3	-81.3
11	G	-75.8	174.3	52.1	118.2	-178.9	-100	-114	-78.9
12	C	---	---	55	134.7	-168.3	-87.9	-119.9	-80.4

Table 7. Chemical shifts from dd04 D_2O spectra taken at 25°C.

Residue Number	H1'	H2	H2'	H2"	H3'	H4'	H5	H6	H71	H72	H73	H8
1	5.77425		2.00073	2.4081		4.70443	5.92019	7.64927				
2	5.9793		2.61394	2.80457	4.97037	4.34424						7.95895
3	5.52275		1.80658	2.19485	4.96826	4.08396		7.04516	1.57336	1.57336	1.57336	
4	5.4122		2.66673	2.7245	4.98224	4.31547						7.82714
5	5.99988	7.24226	2.68494	2.91963	5.05562	4.44539						8.10251
6	6.1396	7.60565	2.55265	2.91502	4.99617	4.45402						8.10019
7	5.88983		1.97267	2.55397	4.80605	4.19733		7.0996	1.26331	1.26331	1.26331	
8	6.10208		2.16606	2.55939	4.89728	4.20162		7.37662	1.52685	1.52685	1.52685	
9	5.66401		2.05157	2.42086	4.86898		5.64226	7.46479				
10	5.8406		2.64364	2.68908	4.98242	4.35864						7.91497
11	5.77658		1.90219	2.33562	4.81149		5.40661	7.32644				
12	6.1526		2.60608	2.36649	4.80956	4.66727						7.94093

Table 8. Helical parameters for dd04.

Local base pair parameters							
Step Number	**bp**	**Shear**	**Stretch**	**Stagger**	**Buckle**	**Propeller**	**Opening**
1	C-G	0.36	-0.1	0.05	9.78	-15.72	-0.46
2	G-C	-0.51	-0.14	0.19	4.33	-7.56	0.17
3	C-G	0.45	-0.09	0	4.41	-12.71	0.32
4	G-C	-0.43	-0.14	0.05	5.21	-16.55	0.83
5	A-T	0.13	-0.02	-0.24	2.75	-17.47	-1.77
6	A-T	0.15	-0.01	-0.12	0.54	-17.03	0.17
7	T-A	-0.15	-0.01	-0.12	-0.59	-17.05	0.38
8	T-A	-0.09	-0.04	-0.27	-2.94	-17.28	-1.02
9	C-G	0.46	-0.13	0.08	-9.32	-12.12	-0.54
10	G-C	-0.46	-0.1	0.01	-4.65	-13.67	0.43
11	C-G	0.49	-0.13	0.18	-4.38	-6.89	0.14
12	G-C	-0.34	-0.1	0.07	-7.91	-14.4	-0.58
Local base pair steps parameters							
Step Number	**step**	**Shift**	**Slide**	**Rise**	**Tilt**	**Roll**	**Twist**
1	CG/CG	-1.12	-0.31	3.27	-5.22	8.63	35.14
2	GC/GC	0.4	-0.57	3.29	3.01	2.5	36.65
3	CG/CG	0.24	-0.89	3.15	-0.31	15.4	25.06
4	GA/TC	-0.44	-0.32	3.33	0.99	4.86	34.06
5	AA/TT	-0.44	-0.03	3.2	-2.58	4.15	41.21
6	AT/AT	0.02	-0.67	3.24	0.2	-1.79	30.76
7	TT/AA	0.4	-0.01	3.2	3.1	3.94	40.39
8	TC/GA	0.16	-0.38	3.42	-1.38	5.83	37.11
9	CG/CG	0.08	-0.54	3.1	0.38	13.55	24.85
10	GC/GC	-0.31	-0.54	3.28	-2.82	2.21	37.38
11	CG/CG	1.16	-0.3	3.24	4.86	7.97	34.46
Local base pair steps helical parameters							
Step Number	**step**	**X-disp**	**Y-disp**	**h-Rise**	**Incl.**	**Tip**	**h-Twist**
1	CG/CG	-1.71	1.05	3.24	13.94	8.43	36.52
2	GC/GC	-1.25	-0.23	3.26	3.97	-4.77	36.85
3	CG/CG	-4.72	-0.52	2.24	31.94	0.64	29.35
4	GA/TC	-1.31	0.89	3.24	8.24	-1.68	34.41
5	AA/TT	-0.48	0.36	3.2	5.87	3.65	41.49
6	AT/AT	-0.92	0	3.27	-3.38	-0.38	30.81
7	TT/AA	-0.44	-0.23	3.21	5.67	-4.47	40.69
8	TC/GA	-1.39	-0.44	3.32	9.09	2.15	37.57
9	CG/CG	-3.97	-0.09	2.48	28.9	-0.8	28.25
10	GC/GC	-1.12	0.11	3.26	3.44	4.38	37.54
11	CG/CG	-1.65	-1.18	3.22	13.16	-8.02	35.67

Table 9. Torsion angles for dd04.

	Strand	I							
Step Number	**Base**	**Alpha**	**Beta**	**Gamma**	**Delta**	**Epsilon**	**Zeta**	**Chi**	**Epsilon-Zeta**
1	C	---	---	55.3	135.5	-158.3	-135.7	-109.6	-22.6
2	G	-73.4	-179.3	52.5	144.6	-179.8	-86.2	-106.8	-93.6
3	C	-65.2	170.9	55.4	109.8	-169.4	-83.3	-120	-86.1
4	G	-70.9	166.3	54.1	97.2	177.1	-90.7	-120.9	-92.2
5	A	-62.9	173.5	56.3	135.8	-174.6	-117.7	-111.6	-56.9
6	A	-67.3	-172.2	53	140.8	175.3	-90.5	-104	-94.2
7	T	-64.2	170.7	59.6	113.7	-178	-91.8	-124.5	-86.2
8	T	-64	173.5	57.5	129.8	-179.1	-94.3	-109.8	-84.8
9	C	-69.4	178.4	54.8	118.8	-171.8	-90.9	-117	-80.9
10	G	-71.2	172.8	52.3	116.9	166.4	-93.6	-111.4	-100
11	C	-64.6	-169.3	53.6	137	-174.2	-91.9	-112.4	-82.3
12	G	-70.1	171.4	53.1	133.8	---	---	-108.8	---
	Strand	**II**							
Step Number	**Base**	**Alpha**	**Beta**	**Gamma**	**Delta**	**Epsilon**	**Zeta**	**Chi**	**Epsilon-Zeta**
1	G	-70.2	172.9	53.4	134	---	---	-109	---
2	C	-65.8	-170.4	53.3	136.6	-175.2	-92.8	-112.6	-82.4
3	G	-71.2	177.5	51.2	120	170.1	-95.2	-113.1	-94.7
4	C	-67.3	167.3	56.9	108.5	-174.2	-90.7	-121.9	-83.5
5	T	-64.3	174.6	56.9	127.8	-174.1	-91.5	-110.2	-82.6
6	T	-63.7	171.5	59.6	115.2	-178.1	-91.6	-123.9	-86.5
7	A	-67.5	-167.4	52	140.1	173.4	-90.1	-102.6	-96.5
8	A	-65.8	-176	54.8	136.3	177.4	-107.7	-107.9	-74.9
9	G	-71.2	171.1	52.8	115.1	170.9	-95.4	-107.3	-93.7
10	C	-65.4	171.6	56	106.3	-169	-83.3	-122.1	-85.7
11	G	-73.1	-177.8	52.3	144.1	179.1	-87.3	-105.9	-93.6
12	C	---	---	55.5	136.4	-159.9	-133.6	-110.3	-26.3

Table 10. Chemical shifts from dd05 D_2O spectra taken at 25°C.

Residue Number	**H1'**	**H2**	**H2'**	**H2"**	**H3'**	**H4'**	**H5**	**H6**	**H71**	**H72**	**H73**	**H8**
1	5.74611		1.96208	2.39232	4.68371	4.05066	5.88981	7.62096				
2	5.87299		2.63503	2.70653	4.94365	4.32345						7.92915
3	5.61071		1.83428	2.28265		4.09735	5.35988	7.23737				
4	5.41396		2.63542	2.70681	4.96665	4.28977						7.82812
5	5.96806	7.2306	2.67689	2.90742	5.04038	4.43787						8.10039
6	6.14204	7.57557	2.54982	2.90877	4.98576	4.44539						8.10148
7	5.88994		1.96105	2.54077		4.18394		7.08546	1.24434	1.24434	1.24434	
8	6.08203		2.06986	2.58021	4.87678	4.18176		7.38733	1.51965	1.51965	1.51965	
9	5.61383		2.03881	2.34683	4.849	4.1721		7.24825	1.71302	1.71302	1.71302	
10	5.80349		2.62153	2.66134	4.95921	4.3436						7.86466
11	5.75832		1.88094	2.32104		4.13335	5.41072	7.31561				
12	6.13366		2.59196	2.33821	4.65435	4.14359						7.92319

Table 11. Helical parameters for dd05.

Local base pair parameters

Step Number	bp	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1	C-G	0.25	-0.08	0.13	4.77	-12.36	-0.77
2	G-C	-0.52	-0.15	0.19	1.94	-6.82	-0.32
3	C-G	0.45	-0.14	0.33	-5.85	-6.72	-0.85
4	G-C	-0.49	-0.12	0.01	3.72	-17.25	1.43
5	A-T	0.13	-0.02	-0.2	1.6	-15.99	-0.05
6	A-T	0.13	-0.01	0.02	0.12	-15.27	-0.39
7	T-A	-0.13	-0.01	0.02	-1.01	-15.21	-0.5
8	T-A	-0.14	-0.02	-0.19	-2.94	-15.9	-0.29
9	C-G	0.48	-0.12	0.01	-4.84	-16.85	1.29
10	G-C	-0.45	-0.14	0.34	5.53	-6.62	-0.95
11	C-G	0.51	-0.15	0.19	-2.07	-6.31	-0.42
12	G-C	-0.24	-0.08	0.14	-4	-12.12	-0.8

Local base pair step parameters

Step Number	Step	X-disp	Y-disp	h-Rise	Incl.	Tip	h-Twist
1	CG/CG	-0.27	-0.76	3.22	-1.95	7.76	27.38
2	GC/GC	0	-0.47	3.46	-0.88	1.54	43.97
3	CG/CG	0.07	-0.83	2.95	2.89	11.5	24.51
4	GA/TC	-0.63	-0.19	3.28	-0.87	3.89	37.87
5	AA/TT	-0.2	-0.3	3.19	-2.82	1.66	37.41
6	AT/AT	-0.05	-0.72	3.25	-0.02	-1.64	32.91
7	TT/AA	0.17	-0.26	3.2	2.86	1.43	38.08
8	TC/GA	0.62	-0.11	3.27	0.86	3.86	38.29
9	CG/CG	-0.08	-0.73	2.95	-2.99	11.17	24.53
10	GC/GC	0.03	-0.43	3.46	0.92	1.61	43.92
11	CG/CG	0.34	-0.73	3.2	2	7.63	27.26

Local base pair steps helical parameters

Step Number	step	X-disp	Y-disp	h-Rise	Incl.	Tip	h-Twist
1	CG/CG	-3.23	0.12	2.91	15.97	4.02	28.5
2	GC/GC	-0.78	-0.08	3.45	2.05	1.17	44.01
3	CG/CG	-4.23	0.46	2.33	25.28	-6.35	27.19
4	GA/TC	-0.79	0.86	3.26	5.98	1.33	38.07
5	AA/TT	-0.67	-0.05	3.19	2.58	4.38	37.55
6	AT/AT	-0.99	0.08	3.28	-2.9	0.04	32.95
7	TT/AA	-0.58	0.1	3.19	2.18	-4.37	38.21
8	TC/GA	-0.65	-0.83	3.26	5.86	-1.3	38.48
9	CG/CG	-3.98	-0.49	2.39	24.61	6.59	27.08
10	GC/GC	-0.73	0.05	3.45	2.16	-1.23	43.96
11	CG/CG	-3.17	-0.25	2.91	15.79	-4.14	28.35

Table 12. Torsion angles for dd05.

	Strand	I							
Step Number	**Base**	**Alpha**	**Beta**	**Gamma**	**Delta**	**Epsilon**	**Zeta**	**Chi**	**Epsilon-Zeta**
1	C	---	---	54.9	137.4	-174.8	-93.8	-109.9	-81
2	G	-73.5	177.5	52.7	128.5	169.9	-96.6	-112.3	-93.5
3	C	-68.7	-164.3	51.8	136.6	-178.6	-89	-110.2	-89.6
4	G	-70.1	177	51.9	119.6	173.3	-102.1	-107.3	-84.6
5	A	-67.2	-171.3	54	135.5	169.9	-95.3	-104.4	-94.8
6	A	-68.1	-170.9	54	133.5	173.9	-92.9	-108.4	-93.2
7	T	-65.4	176.4	59.2	119.7	-179.2	-93.6	-121.4	-85.6
8	T	-64.8	173.8	58.9	122.9	-174.1	-89.2	-116.4	-84.9
9	C	-65.5	167	56.5	106.4	-172.8	-86	-123.4	-86.8
10	G	-68.6	173.5	52.8	122	169.4	-97.5	-111.3	-93.1
11	C	-66.4	-166.9	52.4	135.5	-176.4	-89.4	-110.8	-87
12	G	-68.5	169.9	54	127.6	---	---	-117.1	---

	Strand	II							
Step Number	**Base**	**Alpha**	**Beta**	**Gamma**	**Delta**	**Epsilon**	**Zeta**	**Chi**	**Epsilon-Zeta**
1	G	-68.8	171.1	53.8	128.5	---	---	-117	---
2	C	-66.5	-166.7	52.3	135.8	-177	-89.7	-111.1	-87.3
3	G	-68.5	173.7	52.7	122.1	169.8	-97.9	-112.4	-92.3
4	C	-65.7	166.7	56.2	105.7	-173	-86	-124	-87
5	T	-64.5	172.3	58.9	121.4	-173.6	-88.8	-117.2	-84.8
6	T	-65.4	175.1	59.3	117.5	-178.3	-92.6	-122.4	-85.7
7	A	-68	-170	53.9	134.3	174	-91.6	-107.7	-94.4
8	A	-67.1	-171.4	54.3	136.6	169	-95.1	-103.5	-95.9
9	G	-69.8	175.9	52.2	119.1	172.9	-102.3	-106.4	-84.8
10	C	-68.4	-164.5	51.9	136.1	-177.5	-88.8	-110.2	-88.7
11	G	-73.6	178.1	52.6	128.7	169.8	-96.8	-111.9	-93.4
12	C	---	---	54.9	137.4	-175.1	-94.4	-109.7	-80.7

Table 13. Chemical shifts from dd06 D_2O spectra taken at 25°C.

Residue Number	**H1'**	**H2**	**H2'**	**H2"**	**H3'**	**H4'**	**H5**	**H6**	**H71**	**H72**	**H73**	**H8**
1	5.7708		2.00748	2.4079	4.70683	4.06425		7.65166				
2	5.98145		2.62091	2.80306	4.97607	4.35244						7.96389
3	5.54039		1.81815	2.22788				7.05081	1.59363	1.59363	1.59363	
4	5.41678		2.66547	2.72428	4.99175	4.32012						7.84084
5	5.99745	7.26292	2.69867	2.93799	5.0699	4.46109						8.12735
6	6.15854	7.59317	2.57283	2.93362	5.01007	4.46762						8.12761
7	5.90994		1.98298	2.56566	4.84159	4.20406		7.10671	1.27167	1.27167	1.27167	
8	6.11051		2.10644	2.60849	4.91012	4.20507		7.41375	1.54732	1.54732	1.54732	
9	5.62377		2.07051	2.37067	4.88293	4.11691		7.28583	1.75296	1.75296	1.75296	
10	5.8337			2.67415	4.98942	4.37283						7.90547
11	5.79855		1.90388	2.34435			5.41324	7.33217				
12	6.15294		2.60393	2.36964	4.67557	4.16359						7.9428

Table 14. Helical parameters for dd06.

Local base pair parameters							
Step Number	**bp**	**Shear**	**Stretch**	**Stagger**	**Buckle**	**Propeller**	**Opening**
1	C-g	0.27	-0.08	0.1	3.44	-12.89	-0.23
2	G-C	-0.52	-0.15	0.16	3.9	-5.1	-0.39
3	C-G	0.52	-0.11	0.06	0.9	-9.66	0.32
4	G-C	-0.35	-0.1	-0.01	4.32	-15.73	1.89
5	A-T	0.22	0.02	-0.24	3.76	-14.74	-2.2
6	A-T	0.15	-0.01	-0.14	1.25	-18.39	-0.27
7	T-A	-0.15	-0.02	-0.16	-0.42	-18.94	-0.45
8	T-A	-0.24	0.03	-0.25	-4.23	-15.51	-3.34
9	C-G	0.35	-0.1	0	-5.48	-15.03	1.73
10	G-C	-0.52	-0.11	0.08	-0.31	-9.2	0.15
11	C-G	0.49	-0.14	0.15	-2.98	-5.15	-0.13
12	g-C	-0.21	-0.07	0.21	1.24	-6.84	-0.9
Local base pair step parameters							
Step Number	**Step**	**X-disp**	**Y-disp**	**h-Rise**	**Incl.**	**Tip**	**h-Twist**
1	CG/Cg	-0.33	-0.65	3.16	-1.89	9.55	26.07
2	GC/GC	0.14	-0.74	3.39	1.84	1.9	40.84
3	CG/CG	0.59	-0.67	3.11	3.15	13.88	22.36
4	GA/TC	-1.12	0.16	3.23	-1.46	4.1	40.13
5	AA/TT	-0.4	-0.2	3.25	-2.01	2.84	40.11
6	AT/AT	-0.03	-0.44	3.22	0.06	-1.25	29.86
7	TT/AA	0.31	-0.18	3.26	1.93	3.32	40.29
8	TC/GA	1.16	0.33	3.24	1.65	4.33	40.53
9	CG/CG	-0.57	-0.53	3.07	-3.06	13.64	22.11
10	GC/GC	-0.08	-0.64	3.38	-1.32	1.74	40.79
11	Cg/CG	0.32	-0.71	3.09	0.49	6.09	26.63
Local base pair steps helical parameters							
Step Number	**step**	**X-disp**	**Y-disp**	**h-Rise**	**Incl.**	**Tip**	**h-Twist**
1	CG/Cg	-3.51	0.27	2.76	20.3	4.01	27.8
2	GC/GC	-1.27	0.01	3.36	2.71	-2.63	40.92
3	CG/CG	-4.81	-0.53	2.35	31.98	-7.25	26.46
4	GA/TC	-0.23	1.45	3.27	5.95	2.12	40.35
5	AA/TT	-0.61	0.35	3.24	4.13	2.92	40.25
6	AT/AT	-0.6	0.07	3.24	-2.41	-0.12	29.89
7	TT/AA	-0.63	-0.23	3.25	4.81	-2.79	40.47
8	TC/GA	-0.01	-1.48	3.3	6.22	-2.37	40.78
9	CG/CG	-4.54	0.5	2.4	31.79	7.15	26.11
10	GC/GC	-1.11	-0.03	3.35	2.5	1.89	40.84
11	Cg/CG	-2.9	-0.57	2.86	13	-1.05	27.31

Table 15. Torsion angles for dd06.

	Strand	I							
Step Number	**Base**	**Alpha**	**Beta**	**Gamma**	**Delta**	**Epsilon**	**Zeta**	**Chi**	**Epsilon-Zeta**
1	C	---	---	61	103.1	-178.2	-92.1	-119.3	-86.1
2	G	-68	173.5	55.5	127.3	177.8	-97.6	-117.1	-84.6
3	C	-64.5	-178.9	55.7	127.6	-166.5	-81.2	-117.8	-85.3
4	G	-73.5	167.2	50	127.3	-153	-155.9	-100.6	2.9
5	A	-66.1	157.2	57	139.7	173.3	-98.9	-113.8	-87.8
6	A	-67.1	-163.9	52.2	137.5	178.9	-87.9	-105.3	-93.2
7	T	-63.5	166.8	58.5	113.2	179	-92.7	-121.8	-88.3
8	T	-63.9	177.2	56.6	126.6	-173.3	-89.8	-112.1	-83.5
9	C	-64.7	165.6	57.8	103.8	-173.7	-88.7	-121	-85
10	G	-68.7	174	52	127.1	169.1	-96.6	-111.1	-94.3
11	C	-66.3	-166.5	53.1	133.3	-172.6	-90.7	-112.4	-81.9
12	g	-71.9	170.9	51.9	131.2	---	---	-110	---
	Strand	**II**							
Step Number	**Base**	**Alpha**	**Beta**	**Gamma**	**Delta**	**Epsilon**	**Zeta**	**Chi**	**Epsilon-Zeta**
1	g	-71.7	170.5	51.7	131.5	---	---	-109.9	---
2	C	-67.1	-165	52.4	133.3	-172.3	-90.2	-112.8	-82.1
3	G	-69.4	174.6	51.8	127.1	169.3	-96.3	-112.4	-94.4
4	C	-64.3	165.9	57.6	104.4	-173.6	-89.4	-121.7	-84.2
5	T	-63.9	177	56.8	126.5	-174.3	-89.8	-112.2	-84.5
6	T	-63.7	166.9	58.4	113.6	178.9	-92.5	-121.5	-88.6
7	A	-67.7	-164.1	52.2	137.8	178.9	-87.5	-105.5	-93.6
8	A	-66.1	152.7	56.5	140.1	173.7	-98.6	-113.8	-87.7
9	G	-74.1	166.8	50.1	127.8	-148.6	-161	-98	12.4
10	C	-63.4	-178.9	55.9	127.9	-165.6	-81.2	-117	-84.4
11	G	-72.6	175.5	54	126.2	176.2	-97.9	-115.1	-85.9
12	C	---	---	55	133.7	-173	-94.7	-115.8	-78.3

Table 16. Chemical shifts from dd08 D_2O spectra taken at 25°C.

Row Labels	**H1'**	**H2**	**H2'**	**H2"**	**H3'**	**H5**	**H6**	**H71**	**H72**	**H73**	**H8**	**P**
1	5.76493		1.96953	2.40736	4.70072	5.92119	7.63251					-4.00704
2	5.88909		2.65046	2.72252	4.96684						7.96021	-4.19329
3	6.0166		1.84534	2.595	4.85751	5.39741	7.38215					-3.86092
4	5.14408		3.34432	2.38815	4.85392							-3.62061
5	5.99607	7.23198	2.7002	2.92828	5.06151						8.19956	
6	6.16747	7.61197	2.56536	2.93081	5.00414						8.13321	
7	5.92019		1.98018	2.55576	4.8352		7.10671	1.26667	1.26667	1.26667		
8	6.1025		2.10332	2.59877	4.89334		7.39713	1.52375	1.52375	1.52375		
9	5.65334		2.03405	2.34513	4.85602		7.25068	1.72909	1.72909	1.72909		-3.95643
10	5.83349		2.65983		4.96905						7.85825	-4.18064
11	5.76798		1.88219	2.33298	4.79676	5.42977	7.31823					-4.07076
12	6.15658		2.60225	2.36107	4.66706						7.93653	

Table 17. Helical parameters for dd08.

Local base pair parameters							
Step Number	**bp**	**Shear**	**Stretch**	**Stagger**	**Buckle**	**Propeller**	**Opening**
1	C-G	0.3	-0.1	-0.03	1.56	-15.23	0.23
2	G-C	-0.3	-0.12	0.08	-4.33	-16.4	-1.05
3	C-G	0.46	-0.13	-0.06	-2.78	-8.49	-0.35
4	G-C	-0.45	-0.1	-0.02	7.07	-14.59	2.97
5	A-T	0.01	-0.06	-0.27	13.02	-22.22	0.21
6	A-T	0.35	-0.01	-0.04	2.03	-16.52	-0.52
7	T-A	-0.36	-0.01	-0.04	-1.86	-16.57	-0.7
8	T-A	-0.02	-0.06	-0.28	-12.69	-22.27	0.01
9	C-G	0.44	-0.1	0	-6.98	-13.99	2.6
10	G-C	-0.44	-0.12	-0.05	3.49	-8.5	-0.31
11	C-G	0.31	-0.12	0.06	5.03	-16.33	-1.19
12	G-C	-0.33	-0.09	-0.01	-1.99	-16.15	0.76
Local base pair step parameters							
Step Number	**step**	**Shift**	**Slide**	**Rise**	**Tilt**	**Roll**	**Twist**
1	CG/CG	-0.08	-1.98	3.42	0.01	13.14	25.83
2	GC/GC	-0.53	-0.49	3.25	0.9	0.95	38.58
3	CG/CG	0.43	-0.54	3.13	-4.47	12.3	21.33
4	GA/TC	-0.26	0.15	3.23	1.96	4.81	31.68
5	AA/TT	-0.46	0.09	3.37	-5.26	8.94	38.55
6	AT/AT	-0.02	-1.63	3.36	-0.11	1.57	33.71
7	TT/AA	0.41	0.11	3.37	5.24	9.02	38.58
8	TC/GA	0.24	0.22	3.24	-2.12	5.19	31.56
9	CG/CG	-0.39	-0.48	3.11	4.46	12.11	21.2
10	GC/GC	0.54	-0.49	3.25	-0.58	0.92	38.48
11	CG/CG	0.19	-2.03	3.44	-0.31	13.11	25.66
Local base pair steps helical parameters							
Step Number	**step**	**X-disp**	**Y-disp**	**h-Rise**	**Incl.**	**Tip**	**h-Twist**
1	CG/CG	-6.6	0.16	2.17	27.27	-0.01	28.93
2	GC/GC	-0.86	0.92	3.23	1.44	-1.36	38.6
3	CG/CG	-4.56	-2.2	2.34	29.89	10.86	24.98
4	GA/TC	-0.58	0.82	3.2	8.74	-3.55	32.09
5	AA/TT	-0.95	0.04	3.34	13.26	7.8	39.88
6	AT/AT	-3.07	0.01	3.28	2.71	0.18	33.75
7	TT/AA	-0.93	0.02	3.34	13.36	-7.75	39.92
8	TC/GA	-0.53	-0.82	3.21	9.45	3.87	32.05
9	CG/CG	-4.44	2.13	2.36	29.64	-10.92	24.78
10	GC/GC	-0.86	-0.89	3.23	1.4	0.88	38.5
11	CG/CG	-6.74	-0.44	2.16	27.36	0.65	28.77

Table 18. Torsion angles for dd08.

	Strand	I							
Step Number	**Base**	**Alpha**	**Beta**	**Gamma**	**Delta**	**Epsilon**	**Zeta**	**Chi**	**Epsilon-Zeta**
1	C	---	---	-63	142.1	-176.5	-92	-113.9	-84.5
2	G	-75.1	-176.3	51.7	113.1	175.4	-98.5	-122	-86.1
3	C	-66.5	-172.3	56.5	132.3	-171.5	-87.7	-106.5	-83.8
4	G	-69.4	167.4	49.3	133.2	-78.8	-170.4	-101.2	91.6
5	A	-68	118.8	36.5	123.4	-152.9	-159.3	-86.3	6.4
6	A	-65.5	148.9	57.6	138.7	174.1	-92.5	-121.2	-93.4
7	T	-67.4	-168.9	55.1	127.8	-166	-88.7	-121.9	-77.3
8	T	-73.7	166.4	56	101.9	-173.5	-85.4	-120.9	-88.1
9	C	-67.3	165.8	56.6	100.3	-175.1	-84.7	-127.4	-90.4
10	G	-64.4	177	55.9	126.9	175.9	-96.2	-104.8	-87.9
11	C	-68.2	175.4	57.1	134.6	178.7	-95.6	-113.3	-85.7
12	G	-72.8	-171	50.1	132.2	---	---	-117	---
	Strand	**II**							
Step Number	**Base**	**Alpha**	**Beta**	**Gamma**	**Delta**	**Epsilon**	**Zeta**	**Chi**	**Epsilon-Zeta**
1	G	-73.6	-171	50	133.4	---	---	-115.2	---
2	C	-68.6	176.6	57	135.3	179.7	-97.2	-113	-83.1
3	G	-64.9	177.9	55.7	127.6	175.3	-96.9	-104.6	-87.8
4	C	-66.8	165.5	56.8	100.2	-175.1	-84.7	-128.5	-90.4
5	T	-73.5	166.1	56	101.7	-174.2	-85.5	-121.6	-88.7
6	T	-67.3	-168.9	55.1	127.8	-166.3	-88.3	-121.8	-78
7	A	-65.3	148.4	57.8	138.5	174.1	-92.4	-121.3	-93.5
8	A	-67.6	118.3	36.4	123.5	-153.1	-159.2	-85.8	6.1
9	G	-69.4	167.3	48.8	132.8	-78.5	-171.3	-100.3	92.8
10	C	-66.5	-171.5	56.1	132.1	-171	-87.5	-106.6	-83.5
11	G	-74.1	-175.3	51.7	114	175.2	-97.9	-122.4	-86.9
12	C	---	---	55.9	134.4	-177.9	-91.7	-115.2	-86.2