NMR structural studies of mismatched DNA base pairs and their interaction with *E. coli* MutS protein

A thesis submitted to The University of Manchester for the Degree of PhD in the Faculty of Physical Sciences and Engineering

2010

Tony Cheung

School of Chemistry

Table of contents

Abstract

Escherichia coli MutS is a DNA binding repair protein (97 kDa, monomer) and its biological significance arises from its recognition of mismatches which occur as errors during DNA replication. Mismatches and mutagenic bases represent a fascinating and diverse range of shapes and sizes and it is not obvious how a single protein (MutS) can recognise such molecular diversity against a huge background of canonical Watson-Crick base pairs.

In this project, the structure of a 17mer mismatch GT DNA was carried out using NMR spectroscopy to identify the differences caused by the introduction of a non-canonical base pair on helical structure. The resulting structure was B-form in conformation and local helical distortions were observed about the GT mispair due primarily to its sheared orientation. The effect of mismatch orientation, sequence context and oligonucleotide length on mismatch stability was also investigated using UV absorbance melting and NMR spectroscopy. The results showed that substitution of a TG mispair for a GT mispair was accompanied by a small drop in melting temperature. It was also discovered that sequences in which purine-purine or pyrimidine-pyrimidine stacking occurred induced additional stability of the mismatch resulting in a higher melting temperature of the duplex.

Affinity of mismatch GT DNA and its mismatch orientation, sequence context and length analogues for MutS was investigated by monitoring changes to the chemical shifts and linewidths of imino protons resonances during NMR titration. The results showed that MutS displayed higher affinity towards sequences which involved better stacking between the flanking base pairs and the GT/TG mispair.

Analogous NMR structural investigations of 6-thioguanine modified 13mer GC DNA and its oxidised derivatives have been successfully carried out. The NMR structure was successfully determined of the former and the results obtained showed the effect on helical structure induced by the substitution of a different DNA lesion.

Although the crystal structures of MutS bound to DNA mismatches have been known for a number of years, the analogous crystal structures of uncomplexed *apo* MutS have not been determined to date. Consequently, vital structural knowledge on the large change in conformation of MutS upon binding to the DNA mismatch is seriously lacking. We have successfully isolated the structurally and functionally important NTD of *E. coli* MutS and its labelled (^{13}C, ^{15}N) analogues and have shown that it is endowed with a stable, tertiary structural fold and well suited to NMR structure determination. This is exemplified by the assignments of several backbone amide and side chain resonances using isotope aided 3D NMR techniques.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Dedication

For Wawa,

I just want to simply say thank you. Thank you for all the support you have given me throughout the duration of my PhD and for never losing faith with me. You have always been and will continue to be the one person who will inspire me to aim higher and go further in life and for that amongst many other reasons, I will always love you.

Acknowledgements

I would like to thank my supervisor Dr. Vasudevan Ramesh for his support and guidance throughout my project and my colleagues John King and Misbah Nareen for their helpful discussions and proofreading. I would also like to extend my thanks to my other colleagues Usman Rasul, Sadia Mohammed, Misbah Ghafoor for their support and discussions and Clive Raynor for assistance in proofreading.

Many thanks to Tom Frankiel, Geoff Kelly and Alan Oregioni of the National Institute of Medical Research (NIMR), Mill Hill, for their assistance and support in use of NMR facilities.

For access to the Bruker 1000 MHz spectrometer, I would like to thank Moreno Lelli of the Centre de RMN à Très Hauts Champs in Lyon, France and for providing access to the NMR facilities at the School of Chemistry, University of Manchester, I thank Roger Speak.

I thank Yao-Zhong Zu (Open University) for the kind supply of 6-thioguanine and 6-sulphonateguanine modified DNAs and also Franck Vendeix, Massimiliano Zampini and Finbarr Hayes for the kind supply of normal and isotopically enriched WT-MutS and NTD-MutS proteins.

Finally, I would like to express my gratitude to my family for their continued support and encouragement during the course of this project.

Symbols and Abbreviations

Symbols

Å	Angstrom
A_D	Absorbance of double strand
A_S	Absorbance of single strand
B_0	External magnetic field
B_1	Local magnetic field
bp	Base pair
C_T	Total molar concentration of strands
[D]	Concentration of double strand
d_x	x-displacement
d_y	y-displacement
Da	Daltons
D_x	Shift
D_y	Slide
D_z	Rise
eq.	Equivalent
F_1	First frequency dimension
F_2	Second frequency dimension
F_3	Third frequency dimension
f_D	Fraction of strands when a double stranded oligonucleotide exists in equilibrium.
h	Planck's constant (6.626×10^{-34} J s)
Hz	Hertz
I	Nuclear spin quantum number
J_{xy}	Coupling constant between two nuclei x and y
K	Kelvin
K	Equilibrium constant in van't Hoff relationship
kDa	Kilo Dalton
kHz	Kilohertz
M	Molar
M_z	Bulk magnetization on Z-axiz
mM	Millimolar
OD	Optical density
A_{260}	Absorbance at a wavelength of 260nm
P	Pseudorotation angle
ppm	Parts per million
ps	Picoseconds
r	Distance between spins i and j in the NOE relationship
r_i	NOE distance of peak being considered in a NOESY spectrum
r_{ref}	NOE distance of reference peak considered in NOESY spectrum
R	Ideal gas constant (8.314472 J K^{-1} mol^{-1})
[S]	Concentration of single strand
S_i	Intensity of peak being considered in NOESY spectrum
S_{ref}	Intensity of reference peak being considered in NOESY spectrum

S_x	Shear
S_y	Stretch
S_z	Stagger
SL_{mix}	Spin lock mixing time
t	Time (s)
t_1, t_2, t_3	Evolution or detection period in NMR pulse sequence
T	Temperature
T_1	Longitudinal relaxation time
T_2	Transverse relaxation time
T_m	UV melting temperature
W_0	Zero quantum transition
W_2	Double quantum transition
α, β, γ, δ, ε, ζ	Sugar phosphate backbone dihedral angles
ν_0, ν_1, ν_2, ν_3, ν_4	Deoxyribose sugar ring dihedral angles
ν	Frequency
μ	Nuclear magnetic moment
μL	Microlitre
γ	Gyromagnetic ratio
τ	Delay
τ	Tilt
τ_0	Null delay
τ_c	Correlation time
τ_m	Mixing time
δ	Chemical shift
ψ	Amplitude
η	Inclination
θ	Tip
κ	Buckle
ρ	Roll
σ	Opening
ω	Propeller twist
Ω	Helical twist
ΔH	Enthalpy change
ΔS	Entropy change
ΔG	Gibbs free energy

Abbreviations

A	Adenine
A	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
ATP	Adenosine TriPhosphate
BMRB	Biological Magnetic Resonance Bank
C	Cytosine
C	Cysteine
CI-MS	Chemical Ionisation-Mass Spectrometry
COSY	COrrelated SpectroscopY
CPD	Composite Pulse Decoupling
CPMG	Carr-Purcell-Meiboom-Gill
Cys	Cysteine
D	Aspartic acid
DNA	DeoxyriboNucleic Acid
DQF-COSY	Double Quantum Filtered-COrrelation SpectroscopY
E	Glutamic acid
E. coli	*Escherichia coli*
ESI-MS	ElectroSpray Ionisation Mass Spectrometry
F	Phenylalanine
FID	Free Induction Decay
FT	Fourier Transform
G	Guanine
G	Glycine
G^{SO2}	Guanine sulphinate
G^{SO3}	Sulphonateguanine
Glu	Glutamic acid
Gln	Glutamine
Gly	Glycine
GO	Geometry Optimisation
H	Histidine
His	Histidine
HNPCC	Hereditary Non-Polyposis Colon Cancer
HSQC	Heteronuclear Single Quantum Correlation
HTH	Helix-Turn-Helix
I	Isoleucine
Ile	Isoleucine
INEPT	Insensitive Nuclei Enhanced by Polarisation Transfer
K	Lysine
L	Leucine
Leu	Leucine
Lys	Lysine
M	Methionine
m/z	Mass to charge ratio
mRNA	Messenger RNA
MALDI	Matrix-Assisted Laser Desorption/Ionisation
Met	Methionine

MLEV-17	A broadband composite pulse decoupling scheme
MMR	Methyl-directed Mismatch Repair
MSH	Human MutS
N	Asparagine
nm	Nanometre
NMR	Nuclear Magnetic Resonance
NTD	N-Terminal Domain
NDB	Nucleic acids DataBase
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect SpectroscopY
P	Proline
PDB	Protein Data Bank
PFG	Pulsed Field Gradient
Phe	Phenylalanine
Pro	Proline
Pu	Purine
Py	Pyrimidine
Q	Glutamine
R	Arginine
RCSB	Research Collaboratory for Structural Bioinformatics
rf	RadioFrequency
RMSD	Root Mean Square Deviation
RNA	RiboNucleic Acid
S	Serine
Ser	Serine
SEDUCE-1	A broadband composite pulse decoupling scheme
SDS-PAGE	Sodium Dodecyl Sulphate-PolyAcrylamide Gel Electrophoresis
Taq	*Thermus aquaticus*
tRNA	Transfer RNA
T	Thymine
T	Threonine
TG	Thioguanine
Thr	Threonine
TOCSY	Total Correlation Spectroscopy
Trp	Tryptophan
Tyr	Tyrosine
UV	Ultra-Violet
MD	Molecular Dynamics
RDC	Residual Dipolar Couplings
U	Uracil
V	Valine
Val	Valine
W	Tryptophan
WALTZ-16	A broadband composite pulse decoupling scheme
WATERGATE	WATER suppression by GrAdients-Tailored Excitation
WT	Wild Type
Y	Tyrosine
5-FU	5-Fluorouracil

Index of figures

Index of tables

Chapter 1 - Introduction

1.1 Structure and function of DNA[1-4]

In nature there are two types of nucleic acids and these are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA plays a primary role as it is the carrier of genetic information in all cellular organisms. Without DNA, genetic material cannot be conserved so it is a key molecule in all cellular functions. Two other important roles of DNA are that it replicates itself during cell division and also directs its transcription to form messenger RNA (mRNA). The mRNA in turn is translated by the ribosome and transfer RNA (tRNA) to synthesise a polypeptide of a defined sequence.

The basic building blocks of a nucleic acid molecule are known as nucleotides which contain three components to their structure: a heterocyclic aromatic base, a pentose sugar and a phosphate group. In DNA, there are four bases and these can be substituted monocyclic pyrimidines or bicyclic purines. The pyrimidine bases are cytosine (C) and thymine (T) and the purine bases are adenine (A) and guanine (G) (Figure 1.1).

Nucleotides can covalently link together to form long chains which are known as polynucleotides; covalent bonding occurs between the phosphate and sugar groups (Figure 1.2). For any one nucleotide, the phosphate group attached at the 5' position of the sugar ring in linked to the hydroxyl group at the 3' position on the following nucleotide. Each phosphate-hydroxyl bond is known as a phosphodiester linkage.

With the exception of the first and last nucleotides in a polynucleotide chain, which have free 5' phosphate and 3' hydroxyl groups respectively, all of the 5' and 3' groups are involved in phosphodiester linkages.

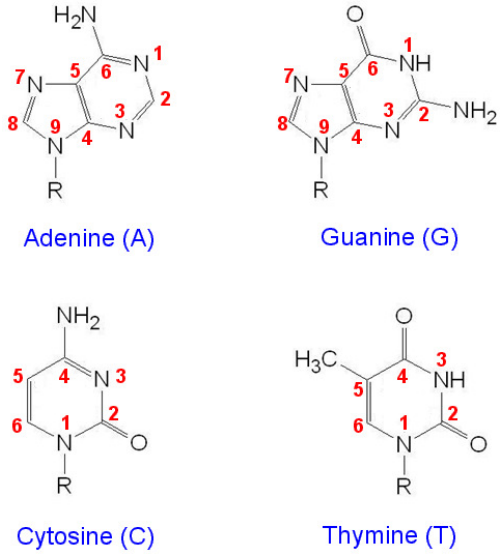


Figure 1.1. Structure of purine (A, G) and pyrimidine (C, T) bases present in DNA molecules. The common numbering scheme of the atoms is shown for each base. The R group represents bond connectivity to the sugar-phosphate group which is discussed later[1].

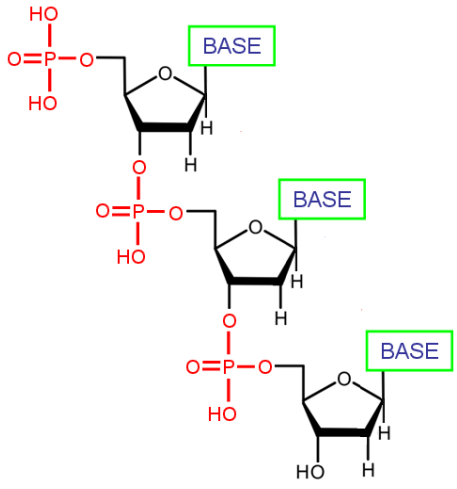


Figure 1.2. Illustration of a polynucleotide chain. The three individual nucleotides are joined by a series of phosphodiester linkages shown in red. The other components of nucleotide structure, pentose sugar and heterocyclic aromatic base, are shown in black and green respectively[1].

Base pairs form according to complementary pairing rules. These state certain associations are preferred between purine and pyrimidine bases. Each base pair is formed through hydrogen bonds between bases which align on opposing polynucleotide strands. Complementary base pairing only allows G to pair with C and A to pair with T.[3] GC and AT base pairing involves three and two hydrogen bonds respectively and these base pairs are often known as canonical or Watson-Crick base pairs (Figure 1.3). Canonical base pairs not only convey stability to the double helical structure of DNA but also align the functional groups in order to interact with other molecules such as proteins.

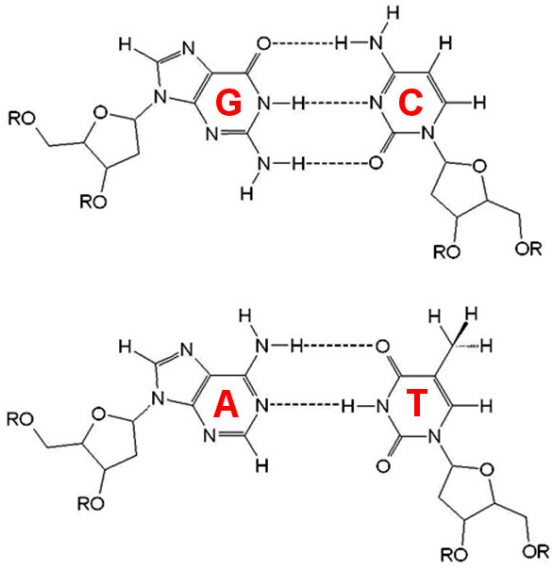


Figure 1.3. Illustration of GC (Top) and AT (Bottom) canonical Watson-Crick base pairs. The R groups denote the attachment sites for the phosphorus atoms at 5' and 3' hydroxyl groups of the sugar ring[1].

Stacking between adjacent bases is caused by interaction of π-orbitals between the aromatic rings of bases. This interaction contributes to the overall stability of nucleic acid structure and base stacking between purine-purine bases are commonly found to be the most stable, followed by pyrimidine-purine and pyrimidine-pyrimidine interactions and this has implications in mismatch DNA repair.[1,2]

The sugar-phosphate group has equal importance as it forms the backbone of DNA molecules. In nucleic acids, there are two types of sugar rings; deoxyribose and ribose. Deoxyribose is present only in DNA (Figure 1.4) whereas ribose is only present in RNA.[1]

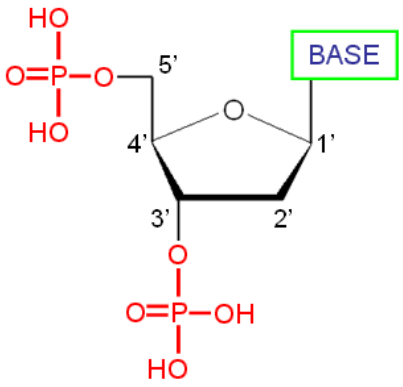


Figure 1.4. Structure of deoxyribose sugar-phosphate group.[2]

The five-membered deoxyribose sugar pucker determines which conformation the DNA exists in. B-DNA has a $C_{2'}$-*endo* pucker and A-DNA has a $C_{3'}$-*endo*; these are shown in Figure 1.5. Z-DNA is slightly more complex as the purines and pyrimidines each adopt different pucker conformations. The purines all have a $C_{3'}$-*endo* sugar pucker and the pyrimidines all have $C_{2'}$-*endo* sugar puckers. The presence of both sugar puckers in Z-DNA is why it is has a dinucleotide repeating unit.[2]

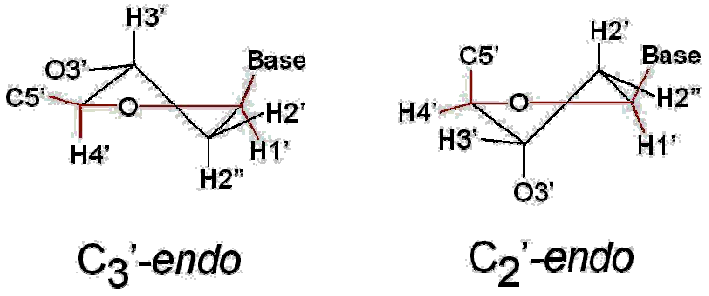


Figure 1.5. Illustration of sugar pucker conformations in nucleic acids; $C_{3'}$-*endo* (RNA) and $C_{2'}$-*endo* (DNA).[2]

Nucleotide conformation is also dependent on a number of dihedral angles which arise from the sugar-phosphate group (Table 1.1 and Figure 1.6). α, β, γ, ε and ζ define the conformation of the phosphate backbone only, whereas δ describes the phase amplitude and is both ribose sugar and phosphate backbone dependent. χ is known as the glycosidic angle and this defines the position of the base itself with respect to the ribose sugar ring.

Five dihedral angles exist which characterise the sugar pucker, ν_0 to ν_4 (Table 1.1 and Figure 1.7) but furthermore, the sugar conformation can also be defined by the pseudorotation phase angle (P) and amplitude (φ) of δ and χ.

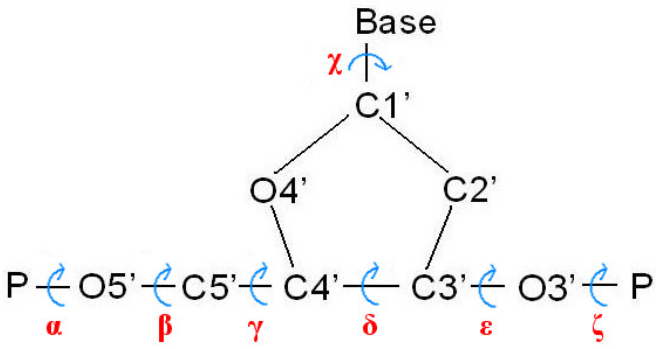


Figure 1.6. Backbone dihedral angles which define the conformation of the nucleotide and orientation of the phosphate backbone in nucleic acids.[2]

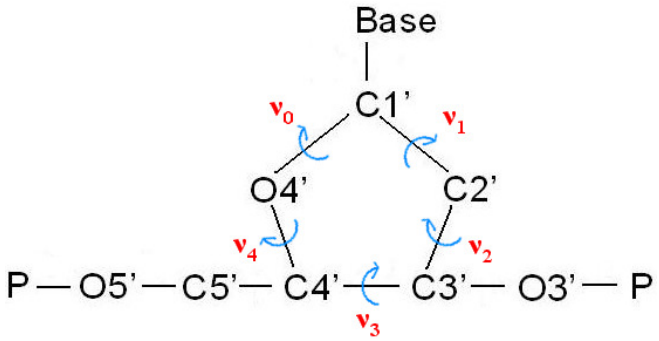


Figure 1.7. Pentose sugar ring dihedral angles which define the sugar pucker in nucleic acids.[2]

Dihedral angle	Defined by atoms
α	$O3'_{i-1}$ - P_i - $O5'_i$ - $C5'_i$
β	P_i - $O5'_i$ - $C5'_i$ - $C4'_i$
γ	$O5'_i$ - $C5'_i$ - $C4'_i$ - $C3'_i$
δ	$C5'_i$ - $C4'_i$ - $C3'_i$ - $O3'_i$
ε	$C4'_i$ - $C3'_i$ - $O3'_i$ - P_{i+1}
ζ	$C3'_i$ - $O3'_i$ - P_{i+1} - $O5'_{i+1}$
χ (Pu)	$O4'_i$ – $C1'_i$ – $N9_i$ – $C4'_i$
χ (Py)	$O4'_i$ – $C1'_i$ – $N1_i$ – $C2'_i$
ν_0	C4' - O4' - C1' - C2'
ν_1	O4' - C1' - C2' - C3'
ν_2	C1' - C2' - C3' - C4'
ν_3	C2' - C3' - C4' - O4'
ν_4	C3' - C4' - O4' - C1'

Table 1.1. Backbone and dihedral angles in nucleic acids listed with the atoms by which they are defined. i, i-1 and i+1 correspond to the atoms on the same, previous and following nucleotide respectively.[2]

The double helix structure of DNA was elucidated by James Watson and Francis Crick in 1953[4] and the discovery is often regarded as one of the greatest scientific achievements in history. The determination of the DNA structure meant significant advances were possible in the field of genetics.

The structure of the DNA duplex itself is very well ordered, however, it adopts many conformations of which the main forms are B-DNA, A-DNA and Z-DNA.

Differences in the structural features distinguish which form of DNA is presented at the end of this section (Table 1.2).

1.1.1 B-DNA[1-5]

The B-form (Figure 1.8) is the most common conformation adopted by DNA and originally discovered using X-ray diffraction through the analysis of DNA fibers in the presence of Na^+ ions at 92% humidity.

B-DNA is a very ordered molecule with a number of distinct structural features. Each molecule consists of two polynucleotide chains which twist to form a right-handed helix with a strand diameter of approximately 20 Å. The duplex winds about a common axis and in such a way that it is impossible to separate the polynucleotide chains without first unwinding the helix.

The polynucleotide chains have both 5' and 3' ends and when a double helix forms they combine in an anti-parallel manner. Thus, the 5' end only joins with a 3' end and similarly a 3' end only joins with a 5' end during replication such that the two strands run in opposite directions.

If a helix is viewed down its central axis, one would see the sugar-phosphate backbone on the outer periphery of the DNA molecule whilst the bases occupy the central core of the helix. In this conformation, extra stability is gained because the repulsion caused by opposing sugar-phosphate groups is minimised. The base pairs are placed so that they lie in a plane, which is almost perpendicular to the helix itself.

In an ideal B-DNA molecule, the helical axis approximately passes through the centre of each base pair. This causes the formation of two exterior grooves, major and minor groves, in the helix structure. The way that the base pairs are aligned also allows base stacking to occur which stabilises the double helix structure even further.

1.1.2 A-DNA[1, 2, 5, 6]

A-form DNA initially appears to be very similar to B-DNA but in fact the bases are tilted 20º more towards the helix. This gives A-form DNA a much wider and flatter appearance with a deep major groove and a very shallow minor groove (Figure 1.8). It was first discovered, in a way similar to B-DNA, by using X-ray diffraction and is formed when the relative humidity drops to 75%. When these conditions are present, a reversible transformation occurs which converts B-DNA to A-DNA.

1.1.3 Z-DNA[1, 2, 5, 7]

Z-DNA is different from conventional B- or A-DNA as it is a left-handed helix. This means that unlike B-DNA, which winds in a clockwise manner, Z-DNA winds in an anti-clockwise manner. The standard convention is that Z-DNA resembles a left-handed drill piece with a deep minor groove and no apparent major groove (Figure 1.8).

Z-form DNA is not known to naturally occur in living cells as a major conformation of DNA although it is biolgically active. However, Z-DNA does form in sequences which have alternating purines and pyrimidines e.g. d(CG). This is because the backbone of a Z-DNA helix has nucleotides which alternate between *syn* and *anti*

conformations where purines are more stable in the *syn* conformation. To convert B-DNA to Z-DNA, a number of bases must flip so that they are upside down relative to their positions in B-DNA; this is brought about by the *anti* to *syn* conformational change.[4-6]

Studies carried out on Z-DNA are rendered difficult because it does not exist as a stable double helix. However, it is known that at high salt concentrations the Z-DNA conformation is stabilised as the counter ions reduce the electrostatic repulsions between the phosphate groups in complementary strands.[8] One way of promoting Z-DNA formation is by methylation at the C5 position of cytosine residues. The addition of this methyl group promotes Z-DNA formation because it is less exposed to the surrounding solvent molecules.[9]

Z-DNA was thought to have had no biological significance in the past, but recent discoveries of proteins that bind to Z-DNA in which the protein binds with high affinity and specificity provide strong evidence for the biological roles of Z-DNA e.g. the poxvirus protein E3L and the RNA modifying enzyme ADAR1.[10, 11]

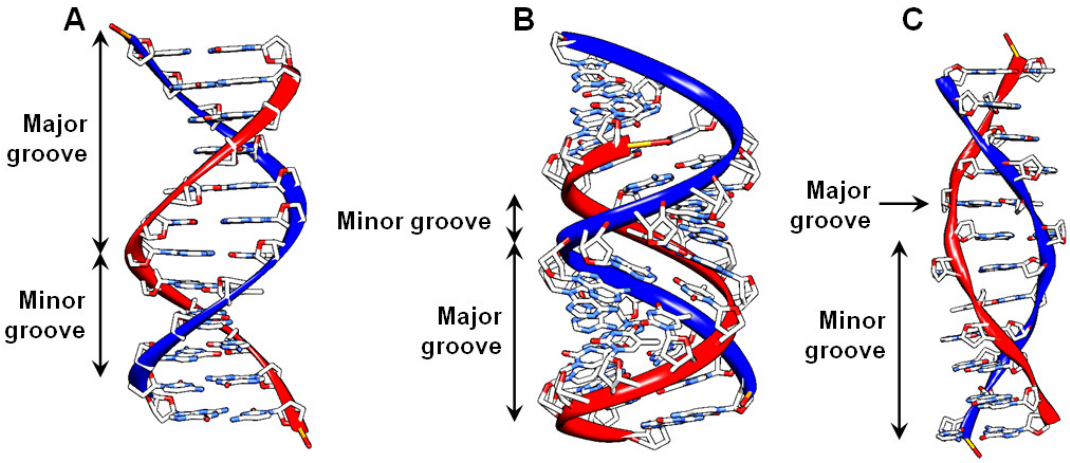


Figure 1.8. Molecular models illustrating the three main conformations of DNA; B-form (A), A-form (B) and Z-form (C).

Helix Parameter	B-DNA	A-DNA	Z-DNA
Helix sense	R	R	L
Sugar pucker	$C_{2'}$-*endo*	$C_{3'}$-*endo*	$C_{3'}$-*endo* (*syn*)
Residues per turn	10	11	12
Helical twist per base pair (°)	36.0	32.7	-9.0, -51.0
Rise per base pair (Å)	3.3 - 3.4	2.9	3.7
Base tilt (°)	-6.0	20.0	-7.0
Major groove width (Å)	11.7	2.7	8.8
Minor grove width (Å)	5.7	11.0	2.0
Major groove depth (Å)	8.8	13.5	3.7
Minor groove depth (Å)	7.5	2.8	13.8

Table 1.2. Summary of main helical parameters showing comparison between B-, A- and Z-DNA conformations.[3]

1.2 Mismatches in DNA and the mismatch repair pathway (MMR)

Mismatches arise in the DNA replication process, which although highly accurate, produces errors leading to mutations. Any damage which can potentially affect the genetic material in DNA can be classed as DNA damage.[12,13] Mismatched bases must be repaired by locating the correct base in the mispair before replacement with the correct base. This ensures the fidelity of the genetic material is maintained and carried forward.

1.2.1 Structure of mismatches

Mismatches are base pairs which do not follow the Watson-Crick base pairing rules and are sometimes known as non-canonical (Figure 1.9). The presence of mismatches distorts the conformation of the DNA helix making the structure less stable. Significant reorientation of the phosphate backbone compensates for the loss in stability.

Since mismatch base pairs are non-canonical, it is a common assumption that they can only form between the four bases associated with DNA. On the contrary, the formation of mismatch base pairs may occur between the four common bases and chemically modified bases such as 6-thioguanine (6-TG), O^6-methylguanine (O^6meG), 5-hydroxyuracil, 5-flurouracil (5-FU) and N^7-methylguanine (N^7meG). The inclusion of modified bases will cause an even greater degree of helical distortion as their base attached functional groups are more bulky resulting in greater steric hindrance.[13-19]

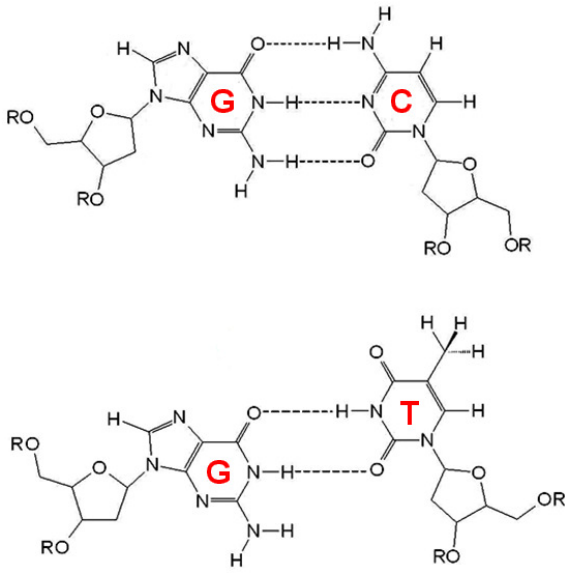


Figure 1.9. Comparison between canonical GC (Top) and mismatched GT (Bottom) DNA base pairs. The R groups denote the attachment sites for the phosphorus atoms at 5' and 3' hydroxyl groups of the sugar ring.

1.2.2 Significance of mismatch recognition and repair

The recognition of mismatches is very important in order to regulate the genome. Errors in the mismatch repair (MMR) pathway are normally a pre-cursor to Hereditary Non-Polyposis Colon Cancer (HNPCC). Therefore, structural insights into MMR are of great biomedical importance.[20-23]

As mentioned previously, the best understood MMR pathway exists in *E. coli*, although MMR pathways have been discovered in nearly all organisms from bacteria to humans. Those found in prokaryotic organisms form dimers whilst in contrast MutS in eukaryotic organisms may contain multiple homologue genes.[24] By using the *E. coli* pathway as a model, more can be understood about the human MMR process. Advancements in this MMR pathway could potentially lead to anti-cancer

treatments as more becomes known about the structure and mechanism of action of MutS.

One of the latest developments concerning the real-life application of MutS is early detection of mismatches by analytical techniques and sensors. Cho *et al.* reported a method of detecting mismatches by a fluorescent technique that involves labelling the dimer interface of MutS with a fluorophore.[25] Thus, it is possible to detect mismatches by monitoring any changes to the level of emission. As the level of emission decreases, this corresponds to an increase in the concentration of mismatches.

Another example of a sensor was constructed by Sachadyn *et al.* which combined MutS protein with an enzyme, β-galactosidase. This combination allowed the functionality of recognition coupled with the benefits of enzymatic activity, which enabled simple colormetric detection.[26,27]

More recently a new electrochemical detection method using MutS to detect single nucleotide polymorphisms (SNPs) has been developed by Chen *et al.*[28] involving electrical impedence spectroscopy. The method published used MutS which has been immobilised onto a gold electrode to recognise single base mutations in heteroduplex DNA. After allowing for the immobilised MutS to bind to the DNA mutations, it is then immersed in a solution. The solution contains methylene blue which intercalates specifically to the DNA and acts as an external redox indicator. The electrochemical current of methylene blue is then measured which gives an indication of the concentration of mismatched DNA which has been adsorped onto

the electrode surface. The method is very sensitive as the affinity for canonical DNA is much lower compared to DNA containing mutations and combines high sensitivity with low cost and rapid detection.

Another important feature of MutS is that it is thought to affect the degree of repair of DNA-anticancer drug complexes in cells when used to treat cancers. An example of where this occurs is in cisplatin,[29] which is an anti-cancer drug, used to primarily treat testicular and ovarian cancers. The mode of action here is that cisplatin binds to the DNA and this causes lesions, which are the cause of cytoxicity in cells by cisplatin which can ultimately lead to cell death. MutS has the ability to bind to these lesions and trigger cell death hence stopping tumour growth.[29-35]

So as the above discussion highlights, the biomedical applications of MutS are wide and arise from a number of sources. These real-life benefits can only be brought about through better understanding of the MMR pathway, both in prokaryotes and eukaryotes.

1.2.3 Mechanism of Mismatch Repair (MMR)

There are three main pathways in which DNA damage can be repaired of which mismatch repair (MMR) is one. MMR is an evolutionarily conserved method of DNA repair and is an example of specific single strand repair.[22] It repairs base errors generated during DNA replication which escape proofreading processes after replication. This is often considered as the main source of DNA damage.

Substitution of bases resulting in a mismatch is the most common type of base mutation. There are two types of base substitution which can result in a mismatch, the first is Transition, which is substitution of a purine or pyrimidine by another e.g. A substitutes for a G in a GC base pair. The second type of substitution is Tranversion, which is the substitution of a purine with a pyrimidine or a pyrimidine with a purine.[20,21]

The MMR pathway reduces errors in the DNA sequence by up to a magnitude of 1000 fold and there are two key steps to the process itself. Figure 1.10 shows the MMR process in *E. coli*, which is the best understood of all MMR pathways.

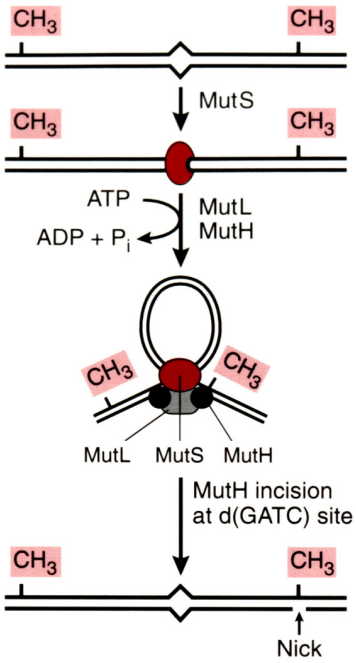


Figure 1.10. Scheme illustrating the methyl-directed mismatch repair pathway in *E. coli*.[20, 30]

The initial recognition of a mismatch by MutS is the key interaction which triggers the subsequent steps in the MMR pathway. If this step does not occur, MMR will not proceed. Recognition by MutS is followed by interaction with MutL and MutH

in an ATP-dependent process. MutH nicks the DNA at the methyl group in the daughter strand at a hemimethylated GATC site and endonucleases strip the daughter strand of all bases at the mismatch site between the methyl groups. *In vitro* studies[23] have shown that MMR is a bidirectional process thus MutH can nick on either side of the mismatch. The correct sequence is then synthesised using DNA polymerase and DNA ligase.[12, 30, 31, 36]

1.2.4 The effect of mismatch orientation, sequence context and oligonucleotide length on mismatch repair[37-39]

The affinity of MutS repair protein for DNA mutations is well known and as shown later in section 1.4, it binds well to a large number of different mutations and that it interacts with greater affinity to mispairs which are in a GC rich environment. However, the effect of changing the orientation of the mismatch or changing the sequence context of neighbouring base pairs has not been investigated to a great extent.

A study by Mazurek *et al.* did provide some experimental evidence regarding whether certain mismatches were more stable when they are flanked by specific base pairs.[37] Mazurek took a number of mismatches for example, GT, CA, GG and GA and measured the mismatch binding kinetics when the DNA was allowed to interact with human MutS protein. The sequence context was changed so that there would be two 3' end purines or two 3' end pyrimidines for each mismatch. The study showed that when a mismatch base pair is immediately flanked by two symmetric 3' end purines, human MutS ATPase activity was increased with the greatest effect

being observed for a GT mismatch, K_D (GT) = 1.9 x 10^{-9} M when the sequence context is two 3' purines. In contrast, in the presence of two 3' end pyrimidines, the ATPase activity of human MutS was reduced, K_D (GT) = 4.6 x 10^{-9} M. The data given by Mazurek supports the theory that mismatches are more stable in GC rich duplexes however, the paper does not indicate how the stability relates to structure.

A second study carried out by Chung *et al.*[38,39], investigated how the stability of DNA microsatellites are affected by varying the microsatellite length and sequence context. A characteristic feature of cells which are MMR deficient is microsatellite instability, which is used as a diagnostic marker for reduced MMR activity in tumour cells.

Microsatellites are effectively short repeating DNA sequences, ranging from 1-6 nucleotide units long which are located in non-coding regions of the genome. However, in a number of instances, these microsatellites are located in coding regions of particular genes and if the MMR pathway is defective frameshift mutations occur within these genes. Over a short period of time these mutations build up resulting in eventual loss of protein function. This type of microsatellite instability is what occurs during the intial development of many colorectal cancers including Hereditary Non-Polyposis Colon Cancer (HNPCC). The results showed that the frequency of mutation when MMR is defective is much higher for microsatellites which are longer. The rate of mutation also seemed to be affected by the sequence context of the DNAs which are flanking the gene.

1.3 6-Thioguanine (6-TG) modified DNA

6-Thioguanine (6-TG) like a number of chemically modified bases has potentially important biomedical and pharmacological uses. More specifically, 6-TG is an important molecule due to its antileukaemic properties.[40] Its cytotoxicity effect is brought about through its incorporation into DNA as 2'-deoxy-6-thioguanosine and although it is widely used, not much is known regarding how this incorporation affects the DNA helix from a structural and functional perspective. Consequently, the mechanism as to why cytotoxicity results from incorporation into DNA is still relatively unknown.[41]

1.3.1 Structure of 6-thioguanine DNA

The structure of 6-TG is shown in Figure 1.11 which clearly shows the single structural difference, which is the substitution of the O6 atom in guanine with an S6 atom. A simple comparison with a GC base pair shows canonical base pairing is conserved even when guanine is substituted with 6-TG (Figure 1.12).[41]

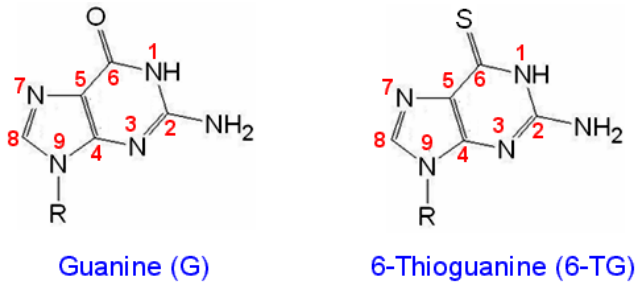


Figure 1.11. Comparison of guanine and 6-thioguanine bases clearly showing the substitution of oxygen with sulphur at the C6 position.[41]

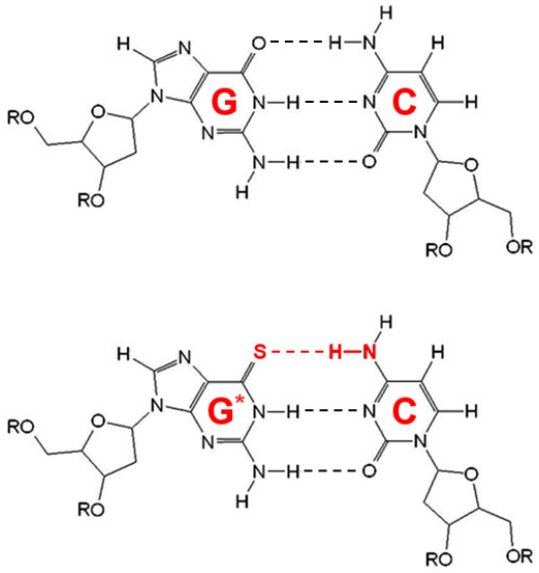


Figure 1.12. Comparison in structure between canonical GC (Top) and 6-thioguanine modified G*C (Bottom) base pairs. The R groups denote the attachment sites for the phosphorus atoms at 5' and 3' hydroxyl groups of the sugar ring.

Sulphur when compared to oxygen is a poor hydrogen bond acceptor and this is observed in a 6-TG substituted GC base pair. The weakening of the hydrogen bond between S6 on 6-TG and the hydrogen bonded amino proton on the opposing cytosine translates in approximately a 10º opening of the base pair towards the major groove (Figure 1.13).[40]

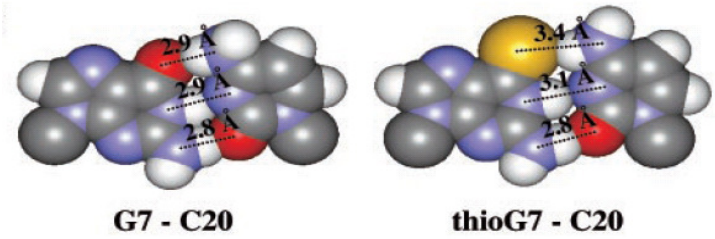


Figure 1.13. Comparison between normal GC (Left) and 6-thioguanine modified GC (Right) base pair hydrogen bond lengths.[40]

The extent of the weakening of the hydrogen bond was investigated by Guerra *et al.*[42] where *ab initio* theory was used to calculate the bonding energies and charge distribution of AT and GC base pair mimics in which the O•••H-N hydrogen bonds were modified to S•••H-N hydrogen bonds. A number of mimics were used where the number of sulphur substitutions varied. Geometries and energies of the base pair hydrogen bonds were obtained at the BP86/TZ2P level of a generalized gradient approximation (GGA).

The study showed that the substitution of sulphur in AT base pairs caused a decrease in the interaction energy of the S4•••H61-N hydrogen bond (ΔE_{int} = -11.1 kcal mol^{-1}) when compared to a Watson-Crick O4•••H61-N hydrogen bond (ΔE_{int} = -13.0 kcal mol^{-1}). This change was also accompanied by an increase in the bond length from 2.85Å to 3.35Å most likely caused by the sulphur atom being larger than oxygen. The increased S4•••H61-N bond length also caused the second hydrogen bond (N1•••H3) to also increase by 0.05Å. Substitution of the thymine O2 atom, which is not normally involved in base paired hydrogen bonding, also increased to its neighbouring N1•••H3 hydrogen bond to 2.84 Å. Although this difference is small compared to the equivalent Watson-Crick length (2.80Å), it is important to note that it led to the compression of the O4•••H61-N hydrogen bond. This effect was not observed to the O2•••H21 bond length when investigating the effect of S4•••H61-N which implies that changes accompanying the substitution of sulphur are not independent on its position.

For GC base pairs, similar results were observed when compared to AT base pairs. Substitution of one oxygen was accompanied by a decrease in the interaction energy

of approximately 2 kcal mol^{-1}. When both oxygens involved in base pair hydrogen bonds have been substituted, the difference is approximately 5 kcal mol^{-1} when compared to a Watson-Crick GC base pair. Bond lengths were also changed in the presence of sulphur but in contrast to AT base pairs, elongation in the top or bottom hydrogen bond resulted in compression of the hydrogen bond at the opposite end of the base pair. S6•••H41-N (3.22Å) change the middle (H1•••N3) and third (O2•••H21) by approximately +0.2Å and -0.1Å respectively and in contrast S2•••H21-N changes the length of H1•••N3 and O6•••H41 by +0.4Å and -0.05Å respectively.

The above discussion shows that bond lengths do change when a single sulphur substitution is carried out to both AT and GC base pairs however, the base pair itself remains planar; this was no longer observed by Guerra *et al.* when both the oxygen atoms were substituted. It was found for both base pairs that the geometry of the base pair adopted a more twisted conformation to compensate for the steric hindrance caused by the two sulphur atoms.

Charge distributions were obtained using the Voronoi deformation density (VDD) method. Changes in the charge of sulphur were dependent on the position at which it was substituted. But it was found that in all cases the sulphur atom held a smaller charge when compared to oxygen which is expected as sulphur is less electronegative when compared to oxygen.

A second theoretical study carried out by Villani[43] also showed similar results to those published by Guerra *et al.*[42] Villani showed that substitution of oxygen by

sulphur elongated the corresponding hydrogen by approximately 0.5Å and that this directly impacts the middle hydrogen bond by increasing its length by ~0.1Å and ~0.2Å in AT and GC base pairs respectively. It was also shown that sulphur substitution casues distortion to the planarity of the base pair and that a double substitution of sulphur results in a non-planar base pair system.

Vilaani was able to give a more detailed analysis of the how the bond energies and atomic charges changed following substitution of oxygen by sulphur and it was discovered that AT and GC base pairs behaved very differently.

For AT base pairs, it was observed that following sulphur substitution, the bond energy of the S4•••H61-N had increased whereas the bond energy of the N1•••H3 hydrogen bond has reduced. The charge distribution remains similar in the N1•••H3 but the addition of sulphur results in the S4•••H61-N bond being less polar which correlates to the findings given by Guerra *et al.* Additional substitution of the O2 atom for S2 results in an ulterior reduction in the bond energy of the N1•••H3.

In contrast for GC base pairs, substitution of oxygen to sulphur cause the bond energy to increase for all three hydrogen bonds. However, when two sulphur substitutions have been made, the bond energies of all three hydrogen bonds are reduced. The atomic charge of the sulphur in a single substituted base pair is similar to that of oxygen and this does not change.

Previous studies have shown that the effect of sulphur substitution remains local to its position but the clearest change occurs to the base pair in which the sulphur is involved in. Weakening of hydrogen bonding may extend to the hydrogen bond between the imino proton (H1) of 6-TG and N3 on the opposing cytosine. The effect on the imino proton chemical shift is that it appears further highfield (Figure 1.14).

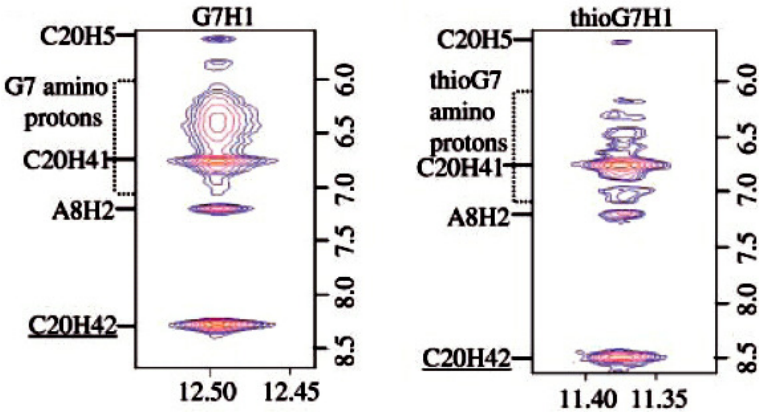


Figure 1.14. 600MHz expansion of imino region showing cross peaks to neighbouring atoms for normal GC (Left) and 6-TG modified GC (Right) base pairs.[41]

NMR variable temperture experiments carried out by Bohon *et al.*[44] showed that the imino proton resonances for the 6-TG bases only broadened at slightly lower temperatures compared to those in canonical base pairs which highlight the stability of the 6-TG base and these changes were highly localised to the 6-TG base.

1.3.2 Significance of 6-thioguanine DNA[40, 44, 45]

6-Thioguanine is significant since it is widely used as an antileukaemic agent yet its mechanism for action is not fully understood. Even though the structure is similar to that of normal guanine, there are significant changes in the base pair dynamics and physical properties of the molecule itself. Weakening of hydrogen bonding causes an 80 fold decrease in the lifetime of a 6-TG modified GC base pair.[40] The decrease in base pair stability was supported by UV absorbance melting studies which showed a 6ºC decrease in the melting temperature upon replacement of the oxygen atom by sulphur at the 6-position. The observation of faster exchange between the 6-TG imino proton and water is also consistent with these findings.

The difference in physical and chemical properties of 6-TG meant that its involvement in DNA-protein interactions has warranted investigation. It is known that other chemotherapeutic agents such as cisplatin carried out their mode of action by significantly changing the structure of the target DNA; this is not the case with 6-TG. The similarity to normal guanine means that changes caused to the DNA structure are localised.

An important example is that of the mismatch repair system in DNA. In mismatch repair, non-canonical base pairs are repaired after replication by a repair protein. However, a recent study has highlighted that when 6-TG modified base pairs are present, irrespective of whether the base pair was canonical or not, binding occurred. More recently, the complex has been shown to have a preference to bind to 6-TG substituted base pairs, the reason for this could lie in the unusual base pair dynamics

of 6-TG. The greater ionic radius of sulphur (1.85Å) compared to oxygen (1.40Å) may cause strain in the local helical conformation resulting in weaker hydrogen bonding and a slight opening of the DNA. This combination may mean that groups involved in base pairing as more likely to be exposed to the surrounding solvent, providing a more accessible target for binding.

The lower electronegaivity exhibited by sulphur may also be crucial in how its biological activity is brought about as this suggests that sulphur may be involved in different types of molecular interactions compared to oxygen. As the studies by Guerra and Villani have shown the atomc charges are highly influenced by the position of the substituted sulphur. This further highlights the importance that base pair energetics and charge distribution, alongside structural change, may play in further understanding how the effect of modified bases, such as 6-TG, may be brought about.

1.4 *E. coli* and human MutS repair protein

1.4.1 The role of MutS in MMR

MutS is a key protein involved in the recognition of DNA mismatches, which is the critical primary event in the repair mechanism. Studies by Wagner *et al.*[46] showed that the affinity of MutS to a number of mismatches in DNA varies with GT and TG mispairs being recognised best (Figure 1.15) and CC being recognised with the least efficiency. From this it can be noted that the efficiency of repair is inversely related to the degree of base pair distortion in a mismatch. GT mispairs cause the least distortion and are repaired with the highest efficiency and in contrast CC mispairs are known to cause the most helical distortion out of the eight base-base mispairs.

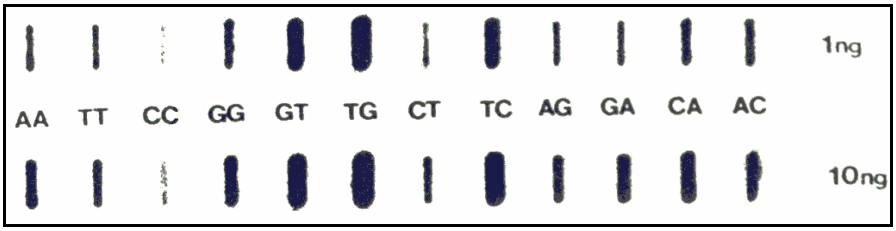


Figure 1.15. Binding affinity of immobilised *E. coli* MutS with various mismatches, 30mer sequence.[46]

However, mismatches are recognised with higher affinity if they are present in GC rich duplexes. This work by Wagner was in support of an earlier study in which the hypothesis was that the repair of a mismatch is strongly influenced by the sequence of the duplex DNA.[47]

Damage recognition is initially performed by MutS homologues, which bind to either the mismatched or unpaired bases; this triggers the MMR process, which repairs the damage. The second stage in the recognition process ensures that the mismatched base is only repaired in the newly synthesised strand and not the parent strand. In other words the second recognition step allows strand discrimination. The process itself is usually known as "long-patch" repair because the strand recognition can occur over a long distance away from the mismatch itself, up to 1000 base pairs away. Following recognition MutS activates MutH (an endonuclease) in an ATP-dependent process with the aid of MutL before repair of the mismatch base is carried out.[29,36,48,49]

The binding of MutS to a mismatch protects about 22 base pairs of DNA from digestion and it has not been shown that any other repair proteins does this, thus illustrating the importance of MutS in mismatch recognition. The exact mechanism of how the mismatch is recognised and repaired is still not understood in three dimensional structure terms.

The change in conformation of the GT base pair is quite surprising since normally this exists as a wobble pair whereas in a mismatch DNA-MutS complex it has a different conformation. A change in sugar pucker conformation at the mismatch site allows possible hydrogen bonding to the mismatch pair by MutS; this will be discussed in Section 1.4.2.[48,49]

More recent research on MutS has revealed some structural features and functions which may be of significance. A recent study of the kinetics of mismatch binding

with *Taq* MutS showed that in the presence of ATP and a mismatch, the mismatch is released by MutS at a slower rate. When no mismatch is present ADP-bound MutS engages in interactions with DNA scanning base pairs for errors. The study also showed that in the absence of ATP, the MutS protein loses its activity as a recognition molecule.[50] ATP binding has been shown to induce conformational change at the mismatch site but this change may diffuse along the DNA. This shows that ATP is crucial in mismatch repair as it not only drives MutL activity, but its binding to MutS and hydrolysis is required in mismatch recognition.[51-53]

1.4.2 X-ray crystal structures of MutS

Studies carried out by X-ray crystallography reveal that mismatch recognition occurs at the NTD. The first *E. coli* MutS-DNA complex solved by X-ray crystallography revealed the possible mechanism of action by which mismatches are repaired. The complex contained a GT mismatch as this is a very stable mismatch and one with which MutS has the highest affinity towards.[46, 49]

Prokaryotic MutS forms homodimers which can recognise a variety of small insertion/deletion loops (IDLs), mismatched base pairs as well as a number of DNA substrates. Eukaryotic MutS in contrast form a number of heterodimers, which have a greater range of recognition in terms of substrates.[49]

Its structure has been established by X-ray crystallography and this revealed a dimer structure (Figure 1.16). Although the dimer appears to be very ordered it is actually

asymmetric and when bound it is known that only one of the monomers binds to the DNA and carries out the mismatch repair process.

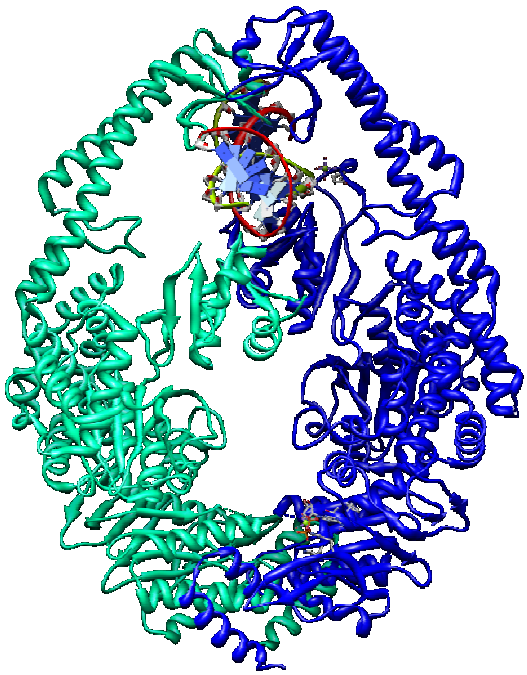


Figure 1.16. X-ray crystal structure of *E.coli* MutS bound to mismatched GT DNA. The two monomers of MutS are shown in cyan and dark blue respectively. The phosphate backbone of the GT DNA are shown in red and yellow.[49]

Each monomer is approximately 97 kDa in size and has 5 distinct domains (Figure 1.17). The most important domain is the mismatch binding domain, which with the clamp domain forms the mismatch binding site on the MutS structure. The mismatch domain binds directly to the mismatch whilst the clamp domain holds the MutS molecule in place by binding with the DNA backbone. Due to their function these two domains are actually mobile and so are disordered in the X-ray crystal structure.

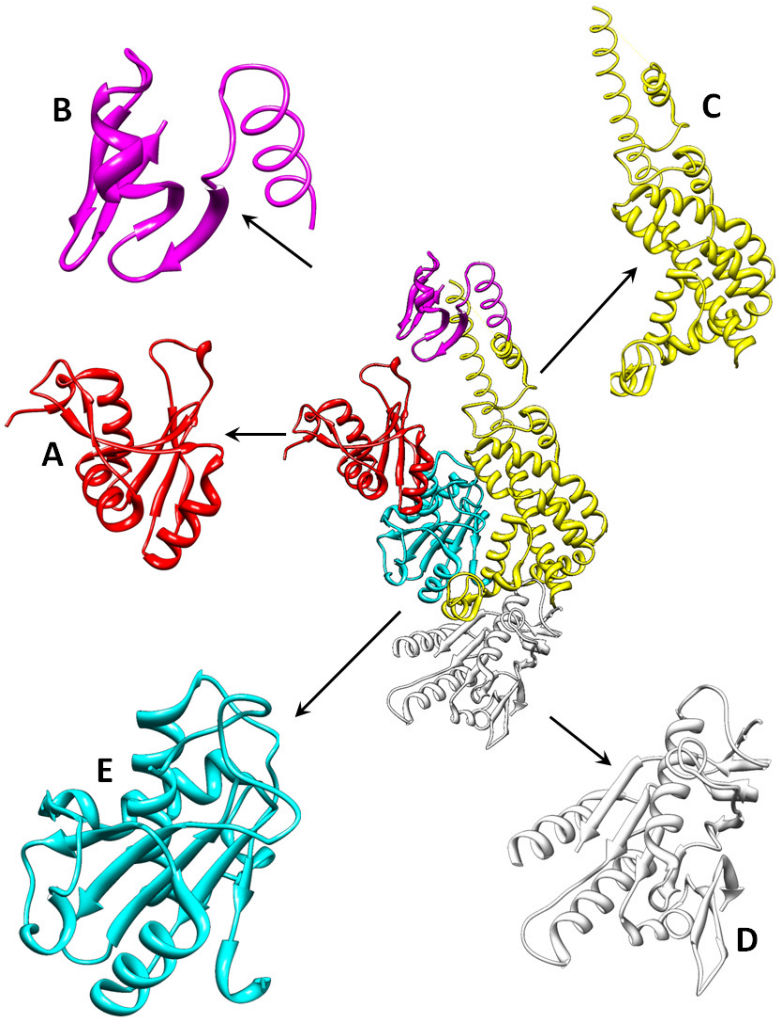


Figure 1.17. X-ray crystal structure of *E. coli* MutS monomer, outlining domains; mismatch binding domain (A), clamp domain (B), lever domain (C), ATPase domain (D) and the connector domain (E).[49]

The ATPase and Helix-Turn-Helix (HTH) domains make up the C-terminus of MutS. It has been found that the structure of MutS changes upon binding with DNA by inducing a change in the ATPase domain and by doing so may cause slight rotation of the dimer itself.

The core domain is also of importance as it provides a structural bridge between the mismatch and clamp domains in order to ensure that both function correctly. The

core domain acts more directly with the clamp domain compared to the mismatch to which the core acts indirectly on through the connector domain.[48,49]

Interaction with mismatched DNA only utilises one of the mismatch domains from the two MutS monomers. The other mismatch domain of the other monomer actually is not involved in the binding of the mismatch but is essential in stabilising DNA binding.

Poor base stacking causes a change in the helical conformation in the local area and the change in conformation is believed to be a possible trigger for recognition by MutS proetin.[54] Upon binding to *E. coli* MutS, the DNA kinks sharply towards the major groove by approximately 60° at the site of mismatch. This is accompanied by a change in the local conformation resulting in several deoxyribose sugar rings adopting a C_3'-*endo* conformation. Figure 1.18 shows the interactions which occur between mismatched GT DNA and *E. coli* MutS which also highlights which residues undergo a change in sugar pucker. The DNA becomes wider at the minor groove, which narrows its complementary major groove. Widening of the minor groove at the mismatch creates a point of entry allowing an amino acid residue to come into the DNA helix and stabilise the mismatch bases.

It is a logical hypothesis to think that the bending of the mismatch DNA would lead to specificity for binding by MutS. However, AFM studies[23] have shown that even when bound to MutS protein, unbent DNAs still occur *in vivo*. This means that kinking of the DNA helix is not enough to assign mismatch DNAs specificity by

MMR. As a result of this finding, the removal and repair of a DNA lesion by MMR is likely to involve multiple changes in the conformation of both DNA and protein.

X-ray crystal studies[49,55] have found that the recognition itself occurs over an area of 1.25Å on the minor groove. There are three main residues involved in the *E. coli* MMR process. These are Phe36, Glu38 and Asp35. With a GT mismatch, Phe36 comes into the DNA structure and base stacks on the mispaired T residue, which stabilises the mismatch base pair. Base recognition by Glu38 and Asp35 is the next stage in MMR. The contact to the mismatch itself involves a Phe-X-Glu motif and it is found that this is conserved in prokaryotic and eukaryotic mismatch repair pathways.

Although the Phe36 plays a critical in stabilising the mismatch by engaging in stacking interactions, a key role is also played by Glu38. As mentioned previously MMR is carried out only on the daughter strand allowing strand discrimination. However, at the point of recognition Glu38 also discriminates between canonical and mismatched DNAs as its forms a key hydrogen bond contact with the mismatched base. Consequently, if a DNA mutation is not present, Glu38 cannot form this hydrogren bond contact and subsequently, the recognition process is compromised. The mechanism described above for *E. coli* MutS and mismatch GT DNA is shown in Figure 1.19.[49,55-58]

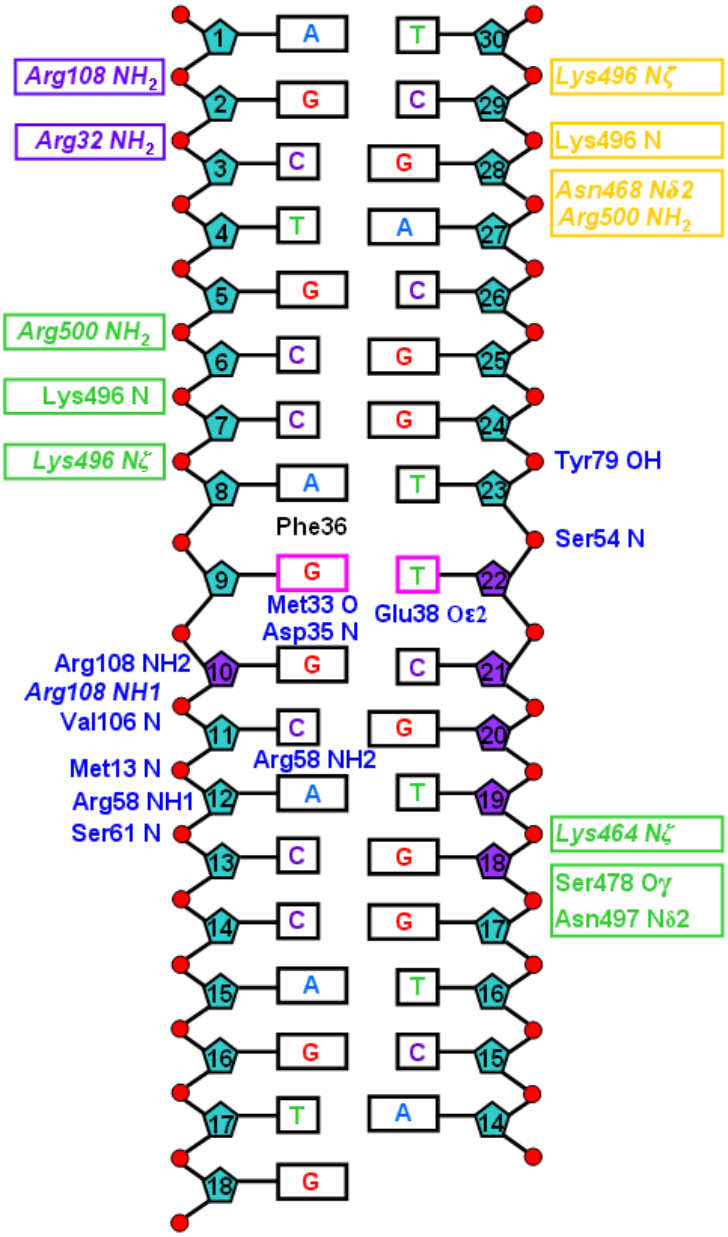


Figure 1.18. Illustration showing types of hydrogen bond interactions and salt bridges (italics) present in *E. coli* MutS - mismatch GT DNA complex. Interactions are distinguished by domain where blue, green, purple and yellow correspond to the NTD, lever, connector and clamp domains respectively, with the key Phe36 is shown in black. Non-specific contacts which take place are boxed and residues which undergo a change in sugar pucker are indicated in purple.[49]

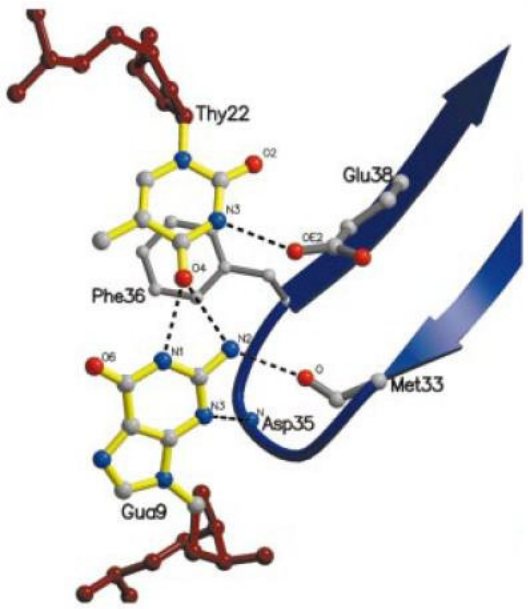


Figure 1.19. Mismatch recognition in *E. coli* MutS showing hydrogen-bonding at site of mismatch.[49]

The presence of the Phe36 has been shown to be very important as this residue base stacks with the subsequent mismatched base and stabilises the MutS-DNA complex. It has been shown that through substitution of Phe36 with a non-aromatic group the recognition site becomes defective in mismatch repair[56,57]. Since Phe36 is the key residue other mutations can be carried through amino acid substitution at this point. The purpose would be to determine what amino acid structural characteristics are important for it to engage in recognition activity as observed with Phe36.

More recently, X-ray crystal structures of human MutS bound to DNA substrates have been elucidated (Figure 1.20). Human MutS is 260 kDa in size, which is significantly bigger in molecular size than *E. coli* MutS, which is approximately 200 kDa. Similarly to *E. coli* MutS, human MutS recognises mispairs and insertion/deletion errors, but in addition it is capable of recognising damage caused by anti-cancer agents such as O^6meG and cisplatin.[59]

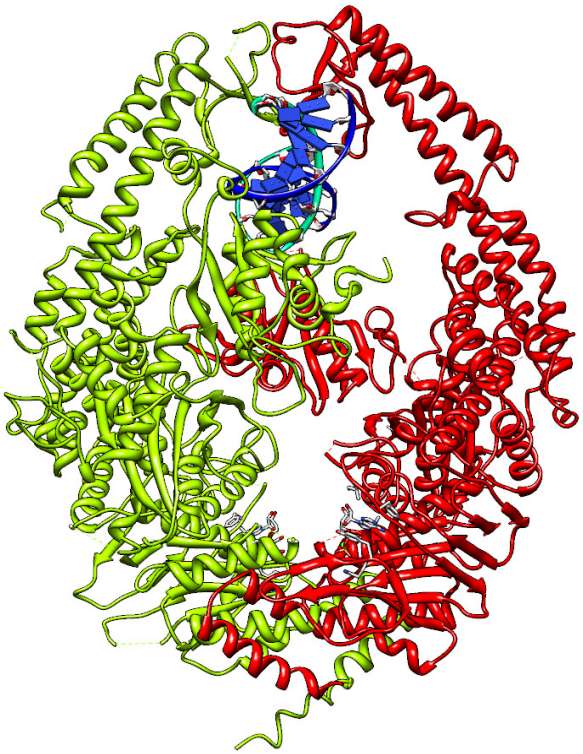


Figure 1.20. X-ray crystal structure of human MutS bound to mismatched GT DNA. The two monomers of MutS, MSH2 and MSH6, are shown in red and yellow respectively and the GT DNA is shown in blue.[59]

From the perspective of structure, one very significant difference between the two is that human MutS exists as a heterodimer (MSH2, MSH6). This means the molecule exists as a dimer but the constituent monomers are different in both domain conformation and function. This contrasts the structure of *E. coli* MutS which exists as a homodimer molecule. Similar to *E. coli*, each monomer can be divided into five main domains and these are shown in Figure 1.21.[59]

Although prokaryotic and eukaryotic MutS may have a lot of the same functionality, sequence homology between the two is very low in conserved regions of both monomers (<25%). However, it has been recently shown that the domains with the highest percentage of conservation are the NTD (14.8%) and the ATPase domain

(17.7%).[60] These values emphasise the importance of these two domains in the functionality of MutS and consequently, the recognition of mismatches.

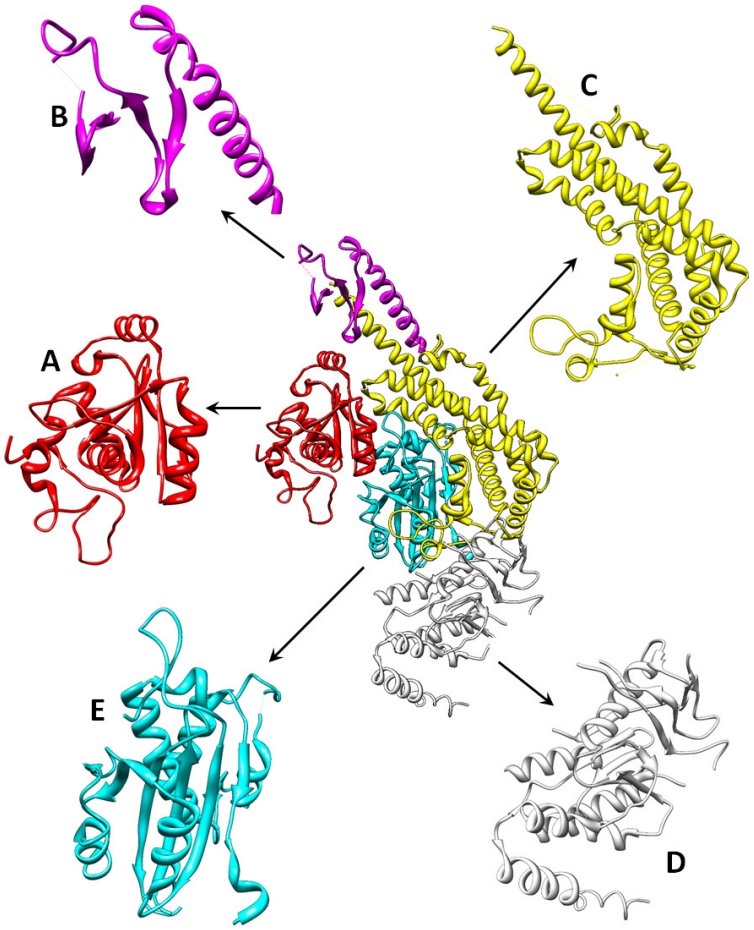


Figure 1.21. Functional domains within MSH6 monomer structure; mismatch binding domain (A), clamp domain (B), lever domain (C), ATPase domain (D) and the connector domain (E).[59]

When bound to a mismatched GT DNA substrate, only MSH6 binds specifically to the mispaired bases; Figure 1.22 outlines the many sites at which protein-DNA binding may occur. The Phe-X-Glu mismatch binding motif is conserved for MSH6, but whereas in *E. coli* MutS, Phe36 was the residue which binds specifically to the mismatched base, in MSH6, this role is performed by Phe432. Other key residues which lock the conformation of the binding site during mismatch recognition are

Glu434, Met459 and Val429 (Figure 1.23, A). The combination of the hydrogen bonding described here widens the DNA at the minor groove causing the mismatched T residue to tilt towards the mispaired G. The interactions and changes in conformation observed for MSH6–mismatched DNA binding is very similar to those observed for *E. coli* MutS, indicating conservation of the DNA binding mechanism in MutS homologs. Even in the presence of different DNA substrates such as modified O^6meG, the bound complex remains very similar in structure when compared to the MSH6–mismatched GT DNA complex (Figure 1.23, C).[59]

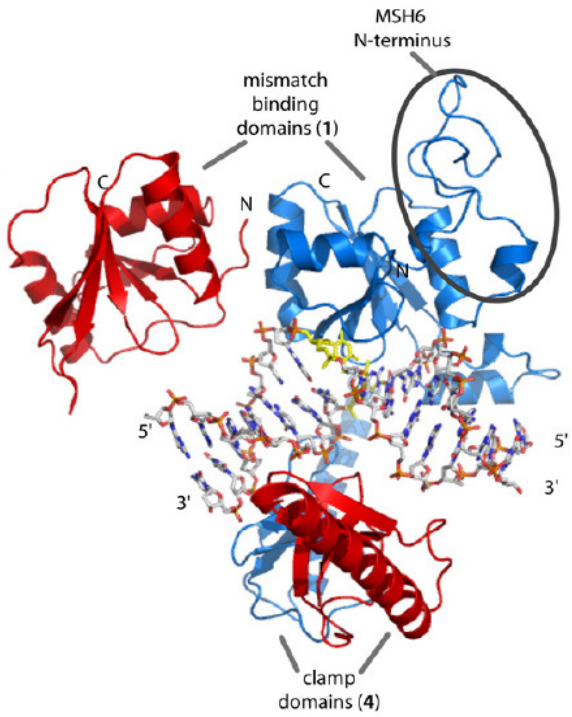


Figure 1.22. Illustration of domains in human MutS which are actively involved in DNA binding during mismatch recognition.[59]

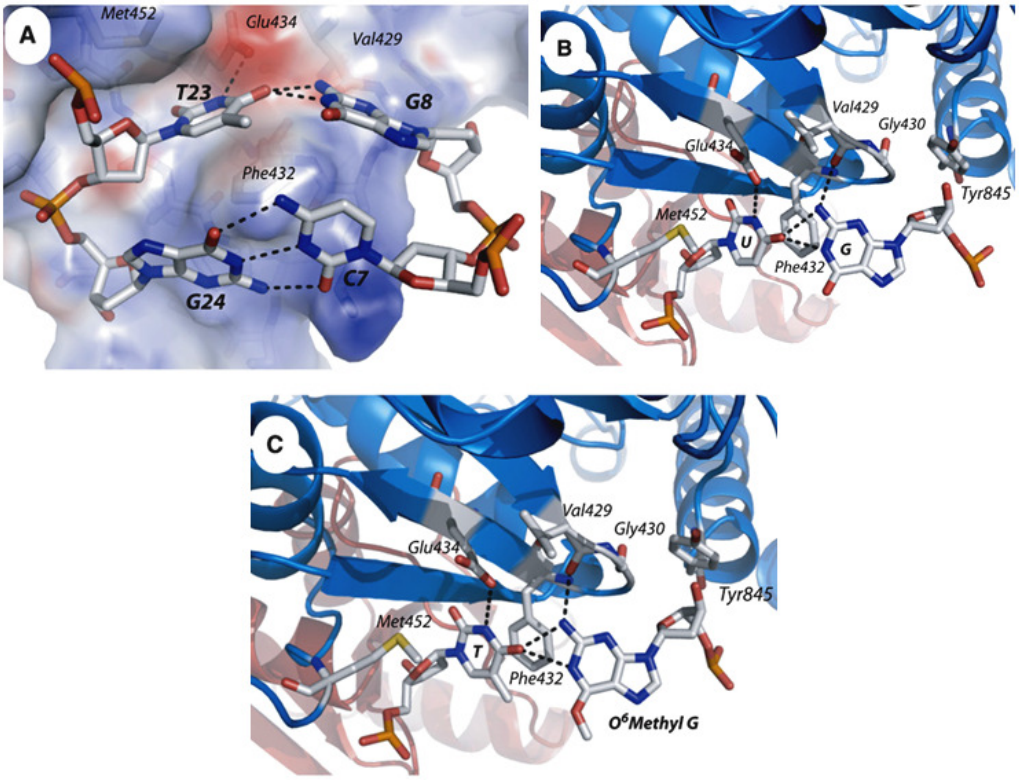


Figure 1.23. Recognition by MSH6 of non-canonical base pairs where (A) GT mismatch, (B) GU mismatch and (C) O^6-methylG.T mismatch. Key amino acids involved during recognition and hydrogen-bonding interactions are shown.[59]

Although it is shown that MSH6 binds to a number of DNA lesions, this does not mean that their complexes will all be similar. Since MSH6 binds to the mispair base via the Glu434 residue, one has to remember that Glu434 is negatively charged in the Phe-X-Glu binding motif. Consequently, some DNA lesions, such as TC mispairs, may prefer a single conformation of the MutS heterodimer.

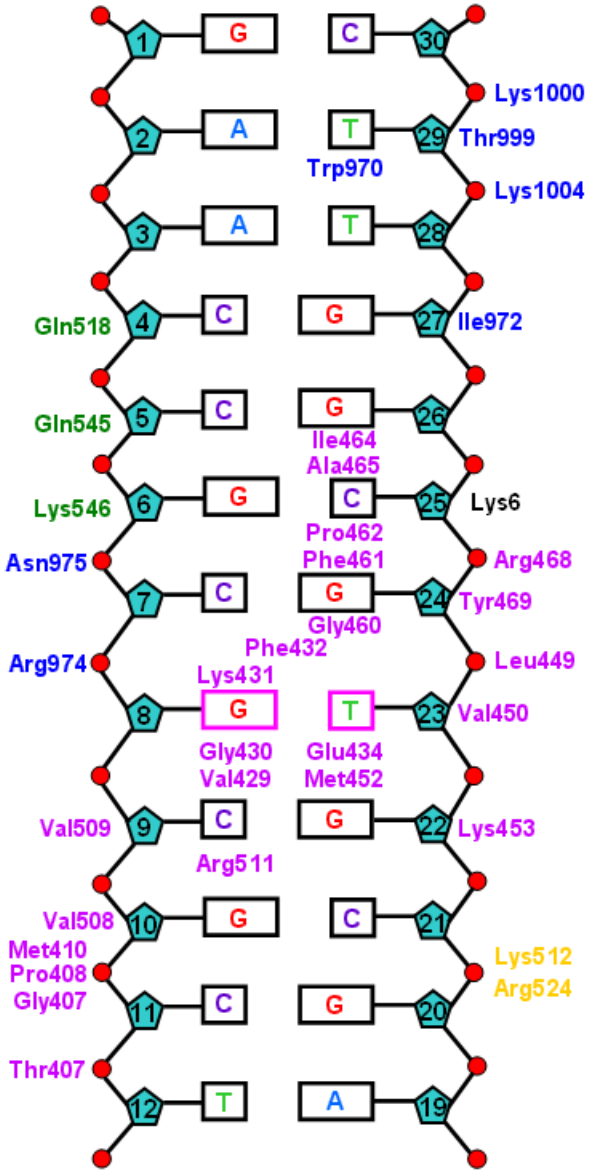


Figure 1.24. Illustration showing types of hydrogen bond interactions present in the crystal structure of human MutS - mismatch GT DNA complex. Interactions are distinguished by domain where pink, blue, green and black correspond to MSH6 domain 1, MSH6 domain 4, MSH2 domain 4 and MSH2 domain 1 respectively.[59]

1.5 Biophysical techniques

Although NMR spectroscopy was the main technique used in this project, other biophysical techniques were used to characterise the DNA and protein molecules being studied. UV absorbance melting was used to characterise the stability of DNA molecules and mass spectrometry was used to characterise isotopically normal and enriched samples of NTD-MutS protein.

1.5.1 UV absorbance melting[2,61]

UV absorbance melting studies were performed to determine the stability of the various DNA duplexes in terms of thermodynamic parameters. This is possible as a change in UV absorbance can be used to monitor the transition between native (duplex form) and denatured (single strand) nucleic acid states. The transition can be induced by increasing temperature and the point at which there is an equilibrium between native and denatured states is termed the melting temperature (T_m). A sudden change in the absorption is a characteristic of the stacked bases in DNA becoming unstacked as the DNA begins to unfold.

If the nucleic acid only has two states, a native duplex and a denatured state, the thermodynamics of the duplex can be calculated using the Gibbs equation (Eqn. 1.5.1). The terms ΔG, ΔH and ΔS correspond to the Gibbs energy, enthalpy and entropy respectively. R is the gas constant and has a value of 8.315, K is the equilibrium constant and T corresponds to the temperature measured in Kelvin.

$$\Delta G = - RT \ln K = \Delta H - T\Delta S \qquad (1.5.1)$$

The derivative of the melting curve measures the change in absorbance with respect to the last measurement taken. Plotting the derivatised data gives the melting temperature for the molecule which is observed as a sharp peak. For molecules which follow a two state model a single T_m value is observed but for more complex systems or molecules, it is possible to observe many phases of melting where each one has a T_m value.

To determine thermodynamic parameters from the UV melting data, the van't Hoff relationship is used (Eqn. 1.5.2). By plotting ln K against the reciprocal of the temperature one would get a straight line from which ΔH and ΔS can be extracted. Subsequent substitution of ΔH and ΔS in to the Gibbs equation (Eqn. 1.5.1) will give a value for ΔG.

$$\ln K = - \frac{\Delta H}{R}\left(\frac{1}{T}\right) + \frac{\Delta S}{R} \qquad (1.5.2)$$

1.5.2 Mass spectrometry[62-64]

Mass spectrometry is an analytical technique which is used to calculate the mass of the molecule being studied. It achieves this by first ionising and fragmenting the molecule by bombardment with an electron beam. This creates a number of charged molecules or fragments of the original molecule, each of these at this point possess a mass to charge ratio (m/z). The positive ions are then accelerated by an electric field before they are measured by the detector and sorted in order of their m/z ratios. The analysis carried out can be both qualitative and quantitative.

A mass spectrometer consists of three main components, an ion source which converts the molecule being analysed into ions, a mass analyser in which the ions are sorted through use of electromagnetic fields and a detector which collects the ions and indicates the relative abundance of each ion type.

The type of ion source is very important in mass spectrometry as different types of ion sources will allow for different types of molecules or compounds to be analysed. For example, electron ionisation and chemical ionisation are common techniques used in the analysis of vapours and gases. In this instance, the molecule studied is ionised via collisions where the molecule is bombarded by ions of a regagent gas in the ion source. In contrast, electrospray ionisation (ESI) and matrix-assisted laser desorption/ionisation (MALDI), are two techniques commonly used for the analysis of liquid and biological samples such as proteins or nucleic acids. In ESI, a liquid containing the molecule studied is dispersed by electrospray into a fine aerosol and in MALDI, a laser is directed at crystals on a target site, the matrix allows itself to

absorb the energy from the laser and becomes charged. This charge is then transferred onto the molecule being studied without it being subject to the energy of the laser source. MALDI is known as a soft ionisation technique as it allows the analysis of molecules which are particularly fragile and thus more likely to fragment easier when using more standard ionisation techniques.

1.6 NMR spectroscopy[65-71]

Nuclear magnetic resonance (NMR) spectroscopy is a powerful analytical technique used in structural and dynamic studies of biologically significant molecules such as nucleic acids and proteins. To achieve this, nuclei with non-zero spins interact with an external magnetic field to produce signals which are dependant on their local chemical/magnetic environments. A description of the basic principles and key experimental parameters are described in the following sections.

1.6.1 Basic principles of NMR

The property of nuclear spin (I) is fundamental to the NMR phenomenon and may have values which are multiples of ½, for example, 1/2, 1, 3/2 and 2. Values of nuclear spin are dependent on the filling of nuclear shells and so various nuclei have different spin values. Nuclei with spin I = 0 are termed NMR inactive or NMR silent as they have no magnetic moment. Consequently, they cannot exhibit nuclear magnetic resonance. A list of common spin active nuclei used in NMR is shown in Table 1.3.

Nuclei	Spin	Natural abundance (%)	γ (rad gauss^{-1} s^{-1})
^{1}H	½	100	26.75
^{2}H	1	0.01	41.06
^{13}C	½	1.10	6.72
^{15}N	½	0.37	-27.13
^{31}P	½	100	108.39
^{19}F	½	100	251.82

Table 1.3. List of common nuclei used in NMR spectroscopy.[65]

The interaction of the nuclear magnetic moment (μ) with the external magnetic field (B_0) causes a splitting in the energy of the spin states. When no magnetic field is present there are a number of spin states but these appear degenerate as all of the possible electron configurations have the same energy. When a magnetic field is introduced, the degeneracy state is split into its individual energy states; the number of energy states is governed by the 2I + 1 rule. In a spin ½ nucleus, there exists two energy states, a low energy state (α) and a high energy state (β) which obey the Boltzmann distribution (Figure 1.24).

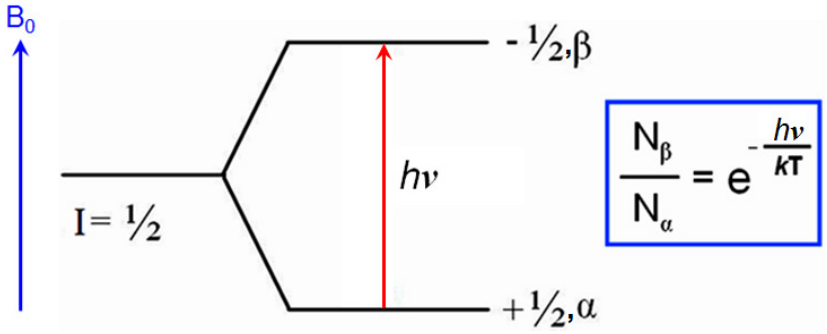


Figure 1.25. Energy level diagram for spin ½ nucleus showing splitting of the degenerate spin state into higher and lower energy states upon introduction of an external magnetic field B_0.

At thermal equilibrium, the Boltzmann distribution governs that there is a slight excess of spins in the lower energy state. The population excess in this state determines the probability of a transition of spins into the upper energy state.

Transitions occur upon irradiation, normally achieved using a radiofrequency (rf) pulse, B_1, which exactly matches the energy gap ($h\nu$). This is known as the resonance condition and the frequency at which the rf pulse is applied is termed the Larmor frequency. The Larmor frequency, at which the states absorb energy, is directly dependent on the gyromagnetic ratio and strength of the external magnetic field (Eqn. 1.6.1).

$$\nu = \frac{\gamma}{2\pi} B_0 \qquad (1.6.1)$$

The absorption of energy can be detected as an emission signal which gives rise to a resonance signal. The energy gap between energy states is different for each type of nucleus therefore each one requires a different resonance frequency.

The gyromagnetic ratio is a proportionality constant which differs for each nucleus and is a measure of the strength of the nuclear magnetic moment. Consequently, the detection sensitivity during an NMR experiment depends on the γ value of the nuclide. The larger the value of γ, the more sensitive a nuclei is and consequently, the easier it is to observe.

1.6.2 Experimental parameters of NMR

When one carries out an NMR experiment, there are a number of important parameters which must be considered and these are chemical shift, line intensity, scalar coupling, relaxation and NOE, these will be discussed in the following sections.

1.6.2.1 Chemical shift

If one considers a molecule with a range of proton environments, the chemical shift (δ) defines the position of each NMR signal corresponding to these protons, with respect to the rf axis. The NMR signals are spread over a range as each proton has a different local electronic environment. Consequently, each is shielded to a different extent when placed in an external magnetic field (B_0). The chemical shift is measured with respect to a reference frequency, and has units of parts per million (ppm). This relationship is shown in Eqn. 1.6.2

$$\delta\,(\text{ppm}) = \frac{\nu - \nu_{ref}}{\nu_0} \times 10^6 \qquad (1.6.2)$$

1.6.2.2 Line intensity

Line intensity of NMR resonances is directly proportional to the molar concentration of the sample. Therefore, a higher concentration of sample is preferred as this will increase the resonance line intensity relative to the baseline noise. However, the strength of the external magnetic field strength also influences the line intensity as it increases the population difference between the spin states due to Boltzmann distribution. Consequently, there will be a greater number of spins in the lower energy state thus increasing the probability of absorption. Resonance signals have different intensities and relative intensities can be calculated by means of integration.

1.6.2.3 Scalar coupling

Also known as *J*-coupling, this arises from magnetisation transfer through chemical bonds between nuclei in a spin system, e.g. A-B-C-D. Coupling of two spins is influenced by bonded molecular electrons on the localised magnetic field produced by the two nuclei. When two nuclei show coupling, signal splitting is observed. The distance between the peaks in a splitting pattern due to *J*-coupling is given by the coupling constant, *J*, which has units of Hertz (Hz).

Vicinal coupling constants depend on the dihedral angle between the coupled protons and this is given by the Karplus equation (Eqn. 1.6.3) where *J* is the coupling constant, θ is the dihedral angle and X, Y and Z are values which depend on the bond type.

$$^{3}J = X\cos^{2}\theta + Y\cos\theta + Z \qquad (1.6.3)$$

1.6.2.4 Relaxation time

The population difference between the α and β states corresponds to bulk magnetisation. At thermal equilibrium, precessing lies along the direction of the external magnetic field, B_0 and the magnetisation in this direction is denoted M_0. Nuclei which align with the external magnetic field are normally said to be in the "Z-axis", therefore the magnetisation of the Z-axis (M_z), is in fact equal to M_0.

Applying a radiofrequency pulse to the system rotates the magnetisation about the X-axis. At this point the magnetisation lies along the Y-axis. The time it takes for the magnetisation to return to equilibrium, i.e. magnetisation along the Z-axis is known as the longitudinal relaxation time or T_1 (Figure 1.26).

T_1 however, is not the only type of relaxation that can occur. Transverse relaxation time or T_2 also exists and describes the process by which magnetisation is lost in the X-Y plane. In T_2, the bulk magnetisation is rotated into the XY plane where some spins will experience a larger magnetic field and these spins start to precess faster than the Larmor frequency. This causes the spins to fan out along the XY plane until no net magnetisation is present. The time taken for the magnetisation to return to equilibrium in this instance is known as T_2 (Figure 1.26).

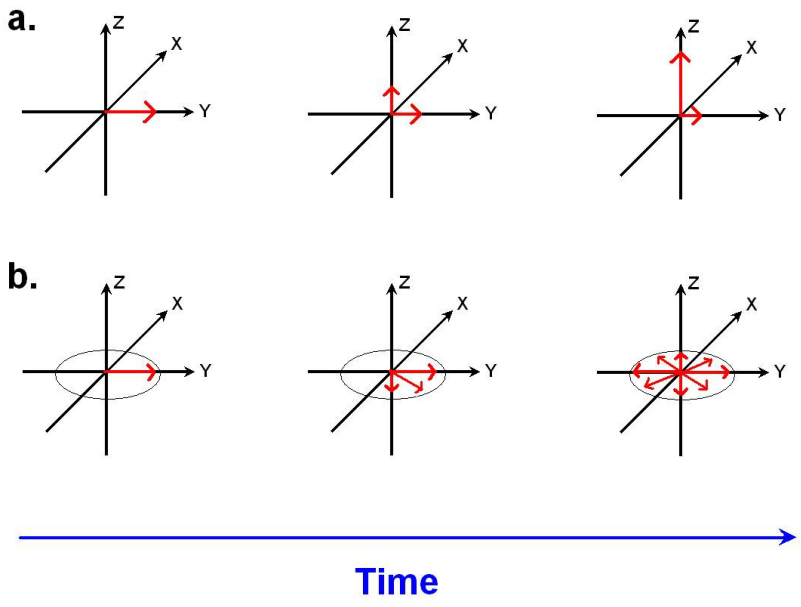


Figure 1.26. Vector diagram showing a. Longitudinal relaxation and b. Tranverse relaxation.[68]

Equilibrium cannot be achieved along the Z-axis unless magnetisation on the Y-axis is completely removed. For this reason, one would expect T_1 to always be greater or equal to T_2.

The value of T2 is also related to the resonance linewidth. The linewidth of resonance signals is important in NMR spectroscopy as greater resolution of peaks will result in a more complete and accurate assignment of resonances. The linewidth is normally measured at half the peak height and in high-resolution NMR spectroscopy this must be as small as possible to produce well resolved peaks.

1.6.2.5 Nuclear Overhauser effect (NOE)[66,69,71]

NOE is a phenomenon which results from the cross relaxation between spins through three dimensional space. NOE coupling is more commonly known as dipolar coupling and unlike scalar couplings which are in the Hz range; dipolar couplings are usually in the kHz range. The irradiation of one proton (H_A) can increase the intensity of a neighbouring proton (H_B).

To explain the NOE effect, a homonuclear two spin system (AX) will be considered. In this instance it will be assumed that the two spins are not scalar coupled with each other, i.e. both are in a singlet state. Figure 1.27 shows the four stages leading to the NOE effect; Equilibrium, Saturation, Excitation and Relaxation. Each spin can exist in an α state (↑) or a β state (↓) and the four energy states (N_1 – N_4) correspond to the four spin arrangements that can exist. The absorption frequencies are equal for both A transitions but they are also equal for both X transitions resulting in two signals.

When an NOE signal wants to be observed one of the signals is saturated so that the resulting effect is observed in the second signal. This takes place in the excitation step. By saturating the A signal, the spin populations are made equal between N_2 and N_4 and also between N_1 and N_3.

Once the excitation step has been carried out, the energy states are as shown in the relaxation stage. The transfer of spin polarisation from one population to another normally occurs via single quantum transitions (W_1). Since $N_1 = N_3$ and $N_2 = N_4$, this means no transitions can take place between these levels but transitions between $N_1 \rightarrow N_2$ and $N_3 \rightarrow N_4$ are possible so absorption will still occur. Hence, by removing the A signal you will increase the intensity in the X signal.

The size of the NOE signal is dependent on the difference between the spin populations. After excitation, the system has to return to equilibrium via W_0 and W_2 relaxation transitions. Positive NOEs arises from double quantum transitions, $W_{2,}$ and cross-relaxation through zero quantum transitions, W_0, lead to negative NOEs. W_0 and W_2 involve the inversion of two spin populations and are known as cross relaxation pathways. Single quantum relaxation transitions, W1, also occur but do not produce NOEs as they only involve the inversion of one spin.

Since dipolar coupling is observed as through space interactions, it is distance dependent and gives information about the three-dimensional structure of the molecule. Normally, dipolar couplings are observed if the distance between the two interacting nuclei is < 5Å. If the distance is greater than 5Å, the NOE signal is too small to be observed.

The distance is not the only factor which affects the NOE signal; the correlation time, τ_c, is also very important. Correlation time is defined at the time it takes a molecule to rotate one radian and since the rate of rotation is affected by molecular size and shape, it would directly influence the number of NOE signals observed.

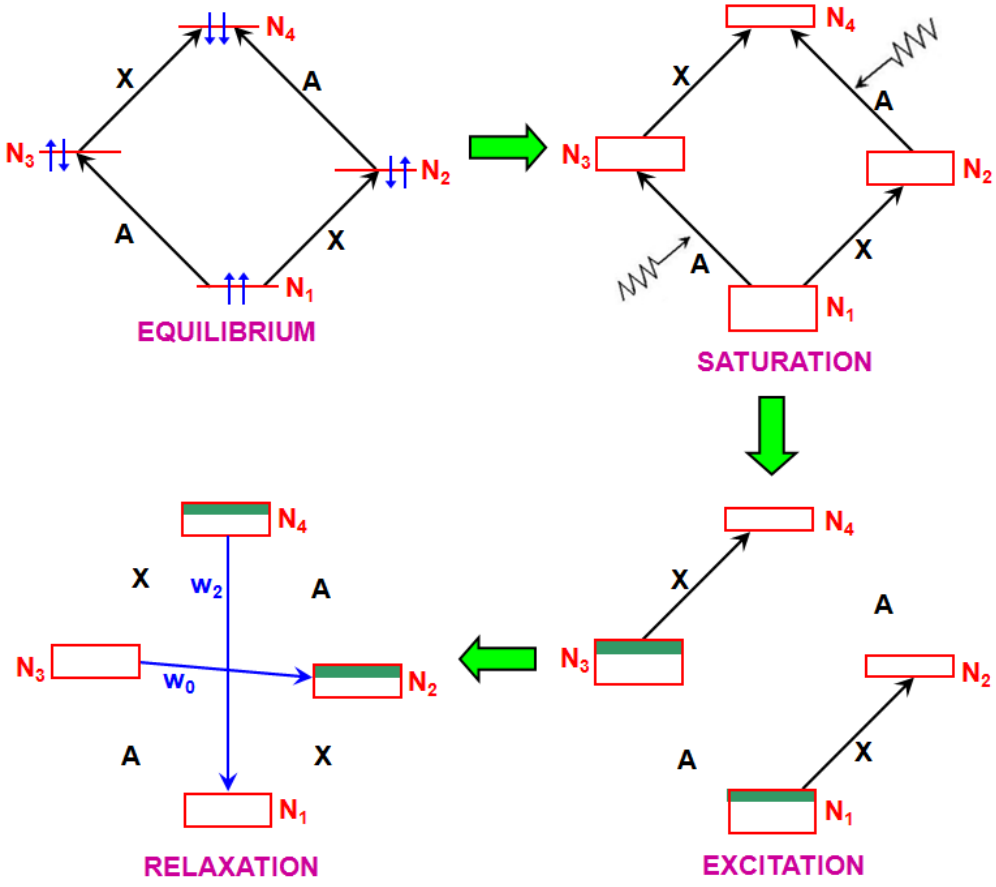


Figure 1.27. The NOE effect as observed in a two spin system, AX, energy level diagram. At equilibrium the population difference between the four energy levels is the same and intensities of A and X are equal. Saturation of the A causes the populations to be equalised between N_1 to N_3 and N_2 to N_4. W_0, W_1 and W_2 are the possible processes that occur to return equilibrium for A but these may also affect population of the α and β states of X. Relaxation along W_2 creates a positive NOE enhancement and that along W_0 creates a negative NOE enhancement.[71]

1.6.3 NMR rate processes[67,69,71]

Chemical exchange processes are important in NMR spectroscopy as they affect the linewidths and peak intensity of resonances. Consequently, this phenomenon can be used to investigate the dynamics in a given system. If at equilibrium a molecule adopts two conformations, A and B, a different set of resonances will be observed for each, ν_A and ν_B.

$$\mathbf{A} \underset{k_B}{\overset{k_A}{\rightleftharpoons}} \mathbf{B}$$

As previously noted, the chemical environment has a direct effect on the chemical shift of a nucleus. However, the peak positions of resonances are also affected by the rate of interconversion between A and B, k_A and k_B. Both k_A and k_B are concentration independent since the conformational changes at equilibrium occur within the same molecule.

In systems where the interconversion rate is slow on the NMR timescale, a separate peak is observed for each conformation (Figure 1.28). The relative intensity of each peak depends on the molar concentrations of conformations A and B, at equilibrium. In slow exchange processes, k_A and k_B are much less than $\Delta\nu$, where $\Delta\nu$ is the difference in chemical shifts measured in Hz.

As the interconversion rate increases it will eventually reach a point where k_A and k_B will be approximately equal to $\Delta\nu$; this is known as intermediate exchange. During

intermediate exchange, there is a broadening of the two peaks for each conformation brought about by a partial averaging of the two chemical shifts (Figure 1.28).

In fast-exchanging systems, k_A and k_B are significantly greater than $\Delta\nu$. The effects on the different chemical shifts are averaged resulting in one peak observed corresponding to both conformations (Figure 1.28). The position of the single peak will depend on the ratio between the concentrations of the two conformations at equilibrium.

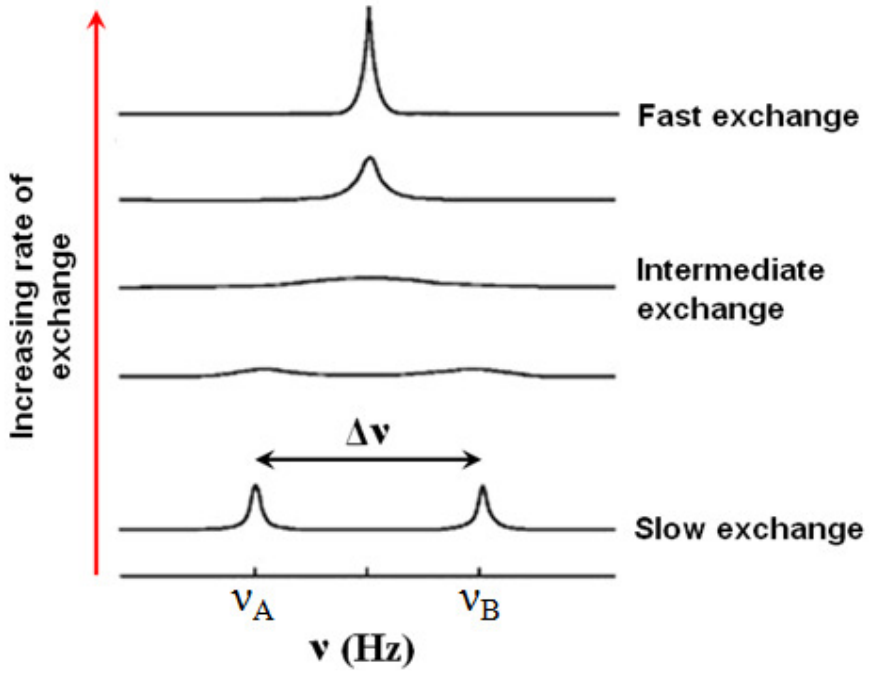


Figure 1.28. Illustration outlining the differences in linewidth and peak intensity between slow, intermediate and fast NMR exchange rate processes for two conformations, A and B, at equilibrium.

1.6.4 Two dimensional NMR spectroscopy (2D NMR)[66,69,71]

The main difference is that in 2D experiments the sample FID signal received by the detector is recorded as a function of two time variables, t_1 and t_2. The advantage in using 2D experiments is that it resolves overlapped signals into a second dimension. Furthermore, additional information is gained about the coupling within the molecule being analysed, which would not be evident in the equivalent 1D spectrum.

All 2D NMR experiments have the same basic format which includes four stages; preparation, evolution, mixing and detection (Figure 1.29). The preparation step prepares the spins by hitting them with a single high powered rf pulse of fixed duration. This is followed by the evolution step, where the magnetisation is allowed to evolve.

Evolution is the key step in generating the second dimension. During the evolution period, the sample signal is recorded at successive, incremental time delays. The data is recorded at $t_1 = 0$ and the pulse sequence is allowed to proceed; this generates an FID which is recorded. The system is then allowed to return to equilibrium before the process is repeated with $t_2 = t_1 + \Delta t$, where Δt is known as the incremental time.

The pulse sequence is carried out again and the signal recorded is stored separately from the first FID. This process is repeated many times over until enough data is collected, with Δt increasing incrementally with every additional run, i.e. $2\Delta t$, $3\Delta t$, $4\Delta t$,…….$n\Delta t$. It is very common that the number of Δt increments to be 300-500. As t_1 increases from zero, the signal obtained also increases because the

magnetisation has been allowed to develop, reaching a maximum if enough time is allowed to evolve it through an angle of 90º.

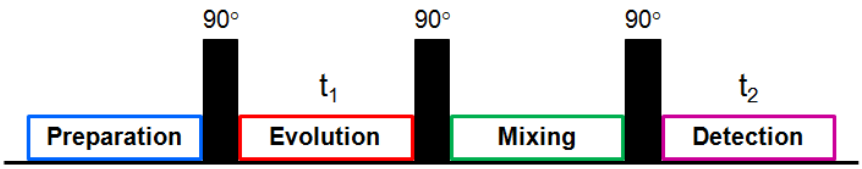


Figure 1.29. General pulse sequence for 2D NMR experiment.

Following evolution is the mixing step, which consists of either a single pulse or a series of pulses of fixed duration. The mixing step is important as the time duration chosen depends a lot of the nature and properties of the molecule. This is more significant in dipolar coupling experiments like NOESY experiments than for scalar coupling methods like TOCSY. Both t_1 and t_2 are Fourier transformed to produce a spectrum which has two frequency dimensions (F_1 and F_2).

1.6.5 Three dimensional NMR spectroscopy (3D NMR)[70]

As molecular size increases one is faced with three problems in NMR spectroscopy; the number of protons, length of correlation time and resonance linewidth all increase proportionally with size. This leads to overlapping of resonances and lower sensitivity in the case of experiments which rely on scalar coupling.

If resonances are unresolved in 2D spectra, it is possible to expand the data acquisition into a third dimension (Figure 1.30). This is important as it increases the dispersion of peaks thus simplifying the spectrum, as each plane in the 3D data only contains a few peaks.

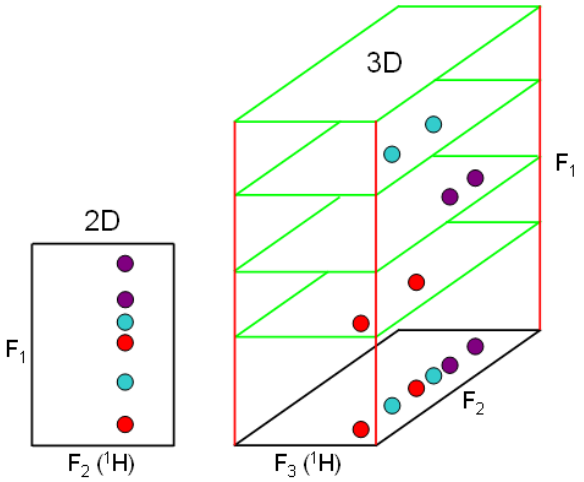


Figure 1.30. Scheme showing the resolution of resonances by expanding into a third dimension in NMR spectroscopy.[70]

Most 3D NMR techniques transfer the magnetisation through one bond (^1J) or two bond (^2J) couplings. This is more efficient than ^3J coupling as the transfer times between spins are shorter during relaxation and the sensitivity of the experiment is also increased. A basic 3D pulse sequence is similar to a 2D pulse sequence only with an additional evolution and mixing period before the detection step (Figure 1.31).

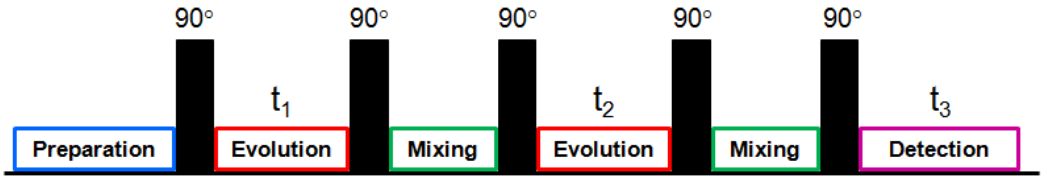


Figure 1.31. General pulse sequence scheme for a 3D NMR experiment.

3D NMR techniques are divided into two main types of experiment, heteronuclear edited and triple resonance (Figure 1.32). Heteronuclear edited experiments use coupling to heteronuclei to help resolve overlapping resonances. The cross peaks between ^{1}H spins are resolved based on the chemical shifts of their directly attached

heteroatom. Examples of heteronuclear edited techniques are HSQC-TOCSY and HCCH-TOCSY.

Triple resonance experiments are very different to heteronuclear edited methods as the 2D part of the experiment does not have to be a ^{1}H-^{1}H correlation. They are referred to as triple resonance because three different nuclei are correlated e.g. ^{1}H, ^{15}N and ^{13}C. Due to the simplicity of the data interpretation, triple resonance experiments are preferred as each frequency dimension in the spectrum only has one resonance signal which significantly reduces the issue of resonance overlap. The only drawback with triple resonance experiments is that a doubly labelled (^{15}N, ^{13}C) sample is required.

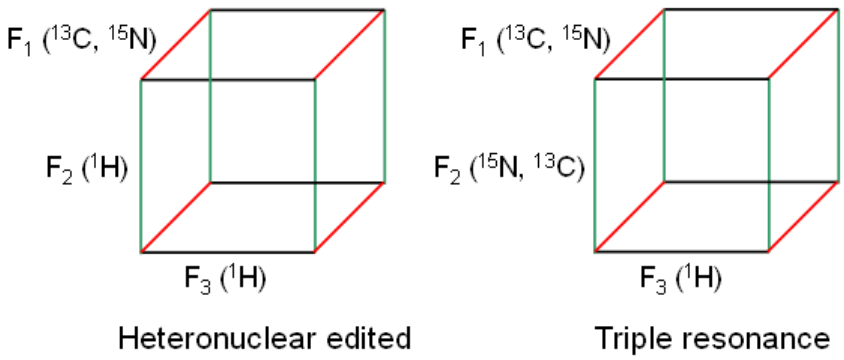


Figure 1.32. Outline scheme illustrating the difference between heteronuclear edited and triple resonance 3D NMR experiments.

1.7 NMR structure determination

NMR is already widely established as one of the leading techniques in the structure determination of macromolecules and study of interaction and dynamics in biologically significant systems. In the Research Collaboratory for Structural Bioinformatics (RCSB) PDB database[72], there are over 7200 protein and 900 nucleic acid NMR structures deposited to date and in this chapter, examples of DNA and protein NMR structures as well as DNA-protein complexes are highlighted and discussed.

1.7.1 NMR structure determination of DNA

Many examples of NMR structure determination have been carried out for DNA[73,74] and the process is becoming more important in structural biology as DNA is an important substrate in many pathways. Thus, the understanding of the dynamics from a structural viewpoint is of high importance in their study.

Until recently molecular dynamic simulations on DNA have been difficult. The main reason is in the nature of the DNA structure itself. DNA is highly charged and is non-globular in structure both of which are unfavourable in molecular dynamic simulations. As molecular dynamic calculations continue to develop, more complex aspects of nucleic acid structure can be analysed such as flexibility of the duplex.

Liu *et al.*[75] described the NMR structure determination of a 17mer DNA duplex using molecular dynamics with the inclusion of NMR experimental constraints (Figure 1.33). The structure determination yielded structures of A-form and B-form

DNA. The average structure was generated by combination of the two structures sets where the RMSD of the ensemble was calculated to be 0.73Å.

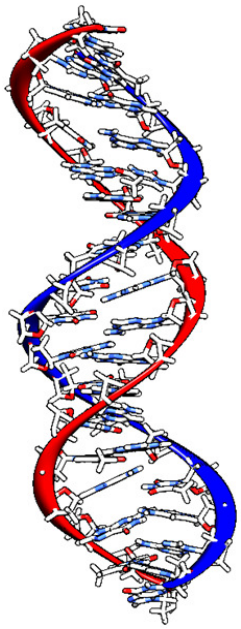


Figure 1.33. Average NMR structure reported by Liu *et al.* of a 17mer DNA duplex (PDB: 1IR5).[75]

One example of structure determination of DNA containing a modified base is presented by Bohon and de los Santos[44] who have carried out a molecular mechanics simulation on 6-TG modified GC and GT DNAs (Figure 1.34). The study investigated the degree of structural change caused by 6-TG and it was found that both duplexes adopted a B-form helix structure. The GT mispair appears as a wobble base pair which is expected due to the change in base pairing. An important observation from the study was that 6-TG modified GT DNA is more stable than 6-TG modified GC DNA, which was shown through NMR variable temperature experiments. The NMR data showed substitution of cytosine to thymine increased the hydrogen bond distances surrounding the sulphur atom increased by an average

of 10% and this caused the 6-TG.T base pair to adopt a slightly non-planar conformation with a buckle of -11º.

The 6-TG modified GC DNA average structure was produced from 29 refined structures with an average RMSD of 0.64Å compared to the 6-TG modified GT DNA average structure which was produced from 25 refined structures with an average RMSD of 0.65Å. The similarity between the two structures produced indicates that the mispairing of 6-TG with thymine is unlikely to cause cell death

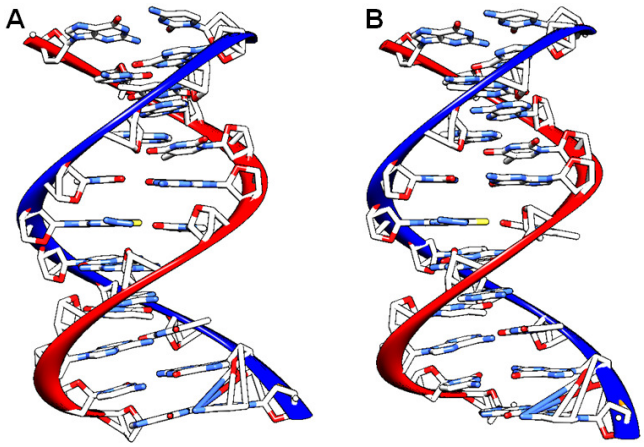


Figure 1.34. (Left) 6-TG modified GC DNA (PDB: 1KB1), (Right) 6-TG modified GT DNA (PDB: 1KBM). Comparison of NMR structures generated from rMD calculations by Bohon and de los Santos.[44]

1.7.2 NMR structure determination of protein

The structure determination of proteins by NMR spectroscopy is much more established. Recently, the NMR assignment of a 723 residue protein by Kay *et al.*[76] demonstrates the difference in progress between protein and nucleic acids. A wide suite of available NMR techniques and the use of auto-assignment programs such as ARIA[77] and TALOS[78], which generate likely dihedral angles for proteins, greatly

assist in obtaining experimental constraints used in structure determination calculations.

One important inclusion in structure determination of proteins is the addition of residual dipolar couplings (RDCs).[79,80] RDCs compliment the use of NOE derived local distance constraints. Whilst NOEs provide short distance constraints, RDCs provide long-range orientational information. To measure RDCs the molecule being studied has to be aligned in a suitable medium. After alignment the orientation of the molecule is fixed so further studies on the protein will result in determination of this fixed conformation.

Roberts *et al.*[81] report the NMR structure determination of MJ1529 which is a methylguanine-methyltransferase protein using a combination of 1D to 4D NMR experiments (Figure 1.35). Incorporation of experimental constraints was done using DYANA and TALOS; structure calculations were carried out using the program ARIA and CYANA.

Initial structure determination by ARIA resulted in a converged C-terminus (RMSD = 0.9 Å) but almost no structure was found for the N-terminus end. The absence of a structure for the N-terminus caused distortion of the C-terminus due to improper contacts between the two. The same structure determination using CYANA gave a better structure for the N-terminus, which was found to have a three stranded β-sheet structure (RMSD = 1.5 Å). RDCs were added after the initial structure determination.

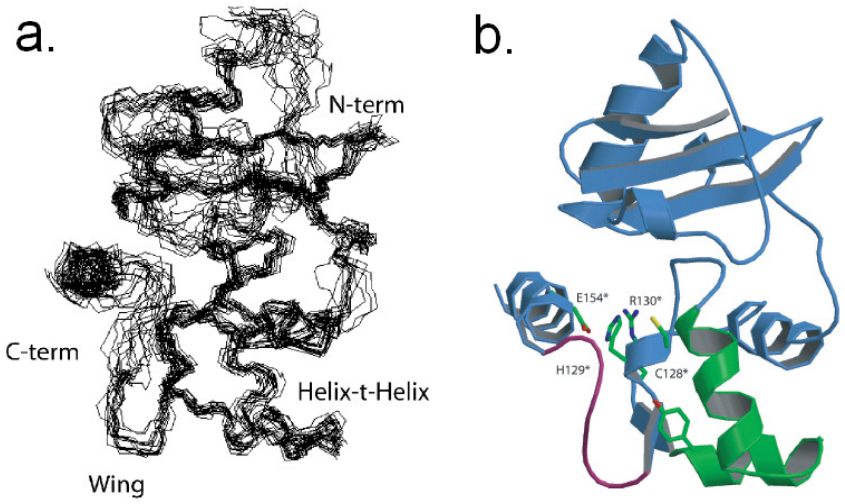


Figure 1.35. NMR solution structure of MJ1529 (PDB:) reported by Roberts *et al.* a. overlay of lowest energy structures of MJ1529, b. ribbon diagram of final structure of MJ1529.[81]

Kay *et al.*[82] report the structure budding yeast chaperone Chz1 bound to histones H2A.Z-H2B also termed the CZBcomplex (Figure 1.36). The structure shows that Chz1 adopts a heterotrimer structure when bound to H2A.Z-H2B and the histone fold was found to be similar to the one observed in the nucleosome. Chz1 stabilises the histone dimer through contact with both positively and negatively charged residues.

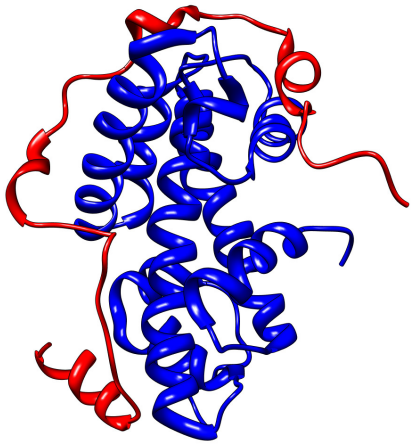


Figure 1.36. NMR structure of CZB complex elucidated by Kay *et al.* (PDB: 2JSS). The Chz1 domain is shown in red which adopts a two α-helical structure connected by a long irregular chain. The histone dimer shown in blue is more structured.[82]

1.7.3 Structure determination of protein-DNA complex

Structure determination of complexes between protein and DNA are important since any binding interaction between two molecules often results in conformational change and a change in the dynamics.

One drawback with the study of protein-DNA complexes is the increased complexity of the spectra which will contain resonances arising from both unbound and bound states. Isotope enrichment has become increasingly important in the study of protein-DNA complexes as it allows conformational studies of the domain in their complex state using isotope edited NMR techniques. With modern techniques, it is possible to selectively label particular domains in a complex.

The development of suitable forcefields has also provided a major challenge when dealing with molecular dynamic simulations of protein-DNA complexes in the past. Currently, a number of forcefields have been created including AMBER,[83] CHARMM[84] and GROMOS.[85]

Halouska *et al.*[86] reported the solution structure of the N-terminal domain of *Pseudomonas putida* protein PpPutA45 and its complex with substrate DNA (Figure 1.37). PpPutA45 is part of a bigger protein which is involved in the oxidation of proline residues to glutamate. PpPutA45 more specifically has a role in DNA binding and this is brought about during its dimerisation. Dimerisation creates an anti-parallel β-sheet which acts as a binding site by integrating with the major groove of the DNA.

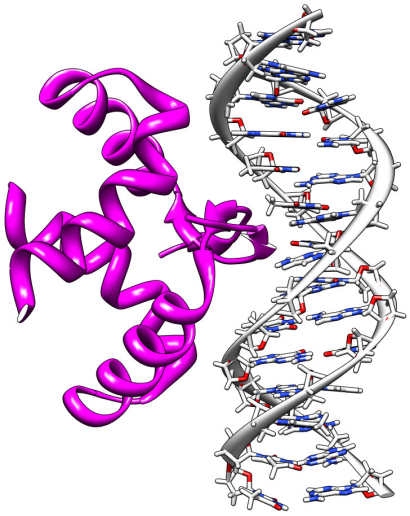


Figure 1.37. NMR structure of PpPutA45 protein domain complexed to substrate DNA (PDB: 2JXI). The protein binds to the major groove of the DNA, where it interacts to a conserved GTTGCA site.[86]

More recently, Cierpicki *et al.*[87] elucidated the structure of a conserved protein domain (CXXC) in the *MLL* gene bound to DNA. The gene is a common target of chromosomal translocations that cause leukaemia and the CXXC domain is conserved in all MLL fusion proteins. The importance of the CXXC domain is that it selectively binds to non-methylated CpG DNA (Figure 1.38).

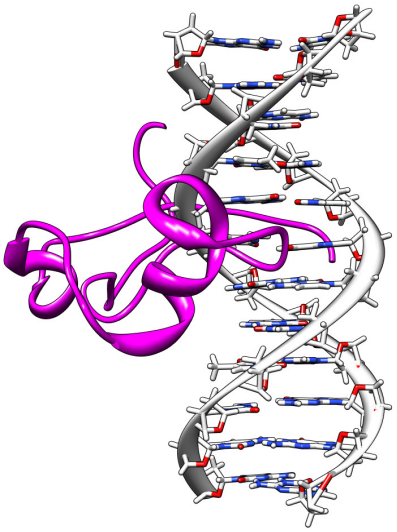


Figure 1.38. NMR structure of CXXC-DNA complex reported by Cierpicki *et al.* (PDB: 2KKF).[87]

1.8 Previous work

The assignment of 9mer GC, GT, CT, O^4-methyl TG, O^4-methyl TA and O^6-methyl GC DNAs d(5'-CCTGXGGAG-3'; 3'- GGACYCCTC-5') has been successfully carried out by NMR.[88] The various DNA mismatches exhibited different levels of thermodynamic stability and exchange dynamics with the solvent. Non-exchangeable proton assignment showed that for some mismatched base pairs, the introduction of the mismatch induced a change in the sugar pucker, causing the sugar ring to fluctuate between $C_{2'}$-*endo* and $C_{3'}$-*endo* conformations. This conformational change was hypothesised to provide an access route for MutS protein during mismatch recognition, which is consistent with the X-ray crystal structure of mismatched GT DNA - *E. coli* MutS complex. However, no NMR structure determination has been carried out for any of the DNAs investigated to date in the group. Consequently, it has not been possible to correlate the changes observed in the NMR spectra to the structure.[88] Recognition of mismatched and modified DNAs by *E. coli* MutS was observed by NMR titration. Specific perturbations to the imino proton resonances gave evidence of selective binding to the mismatch and modified base pairs by MutS.[88]

The expression, isolation and purification of His_6-tagged *E. coli* MutS protein (~ 98 kDa, monomer) was optimised allowing high yield of pure normal and isotopically labelled protein (10 - 15mg per litre of cell culture). Studies were also carried out on limited proteolysis by chymotrypsin which showed that in solution *E. coli* MutS protein was organised into well defined, stable domains. In the presence of mismatch DNA, however, the proteolysis was impeded suggesting conformational changes upon DNA binding.[88]

1.9 Aim of the Project

The overall objective of the project is to understand the role of *E. coli* MutS in the DNA mismatch repair pathway in structural terms. The effect and structural differences that mismatch and modified base pairs cause when substituted into canonical DNA were studied and analysed. The main sequences studied include a 17mer mismatch GT DNA, which mirrored the sequence published in the X-ray crystal structure of the *E. coli* MutS-DNA complex, and its derivative 17mer canonical GC DNA which will act as a reference structure. All conformational changes induced by the substitution of the GT mispair will be determined by comparison to the 17mer canonical GC DNA. The core technique used will be NMR spectroscopy followed by molecular modelling, although other biophysical methods, such as UV absorbance melting, will be used to characterise the properties of the DNA.

Further to the structure determination of the 17mer mismatch GT DNA, the effect of mismatch orientation, sequence context and sequence length on the local helical conformation and overall mismatch stability will be monitored by UV absorbance melting and NMR spectroscopy.

NMR studies of 6-thioguanine modified 13mer GC DNA and its oxidised derivatives will be carried out, with the NMR structure being determined of the former. The sequence characterised by NMR will be based on that of the 17mer canonical GC DNA. This provides an opportunity to study the conformational change caused by a different DNA lesion.

The results provided from the above studies will give further insight into how the stability of the B-form DNA structure is affected by different DNA lesions and how this is mainfested in conformational change.

The affinity of WT-*E. coli* MutS to the DNA sequences described above will be monitored using NMR spectroscopy by titration with sub-stoichiometric amounts of MutS protein. It is known that MutS has a high affinity for a GT mispair but the effect of mismatch orientation and sequence context on the level of binding is not fully known and this study will give new information on this. The 17mer mismatch GT DNA sample after titration will be investigated by NMR to help identify changes to the DNA conformation induced by protein binding.

Although the crystal structure for the *E. coli* MutS-mismatch GT DNA complex has been determined, at present it is not feasible to carry out an analogous NMR study of this large complex (~220 kDa). However, it is necessary to point out that the *apo* structure of the key DNA binding N-terminal domain (NTD) of MutS has not been determined in previous X-ray crystallography studies. An important aim of this project is to express, isolate and purify NTD-MutS protein in rich and minimal media, for NMR study using isotope edited 3D techniques.

Chapter 2 – Materials and methods

In this chapter the materials used and methods applied in the project are given in necessary detail including sample preparation, NMR experiments and computational modelling techniques.

2.1 Oligonucleotide synthesis

All DNA oligonucleotides used were chemically synthesised, HPLC purified and commercially obtained from MWG-Biotech, Germany with the exception of thio modified DNAs. 6-Thioguanine and 6-sulphonateguanine modified DNAs were kindly supplied by Dr. Y-Z. Xu, Open University. Oligonucleotides were obtained in a lyophilised salt form containing sodium phosphate, which have been purified by HPLC and consequently, no further purification was required prior to their use in NMR.

2.2 Expression, isolation and purification of isotopically normal and enriched full length *E. coli* MutS and NTD-MutS protein

The production of hexahistidine tagged full length *E. coli* MutS protein was carried out by Dr. F. Vendeix and was initially obtained from the *E. coli* K-12 bacterial strain TX3149 containing the plasmid pTX412, supplied by Dr. M. Winkler and Dr. J. Andrews. The resulting MutS protein was purified and isolated using a Ni^{2+} affinity column (Novagen, U.K.).

Using the same plasmid (pTX412) above as a template, 117 codons comprising the 5' end of MutS were amplified by the polymerase chain reaction (PCR). The PCR product was recloned following digestion with NdeI-BamHI and was used in the overproduction of the 117 residue NTD-MutS in *E. coli* BL21 (DE3) which was purified by Ni^{2+} affinity column.

For ^{15}N isotopically labelled NTD-MutS, the *E. coli* BL21 (DE3) cells were grown in minimal media containing $^{15}NH_4Cl$. For ^{13}C, ^{15}N isotopically labelled NTD-MutS, *E. coli* BL21 (DE3) cells were grown in minimal media containing ^{15}N ammonium chloride ($^{15}NH_4Cl$) and ^{13}C glucose. Production of the NTD-MutS and its isotopically labeled derivatives were carried out by Dr. M. Zampini.

2.3 NMR sample preparation

Both DNA and proteins have to be prepared differently for study by NMR. Factors such as buffer salt concentration, sample volume and concentration have to be considered.

2.3.1 DNA samples

In order to prepare DNA duplexes from the synthesised oligonucleotides, they were mixed in equimolar amounts, approximately 100 absorbance units, before being annealed by heating to 90˚C for three minutes followed by slow cooling to room temperature. The samples was then stored at 4ºC.

NMR samples were prepared in 90% $^{1}H_2O$ and 10% $^{2}H_2O$, or 100% $^{2}H_2O$. Samples in $^{2}H_2O$ were lyophilised in order to remove any traces of water from the sample, before addition of high purity (≤ 99.9%) $^{2}H_2O$. Lyophilisation of samples was performed at -60˚C and at a pressure of 6-8 torr. The sequences studied using NMR spectroscopy in this project in addition to those listed in Figure 2.2 and 2.3 are given in Figure 2.1.

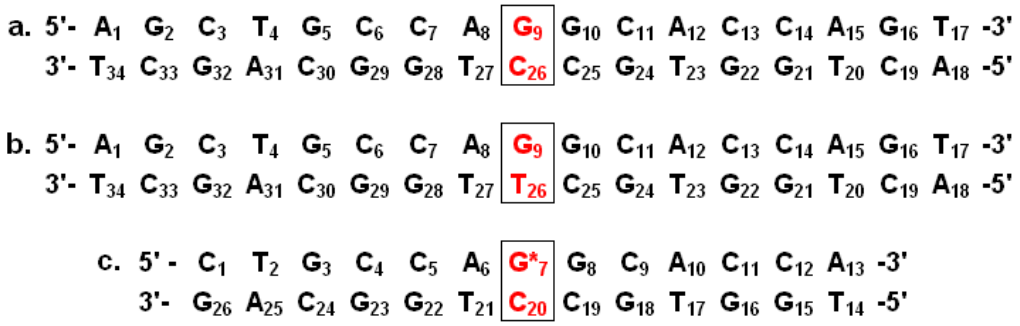


Figure 2.1. DNA sequences studied using NMR spectroscopy, a. 17mer canonical GC, b. 17mer mismatch GT and c. 6-TG modified 13mer GC DNAs. G* indicates the 6-TG base.

2.3.2 Protein samples

Samples of WT-MutS protein were produced in the group by Dr. Franck Vendeix in buffer solution containing 50mM PO_4^{3-} and 100mM NaCl. Samples were stored at -20°C prior to use in NMR titration.

Three samples of NTD-MutS were kindly supplied by Dr. Massimiliano Zampini; an unlabelled sample, a ^{15}N isotopically labelled sample and a ^{15}N, ^{13}C isotopically labelled sample. Further addition of buffer salts was not necessary prior to NMR sample preparation since the protein samples were prepared and isolated in buffer solution already containing 50mM PO_4^{3-} and 100mM NaCl. Although buffer salts

are essential in the stability of the NTD-MutS protein domain, their concentrations were kept to a minimum as high salt concentrations leads to a decrease in the performance of cryogenically cooled probes fitted in NMR spectrometers.

Similar to DNA, NMR samples of NTD-MutS protein were prepared in 90% $^{1}H_2O$ and 10% $^{2}H_2O$ or 100% $^{2}H_2O$ and lyophilisation of samples were carried out using the same methodology in sectin 2.3.1 and stored at 4°C.

2.3.3 DNA-protein complex

DNA-protein complexes were prepared by NMR titration which is discussed later in section 2.9. During titration, the sample volume would increase due to the addition of the titrant. If the volume was beyond 900 μL during titration, the sample would be lyophilised and be reconstituted to 600 μL. Samples were prepared in $^{1}H_2O$ or $^{2}H_2O$ solvent.

2.4 Mass spectrometry

It has been noted in this chapter that a ^{15}N isotopically labeled sample was used to study the NTD-MutS protein domain. In order to determine the level of ^{15}N enrichment in the NTD-MutS sample, mass spectra were measured using electrospray ionisation mass spectrometry (ESI-MS).[89]

The work was conducted on a Waters brand LCT (time of flight) mass spectrometer, using an ESI source. The data was deconvoluted using Waters' Masslynx MaxEnt 1

software. The sample in a ACN/Water mix (0.1% formic acid) was directly infused at a capillary voltage of 2800V. 0.13mg of unlabelled NTD-MutS and 0.40 mg of ^{15}N enriched NTD-MutS were submitted for analysis by ESI-MS and the resulting mass spectra indicated an enrichment level of greater than 95% in the ^{15}N enriched sample of NTD-MutS.

2.5 UV absorbance melting experiments (T_m)

In order to characterise the stability of mismatch DNA, changes to the UV absorbance upon thermal denaturation[61] were measured for mismatch GT DNA and a number of mismatch orientation, sequence context and length analogues. The thermal denaturation of normal GC DNA was also measured as a control. All samples were prepared in 50mM PO_4^{3-} and 50mM NaCl buffers at a pH of 6.2.

2.5.1 Melting curves and first derivative plots

The absorbance of each DNA duplex was monitored at a wavelength of 260 nm (Varian Cary UV Spectrophotometer) over a temperature range of 20 – 90 ºC. The temperature during each sample run was increased at a rate of 1.0 ºC min^{-1}, with the absorbance being recorded at every 1.0 ºC interval. Melting temperatures (T_m) were determined from the midpoint of the slope of the thermal denaturation curves. These were obtained through derivatisation of the normalised absorbance data using KaleidaGraph[90]. Details of the duplex concentrations are given in Chapter 3.3, Table 3.23 and Chapter 3.4, Table 3.25, where the results of the UV absorbance melting studies are presented.

The errors in UV absorbance melting temperature measurements can be best assessed by monitoring the distribution of data points, i.e. location, dispersion and shape of the distribution, in the dA/dT vs. T plots. The location identifies the 'centre' of data distribution, i.e. the value about which most of the data is grouped (mean, median and mode). The reliability of the sampled data mean can be judged by the standard error, SE, given below, where s is the standard deviation and n is the number of data points sampled.[91]

$$\mathbf{SE = s / \sqrt{n}} \qquad (2.5.1)$$

Standard error, under strict conditions, gives an esitimate of the dispersion of the repeated sample means around the true value. Hence, the smaller the SE, the more reliable the sample mean is likely to be of an estimated true value.[91]

The effects of two factors on GT mismatch stability were investigated. Firstly, the change in UV absorbance with decreasing oligonucleotide length was also studied using the sequences shown in Figure 2.2 (a-e). Secondly, UV absorbance changes in mismatch orientation and sequence context analogues were also measured. The sequences studied in this part of the investigation are shown in Figure 2.3 (a-e).

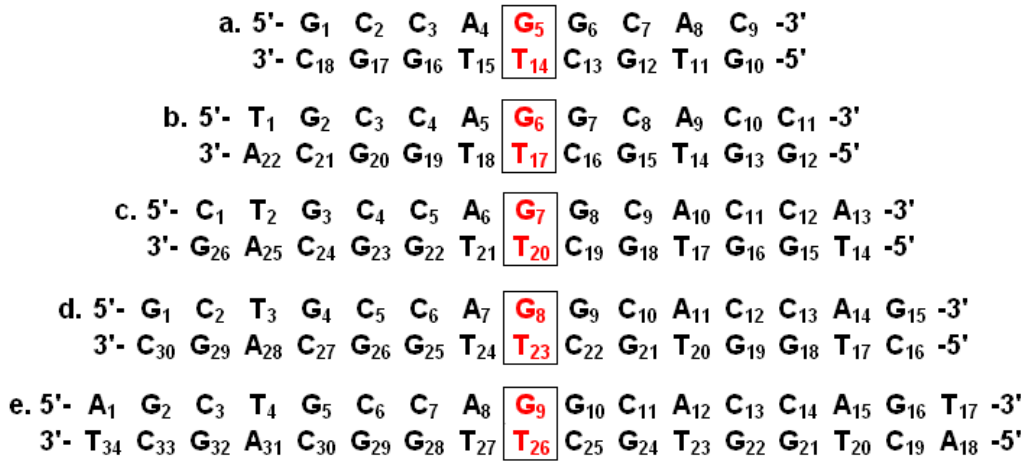


Figure 2.2. DNA sequences used in UV absorbance melting studies investigating effect of oligonucleotide length on mismatch stability where a. 9mer GT, b. 11mer GT, c. 13mer GT, d. 15mer GT and e. 17mer GT DNAs.

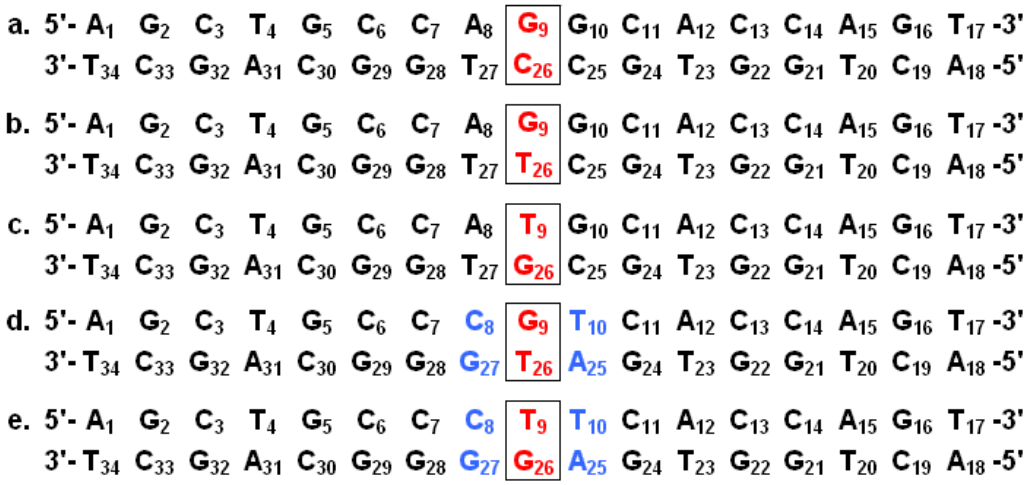


Figure 2.3. DNA sequences studied by UV absorbance melting to investigate the effect of mismatch orientation and sequence context on the stability of mismatch DNA where a. 17mer GC, b. 17mer GT, c. 17mer TG, d. 17mer GT-FB and e. 17mer TG-FB DNAs. FB refers to a change in the flanking base pairs of the GT/TG mispair and this is indicated in blue.

2.5.2 Calculating thermodynamic values using the van't Hoff relationship

Application of the van't Hoff relationship[92-94] (Eqn. 2.5.2) to the acquired UV absorbance melting data was used to determine the thermodynamic parameters for each data set.

$$\ln K = -\frac{\Delta H}{R}\left(\frac{1}{T}\right) + \frac{\Delta S}{R} \qquad (2.5.2)$$

In order to determine ln *K*, a two-state model has to be assumed for a double stranded oligonucleotide. When a double stranded oligonucleotide exists in equilibrium, the fractions of the strands (f_D) can be calculated (Eqn. 2.5.3).

$$f_D = \frac{A - A_S}{A_D - A_S} \qquad (2.5.3)$$

A is the absorbance at a specific temperature, A_S is the absorbance of the single strands of the duplex and A_D is the absorbance of the duplex. Following successful determination of the f_D value, *K* can be calculated (Eqn. 2.5.4).

$$K = \frac{[S]^2}{[D]} = \frac{2(1 - f_D)^2}{f_D} c_T \qquad (2.5.4)$$

By plotting ln *K* against the reciprocal of the temperature, the value of ΔH can be calculated from the gradient and ΔS from the y-intercept. Subsequent substitution into the Gibbs energy relationship gives a value for ΔG (Eqn. 2.5.5).

$$\Delta G = \Delta H - T\Delta S \qquad (2.5.5)$$

2.6 NMR Instrumentation

NMR data was primarily obtained from three facilities. The first was the National Institute for Medical Research (NIMR), Mill Hill, London, housing Varian Inova 600 and 800 MHz spectrometers, and Bruker Avance 600 and 700 MHz spectrometers. The Bruker 600, 700 and Varian 800 MHz spectrometers were equipped with a 5mm $^{1}H/^{13}C/^{15}N$ triple resonance cryogenically cooled PFG probes. The Varian 600 MHz was equipped with a 5mm $^{1}H/^{13}C/^{15}N$ triple resonance standard PFG probe which could be changed to carry out experiments detecting ^{31}P and ^{19}F.

Access to medium-field strength spectrometers was provided at the School of Chemistry, University of Manchester. Instrumentation used included Bruker Avance 200, 400 and 500 MHz spectrometers all equipped with broadband probes capable of detecting ^{1}H, ^{13}C, ^{15}N and ^{31}P.

More recently, access was granted on the world's first 1000 MHz spectrometer based at the Centre de RMN à Très Hauts Champs in Lyon, France (Figure 2.4). The spectrometer was equipped with a 5mm $^{1}H/^{13}C/^{15}N$ triple resonance cryogenically cooled PFG probe.

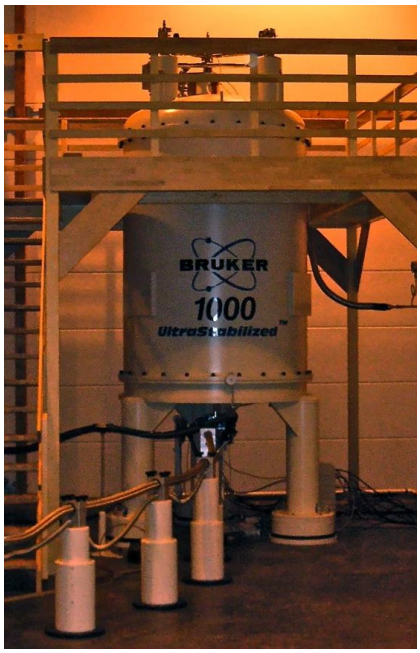


Figure 2.4. The world's first 1000 MHz NMR spectrometer installed at the Centre de RMN à Très Hauts Champs in Lyon, France.

2.7 NMR experiments

Further to one-dimensional data acquisition, two-dimensional homo- and heteronuclear NMR experiments were used to aid assignment of the DNAs and protein domain being studied. Here, a brief description of the main NMR experiments used is given.

2.7.1 Solvent Suppression

Important molecules studied such as proteins and nucleic acids are of biological or biomedical interest and in order to observe all the possible chemical environments in these molecules, they must be studied in water. This ensures that labile protons such as amino and imino protons can be observed. Typically, experiments are carried out

in 90% $^{1}H_2O$ and 10% $^{2}H_2O$. The addition of $^{2}H_2O$ is to maintain a field frequency lock for the sample.

Solvent suppression is required as the concentration of protons in the sample is very small, when compared to that in the solvent. Conventionally, the proton concentration in $^{1}H_2O$ is 110M, whereas the sample only contains millimolar (mM) concentrations. The difference observed here clearly becomes problematic during experimental measurements. Consequently, the goal is to reduce the magnitude of the solvent signal before the NMR signal reaches the receiver. A reduction in the solvent signal means that more sample resonances can be observed. In the experimental work, the main water suppression methods used were pre-saturation and WATERGATE.

Pre-saturation[95] suppresses the water signal by application of a weak, continuous rf pulse prior to the excitation and acquisition steps in an NMR experiment (Figure 2.5). This results in saturation of the water resonance rendering it unobservable. The main disadvantage is that exchangeable resonances may also be suppressed due to saturation transfer but non-exchangeable resonances remain unaffected. It is possible to minimize this drawback by changing conditions such as pH and temperature which reduce exchange of labile protons with water.

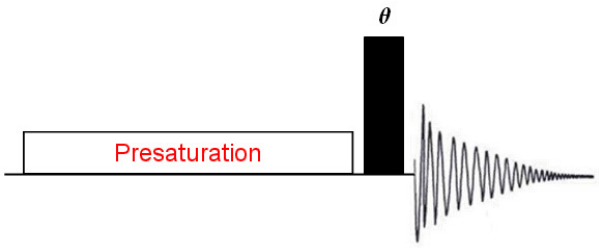


Figure 2.5. Pre-saturation pulse sequence.[95]

Water suppression by gradients-tailored excitation (WATERGATE)[96] (Figure 2.6) differs to pre-saturation as it involves the destruction of the net solvent magnetisation by means of pulsed field gradients (PFGs). This ensures that no magnetisation due to the solvent is observable prior to acquisition. The WATERGATE technique does not carry the disadvantages of pre-saturation and both exchangeable and non-exchangeable resonances can be observed.

An initial $90°_x$ pulse is followed by a G_1-S-G_1 echo scheme which causes the destruction of the solvent magnetisation. The two gradients (G_1) dephase the solvent magnetisation leading to its destruction whereas all other resonances are refocused by the spin-echo (S) thus making them observable. Consequently, at the point of acquisition there is no water signal left.

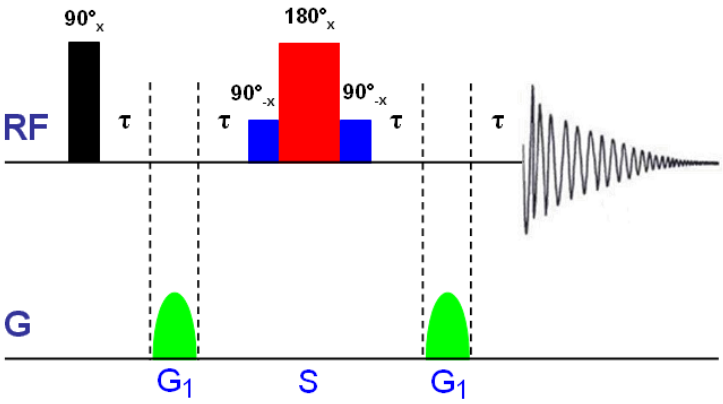


Figure 2.6. Pulse sequence for WATERGATE solvent suppression method. Two selective $90°_{-x}$ pulses are used to rotate the water magnetisation to prevent its rephasing during the second gradient and are placed symmetrically around the non-selective $180°_{x}$ pulse. Four delay periods (τ) are used to allow for gradient recovery.[96]

2.7.2 X {^{1}H} decoupled 1D experiment[71]

With a majority of biological molecules, NMR data recorded for different spin active nuclei can provide important structural and functional information. For example, in DNA, data collected on ^{31}P chemical shifts will give information about the phosphate backbone and in contrast, data collected on ^{15}N chemical shifts can provide information on the base pairing.

Proton decoupling is a non-selective (broadband) method and this involves decoupling of ^{1}H nuclei in order to enhance the signal of X nuclei and simplify the spectrum by removing all ^{1}H-X couplings. The absence of such couplings focuses all the X nuclei into a single sharp and intense line. Experimentally, proton decoupling is achieved by application of a long, low power decoupling pulse during t_2 detection of the X nucleus (Figure 2.7).

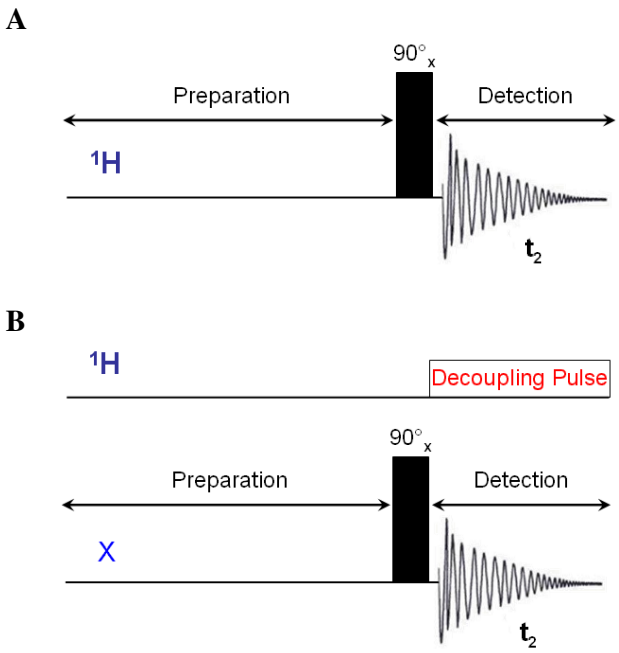


Figure 2.7. Comparison between A. normal ^{1}H-NMR 1D and B. X {^{1}H} decoupled 1D pulse sequences where X represents any spin active heteronucleus.[71]

For 2D experiments which require decoupling, composite pulse decoupling (CPD) is normally used. These are decoupling pulse techniques which include sequences such as MLEV-16 and WALTZ-16. The first CPD sequence was a continuous series of 180º pulses applied to the nucleus that required decoupling. For spin 1/2 nuclei, the 180º pulse inverts the α and β states and this causes rates of precession to be switched. The switch in rates of precession causes refocusing of the spins and if this is done at a fast rate relative to the relevant coupling constant, the coupling is suppressed. Improvements and modifications have been made to this intial method to remove defects and/or to make them specialised for specific purposes.

2.7.2 Correlated Spectroscopy (COSY)

COSY[97,98] is an experiment used to identify protons which are scalar coupled to each other within three bonds (^3J) (Figure 2.8) and two 90° pulses are required which lie either side of the evolution period. The first 90° pulse excites the spin population and creates coherance on each spin which then evolve during t_1. If spins are ^3J coupled to each other, coherence transfer occurs between them during the second 90° pulse, which is the mixing period.

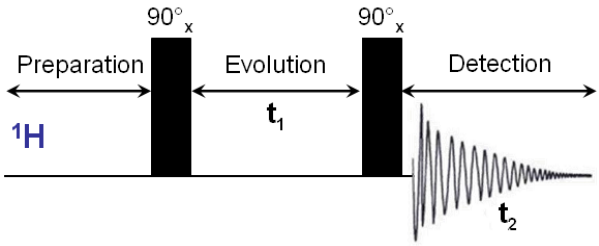


Figure 2.8. Pulse sequence for COSY experiment.

2.7.3 Double Quantum Filtered Correlated Spectroscopy (DQF-COSY)

The DQF-COSY[99,100] experiment (Figure 2.9) is a phase sensitive COSY experiment where all the observed resonances have been filtered using double quantum coherence and appear in an anti-phase manner. By performing the filtration step, which occurs between the second and third 90º pulses, only signals which arise from double quantum coherence are detected thus increasing spectral resolution.

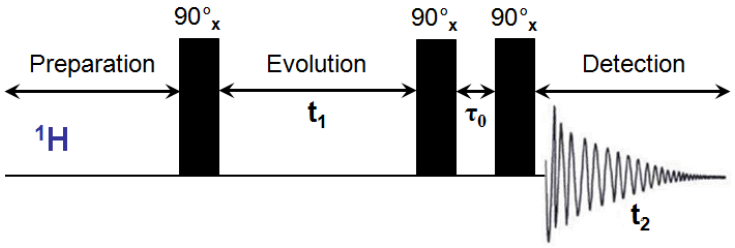


Figure 2.9. Pulse sequence for DQF-COSY experiment. The delay time, τ_0, is of fixed duration and is typically a few microseconds.

A DQF-COSY experiment can be used for identifying scalar (*J*) couplings between protons. A crucial advantage of a DQF-COSY experiment is that the filter also removes singlet resonances along the diagonal. This is particularly advantageous for nucleic acids as correlations between the sugar proton resonances appear quite close to the diagonal and can be seen more clearly.

2.7.4 Total Correlated Spectroscopy (TOCSY)

TOCSY[101,102] (Figure 2.10) is performed to gain information on vicinal and geminal scalar couplings within a spin system. Once magnetisation is transferred into the first nucleus the magnetisation can then be passed onto subsequent nuclei until the spin system ends.

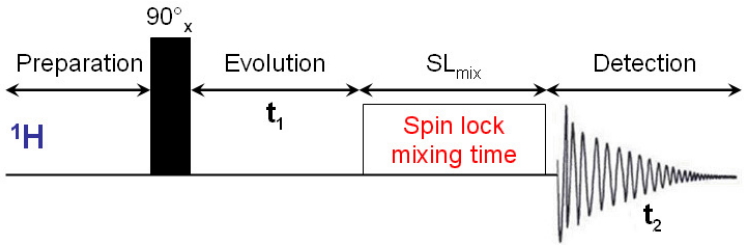


Figure 2.10. Pulse sequence for TOCSY experiment. The spin-lock mixing pulse executes the relayed magnetisation transfer and replaces the single mixing pulse in the COSY experiment.

A spin lock pulse (SL_{mix}) is added after the evolution period and during this time the chemical shifts are averaged out by applying a continuous sequence of pulses. This distributes the magnetisation throughout the entire spin system. The most common spin lock pulse scheme is known as MLEV-17 which is a series of closely spaced 180° pulses which allows homonuclear coupling to evolve whilst refocusing chemical shift evolution.

One can control the number of correlations observed by adjusting the spin lock mixing time and this is one of the advantages of the TOCSY experiment. Shorter mixing times mean that only correlations between adjacent nuclei may be seen whereas longer mixing times leads to more signals being observed as the time of magnetisation transfer is longer.

2.7.5 Nuclear Overhauser Effect Spectroscopy (NOESY)

NOESY[103,104] (Figure 2.11) is used to determine dipolar coupling within a molecule as a result of spins, which are close together in three-dimensional space, experiencing cross relaxation. Cross peaks will be observed if two nuclei are ≤ 5.0Å apart in space. The magnitude of the cross relaxation will depend on the distance between the spins in space and also on the correlation time of the molecules, τ_c.

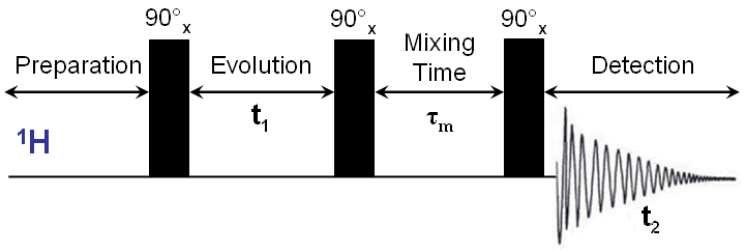


Figure 2.11. Pulse sequence for NOESY experiment.

The NOESY pulse sequence has three 90° pulses and transfer of magnetisation occurs *via* population transfer (Z-magnetisation) as opposed to coherence (XY-magnetisation). To achieve this two 90° pulses are needed during the t_1 step. Following t_1 is the mixing period; here the relaxation of the spins is allowed to occur, where τ_m is the mixing time. A final 90° pulse is applied before the final signal is detected and stored during t_2.

2.7.6 Heteronuclear Single-Quantum Correlation Spectroscopy (HSQC)

HSQC[105,106] (Figure 2.12) is an NMR technique which correlates protons to their directly attached heteroatom. In very basic terms, the HSQC is a double Insensitive Nuclei Enhanced by Polarisation Transfer (INEPT)[107,108] experiment as magnetisation is transferred from proton to $^{13}C/^{15}N$ and then back to the proton again.

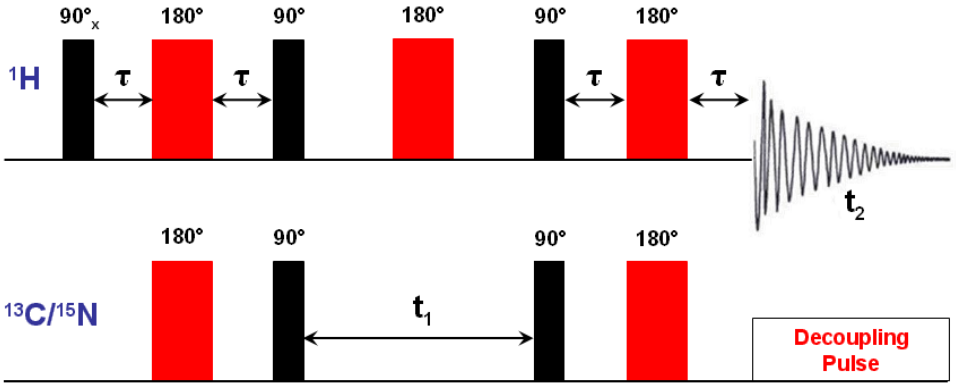


Figure 2.12. Pulse sequence for HSQC experiment. τ represent delay times which are dependent on the value of J_{HX}, typically a value of $1/(^4J)$ is used.

The experiment first begins with proton magnetisation, which is followed by the first INEPT step where proton anti-phase magnetisation is refocused by the double 180º pulses. Following refocusing of the magnetisation, it is transferred to the heteroatom (^{13}C, ^{15}N) where it is allowed to evolve with the chemical shift of the heteroatom. During this period proton coupling is removed by use of a 180º pulse halfway into the t_1 period. Following the evolution step, a reverse INEPT step is carried out which transfers the magnetisation back to the proton. When the magnetisation is back on the proton, it is refocused by the two 180º pulses. Magnetisation of the proton is detected during the detection time (t_2).

2.7.9 CPMG-HSQC-NOESY[109]

^{31}P resonances can provide important information about the phosphate backbone of DNA such as dynamics. However, carrying out ^{1}H-^{31}P NMR experiments prove to be uninformative as phosphorus resonances give very broad linewidths. One way to tackle this is to apply a Carr-Purcell-Meiboom-Gill (CPMG) pulse train, which is a series of closely linked 180° pulses applied during periods of coherence transfer between the phosphorus and the scalar coupled H3' and H5'/5'' protons. This results in enhanced sensitivity due to optimum coupling refocusing between nuclei in the phosphate backbone. The NOESY element of this experiment allows the identification of other sugar and base protons which are close in three-dimensional space. The CPMG-HSQC-NOESY pulse sequence is shown in Figure 2.13.

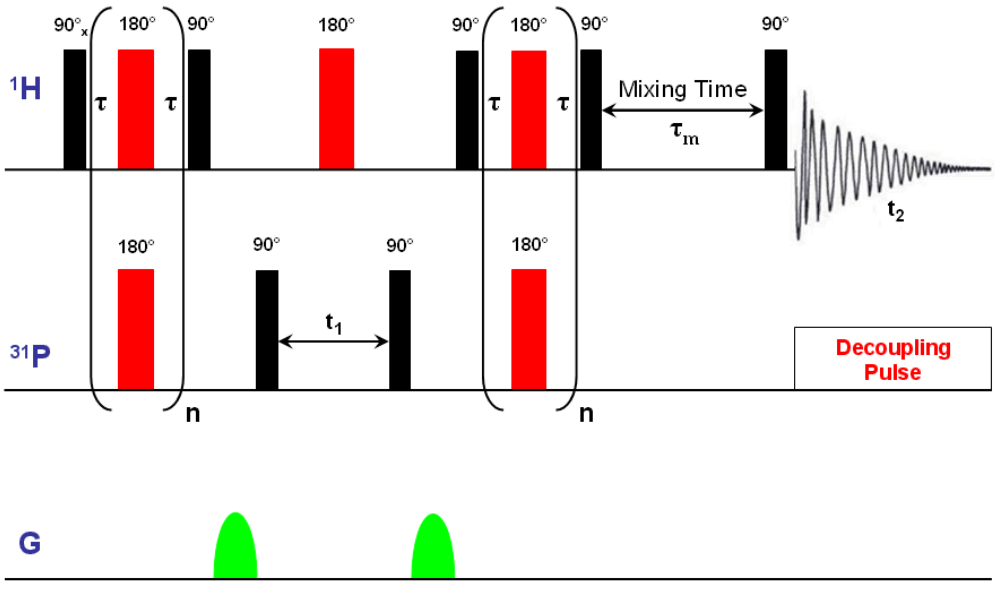


Figure 2.13. Pulse sequence for ^{1}H-^{31}P CPMG-HSQC-NOESY experiment. τ represent delay times which are set to 100 μS and the mixing period (τ_m) was set to 500 ms.

2.7.10 HNCA[110]

The HNCA experiment is used to correlate the amide proton ($^{1}H^{N}$) and backbone ^{15}N chemical shifts to the $^{13}C_{\alpha}$ chemical shifts in proteins. The outline of the resonances identified in this technique is shown in Figure 2.14.

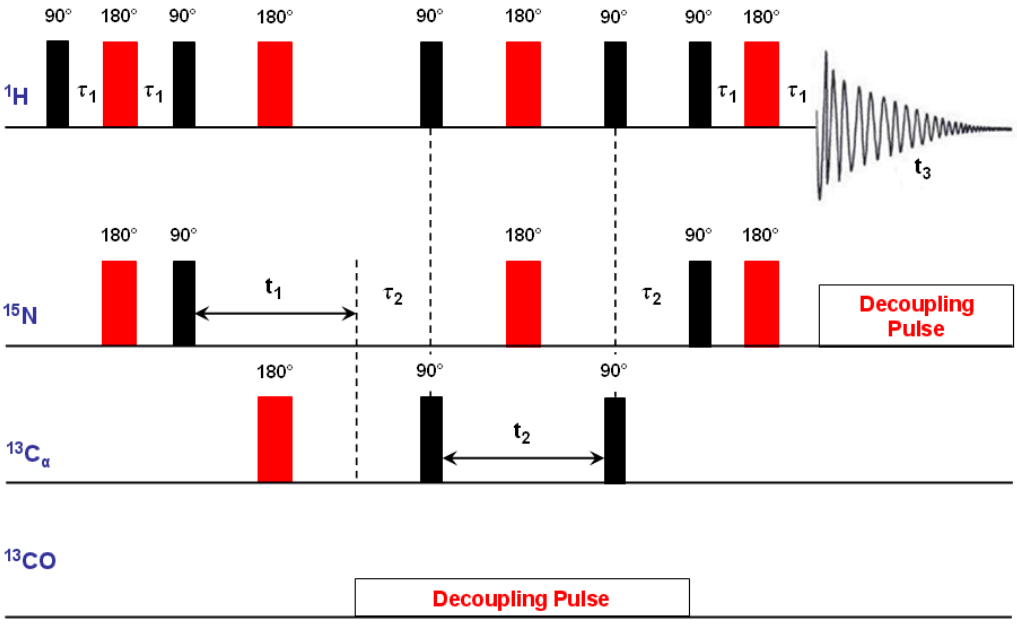


Figure 2.14. Pulse sequence for HNCA experiment. The delay period $2\tau_1$ and τ_2 are typically set to $1/(^{2}J_{NH})$ and $1/(J_{NH})$. ^{15}N and ^{13}CO decoupling was carried out using WALTZ-16.

Correlations to the $^{13}C_{\alpha}$ chemical shifts are established by using the small one bond coupling constant between ^{15}N and $^{13}C_{\alpha}$ which has a value of 7-11 Hz. The one bond coupling described here gives assignments to the intraresidue correlation, but in an HNCA experiment correlations can also be established to the $^{13}C_{\alpha}$ chemical shifts of the preceeding residue. This is achieved by transferring the magnetisation from the ^{15}N nuclei to the $^{13}C_{\alpha}$ *via* interresidue two bond ^{15}N – $^{13}C_{\alpha}$ coupling. The two bond coupling can be as large as 9 Hz in size.

From an assignment perspective, the data is viewed as 2D frequency planes where each plane corresponds to a different ^{15}N chemical shift (F_1). In the 2D plane, the y-axis (F_2) corresponds to a ^{13}C dimension and the x-axis (F_3) corresponds to a ^{1}H dimension.

2.7.11 HNCACB[111]

The HNCACB experiment (Figure 2.15) was developed originally as an alternative to the CBCANH experiment as another approach to correlate the backbone ^{15}N nuclei to the $^{13}C_\alpha/^{13}C_\beta$ nuclei. Compared to the CBCANH approach the HNCACB experiment has two advantages. It is firstly more sensitive but can also be applied to larger proteins, where the $^{13}C_\alpha$ and $^{13}C_\beta$ sites have been perdeuterated. The HNCACB experiment is in a sense an extension of the HNCA experiment as it goes one step further in the sequential assignment of proteins. In an HNCACB experiment, the magnetisation transfer does not stop at the $^{13}C_\alpha$ nuclei, but extends to the $^{13}C_\beta$ nuclei.

Intraresidue correlations to the $^{13}C_\beta$ nuclei are established through a series of small one bond couplings between ^{15}N, $^{13}C_\alpha$ and $^{13}C_\beta$ resonances. Typically the coupling constants for these couplings are $^{1}J_{NH} \approx 91$ Hz, $^{1}J_{NC\alpha} \approx 7\text{-}11$ Hz and $^{1}J_{C\alpha C\beta} \approx 35$ Hz. Similarly to the HNCA experiment, interresidue correlations can also be established to the $^{13}C_\beta$ nuclei of the preceeding residue. This is achieved by transferring the magnetisation from the ^{15}N nuclei to the $^{13}C_\alpha$ *via* the interresidue two bond ^{15}N – $^{13}C_\beta$ coupling, before the magnetisation is transferred to the $^{13}C_\beta$ spin. The two bond ^{15}N – $^{13}C_\beta$ coupling can be as large as 9 Hz in size.

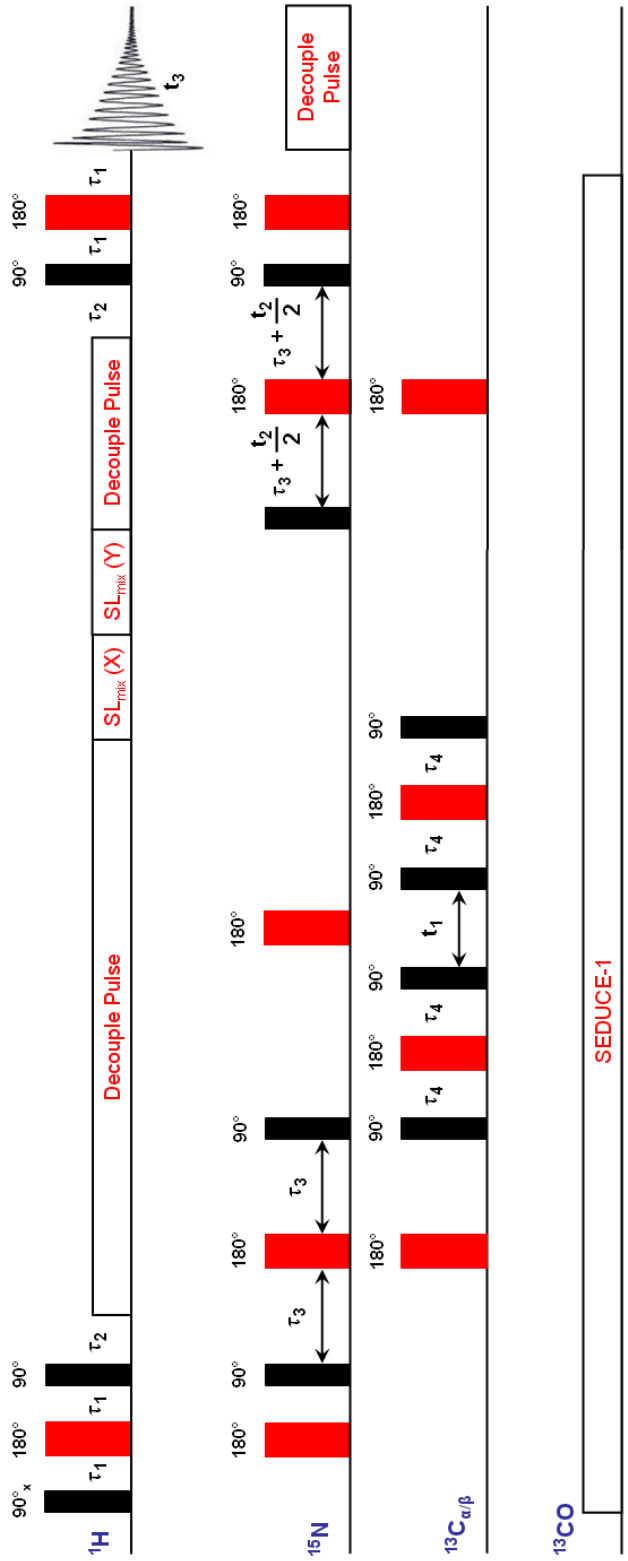


Figure 2.15. Pulse sequence for HNCACB experiment. The SL_{mix} pulses represent ^1H spin-lock pulses with SL_{mix} (X) = 5.0 ms and SL_{mix} (Y) = 3.4 ms. WALTZ-16 decoupling was applied for both ^1H and ^{15}N resonances and SEDUCE-1 ^{13}C homonuclear decoupling was applied to the ^{13}CO resonances. The delay periods τ_1, τ_2, τ_3 and τ_4 were set to 2.4, 5.4,12.0 and 3.55 ms respectively.

From an assignment perspective, the data is viewed as 2D frequency planes where each plane corresponds to a different ^{15}N chemical shift (F_1). In the 2D plane, the y-axis (F_2) corresponds to a ^{13}C dimension and the x-axis (F_3) corresponds to a ^{1}H dimension. At each ^{1}H chemical shift there should be two corresponding ^{13}C chemical shifts, one for $^{13}C_\alpha$ and one for $^{13}C_\beta$ nuclei.

2.7.12 HCCH-TOCSY[112]

HCCH-TOCSY (Figure 2.16) is used to assign aliphatic ^{1}H and ^{13}C resonances in ^{13}C isotopically enriched samples. ^{13}C enrichment allows dispersion of a 2D TOCSY into a third ^{13}C frequency dimension and this is achieved by using three magnetisation transfer steps. Using the HCCH-TOCSY experiment is more efficient than one step ^{1}H magnetisation transfer using ^{1}H – ^{1}H unresolved *J* couplings.

Magnetisation is first transferred from a $^{1}H_\alpha$ nucleus to its directly attached $^{13}C_\alpha$ *via* the one bond $^{1}J_{CH}$ coupling. Following this the magnetisation is transferred to all other ^{13}C nuclei within the same spin system *via* $^{1}J_{CC}$ couplings and finally to the directly attached ^{1}H nuclei. During the second step, magnetisation transfer occurs using an isotropic mixing period of the ^{13}C nuclei. This results in both direct and multiple-relayed transfer along the entire spin system.

The duration of the spin lock mixing period is normally optimised to be 20-35ms which corresponds to the rate of magnetisation transfer between ^{13}C nuclei. When processed the data is viewed as 2D TOCSY planes where the frequency dimensions are ^{13}C ^{1}H and ^{1}H detected in F_1, F_2 and F_3 respectively.

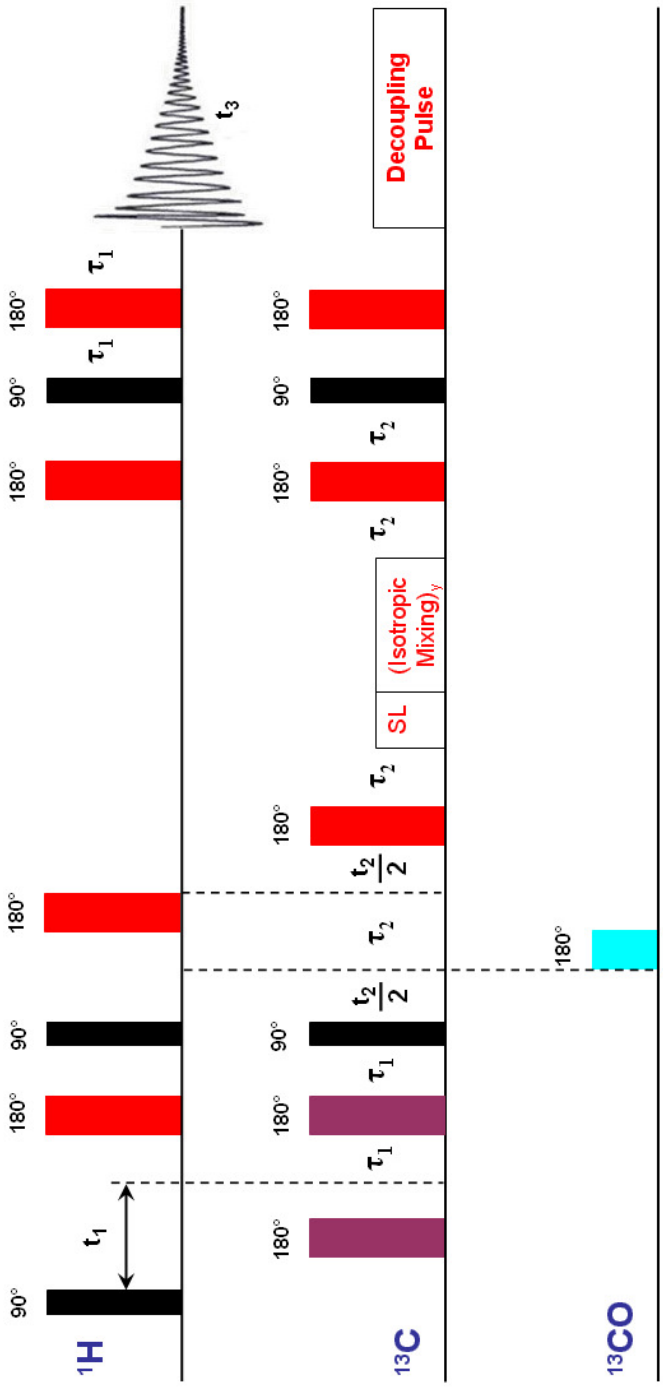


Figure 2.16. Pulse sequence for HCCH-TOCSY experiment. The purple shaded 180° pulses are composite $90°_x$-$180°_y$-$90°_x$ pulses and the ^{13}CO 180° pulse is applied as a DANTE series consisting of thirty-six 5° pulses each with a duration of 10 μS. The delay periods τ_1 and τ_2 are $1/(^4J_{CH})$ and $1/(^6J_{CH})$ approximately in order to allow magnetisation from methane to methylene to methyl resonances. The short spin-lock (SL) period is a trim-pulse which defocuses ^{13}C anti-phase coherence that is not parallel to the effective magnetic field and is normally set to 2 ms in duration. This is followed by the isotropic mixing period which transfers the in-phase ^{13}Cα magnetisation to other ^{13}C nuclei in the spin system via $^1J_{CC}$ couplings (32-40 Hz). A reverse INEPT experiment transfer the magnetisation back onto the directly attached protons prior to detection.

2.8 NMR titration

A simple method for study of complex formation is by NMR titration. Here, the changes in the resonances of the molecule being studied are monitored upon addition of full length WT-*E. coli* MutS protein. 1D ^{1}H-NMR spectra of mismatch DNAs were recorded before and after the progressive addition of small volumes of MutS protein. Titrations were carried out on all mismatched and modified DNAs listed in Figures 2.1 - 2.3. The concentration of WT-MutS added for each titration is given in the corresponding results chapters where the titration data obtained is discussed, i.e. Chapter 3.5, Figures 3.24 – 3.31 for 17mer mismatch GT DNA and its analogues and Chapter 4.3, Figures 4.19 – 4.21 for 6-thioguanine modified 13mer GC DNA and its oxidised derivatives.

The imino proton resonances were the most sensitive to changes during each titration and therefore, the assignment for 17mer mismatch GT DNA and 6-TG modified 13mer GC DNA samples (Chapters 3.1 and 4.1) served as a reference to monitor and observe the formation of mismatch DNA-MutS and modified DNA-MutS complexes. Upon complex formation, the imino protons of the residues involved in binding with MutS gave rise to changes in chemical shifts and line broadening when compared to those of free DNA. This allowed unambiguous identification of the residues which were involved in interaction with MutS protein.

2.9 NMR data acquisition

1D ^{1}H-NMR spectra were usually recorded with 128-512 scans, with the proton transmitter offset (carrier frequency) set to be the same as the $^{1}H_2O$ proton frequency. Experiments in $^{1}H_2O$ solvent were measured at 1-5°C whereas those in $^{2}H_2O$ at 25°C.

For DNA, spectral widths were set to 24-26 ppm and 10-12 ppm, for experiments carried out in $^{1}H_2O$ and $^{2}H_2O$ respectively. Water suppression was used to minimise the large solvent signal. For the NTD-MutS, the spectral widths were set to 12-14 and 8-10 ppm for experiments carried out in $^{1}H_2O$ and $^{2}H_2O$ respectively.

2D NMR experiments were set up with the number of scans set to be 256 or less and 3D experiments typically were set up with 32 scans or less. TD1 and TD2 indicate the number of data points for each time domain and the set values were optimized depending on the experiment type. For example, for a NOESY experiment TD1 and TD2 were typically set with optimum values of 512 and 4096 respectively. The number of scans and data points were carefully chosen to give data with the best sensitivity and spectral resolution within the time available.

As with 1D, in 2D and 3D experiments, the carrier position for ^{1}H was set on the $^{1}H_2O$ proton frequency. Experiments involving the detection of ^{13}C, ^{15}N and ^{31}P were carried out in $^{2}H_2O$ at 25°C and the carrier positions for each were set to 100 ppm, 118 ppm and -3.522 ppm respectively.

Mixing times for the NOESY spectra varied from 150-250 ms, for the TOCSY, it was set to 75 ms and in the ^{1}H-^{31}P CPMG-HSQC-NOESY, the mixing time was 500 ms.

2.10 NMR data processing

1D NMR data was processed using Spinworks,[113] typically using a lorentzian window function, sine solvent filter and line broadening adjustments between 0.5-5.0. Spinworks provides very quick and simple 1D NMR data processing.

2D NMR data was processed using NMRPipe,[114] which is a UNIX based multidimensional NMR data processing software package. Spectra are obtained by first processing of the FID data, followed by phasing and solvent peak suppression method. Data processed by NMRpipe can be checked using NMRdraw, which is a visual interface. The processed data was converted into UCSF format in order to be viewed so that assignments can be made.

2.11 NMR data analysis

The software SPARKY[115] was used to view UCSF files and assign resonances. The completed assignment file was transferred onto the program CcpNMR analysis.[116] This program, similar to SPARKY, allows the viewing and assignment of 2D and higher dimension NMR spectra but also generates NOE distance in an automated manner. This is discussed further in Chapter 2.12.4.

2.12 NMR structure determination

Structure determination is the process of producing a three dimensional refined and validated structure from experimental data through the use of computational simulation using molecular mechanics and molecular dynamics.

To successfully generate a NMR structure a sequence of steps has to be undertaken (Figure 2.17). Experimental constraints are derived from the analysed NMR data before they are submitted to a structure determination program. Following structure calculations, the resulting structures are analysed, checked and the constraints are modified until a reliable structure is produced for detailed conformational analysis.

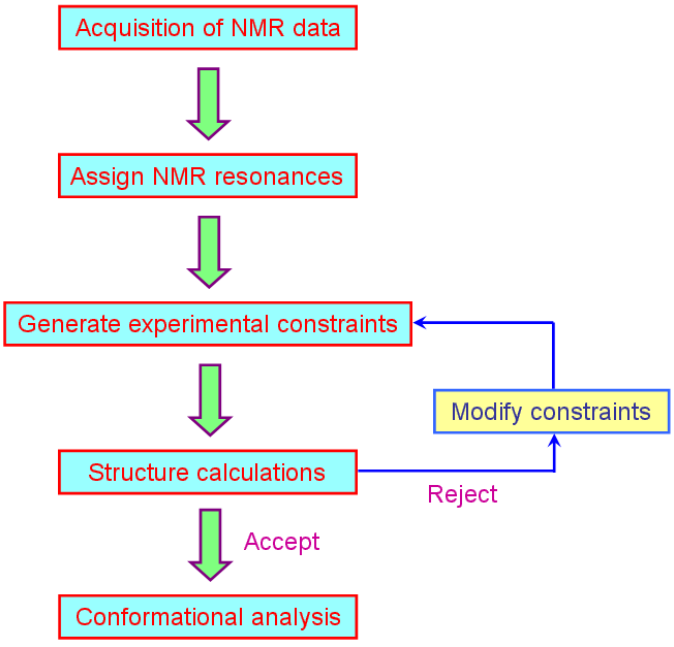


Figure 2.17. Flow diagram outlining the process of NMR structure determination.

2.12.1 Assignment strategy for DNA

The methodology for the assignment of NMR resonances in nucleic acids is quite well established[117-121] and the approach adopted in this project is shown in Figure 2.18.

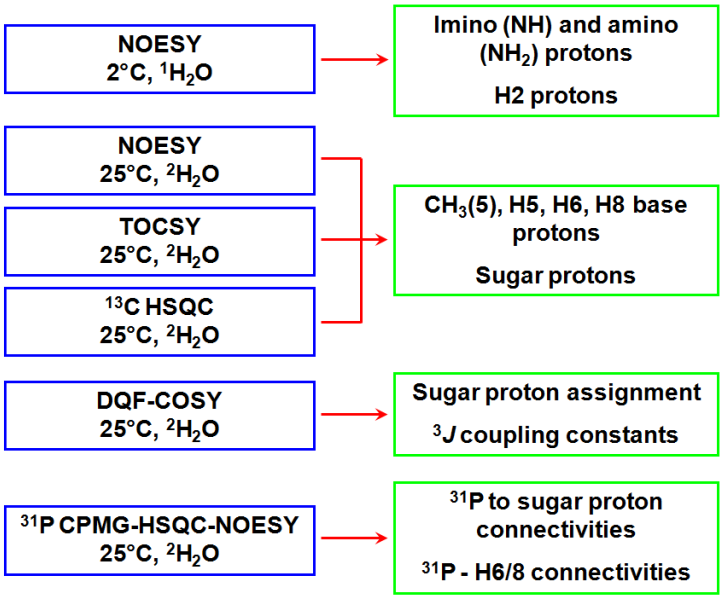


Figure 2.18. NMR experimental strategy towards assignment of ^{1}H, ^{13}C and ^{31}P resonances in nucleic acids.

Spectra measured in ^1H_2O at 2°C identified exchangeable imino (NH) and amino (NH_2) protons whereas spectra measured in ^2H_2O at 25°C identified non-exchangeable deoxyribose sugar and aromatic protons. ^{13}C and ^{31}P resonances are identified in natural abundance heteronuclear spectra measured in ^2H_2O at 25°C.

2.12.2 Assignment strategy for NTD-MutS protein domain

Since the NTD-MutS domain is a protein, the NMR strategy is different to that adopted for DNA. However, the methodology for NMR of proteins is well-established in literature[122-126] and the suite of experiments used is summarized in Figure 2.19.

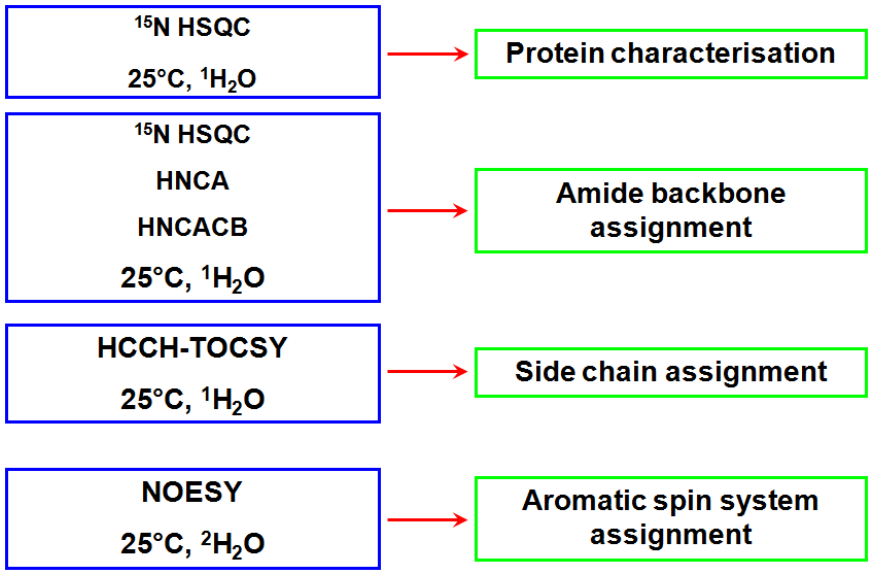


Figure 2.19. NMR experimental strategy towards assignment of ^{1}H, ^{13}C and ^{15}N resonances in proteins.

Due to the high overlap of cross peaks it was necessary to carry out 3D NMR experiments which required a ^{15}N, ^{13}C isotopically labelled sample of NTD-MutS. Both heteronuclear edited and triple resonance 3D experiments were used. Experiments in $^{1}H_2O$ were carried out at 25°C as this gave narrower linewidths and better reflected native condition temperatures. Similar to DNA, resonances could be identified using intra and interresidue correlations.

2.12.3 Molecular modelling[127]

Molecular modelling can be used to generate three-dimensional representations of molecules in order to understand their structure and behaviour under given conditions. Positions of atoms or molecules in the modelled system are represented by Cartesian coordinates (x, y and z) of the atoms or by internal coordinates, in which all the atoms in the system are represented relative to each other.

Molecular dynamics (MD) is a type of computer simulation in which atoms and molecules are allowed to interact for a period of time. This gives an indication of the internal motion of atoms in a molecule or system. The geometry optimisation (GO) of the molecule is normally carried out prior to MD simulations. GO simulations locate the local minima on the potential energy surface rather than the global minimum.

2.12.4 Distance constraints

In NMR structure determination, the most important parameter which can be used is the NOE. The theory of the NOE effect states that the dipolar cross relaxation rate is proportional to the inverse sixth power of the distance between two interacting ^{1}H spins (i and j) (Eqn. 2.13.1).

$$\mathbf{NOE_{ij} \sim 1/r_{ij}^{6}} \qquad (2.13.1)$$

The distance constraints are generated from the very large number of cross peaks observed in the DNA NOESY spectrum by the program CcpNMR analysis. A built-in protocol within CcpNMR analysis automatically generates the distance constraints from assigned ^1H_2O and ^2H_2O NOESY spectra.

A reference NOE distance is set from which all other distances are calculated. This is normally set to a known covalent distance which is unlikely to change in the structure. The default distance function, "standard intensity$^{-1/6}$", was selected and the peak intensities were measured based on integrated peak volumes. This generates a table of peak intensities from the assigned peaks in the NOESY spectrum.

In DNA, the reference distance selected is the intraresidue guanine NH- $NH_2(1)$ proton distance (2.0 Å) for experiments in ^1H_2O and cytosine H5-H6 proton distance (2.4Å), for experiments in ^2H_2O. CcpNMR calibrated the peak intensities based on these reference distances (Eqn. 2.13.2).

$$\mathbf{r_i = r_{ref}\,(S_{ref}/S_i)^{1/6}} \qquad (2.13.2)$$

The upper and lower distance limits are set by CcpNMR based on the average peak intensity. The lowest distance limit is set to 1.7Å, which is the van der Waals radius in atoms. Even though NOE connectivities are usually for distances up to 5Å, the upper distance limit was set to 6.5Å to account for factors such as spin diffusion. Once the distance constraints were calculated they were exported in a format which can be recognised by the Xplor-NIH[128,129] structure calculation program.

2.12.5 Dihedral angle constraints

As mentioned in Chapter 1.1, dihedral angles give information about the sugar pucker, glycosidic dihedral angle and sugar orientation with respect to the backbone. In nucleic acids there are two conformations the sugar ring can be in and these are $C_{2'}$-*endo* and $C_{3'}$-*endo* (Figure 1.5).

The conformation adopted by the sugar ring can be determined from coupling constants ($^3J_{H1'-H2'}$) from DQF-COSY experiments. Large values of 7-11 Hz for this coupling typically demonstrate that the sugar pucker is $C_{2'}$-*endo* and hence B-form DNA in conformation. In contrast, small values of 2-3 Hz indicate a $C_{3'}$-*endo* sugar pucker and hence A-form conformation DNA. Based on the measured $^3J_{H1'-H2'}$ values, dihedral angles were assigned based on standard values outlined by Wüthrich *et al.*[130]

Initial structure calculations using Xplor-NIH were carried out with these standard values with large upper and lower limits. It was found that the calculated values from Xplor-NIH were centred on a narrow range of values. The dihedral angles were modified for later structure calculations based on these ranges. This allowed the calculation to be carried out with narrower upper and lower limits.

2.12.6 Hydrogen bond/planarity constraints

Exchangeable imino protons are involved in base pairing and they appear lowfield in a spectrum due to stacking of the base pairs in the DNA structure which induces an aromatic ring current effect. Hydrogen bond constraints were included upon observation of internucleotide imino to imino (10-15 ppm) and imino to amino (6-9 ppm) exchangeable proton coupling in NOESY spectra which confirmed base pairing. Planarity constraints ensure each base pair remains planar in the DNA molecule. These were included during initial sets of calculations but were removed gradually with the improvement of the NMR structures obtained.

2.12.7 Structure calculations

With any structure determination calculation a starting point has to be created before constraints can be added to the calculation. Requirements for the calculation include a topology file and a structure file. The topology file gives information on the molecule including atom types, masses and charges and bonding whilst the structure file gives information on atom coordinates. These have to be combined to generate a protein data file (psf), which can be input into a structure determination program (Figure 2.20)

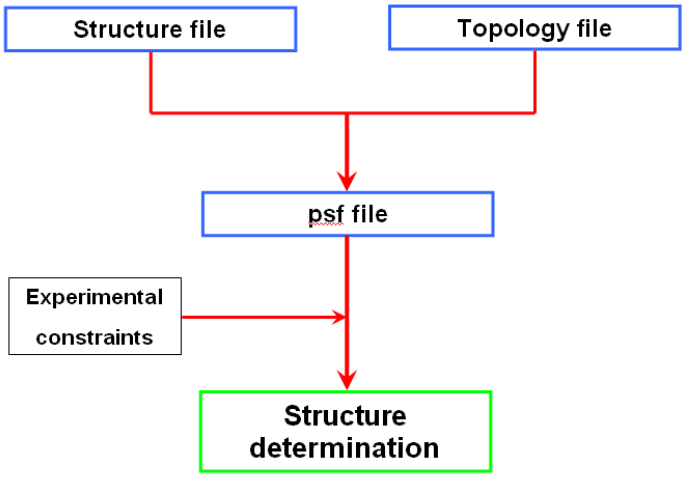


Figure 2.20. Scheme illustrating requirements for structure determination.

The inclusion of experimental constraints can be added in addition to the psf file prior to the structure determination calculations. If experimental constraints are excluded from the calculation, assumptions have to be made in order to restrain the structure and dynamics in some way before the structure calculation is run to avoid production of very random starting structures. For example, for nucleic acids, if canonical base pairs are present, the base pairs can be set to have standard hydrogen bonding constraints.

The program Xplor-NIH was used to carry out all structure determination calculations. Calculations were made to both the 17mer normal GC and mismatch GT DNA and the scheme followed is shown in Figure 2.21.

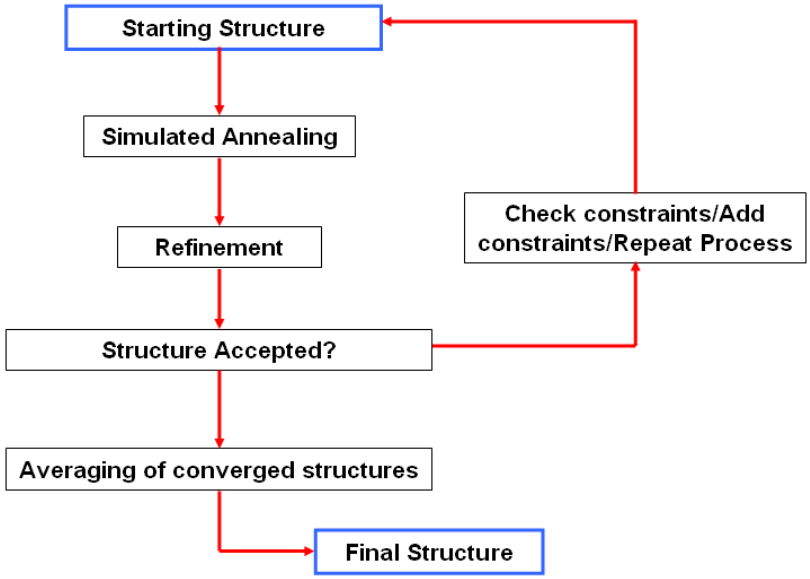


Figure 2.21. Flow chart showing scheme followed for structure determination using Xplor-NIH.

The structure file for the 17mer DNA duplexes was generated using the ".seq" template from Xplor-NIH. The ".seq" file allows one to input the sequence of each strand in the duplex. The nucleotides are inputted as three letter codes, for example, guanine will be GUA. The topology file is generated from combining the two ".seq" files and using the Xplor-NIH protocol "nucleic.top".

NOE distance constraints were included using the "NOE.tbl" Xplor-NIH file. Constraints from NOESY experiments carried out in $^{1}H_2O$ and $^{2}H_2O$ were included as separate files and renamed accordingly, for example, 17GTH2O.

The hydrogen bond constraints were added using the "hbond.tbl" Xplor-NIH file. Where hydrogen bonds are present the "hbond.tbl" file will identify which atoms the hydrogen bond is between, the hydrogen bond distance and also the upper and lower bounds of the hydrogen bond length. In addition to the "hbond.tbl" file, another file

"planarity.tbl" is used to set the base pairs in the duplex to be planar. In the case of mispairs, planarity constraints were omitted from the structure calculation.

Once the sugar pucker has been identified through assignment of H1'-H2' coupling constants, the dihedral angles can be added to the structure calculation using the "tor.tbl" file. Once the Xplor-NIH files have been modified to account for the experimental data, the calculation can be started.

Using Xplor-NIH, 100 structures were generated from an intial extended, random structure using restrained simulated annealing calculations. The van der Waals energy was not included in order to sample more of the conformational space.

Randomisation was followed by simulated annealing, which was carried out on each of the 100 random structures. A long period of restrained molecular dynamics was carried out with the starting temperature set at 3500 K. This provides energy to atoms, which allow them to move randomly through conformational space. Following this, the temperature of the system is reduced slowly by 12.5 K per 0.2 ps in order to find the lowest energy conformations.

The lowest energy structure generated during simulated annealing was put through a refinement step. The main purpose of the refinement step was to reduce the number of violations produced during simulated annealing and aim to generate improved structures with lower RMSD when compared to simulated annealing. The refinement step was set up to produce 100 structures. Structures were rejected if they had a large number of violations and on the basis of high energy and RMSD

when compared to the average of an ensemble of structures. A final structure is selected from a selection of the best structures after refinement based on low energy, low RMSD and a low number of violations.

Structures were viewed and analysed using the VMD-Xplor[131] software package. VMD-XPLOR combines use of XPLOR which is a widely used structure refinement program for NMR structure determination and VMD which is a program developed for the viewing and analyzing of large molecules.

2.12.8 Conformational analysis

The final structure obtained after structure determination can be further analysed to verify if the overall conformation is correct and this can be carried out using the 3DNA software[132]. The 3DNA software analyses the three-dimensional structure of nucleic acids by characterising all base-base interactions and helical character of the base pairs. The calculated values are based on comparison to a standard nucleic acid reference frame[133] (Figure 2.22). These values can be classified in three groups: base pair relationships, base pair step parameters and local helical parameters.

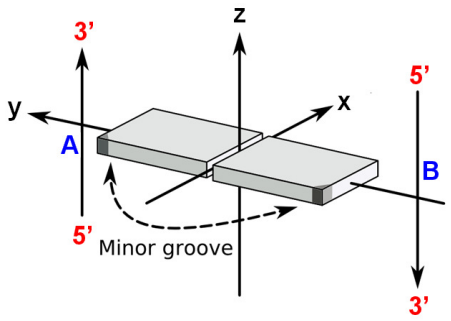


Figure 2.22. Standard reference frame used for determination of parameters in the conformational analysis of nucleic acid structures. A base pair, between polynucleotides A and B, is illustrated by the two rectangular cuboids. The Cartesian axes are shown with respect to the base pair.

Base pair relationships include values of shear (Sx), stretch (Sy), stagger (Sz), buckle (κ), propeller twist (π) and opening (σ) (Figure 2.23). These define the relative position and orientation of the complementary bases in a base pair. Base pair step parameters define the stacking geometry from a local point of reference and include shift (Dx), slide (Dy), rise (Dz), tilt (τ), roll (ρ) and helical twist (Ω) (Figure 2.24). Local helical parameters define the regularity of the helix based on the calculated conformation and include x-displacement (dx), y-displacement (dy), inclination (η) and tip (θ) (Figure 2.25).

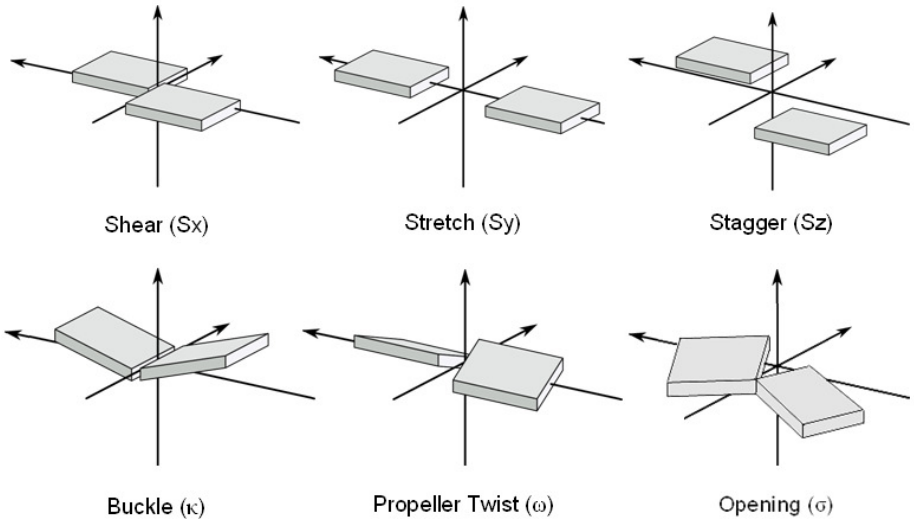


Figure 2.23. Illustration of base pair relationships in conformational analysis of nucleic acid structures.

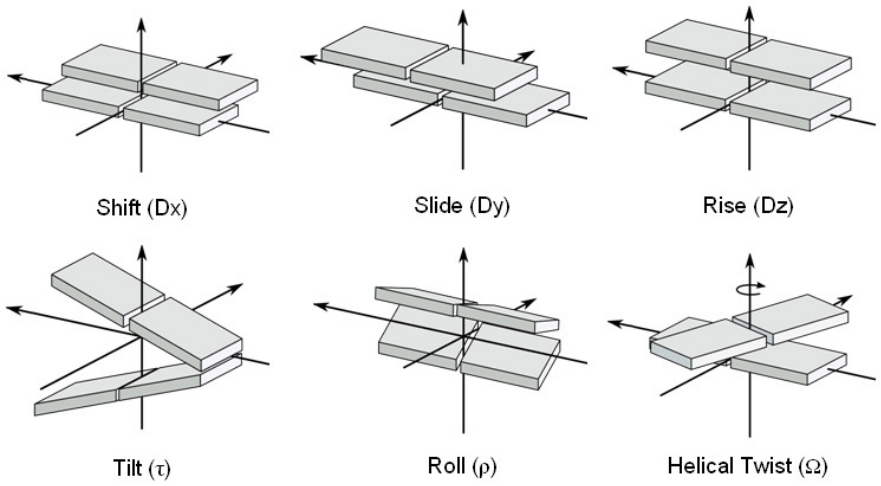


Figure 2.24. Illustration of base pair step parameters in conformational analysis of nucleic acid structures.

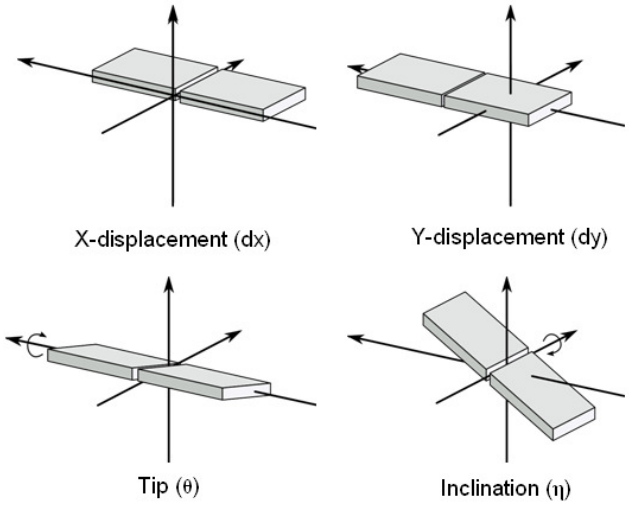


Figure 2.25. Illustration of local helical parameters in conformational analysis of nucleic acid structures.

2.12.9 Deposition of structures in databases

Structures elucidated by NMR spectroscopy can be deposited in three main databases: Nucleic Acid Database (NDB),[134] Protein Data Bank (PDB)[135] and Biological Magnetic Resonance Bank (BMRB).[136]

Another important site to consider is the Research Collaboratory for Structural Bioinformatics (RCSB) PDB[72] which acts as a single depository for structures of large molecules. When submitting a NMR structure, the Cartesian coordinate and constraint data files are processed and released by the RCSB PDB, where upon the structure is assigned a unique ID number. NMR spectral data including chemical shifts and coupling constants are processed and archived by the BMRB. Structures of solely nucleic acids or nucleic acid complexes can be deposited to the NDB for archiving.

Chapter 3 – NMR structure determination of mismatch GT and canonical GC DNAs

The NMR assignment of the ^{1}H, ^{13}C and ^{31}P resonances of 17mer mismatch GT DNA is presented in this chapter, following the methodology outlined in Chapter 2.12.1. Experimental constraints generated from the NOESY data were used to successfully determine the NMR solution structures of these DNA molecules.

Following the determination of the NMR structure of 17mer mismatch GT DNA, the effect of oligonucleotide length, mismatch orientation and sequence context on mismatch stability was investigated using UV absorbance melting.

As a control, the NMR structure 17mer canonical GC DNA was also determined. Conformational analysis using the 3DNA software confirmed the B-DNA helix type adopted by each DNA molecule. Conformational change arising from the GT mispair can be determined by comparison of the helical parameters.

3.1 NMR structure determination and conformational analysis of 17mer mismatch GT DNA

Under aqueous conditions, it can be assumed that a double helical DNA molecule adopts a right-handed B-DNA conformation. Identification and assignment of non-exchangeable proton resonances in DNA is relatively easier to carry out than in RNA, mainly due to a wider dispersion of chemical shift ranges of the former. In RNA, proton resonances are normally observed between 3.5-15 ppm whereas in

DNA resonances can be seen from 0.0-15.0 ppm. The reason for this chemical shift difference is due to the presence of CH_3 methyl groups on thymine residues, which typically resonate between 00.-1.5 ppm and secondly, due to the presence of H2' and H2'' protons at the C2' position in the deoxyribose sugar ring. The substitution of an OH hydroxyl group with a hydrogen results in greater shielding of the protons directly attached to C2', which consequently resonate further highfield. In contrast, all the ribose proton resonances in RNA appear between 3.5-5.0 ppm leading to the problem of signal overlap in this region and hence making assignment more difficult.

In this section, a canonical GC base pair is replaced with a GT mismatch in the 17mer DNA duplex. It is known that canonical base pairing at the mismatch site is disrupted due to this substitution, leading to changes in the local conformation but the extent of this helical distortion is not fully known.

3.1.1 Identification and assignment of exchangeable proton resonances

Imino proton resonances are clearly identified as they appear further lowfield than all other resonances (12.0-15.0 ppm). The lowfield chemical shifts are indicative of good base stacking in the DNA helix. A reliable classification is that imino protons of AT base pairs are typically observed lowfield between 13.0-15.0 ppm while those for GC base pairs are observed further highfield between 12.0-13.0 ppm. This indicates that the base stacking is better in thymine residues compared to guanine.

The effect of the GT mispair on the imino proton chemical shifts was immediately revealed in the 1D ^{1}H-NMR spectrum (Figure 3.1). In contrast to canonical DNA,

peaks were observed between 10.0-15.0 ppm in 17mer mismatch GT DNA. Two clearly resolved peaks were observed in the highfield region and these were identified as the imino proton resonances of the GT mispair. A distortion in the base stacking at the mismatch site is responsible for the highfield shift of the imino protons; G9 was assigned as the resonance at 10.76 ppm and T26 at 11.82 ppm.

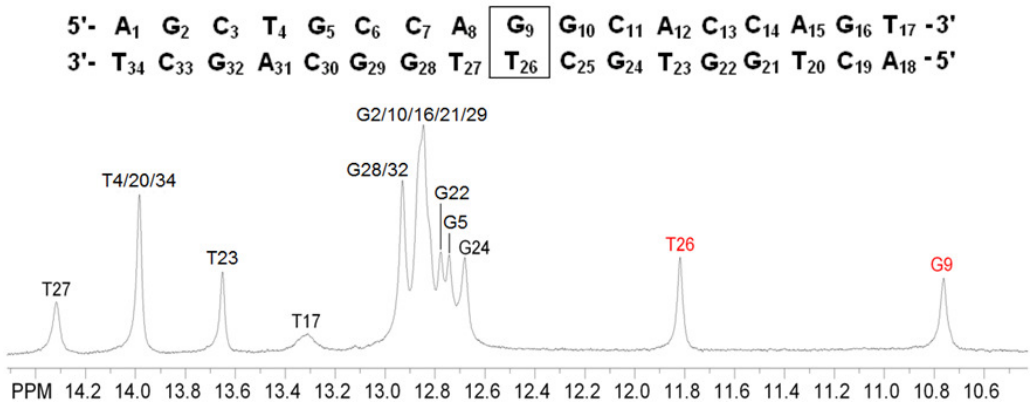


Figure 3.1. 800MHz 1D ^{1}H-NMR spectrum showing imino proton region in 17mer mismatch GT DNA (0.8mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 1°C. The assignment of imino resonances is based on sequential connectivity in the NOESY (τ_m = 250ms) spectrum measured in $^{1}H_2O$.

Assignment of the imino protons was achieved by following the imino-imino proton sequential connectivity observed in the $^{1}H_2O$ NOESY spectrum. These NOEs can be observed along the same strand (intrastrand) and also between strands (cross-strand). Due to internal similarities within this sequence, there was overlap of some imino proton resonances (12.6 – 13.0 ppm).

The highfield shifted resonances of G9 and T26 provided a good starting point for assigning imino-imino NOE connectivities in the NOESY spectrum measured in $^{1}H_2O$. Sequential connectivities were identified from G9 to T34 and G16, only T17 was not identified through these connectivity pathways; sixteen of the seventeen imino proton resonances were successfully assigned in this region. Figure 3.2 shows

the sequential connectivity from G9 to G16. T17 was assigned by elimination in the 1D ^{1}H-NMR spectrum as the broad peak at 13.36 ppm since this was the only unassigned peak in the imino proton region. The broadening of the peak indicates rapid exchange with the water solvent which supports the assignment of T17, as it is a terminal base and is therefore more susceptible to this exchange.

The NOE cross peak between G9 and T26 was intense when compared to other cross peaks in this region giving indication that the two imino protons were very close in space. Furthermore, NOEs were observed between the GT mispair and its flanking base pairs suggesting high stability and good base stacking in the local conformation.

Another important feature was the change in chemical shift of T27 (δ = 14.32 ppm); in 17mer canonical GC DNA, the imino proton of T27 appears at 13.97 ppm (shown in section 3.3). The lowfield shift is caused by better base stacking between T27 and T26. A similar effect was also observed for G32 (δ = 12.83 ppm) giving evidence that the effect caused by the substitution of the GT mispair was not just to the local conformation. All other imino protons were found to resonate at similar chemical shifts compared to 17mer canonical GC DNA.

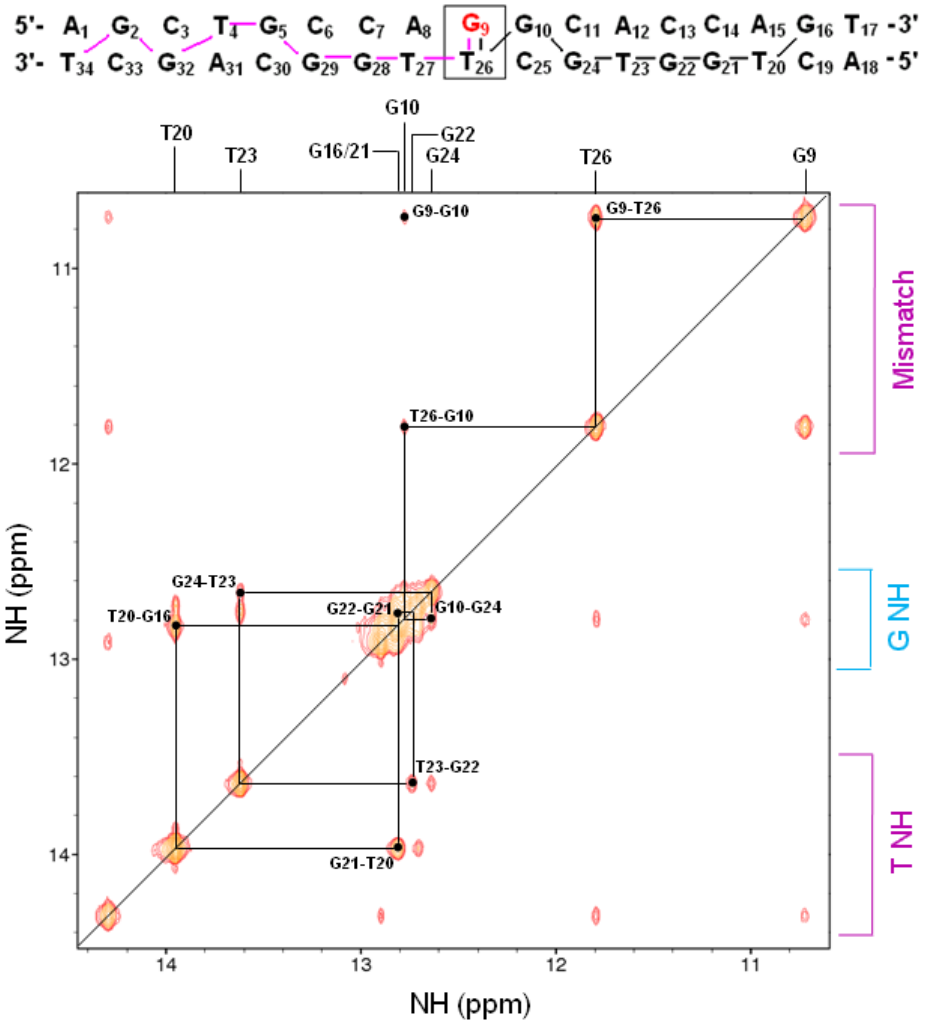


Figure 3.2. 800 MHz NOESY (τ_m = 250ms) spectrum of 17mer mismatch GT DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 1°C. The sequential connectivity stretching from G9 to G16 is indicated by the black trace. The unassigned NOEs in the spectrum correspond to the sequential connectivity stretching from G9 to T34 (shown in Appendix)

Assignment of amino NH_2 resonances

Imino proton assignments based on sequential connectivities were confirmed by assigning NOEs to proximal base paired amino protons. Amino group (NH_2) protons of cytosine, adenine and guanine residues resonate between 5.5 and 9.0 ppm and are overlapped with non-exchangeable aromatic H2, H5, H6 and H8 resonances.

They are distinguished from the non-exchangeable aromatic ring protons by their disappearance in the NMR spectrum measured in ^2H_2O solvent.

Although adenine, cytosine and guanine have amino protons, they can be easily distinguished in NOESY spectrum. Hydrogen-bonded adenine and cytosine amino protons [$NH_2(1)$] resonances are typically observed between 7.6-8.5 ppm (Figure 3.3). NOE cross peaks were observed between imino protons and the amino protons in canonical GC and AT base pairs within the duplex and inter/intrastrand NOE cross peaks aided in the assignment. Non-hydrogen bonded amino protons [$NH_2(2)$] resonate further highfield in this region. For adenine, they typically resonate between 6.0-6.5 ppm and for cytosine between 5.5-6.0 ppm. For guanine residues, both hydrogen and non-hydrogen bonded amino protons resonate between 6.0-7.0 ppm. The hydrogen bonded amino protons were easily assigned as they gave strong NOE cross peaks to the imino protons (shown in Appendix).

The assignment of amino protons showed changes in chemical shifts to a number of resonances. Since the GT mismatch base pair has non-canonical base pairing, the pattern of NOE connectivities is different in the imino-amino region of the NOESY spectrum. G9 is base paired to a thymine residue and not cytosine. Consequently, no NOEs are observed between cytosine amino protons to the G9 imino proton and this absence is clearly observed in the NOESY spectrum measured in ^1H_2O at 1°C. The hydrogen bonded guanine amino proton [$NH_2(1)$] has also shifted highfield (δ = 5.81 ppm) and appears more intense due to the non-canonical base pairing.

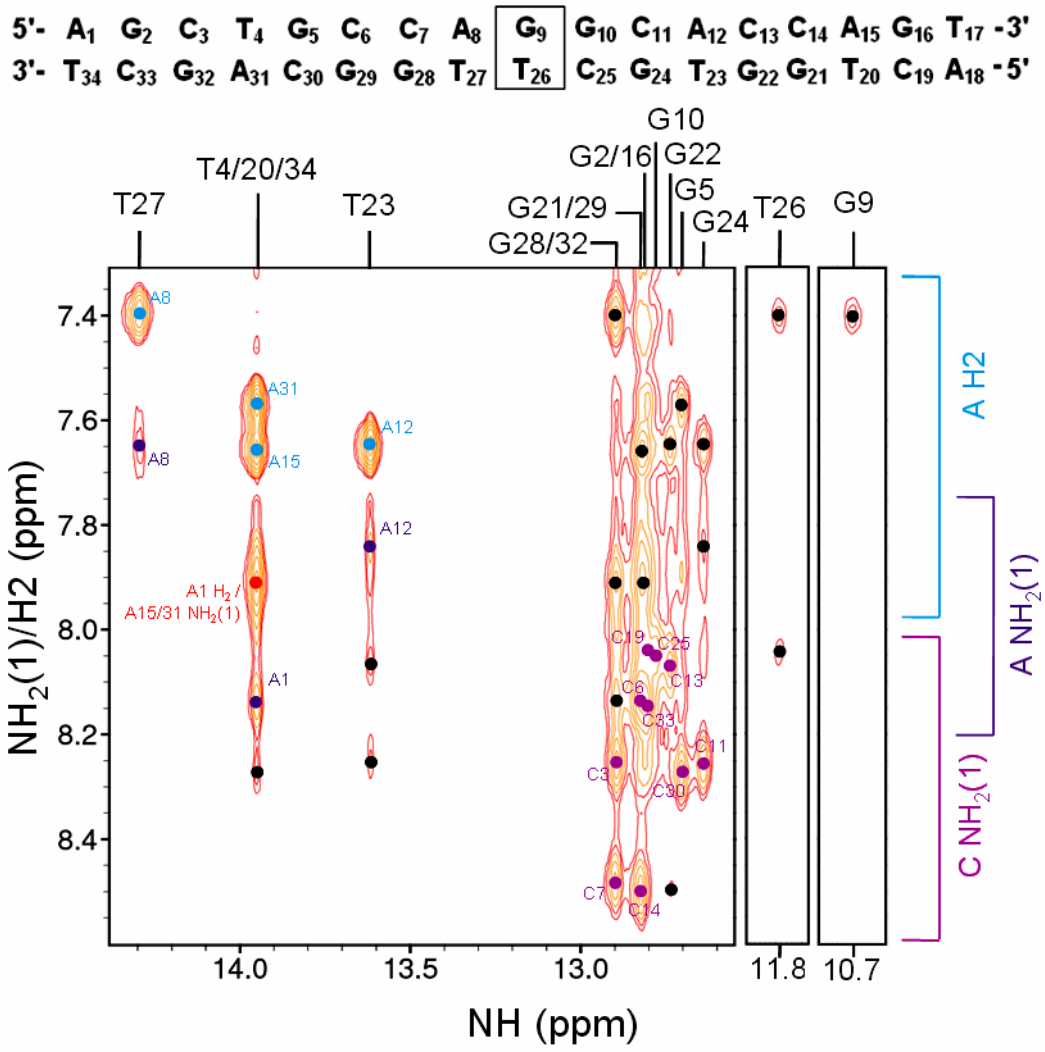


Figure 3.3. 800 MHz NOESY (τ_m = 250ms) spectrum of 17mer mismatch GT DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 1°C. Base pair imino-amino NOE connectivities to adenine H2 and hydrogen bonded cytosine/adenine NH_2 protons are indicated by blue, light purple and dark purple circles respectively. Important intra/interstrand NOE connectivities which confirmed assignment are also highlighted by black circles in this region.

3.1.2 Identification and assignment of non-exchangeable proton and carbon resonances

The effect of the GT mismatch on exchangeable protons was clearly observed in the previous section, however, non-exchangeable protons should also be affected. This section presents the assignment of non-exchangeable protons in 17mer mismatch GT DNA with key differences to the 17mer canonical GC DNA being highlighted.

Assignment of H2 resonances

Non-exchangeable adenine H2 resonances give strong NOEs to their base-paired thymine imino proton resonances and were assigned first in the NOESY spectrum measured in $^{1}H_2O$ (shown in Appendix). The C2 resonances appear well resolved from other regions in the ^{1}H-^{13}C HSQC spectrum (150-160 ppm) and were assigned by correlating with the H2 assignment. Five of the six expected adenine H2 protons were assigned using these NOEs, with only A18 unassigned as no NOE was observed to the thymine imino proton. The A18 H2 resonance was finally assigned in the ^{1}H-^{13}C HSQC spectrum by elimination as one peak remained unassigned, A18 (δ = 8.03 ppm).

With the exception of adenine H2 resonances, all other non-exchangeable resonances were assigned using homonuclear and heteronuclear NMR spectra measured in 100% $^{2}H_2O$ at 25°C.

Assignment of cytosine H5-H6 and thymine CH_3-H6 correlations

Cytosine H5-H6 correlations were identified using DQF-COSY, TOCSY and NOESY spectra. All ten cytosine H5-H6 correlations were identified and gave coupling constants of 7-8 Hz in the DQF-COSY spectrum which was the expected value (Figure 3.4). The position of cytosine H5-H6 cross peaks were consistent to those in 17mer canonical GC DNA (shown in Chapter 3.3) with the exception of C25 which shifted due to the inclusion of a GT mispair [$\Delta\delta$ (H5) = 0.05 ppm and $\Delta\delta$ (H6) = 0.07 ppm].

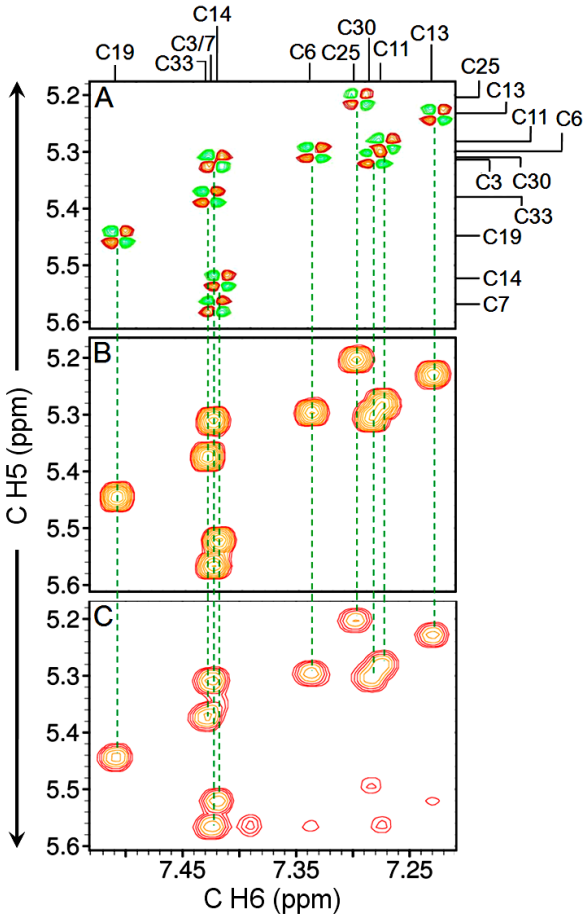


Figure 3.4. 800 MHz spectra stack showing the assignment of cytosine H5-H6 connectivities in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. DQF-COSY, B. TOCSY (SL_{mix} = 75ms) and C. NOESY (τ_m = 250ms).

Correlation of the assigned H5 and H6 resonances allowed the identification of C5 (140-145 ppm) and C6 (140-145 ppm) resonances in the ^{1}H-^{13}C HSQC spectrum (Figure 3.5).

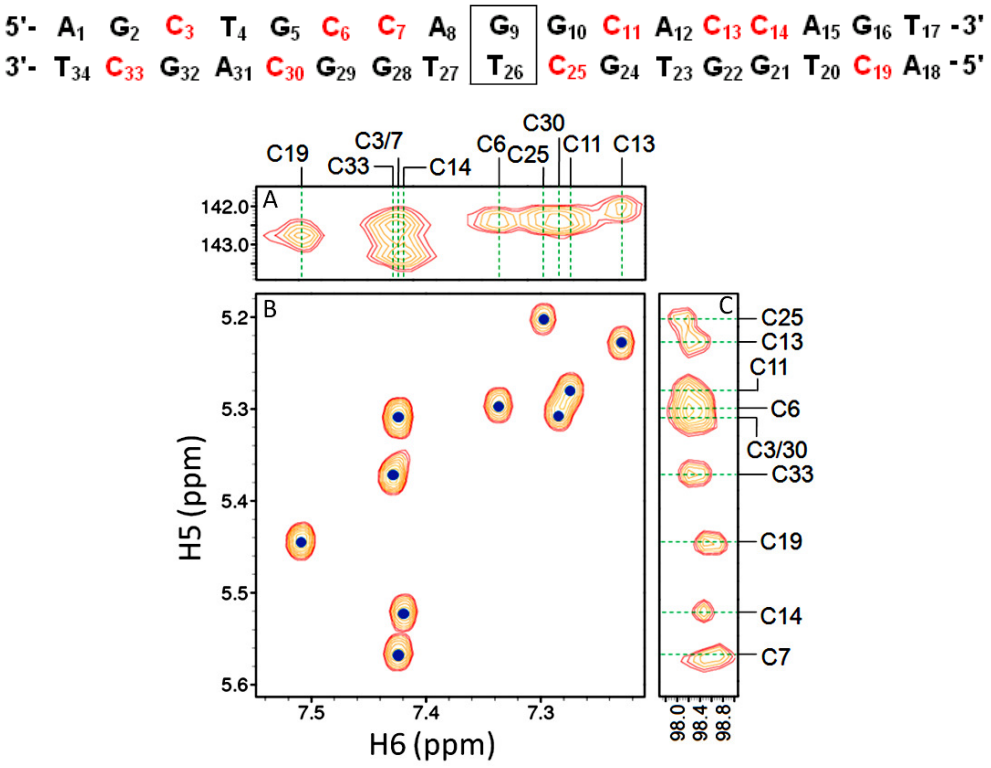


Figure 3.5. 800 MHz NOESY (τ_m = 250ms) spectrum (B) highlighting cytosine H5-H6 correlations in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. 800 MHz ^{1}H-^{13}C HSQC panels show the assignment of the cytosine C6 (A) and C5 (C) chemical shifts.

Thymine CH_3-H6 correlations were assigned using the same methodology but the cross peaks appear weaker as they are further apart in space (2.9Å) compared to cytosine H5-H6 (2.4Å). The mispaired T26 CH_3 resonance appears lowfield in this region (δ = 1.74 ppm) and a change in the chemical shift of the H6 resonance in T27 is also observed ($\Delta\delta \approx 0.1$ ppm). All six thymine CH_3 and H6 resonances were assigned and correlation with the ^{1}H-^{13}C HSQC spectrum identified CH_3 (C5) (0-20 ppm) and C6 (140-145 ppm) resonances.

Assignment of aromatic H5/CH_3 – aromatic H6/8 sequential connectivity

Following the assignment of the directly scalar coupled spin systems in cytosine and thymine, internucleotide NOE connectivities to aromatic protons were determined using the NOESY spectrum. Assignment of cross peaks between cytosine H5/thymine CH_3 protons of a given nucleotide (i), to aromatic H6/H8 protons of the previous nucleotide (i-1) established partial sequential connectivity for the upper and lower strands of 17mer mismatch GT DNA. The results showed that despite the non-canonical base pairing at the GT mispair, this connectivity pathway was observed for C25-G24, T26-C25 and T27-T26 in the NOESY spectrum measured in ^2H_2O at 25°C. This further clarifies that the GT mismatch site is very stable as originally shown by the imino-imino proton sequential connectivity (Figure 3.2).

The observed H5/CH_3-H6/8 sequential connectivity is important as it was a pre-requisite for the identification and subsequent assignment of sugar H1' and further aromatic H6/8 protons in the key H6/8-H1' region which is discussed below.

Assignment of aromatic H6/8 – sugar H1' sequential connectivity

In helical B-form DNA, the sugar pucker is C_2'-*endo* which results in the H1' being close in proximity to H6/H8 (3.6-3.9 Å) and these connectivities can be clearly observed in the NOESY spectrum. The chemical shifts of cytosine and thymine H6 (7.0-7.5 ppm) are distinct from the H8 resonances of adenine (8.0-8.5 ppm) and guanine (7.5-8.0 ppm). In addition, H1' protons resonate between 5.0-6.5 ppm and

this wide dispersion in chemical shifts allows cross peaks to be clearly observed since overlapping of resonances is reduced.

Sequential assignment of non-exchangeable protons was achieved by identifying intra- and internucleotide NOEs between aromatic H6/8 and sugar H1' protons in the $^{2}H_2O$ NOESY spectrum (Figure 3.6). H1' protons for 5' end terminal residues correlate not only to the H6/H8 aromatic proton on the same residue (intranucleotide) but also to the next residue (internucleotide). In contrast, residues at the 3' end of the DNA only show an intranucleotide H6/8-H1' correlation. This quickly identified the terminal residues of the sequence (A1, A18, T17 and T34). Assignment was carried out in the 3' direction from A1 and A18 and all intranucleotide and internucleotide connectivities were successfully identified, leading to a full assignment of all thirty four H6, H8 and H1' resonances. Independent assignment of aromatic H6/8-H6/8 cross peaks in the NOESY spectrum confirmed the chemical shifts of H6, H8 and H1' protons.

It was evident that the substituted GT mispair induced chemical shift changes to resonances of the flanking base pair residues, as was observed with cytosine H5-H6 and thymine CH_3-H6 correlations. There were clear changes to the aromatic H8 resonances of A8 and G9 ($\Delta\delta \approx 0.1$ ppm), however, the sign of the shift differs. With G9, the H8 resonance was shifted highfield suggesting poor base stacking and non-canonical base pairing and in contrast, with A8, the H8 was shifted lowfield indicating stronger base stacking. These shifts are consistent with the T27 imino proton and aromatic H6 proton resonances which have been observed to experience lowfield shifts in the presence of a GT mispair.

Assigning all H6, H8 and H1' protons was important as it established the foundation through which all other non-exchangeable resonances can be assigned.

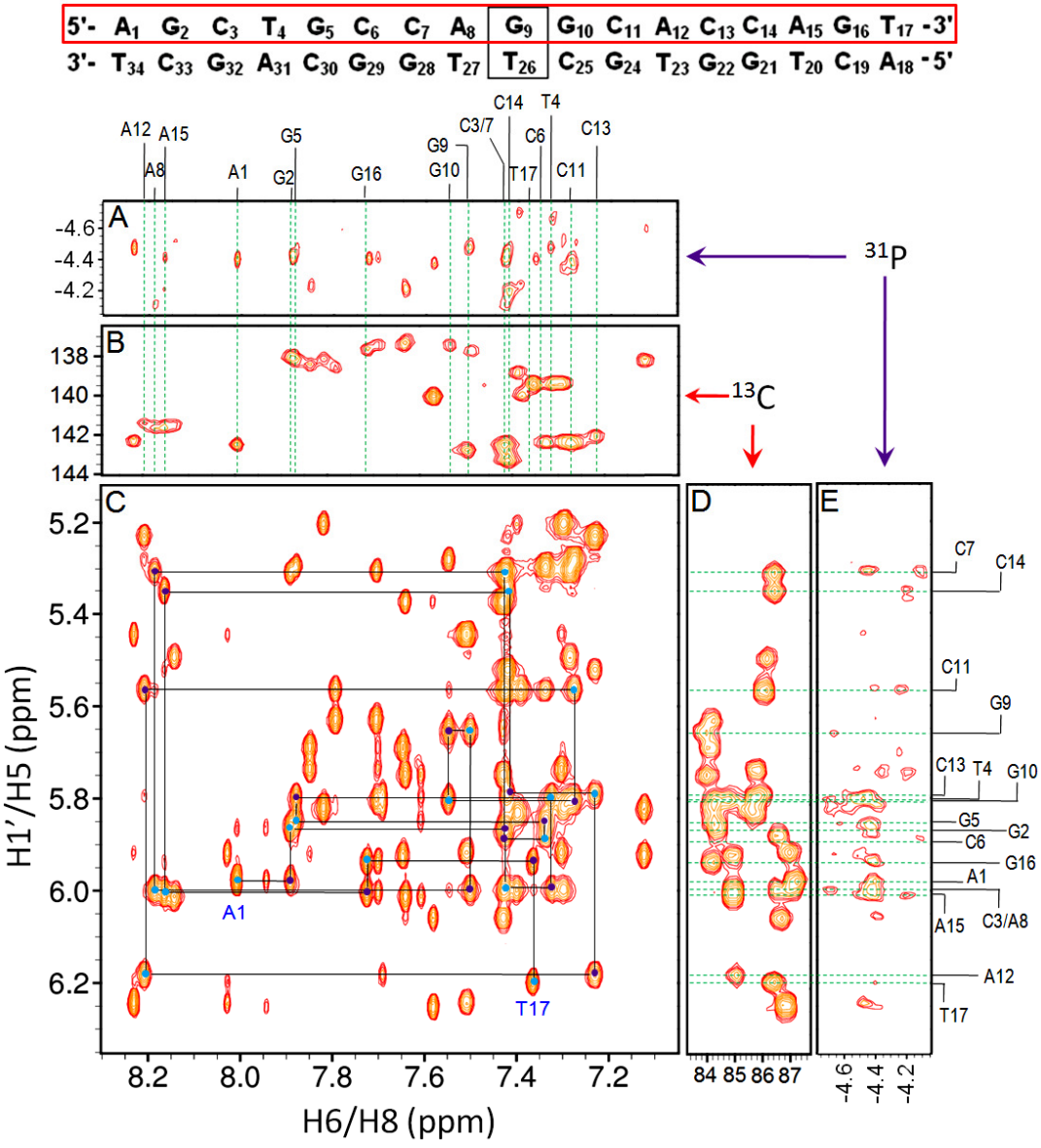


Figure 3.6. 800 MHz NOESY (τ_m = 250ms) spectrum (C) showing H6/8-H1' sequential NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^{2}H2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide connectivities are shown by blue circles and internucleotide connectivities by purple circles. 800 MHz ^{1}H-^{13}C HSQC spectra show the corresponding assignment for C6/C8 (B) and C1' (D) resonances respectively. 600 MHz ^{1}H-^{31}P CPMG-HSQC-NOESY spectra show the corresponding assignment for backbone ^{31}P resonances through H6/H8-^{31}P (A) and H1'-^{31}P (E) NOE connectivities respectively. The unassigned peaks which are seen in this region correspond to the H6/8-H1' sequential connectivity stretching between A18 and T34. These resonances have been assigned and are shown in the Appendix.

Assignment of H2' and H2'' resonances

Unlike RNA, there are two protons at the C2' position and these resonate highfield of all other sugar protons (δ = 1.8-3.0 ppm). H2' and H2'' can be differentiated in the DQF-COSY spectrum as the coupling constant of $^{3}J_{H1'-H2'}$ correlations are larger than $^{3}J_{H1'-H2''}$ correlations (Figure 3.7, A).

Interestingly, it was found during assignment that for both pyrimidines and purines, H2' protons resonated at higher than H2'' protons. All H1'-H2' correlations were identified in the DQF-COSY spectrum on basis of their coupling constants (7-11 Hz). However, since the coupling constant of H1'-H2'' correlations are very small (2-3 Hz), the peaks were difficult to observe in the DQF-COSY spectrum. This problem was not observed in the TOCSY spectrum (Figure 3.7, B) where each H1'-H2'' was seen as an intense peak.

The chemical shifts of H2'/2'' protons were confirmed in the NOESY spectrum (Figure 3.7, C) as both exhibit NOEs to the same H1' resonance for each nucleotide. H1'-H2'' NOESY cross peaks are more intense as the H1'-H2'' distance is shorter than H1'-H2' for all sugar conformations. H2' and H2'' protons were assigned for all nucleotides in 17mer mismatch GT DNA.

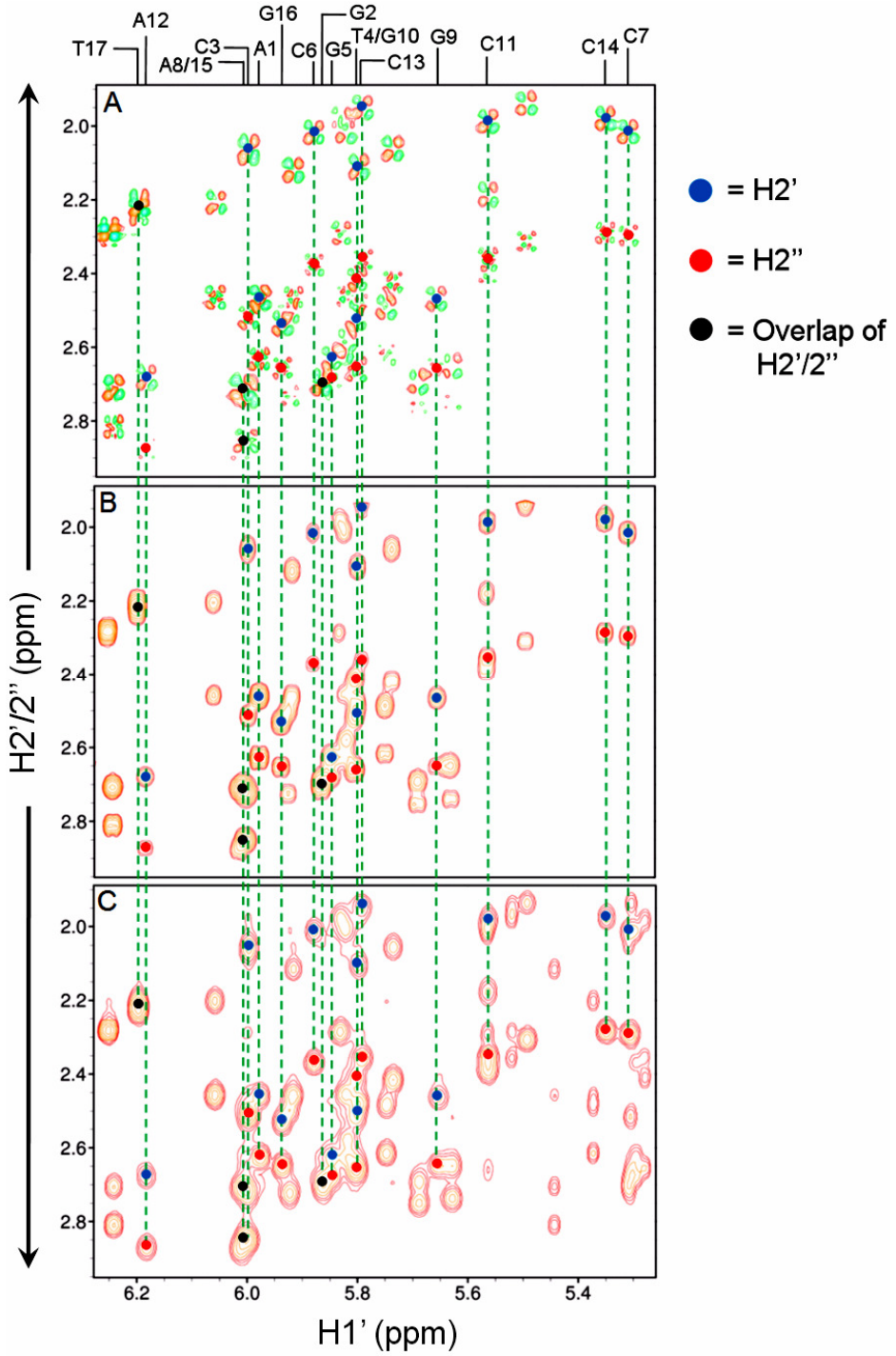


Figure 3.7. 800 MHz spectrum stack showing identification and assignment of H2'/H2'' resonances for A1-T17 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. DQF-COSY, B. TOCSY (SL_{mix} = 75ms) and C. NOESY (τ_m = 250ms). The unassigned peaks which are seen in this region correspond to H1'-H2'/2'' NOEs for nucleotides A18 to T34; these resonances have been assigned and are shown in the Appendix.

All H2' and H2'' resonances were clearly resolved and assigned with the exception of G2, T17 and T34, where the H2' and H2'' resonances were overlapped and were given the same chemical shift. The NOESY spectrum measured in $^{2}H_2O$ for 17mer canonical GC DNA was recorded at 600 MHz (section 3.3.2); in contrast, the NOESY spectrum measured in $^{2}H_2O$, shown in this section, was recorded at 800 MHz. The higher magnetic field allowed greater spectral dispersion of the peaks in this region compared to 17mer canonical GC DNA. Thus, peaks which may have been overlapped are not easy to distinguish from one another. Distance constraints used in the structure calculation should therefore be more reliable.

By distinguishing the different chemical shifts of the H2' and H2'' resonances, H6/8-H2'/2'', NOE connectivities could be assigned which give important distance constraints in structure calculations. NOEs between aromatic H6/8 and sugar H2'/2'' protons can be distinguished by the difference in their intramolecular distances where H6/8-H2' and H6/8-H2'' have values of 2.0-3.6 Å and 3.4-4.5 Å respectively. Consequently, their NOEs will give different peak intensities in the NOESY spectrum (Figure 3.8).

The chemical shifts of H2'/2'' protons assigned above were independently confirmed from their NOEs to H3' protons. The corresponding C2' resonances were identified and assigned in the ^{1}H-^{13}C HSQC spectrum by correlating with the chemical shifts of the H2' and H2'' resonances.

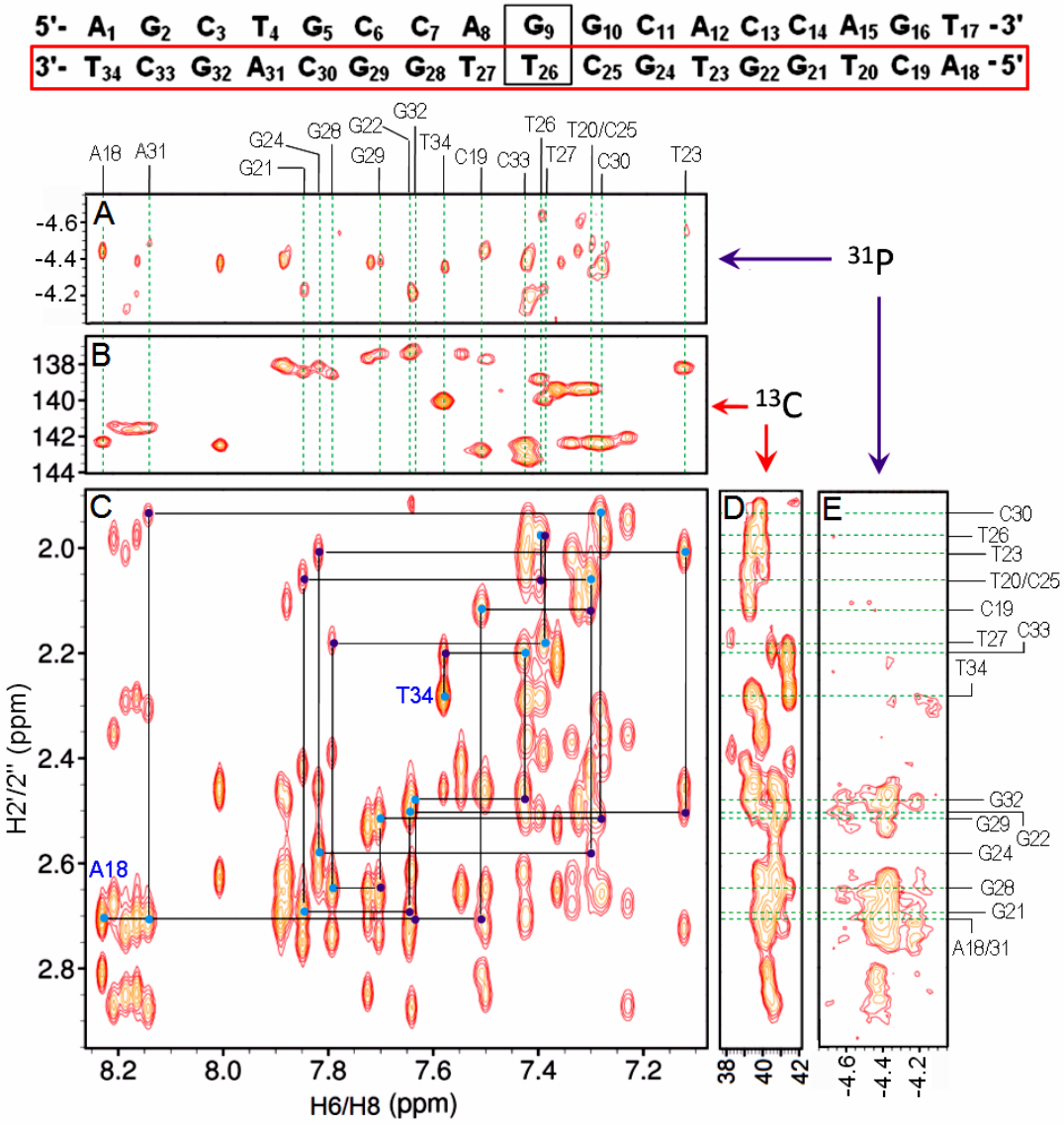


Figure 3.8. 800 MHz NOESY (τ_m = 250ms) spectrum (C) showing H6/8-H2' sequential NOE connectivities between A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide connectivities are shown by blue circles and internucleotide connectivities by purple circles. 800 MHz ^1H-^{13}C HSQC spectra show the corresponding assignment for C6/C8 (B) and C2' (D) resonances respectively. 600 MHz ^1H-^{31}P CPMG-HSQC-NOESY spectra show the corresponding assignment for backbone ^{31}P resonances through H6/H8-^{31}P (A) and H2'-^{31}P (E) NOE connectivities respectively. The unassigned peaks which are seen in this region correspond to the H6/8-H2' sequential connectivity stretching between A1 and T17 and also the H6/8-H2'' sequential connectivity for both strands. These resonances have been assigned and are shown in the Appendix.

To assign the H3', H4' and H5'/5'' sugar resonances, the direct method is to identify H2'-H3', H3'-H4' and H4'-H5' ^{3}J scalar coupled correlations in the DQF-COSY or TOCSY spectrum as these give intense cross peaks. This however proved difficult as these regions displayed a high degree of overlap over a narrow chemical shift range, so distinguishing one resonance from another was not possible.

Assignment of H3' resonances

The H3' resonances were assigned by identifying cross peaks to H1' resonances. This was carried out in the NOESY spectrum as strong impurity peaks in the TOCSY spectrum caused streaking across the H1'-H3' region. H3' protons resonate between 4.5-5.2 ppm, however, a clear divide was noticed in the chemical shift depending on the base type (pyrimidine or purine). H3' protons of pyrimidines typically resonate highfield of 4.9 ppm and those of purines lowfield of 4.9 ppm.

The most immediate observation in the NOESY spectrum was the two intense cross peaks observed at approximately 4.5 ppm. These were assigned as the 3' end terminal residues T17 and T34 and their enhanced intensity was plausible as terminal residues are more dynamic allowing the H1' to be closer to the H3' in space.

Strong NOEs were also observed between aromatic H6/8 and H3' protons, which allowed the H6/8-H3' sequential connectivity to be identified for each strand. This confirmed the assignment of H3' resonances from the H1'-H3' region; all H3' and C3' resonances were also assigned.

Assignment of H4' and H5'/5'' resonances

The assignment of H4' and H5'/5'' resonances proved more difficult as they overlap in the same region of the NOESY spectrum. The H1'-H4' NOE distance is shorter than H1'-H5'/5'' thus giving a more intense cross peak. Fortunately, the H4' resonances for purines (4.3-4.5 ppm) appear at a different chemical shift range to pyrimidines, which overlap with H5'/5'' resonances (4.0-4.3 ppm), making them easier to assign. All H4' resonances were assigned by identifying H1'-H4' NOEs and confirmed by H6/8-H4' NOESY cross peaks. Changes to H4' resonances were restricted to only residues directly flanking the GT mispair with those for A8, T27 and C25 shifting. But for residue G10, the H4' proton remain unchanged indicating that the backbone in the G9-G10 region is stable despite the GT mispair substitution. All C4' resonances were also assigned by correlating with the ^{1}H-^{13}C HSQC spectrum.

H5' and H5'' resonances were assigned similarly to H2'/2'' resonances with H1'-H5'' NOEs being more intense than those for H1'-H5' correlations (Figure 3.9). The same observation applied for NOEs between aromatic H6/8 and sugar H5'/5'' protons where those for H6/8-H5'' appeared more intense.

The H5'/5'' resonances of the A1 and A18 residues were readily identified as they appeared clearly separate from other resonances. The chemical shift change of some H1' resonances, caused by the GT mispair, altered the position of some H5'/5'' resonances with the most striking change being the T26 C5' chemical shift. The ^{1}H-^{13}C HSQC spectrum showed that the C5' resonance was clearly resolved from all

other residues. This implied that the GT mispair affected both the central region of the helix and the sugar-phosphate backbone which should be seen more clearly in the NMR structure of 17mer mismatch GT DNA obtained from Xplor-NIH.

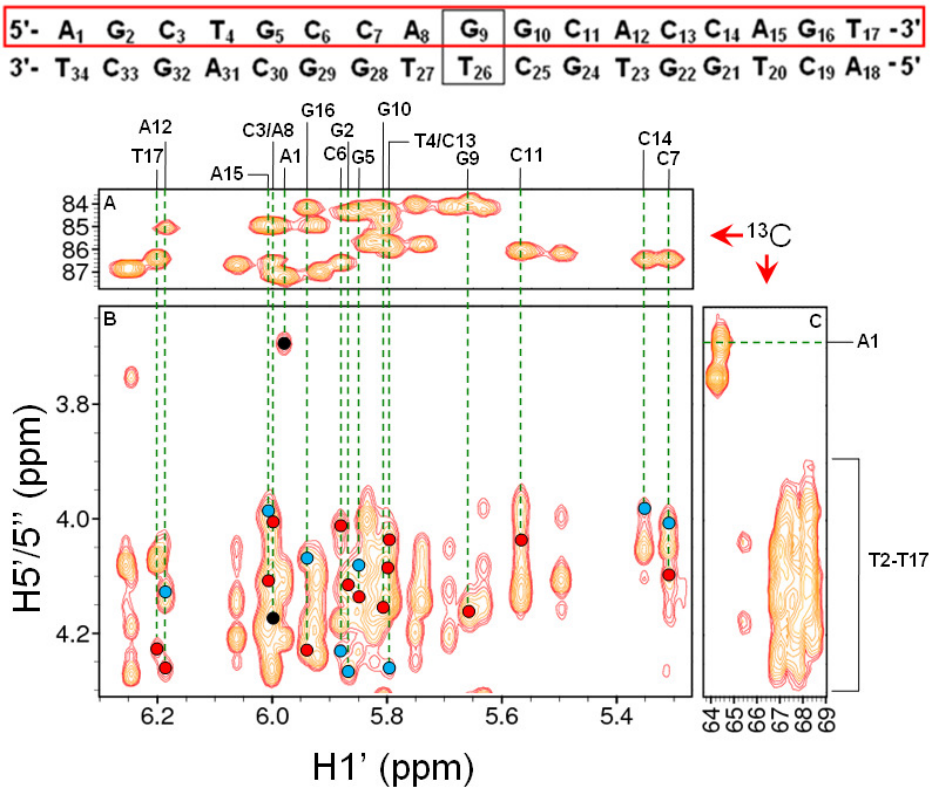


Figure 3.9. 800 MHz NOESY (τ_{mix} = 250ms) spectrum (B) showing the assignment of H1'-H5'/5'' correlations for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. 800 MHz ^{1}H-^{13}C HSQC panels show the corresponding assignment of C1' (A) and C5' (C) resonances. The H5'/5'' resonances of residues T2-T17 are highly degenerate and appear in the heavily overlapped region indicated in the figure. The unassigned resonances in the region correspond to H1'-H5'/5'' for residues A18 to T34. H1'-H4' NOEs for pyrimidines also appear in this region. These have been assigned and are shown in the Appendix.

3.1.3 Assignment of phosphorus resonances

^{31}P resonances were assigned using ^{1}H decoupled 1D ^{31}P-NMR (shown in appendix) and 2D ^{1}H-^{31}P CPMG-HSQC-NOESY experiments. In the ^{1}H-^{31}P CPMG-HSQC-NOESY, the magnetisation transfers initially from H3' and H5'/5'' to the ^{31}P atom in the backbone, so the most intense peaks in the spectrum are the H3'-^{31}P and H5'/5''-

^{31}P HSQC peaks. As the H5'/5'' resonances are heavily overlapped, the ^{31}P resonances were assigned using H3'-^{31}P HSQC peaks (Figure 3.10). A total of twenty five ^{31}P resonances were identified using H3'-^{31}P correlations. With the exception of C6, the remaining eight residues were assigned by identifying the NOEs to aromatic H6/8 and sugar H1' protons.

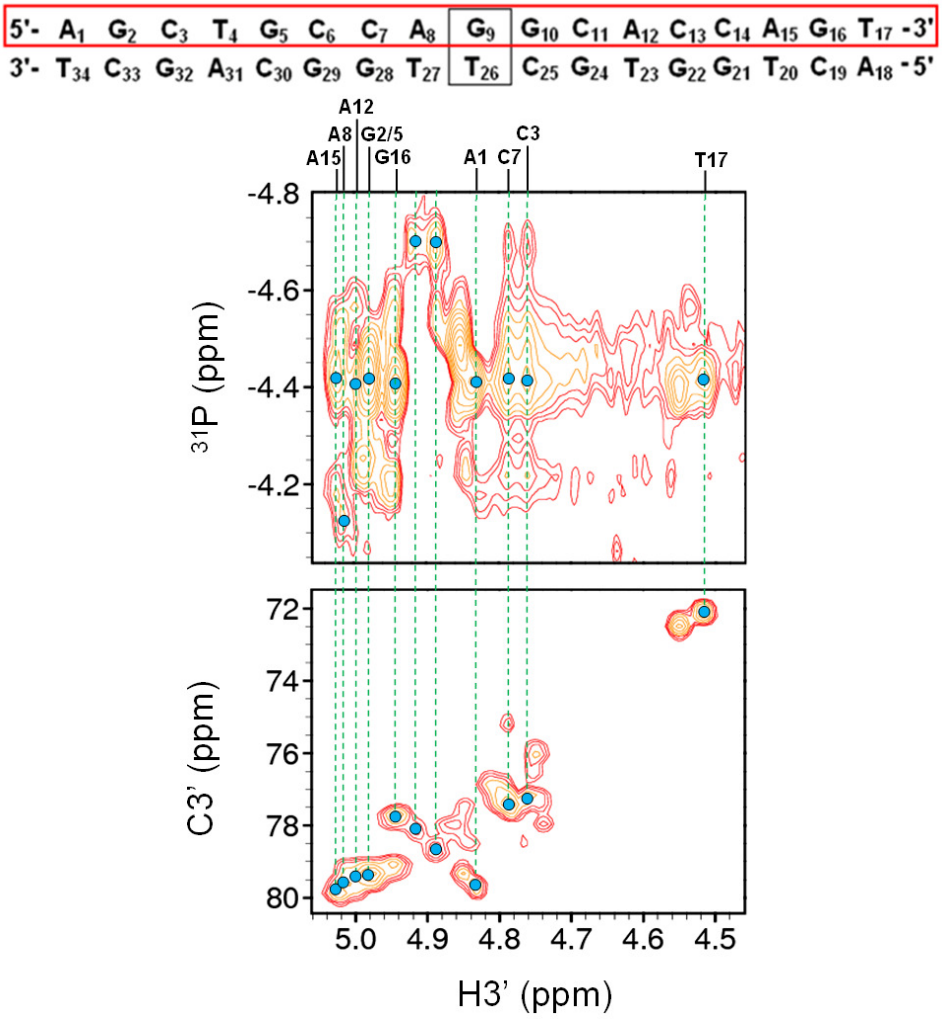


Figure 3.10. (Top panel) 600 MHz ^{1}H-^{31}P CPMG-HSQC-NOESY (τ_{mix} = 500ms) spectrum and (bottom panel) 800 MHz ^{1}H-^{13}C HSQC spectrum. The assignment of ^{31}P resonances is shown by means of H3'-^{31}P HSQC correlations for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. The unaasigned peaks correspond to H3'-^{31}P HSQC peaks for residues A18-T34 whose assignment has not been shown here but is given in the Appendix.

3.1.4 Table of assignments

Based on the identification and assignment procedures described previously in this chapter, the chemical shifts of exchangeable and non-exchangeable proton, carbon and phosphorus resonances following assignment are summarised in Tables 3.1 and 3.2.

Nucleotide Number	Imino H1	Amino $NH_2(1)$	Amino $NH_2(2)$	H2	$CH_3(5)$	H5	H6	H8	H1'	H2'	H2"	H3'	H4'	H5'	H5"
A1		8.14	6.24	7.91				8.01	5.98	2.47	2.63	4.84	4.21	3.69	3.69
G2	12.92	6.69						7.89	5.87	2.71	2.71	4.98	4.41	4.26	4.12
C3		8.25	5.97			5.31	7.43		6.00	2.06	2.51	4.75	4.27		4.17
T4	13.98				1.61		7.33		5.81	2.11	2.48	4.88	4.15		4.09
G5	12.73	6.41						7.88	5.85	2.63	2.68	4.98	4.38	4.08	4.14
C6		8.13	5.65			5.30	7.34		5.89	2.02	2.37	4.99	4.19	4.23	4.12
C7		8.49	5.81			5.57	7.43		5.31	2.01	2.29	4.79	4.05	4.01	4.10
A8		7.65	6.41	7.41				8.19	6.00	2.73	2.85	5.01	4.35	4.17	4.01
G9	10.75	5.81						7.50	5.66	2.47	2.65	4.92	4.34		4.17
G10	12.81	6.52						7.55	5.81	2.42	2.66	4.90	4.32		4.16
C11		8.25	5.78			5.28	7.28		5.57	1.99	2.36	4.82	4.12		4.04
A12		7.84	6.14	7.65				8.21	6.19	2.68	2.88	5.00	4.40	4.26	4.13
C13		8.07	5.79			5.23	7.23		5.79	1.95	2.36	4.89	4.14	4.26	4.04
C14		8.51	5.78			5.53	7.42		5.35	1.98	2.29	4.81	4.06	3.99	
A15		7.91	6.36	7.66				8.17	6.01	2.71	2.85	5.03	4.38	3.98	4.10
G16	12.84	6.85						7.73	5.94	2.53	2.66	4.94	4.37	4.07	4.22
T17	13.36				1.56		7.37		6.20	2.22	2.22	4.52	4.07		4.23
A18				8.03				8.23	6.25	2.71	2.81	4.85	4.27	3.75	3.75
C19		8.02	5.79			5.45	7.51		5.92	2.12	2.47	4.86	4.24	4.19	4.10
T20	13.98				1.62		7.31		5.74	2.06	2.42	4.86	4.14	4.06	4.10
G21	12.85	6.84						7.85	5.69	2.70	2.75	4.99	4.36	4.21	4.12
G22	12.76	6.52						7.65	5.93	2.50	2.73	4.89	4.40	4.11	4.20
T23	13.64				1.33		7.13		5.83	2.02	2.46	4.85	4.17	4.20	4.12
G24	12.67	6.43						7.82	5.82	2.58	2.63	4.94	4.34	4.06	4.12
C25		8.05	5.79			5.21	7.30		6.00	2.07	2.52	4.74	4.25		4.19
T26	11.82				1.74		7.40		5.83	1.99	2.29	4.96	4.00	4.04	4.18
T27	14.32				1.61		7.39		5.57	2.18	2.39	4.84	4.14		3.98
G28	12.92	6.81						7.80	5.63	2.65	2.74	4.98	4.32	4.15	4.08
G29	12.85	6.40						7.71	5.81	2.52	2.66	4.95	4.36		4.15
C30		8.28	5.82			5.31	7.29		5.50	1.94	2.31	4.81	4.11	4.00	4.16
A31		7.91	6.27	7.57				8.14	6.01	2.71	2.88	5.02	4.37	4.00	4.11
G32	12.83	6.57						7.65	5.75	2.49	2.62	4.95	4.35	4.25	4.20
C33		8.25	5.97			5.37	7.43		6.06	2.21	2.46	4.74	4.21	4.05	4.14
T34	13.98				1.74		7.58		6.25	2.29	2.29	4.55	4.08	4.05	4.15

Table 3.1. Summary of chemical shifts for ^{1}H proton resonances in 17mer mismatch GT DNA. Boxes shaded in grey indicate that the absence of the proton in the specified residue and those shaded in blue indicate unassigned protons.

Nucleotide Number	C2	C5	C6	C8	C1'	C2'	C3'	C4'	C5'	^{31}P
A1	154.5			142.8	87.2	39.7	79.7	89.6	64.5	-4.41
G2				138.1	84.3	40.2	79.3	87.5		-4.43
C3		98.2	142.5		87.0	40.0	76.0	85.9		-4.54
T4		14.5	139.4		85.8	39.2	78.0	85.2		-4.70
G5				138.2	84.3	40.8	79.3	87.5		-4.43
C6		98.2	142.5		86.6	40.0	79.3	85.7		
C7		98.7	142.5		86.4	39.4	77.5	85.8		-4.41
A8	153.9			141.7	85.0	40.3	79.3	86.5		-4.13
G9				137.8	83.9	40.5	78.2	86.8		-4.70
G10				137.5	84.8	41.5	78.8	87.1		-4.40
C11		98.2	142.5		86.2	39.8	76.9	85.5		-4.25
A12	155.0			141.4	85.1	40.6	79.3	86.9		-4.41
C13		98.3	142.2		86.0	39.8	78.0	86.1		-4.46
C14		98.5	143.3		86.4	39.4	76.9	85.8		-4.21
A15	154.6			141.6	85.0	40.3	79.9	86.5		-4.41
G16				137.6	84.2	40.8	72.7	87.3		-4.41
T17		14.1	139.5		86.5	41.3	72.1	87.1		-4.41
A18	154.9			142.3	86.8	40.2	79.3	89.4	64.3	-4.49
C19		98.6	142.8		87.0	39.2	78.0	85.9		-4.51
T20		14.7	139.4		85.8	39.1	78.0	86.1		-4.37
G21				138.4	84.1	40.2	79.3	87.4		-4.24
G22				137.5	84.9	41.2	78.8	87.5		-4.52
T23		14.2	138.2		85.6	40.0	77.5	85.7		-4.58
G24				138.1	84.3	40.7	72.7	86.8		-4.58
C25		98.1	142.2		86.6	40.0	77.3	85.9		-4.41
T26		14.4	138.9		85.6	39.4	79.2	85.0	65.4	-4.22
T27		14.5	140.0		86.1	38.3	78.5	86.1		-4.24
G28				138.6	84.2	40.1	79.3	87.1		-4.43
G29				137.4	84.3	40.8	79.1	87.4		-4.42
C30		98.2	142.4		86.2	40.0	76.9	85.5		-4.41
A31	154.2			141.5	85.0	40.6	79.2	86.5		-4.51
G32				137.2	84.0	40.8	79.1	87.2		-4.21
C33		98.3	142.5		86.7	40.5	78.0	85.6		-4.41
T34		14.4	140.0		86.8	41.5	72.5	87.3		-4.40

Table 3.2. Summary of chemical shifts for ^{13}C and ^{31}P resonances in 17mer mismatch GT DNA. Boxes shaded in grey indicate that the absence of the resonance in the specified residue and those shaded in blue indicate unassigned resonances. Boxes shaded in red indicate that the chemical shifts of C5' resonances are highly overlapped. The indicated C5' resonances all appear between 66.5-68.8 ppm.

3.1.5 Determination of experimental constraints for structure calculation

NMR distance constraints were extracted from the $^{1}H_2O$ and $^{2}H_2O$ NOESY experiments (τ_m = 250 ms) measured at 1°C and 25°C respectively, using the CcpNMR analysis program. Distances were grouped in strong (1.5-2.5 Å), medium (2.6-4.0 Å) and weak (4.1-6.0 Å) which are determined by CcpNMR analysis with respect to the reference distance (cytosine H5-H6, 2.4 Å). Exchangeable constraints were given wider error bounds due to potential exchange with the solvent and were calibrated with respect to the guanine imino proton – NH_2(1) NOE distance (2.5 Å).

Dihedral angles constraints were constrained according to the sugar pucker type determined from $^{3}J_{H1'-H2'}$ coupling constants in the DQF-COSY spectrum. α, β, γ, δ, ε, ζ, χ dihedral angles were assigned based on standard B-form values stated by Wuthrich *et al.*[110]. Two sugar ring dihedral angles (ν_1 and ν_2) were also constrained based on standard B-form values.

Hydrogen bond constraints were included in the structure calculation to loosely constrain base pairs which could be identified through imino-imino proton NOE connectivities in the $^{1}H_2O$ NOESY spectrum.

3.1.6 Structure calculation

The number of constraints used in the structure calculation of 17mer mismatch GT DNA was significant (472) and this is similar to the subsequent 17mer canonical GC DNA (shown in section 3.2) despite the substitution of a GT mismatch. A summary is given in Table 3.3.

Constraint Type	No. of constraints
Total NOEs	472
Strong (1.5 - 2.5 Å)	33
Medium (2.6 - 4.0 Å)	419
Weak (4.1 - 6.0 Å)	20
Hydrogen bond	98
Total dihedral angle	301
Helix (α, β, γ, δ, ε, ζ)	199
Glycosidic (χ)	34
Ribose pucker (υ_1, υ_2)	68
NOE constraints per residue	13.9
NOE and dihedral constraints per residue	22.7

Table 3.3. Summary of experimental constraints used in the structure calculation of 17mer mismatch GT DNA carried out with the Xplor-NIH program.

NOE distance constraints were not included for connectivities to H4' and H5'/5'' due to heavy overlapped which would involve much larger error bounds than allowed in the structure calculation. For the same reason, sugar proton-sugar proton NOE connectivities were not included in the structure calculation with the exception of H1'-H2'/2'', H1'-H3' and H2'/2''-H3' as the NOE cross peaks observed for these were well resolved.

From the 100 structures generated from Xplor-NIH, ten were selected based on the criteria of lowest energy and lowest RMSD, from which a final average structure was generated (Figure 3.11). The overlay of the ten best structures (Figure 3.11, A, RMSD = 0.47) shows that the GT mispair adopts two orientations in an overall B-form DNA conformation. The major orientation appears to be where G9 angles up towards the 3' end of the top strand giving the GT mispair a sheared appearance. This observation was extended to the G10-C25 base pair and the A8-T27 base pair appears to remain planar. The NMR structure also confirms the T26 residue base stacking well with T27 which accounts for the T27 imino proton shifting lowfield.

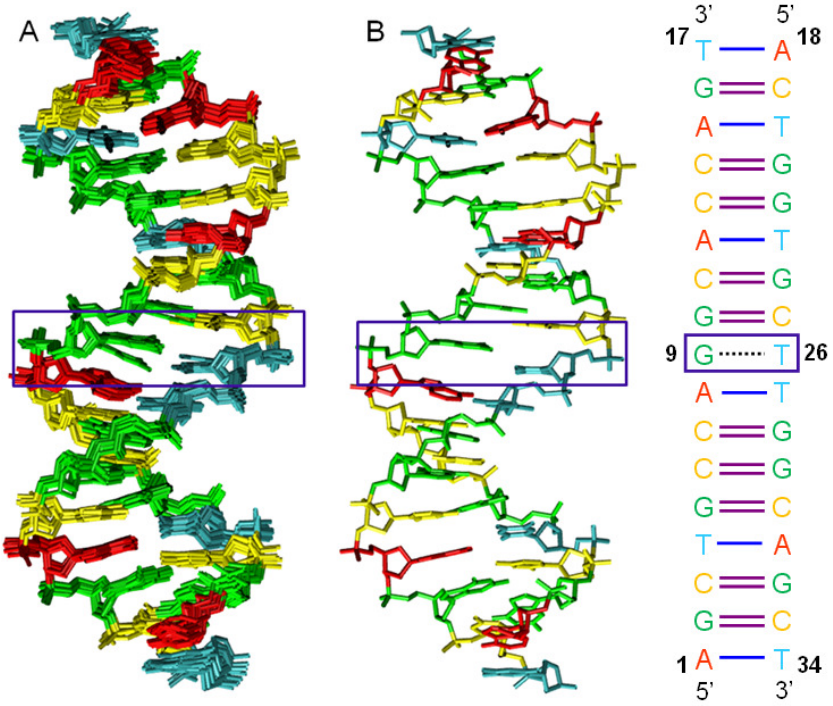


Figure 3.11. Illustration showing an overlay of the ten lowest energy and RMSD structures (A) and the final structure (B) for 17mer mismatch GT DNA calculated from Xplor NIH. Sequence is shown on the right and the position of the G9-T26 mismatch base pair is outlined by the purple box.

A comparison between the NMR structures of 17mer canonical GC and mismatch GT DNA (Figure 3.12) shows the helix appearing to be wider at the GT mispair

which was caused by its sheared orientation. Base stacking is equally good in both structures at the mispair site despite the substitution of the C26/T26 residue and there does not appear to be any major distortions in the sugar-phosphate backbone for either molecule. However, the helix seems to be winding more at the terminal ends of 17mer mismatch GT DNA.

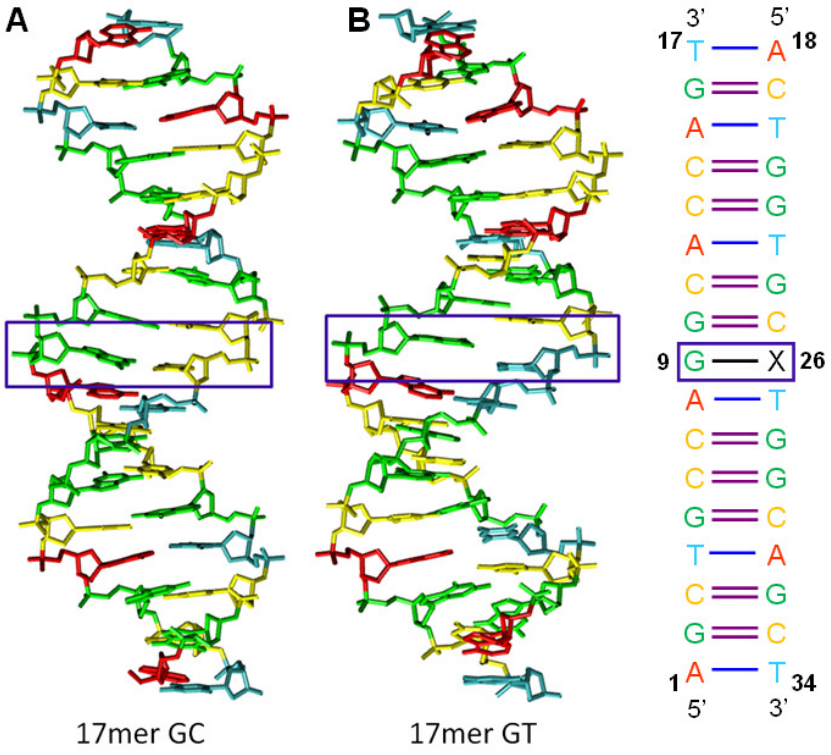


Figure 3.12. NMR structures of 17mer canonical GC DNA (A), 17mer mismatch GT DNA (B) and scheme illustrating the 17mer DNA sequence (C) where X = C/T. NMR structures were obtained using the Xplor-NIH program.

3.1.7 Conformational analysis

Detailed analysis of the helical parameters and backbone and sugar dihedral angles in the final NMR structure was carried out using the 3DNA program. The helical parameters given by 3DNA are split into three sets of values; local base pair parameters, local base pair step parameters and local base pair helical parameters.

Local base pair parameters which describe the spatial arrangement of bases within a base pair and include shear (S_x), stretch (S_y), stagger (S_z), buckle (κ), propeller twist (ω) and opening (σ). Local base pair step parameters describe the relationship between base pairs which are adjacent to each other in space and include shift (D_x), slide (D_y), rise (D_z), tilt ($T\tau$), roll ($P\rho$) and helical twist (Ω). Local base pair helical parameters describes the position of successive base pairs in relation to the local helical axis and include x-displacement (d_x), y-displacement (d_y), inclination (η) and Tip (θ).

The data for local base pair parameters (Table 3.4) confirm that good base pair planarity is observed in the average structure NMR structure for 17mer mismatch GT DNA. The sheared appearance of the GT mispair in the NMR structure is confirmed here with a shear value of -1.35 which is much higher than for all other base pairs. Its values for stretch and stagger parameters confirms that its introduction causes a distortion to the helix conformation. Large buckle (20.82) and propeller twist (-13.49) values to the A8-T27 base pair suggest a widening of the DNA helix which indicates a possible interaction site during mismatch recognition with MutS protein. Another interesting observation was the large change to buckle, opening and

propeller twist values to AT base pairs within the helix. This gives evidence of their lower stability hence their conformations are more easily distorted.

Local base pair step parameters given in Table 3.5 shows good consistency with all base pairs giving values within a very narrow range. The step between A8-T27 and G9-T26 indicates a bigger change in conformation compared to the rest of the helix with values of -1.46 and -7.22 for slide and tilt parameters respectively. A helical twist value of 31.86 also indicates a change in the sugar-phosphate backbone conformation as this is lower than the expected value of 35.6° in canonical DNA.

Table 3.6 illustrating local base pair helical parameters show large inclination values for a number of base pairs signifying that they are less perpendicular to the sugar-phosphate backbone. With regards to the GT mispair and its flanking bases, the inclination value is large (11.09) for the step after the mismatch (step 10) and with respect to the step before the mismatch (step 8), the tip value is high. This gives further evidence of the local conformational distortion caused by the substitution of the GT mispair.

Base Pair	Nucleotides	Shear (S_x)	Stretch (S_y)	Stagger (S_z)	Buckle (κ)	Propeller Twist (ω)	Opening (σ)
1	A-T	-0.06	-0.37	1.51	13.51	0.08	0.36
2	G-C	-0.06	-0.50	-0.93	12.62	-0.33	-6.07
3	C-G	-0.06	-0.35	0.82	-15.77	-2.29	-4.92
4	T-A	-0.31	-0.30	1.26	-9.69	-11.29	-4.08
5	G-C	0.21	-0.13	0.68	12.01	-4.61	-2.00
6	C-G	-0.02	-0.18	1.23	-11.12	0.61	-1.13
7	C-G	-0.04	-0.43	0.77	-2.92	-5.55	-4.97
8	A-T	0.27	-0.05	0.42	20.82	-13.49	3.28
9	G-T	-1.35	-0.52	1.91	5.47	-8.94	4.84
10	G-C	0.10	-0.31	1.20	3.35	-0.59	1.35
11	C-G	0.13	-0.22	0.92	-14.52	-3.86	1.33
12	A-T	-0.02	-0.02	-0.11	4.41	-6.25	1.90
13	C-G	-0.33	-0.28	-0.56	1.44	7.19	-7.12
14	C-G	-0.27	-0.26	0.98	-8.23	-2.67	-3.02
15	A-T	0.04	-0.24	0.28	17.91	-9.06	-3.29
16	G-C	0.04	-0.49	-0.86	3.01	2.37	-3.61
17	T-A	0.14	-0.01	-0.54	-17.35	-8.14	0.33
Average		-0.09	-0.27	0.53	0.88	-3.93	-1.58

Table 3.4. Helical parameters describing local base pair parameters in the 17mer mismatch GT DNA NMR structure. Parameters were calculated using 3DNA software from the average structure shown in Figure 3.11. Boxes shown in yellow indicate large differences compared to the expected value.

Step No.	Step	X-displacement (d_x)	Y-displacement (d_y)	Inclination (η)	Tip (θ)
1	AG / CT	-1.86	2.47	5.09	-19.03
2	GC / GC	-4.13	-1.32	21.65	10.04
3	CT / AG	-2.21	-0.27	7.56	2.73
4	TG / CA	-1.48	-0.07	2.23	-3.51
5	GC / GC	-4.06	-0.70	23.98	3.98
6	CC / GG	-2.20	1.03	6.44	-0.18
7	CA / TG	-1.24	-0.81	-2.52	-5.02
8	AG / TT	-3.20	-2.32	4.81	13.27
9	GG / CT	-1.15	0.54	-2.64	-2.90
10	GC / GC	-2.98	0.29	11.09	-2.71
11	CA / TG	-1.98	0.48	4.22	-7.44
12	AC / GT	-3.02	1.06	9.13	2.56
13	CC / GG	-3.53	-1.21	16.17	8.85
14	CA / TG	-1.65	0.41	4.88	-5.52
15	AG / CT	-2.83	0.51	10.26	-5.83
16	GT / AC	-2.82	-0.20	10.43	-2.57
Average		-2.52	-0.01	8.30	-0.83

Table 3.5. Helical parameters describing local base pair helical parameters in the 17mer mismatch GT DNA NMR structure. Parameters were calculated using 3DNA software from the average structure shown in Figure 3.11. Boxes shown in yellow indicate large differences compared to the expected value.

Step No.	Step	Shift (D_x)	Slide (D_y)	Rise (D_z)	Tilt (Tτ)	Roll (Pρ)	Helical Twist (Ω)
1	AG / CT	-0.42	-0.85	3.63	11.50	3.08	35.84
2	GC / GC	0.16	-1.17	4.48	-6.29	13.56	37.51
3	CT / AG	0.03	-0.96	2.89	-1.62	4.49	34.66
4	TG / CA	0.18	-0.82	2.38	2.16	1.38	35.91
5	GC / GC	0.21	-1.16	3.98	-2.49	15.01	37.53
6	CC / GG	-0.61	-1.01	2.86	0.11	3.82	34.54
7	CA / TG	0.69	-0.85	2.34	3.04	-1.53	35.33
8	AG / TT	0.45	-1.46	3.82	-7.22	2.61	31.86
9	GG / CT	-0.19	-0.92	3.60	1.93	-1.76	38.92
10	GC / GC	-0.01	-1.16	3.91	1.68	6.86	36.28
11	CA / TG	0.00	-1.01	2.37	4.41	2.50	34.61
12	AC / GT	-0.75	-1.26	3.21	-1.47	5.24	33.48
13	CC / GG	0.22	-1.20	3.72	-5.28	9.65	35.29
14	CA / TG	-0.05	-0.81	2.15	3.29	2.91	34.77
15	AG / CT	0.07	-1.05	4.05	3.53	6.21	35.47
16	GT / AC	0.30	-1.07	4.14	1.60	6.48	36.42
Average		0.02	-1.05	3.35	0.56	5.03	35.53

Table 3.6. Helical parameters describing local base pair step parameters in 17mer mismatch GT DNA NMR structure. Parameters were calculated using 3DNA software from the average structure shown in Figure 3.11. Boxes shown in yellow indicate large differences compared to the expected value.

Base pair hydrogen bond lengths (shown in appendix) show that the GT mispair has increased those in the flanking base pairs which accounts for the widening of the helix at the mismatch site which is observed in the NMR structure.

Backbone torsional angles show very little deviation from the input values with all values being within ±25% of the error bound for all residues with the exception of the glycosidic angle, χ. There is greater variation in χ values, in particular for bases flanking the GT mispair. The majority of residues adopt a $C_{2'}$-*endo* sugar pucker with six residues adopting a $C_{1'}$-*exo* sugar pucker. This is very close to $C_{2'}$-*endo* and so the structure itself appears as a B-form DNA molecule. A comparison of the average values for the backbone and sugar ring dihedral angles between the Dickerson 12mer DNA, 17mer canonical GC DNA (analysis presented in section 3.3) and the 17mer mismatch GT DNA are given in Tables 3.7 and 3.8.

The average values for the backbone and sugar ring dihedral are comparable between the two 17mer DNAs. What is surprising is that the values for the 17mer mismatch GT DNA are in fact closer to the ones obtained from the Dickerson 12mer DNA despite not all nucleotides being $C_{2'}$-*endo* in conformation. This further highlights the quality of the NMR data being collected at 800 MHz and not 600 MHz.

Dihedral Angle	Average Value		
	Dickerson 12mer DNA	**17mer canonical GC DNA**	**17mer mismatch GT DNA**
α	-62.8	-64.2	-73.7
β	-169.3	-169.1	-171.2
γ	59.5	48.3	57.7
δ	122.7	137.6	134.4
ε	164.2	178.0	153.8
ζ	-105.0	-99.8	-99.4
χ	-117.3	-110.2	-115.3

Table 3.7. Comparison of backbone dihedral angles between Dickerson DNA (PDB: 1BNA), 17mer canonical GC DNA and 17mer mismatch GT DNA. Values were calculated using 3DNA and for 17mer canonical GC DNA are based on the average structure shown in section 3.3.5. For 17mer mismatch GT DNA, the values are based on the average structure given in Figure 3.11.

Dihedral Angle	Average Value		
	Dickerson 12mer DNA	**17mer canonical GC DNA**	**17mer mismatch GT DNA**
ν_0	-38.7	-22.4	-29.4
ν_1	40.8	32.3	37.0
ν_2	-30.5	-29.6	-29.8
ν_3	17.3	17.6	13.5
ν_4	21.7	2.8	9.9
Amplitude	46.4	32.3	36.2
Phase	127.6	156.5	145.6

Table 3.8. Comparison of sugar ring dihedral angles between Dickerson DNA (PDB: 1BNA), 17mer canonical GC DNA and 17mer mismatch GT DNA. Values were calculated using 3DNA and for 17mer canonical GC DNA are based on the average structure shown in section 3.3.5. For 17mer mismatch GT DNA, the values are based on the average structure given in Figure 3.11.

3.1.8 Comparison of 800 MHz and 1000 MHz data for 17mer mismatch GT DNA

Recently, NMR data of 17mer mismatch GT DNA was acquired on the Bruker 1000 MHz spectrometer based at the Centre de RMN à Très Hauts Champs in Lyon, France. This section will highlight the benefits of increasing the magnetic field strength and compare the NMR structures produced from data collected at each one.

Increasing the spectrometer field strength has two main advantages; a gain in sensitivity and increased spectral dispersion. Figure 3.13 shows the H6/8 – H3' region of the NOESY spectrum measured in $^{2}H_2O$. As explained in section 3.1, H6–H3' NOEs for pyrimidine residues were difficult to observe at 800 MHz. It was thought that this was due to the NOE distance being greater than 5Å. However, for the same sample at 1000 MHz, these appear as intense NOEs so the sensitivity gain was clearly shown which was accompanied by an increase in spectral dispersion.

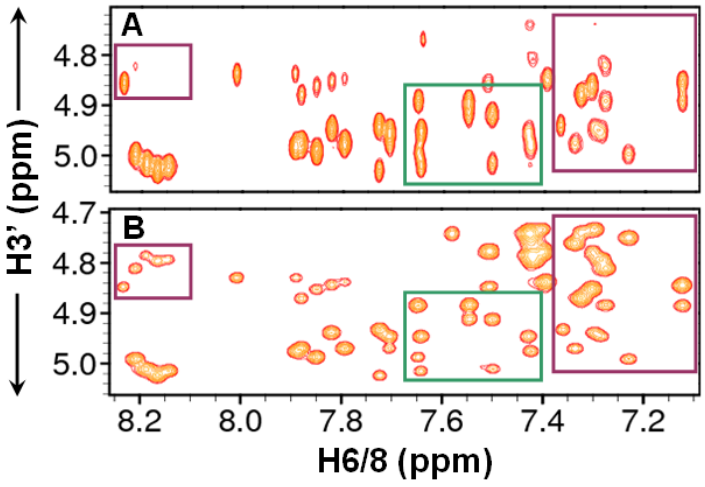


Figure 3.13. Comparison of NOESY spectra recorded at 800 MHz (A) and 1000 MHz (B). Both spectra show the H6/8-H3' NOE connectivity region. The regions highlighted in purple show gain in sensitivity and those in green show the increase in resolution as the peaks are clearly more separated.

Figure 3.14 shows the cytosine H5-H6 region in the NOESY spectrum of 17mer mismatch GT DNA measured in $^{2}H_2O$ at the two field strengths. The overlap of C7, C11, C14 and C30 H5-H6 NOEs was resolved at 1000 MHz. The intraresidue aromatic H6-H1' NOE for C14 also appeared clearly separate from the C3 and C33 H5-H6 NOEs.

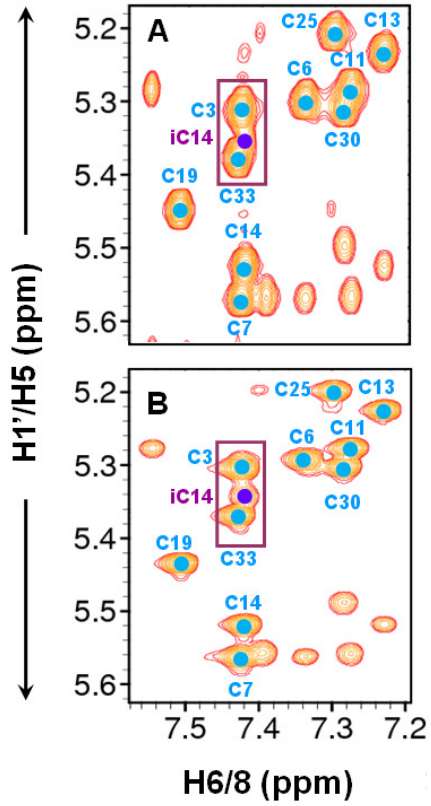


Figure 3.14. Comparison of NOESY spectra recorded at 800 MHz (A) and 1000 MHz (B). The region highlighted shows the cytosine H5-H6 NOE connectivities which are marked in blue. The intraresidue H6-H1' NOE connectivity for C14 is also highlighted in purple.

The increased dispersion and appearance of new peaks at 1000 MHz resulted in more constraints for structure calculation (510). Peaks observed at 1000 MHz also show less overlap so the NOE distances generated would be more precise, these are summarised in Table 3.9.

Constraint Type	No. of constraints
Total NOEs	510
Strong (1.5 - 2.5 Å)	24
Medium (2.6 - 4.0 Å)	452
Weak (4.1 - 6.0 Å)	34
Hydrogen bond	98
Total dihedral angle	301
Helix (α, β, γ, δ, ε, ζ)	199
Glycosidic (χ)	34
Ribose pucker (υ_1, υ_2)	68
NOE constraints per residue	15.0
NOE and dihedral constraints per residue	23.9

Table 3.9. Summary of experimental constraints used in the structure calculation of 17mer mismatch GT DNA carried out with the Xplor-NIH program using NMR data collected at 1000 MHz.

Figure 3.15 compares the average structures for the 17mer mismatch GT DNA using data at the two field strengths. Both structures look similar with the GT mispair adopting a sheared orientation. The inclination of some base pairs appears greater in the 1000 MHz NMR structure and some base pairs appear to be more buckled, which has consequently caused the propeller twist parameter to increase.

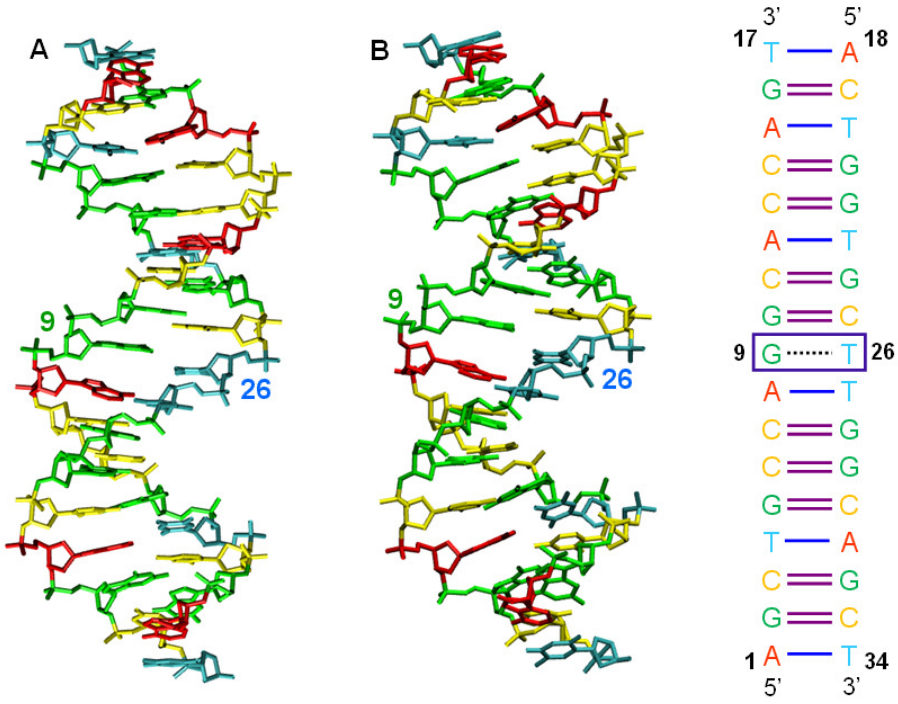


Figure 3.15. Comparison of the 800 MHz (A) and 1000 MHz (B) average NMR structures generated by Xplor-NIH of mismatch 17mered GT DNA. The corresponding sequence is shown on the right and the position of the G9-T26 mismatch base pair is highlighted by the purple box.

Conformational analysis comparison

Although the NMR structure obtained from the 1000 MHz data was not refined to the same level as the 800 MHz structure, the use of better quality distance constraints meant that the values for each parameter remained more consistent with fewer fluctuations. Local base pair parameters (Table 3.10), local base pair helical parameters (Table 3.11) and local base pair step parameters (Table 3.12) are shown in the comparison below.

Helical Parameter	Average Values	
	800 MHz	**1000 MHz**
Shear (S_x)	-0.09	0.09
Stretch (S_y)	-0.27	-0.31
Stagger (S_z)	0.53	0.52
Buckle (κ)	0.88	1.55
Propeller Twist (ω)	-3.93	-10.24
Opening (σ)	-1.58	-2.86

Table 3.10. Comparison of average values for local base pair parameters in 17mer mismatch GT DNA NMR structures obtained from data at 800 MHz and 1000 MHz using 3DNA.

Helical Parameter	Average Values	
	800 MHz	**1000 MHz**
x-displacement (d_x)	-2.52	-2.30
y-displacement (d_y)	-0.01	-0.01
Inclination (η)	8.30	14.04
Tip (θ)	-0.83	0.09

Table 3.11. Comparison of average values for local base pair helical parameters in 17mer mismatch GT DNA NMR structures obtained from data at 800 MHz and 1000 MHz using 3DNA.

Helical Parameter	Average Values	
	800 MHz	**1000 MHz**
Shift (D_x)	0.02	0.00
Slide (D_y)	-1.05	-0.71
Rise (D_z)	3.35	3.21
Tilt ($T\tau$)	0.56	0.01
Roll ($P\rho$)	5.03	8.65
Helical twist (Ω)	35.53	35.08

Table 3.12. Comparison of average values for local base pair step parameters in 17mer mismatch GT DNA NMR structures obtained from data at 800 MHz and 1000 MHz using 3DNA.

The average buckle and inclination values were twice as high at 1000 MHz, which confirms what was seen in the average NMR structure. The propeller twist and opening values were also higher giving evidence that the base pairs were not as planar at 1000 MHz. The average value for slide and roll values also decreased and increased by 30% respectively.

In the 1000 MHz NMR structure, more nucleotides adopted a $C_{1'}$-*exo* sugar pucker giving further evidence that the 800 MHz structure has been more refined. Tables 3.13 and 3.14 compare the backbone and sugar ring dihedral angles respectively.

Dihedral Angle	Average Values	
	800 MHz	**1000 MHz**
α	-73.7	-74.2
β	-171.2	-162.5
γ	57.7	57.2
δ	134.4	134.5
ε	176.3	178.1
ζ	-99.4	-92.7
χ	-115.3	-111.2

Table 3.13. Comparison of average values for backbone dihedral angles in 17mer mismatch GT DNA NMR structures, obtained from data at 800 MHz and 1000 MHz, using 3DNA.

Dihedral Angle	Average Values	
	800 MHz	**1000 MHz**
ν_0	-29.4	-28.8
ν_1	37.0	36.4
ν_2	-29.8	-29.7
ν_3	13.5	13.9
ν_4	9.9	18.3
Amplitude	36.2	35.8
Phase	145.6	146.9

Table 3.14. Comparison of average values for sugar ring dihedral angles in 17mer mismatch GT DNA NMR structures, obtained from data at 800 MHz and 1000 MHz, using 3DNA.

With the exception of β, ζ, and υ_4, all the dihedral angles had similar average values between the 800 MHz and 1000 MHz NMR structures. υ_4 is defined by C3'–C4'–O4'-C1' and so lies along the backbone of the DNA. Consequently, having a larger average value will directly impact the local base pair helical and step parameters particularly inclination.

3.2 NMR structure determination and conformational analysis of 17mer canonical GC DNA

The NMR structure determination of 17mer canonical GC DNA is presented in this section which acts as a reference structure to compare with the NMR structure of 17mer mismatch GT DNA. The approach used was identical to section 3.1 and hence, results are shown in brief in this chapter and figures which are not shown are given in the appendix.

3.2.1 Identification and assignment of exchangeable proton resonances

The imino protons of 17mer canonical GC DNA were all observed between 12-15 ppm in the 1D ^{1}H-NMR spectra recorded in $^{1}H_2O$ (Figure 3.16). Sixteen residues were assigned through imino-imino proton sequential NOEs starting at T27 (δ=13.89); T17 was assigned from an NOE to its base paired NH_2(1) proton. Figure 3.17 shows the sequential connectivity from T27 to G16.

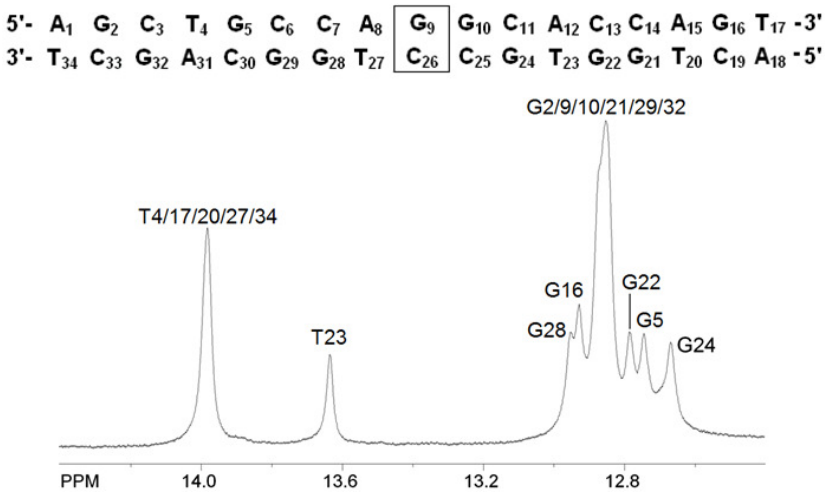


Figure 3.16. 800 MHz 1D ^{1}H-NMR spectrum showing imino proton region in 17mer canonical GC DNA (0.8mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C. The assignment shown is based on their identification in the NOESY spectrum (Figure 3.17).

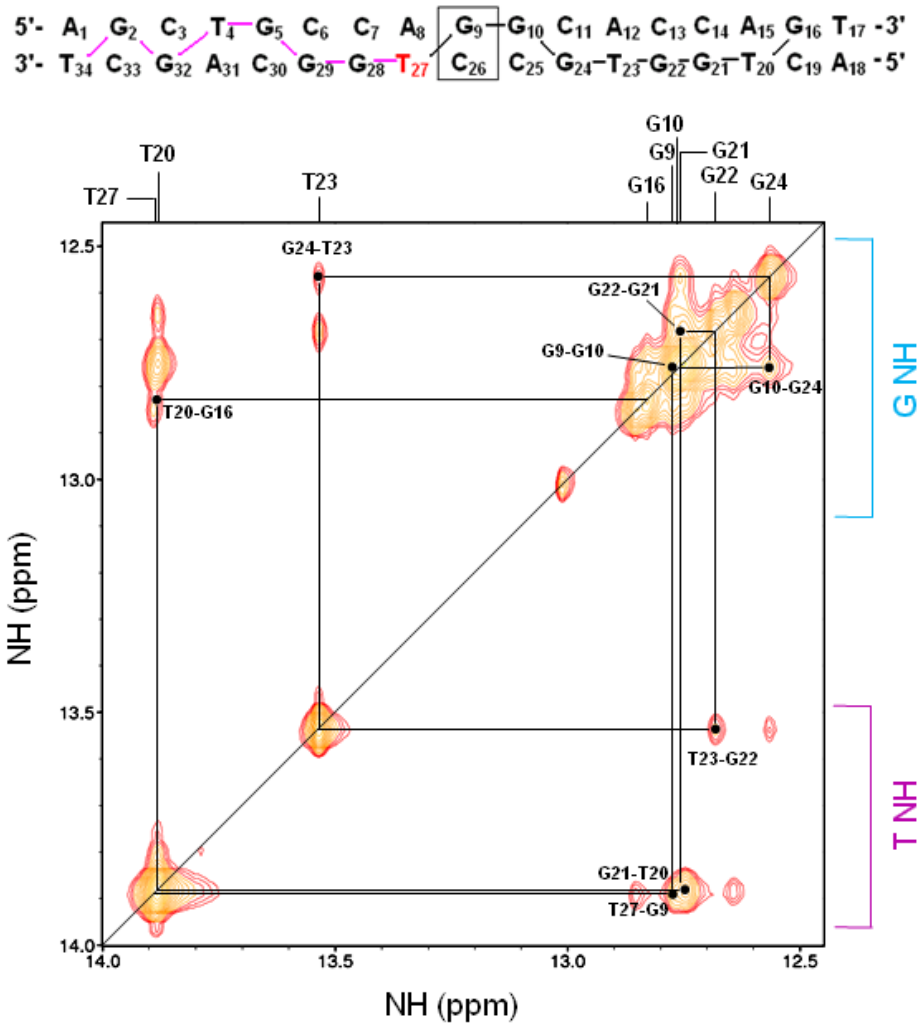


Figure 3.17. 700 MHz NOESY (τ_m = 250ms) spectrum of 17mer canonical GC DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C. The sequential connectivity stretching from T27 to G16 is indicated by the black line. Unassigned peaks in the region correspond to the sequential connectivity stretching from T27 to T34. These resonances have been assigned and the equivalent imino-imino proton sequential connectivity is shown in the Appendix.

All amino proton resonances were identified and assigned for adenine, cytosine and guanine based on their NOEs to their base paired imino protons and cross-strand NOEs to neighbouring nucleotides (shown in Appendix).

3.2.2 Identification and assignment of non-exchangeable proton and carbon resonances

All six adenine H2 resonances were assigned as described in section 3.2.1 through NOEs observed in the NOESY spectrum measured in $^{1}H_2O$ (figure in Appendix). Eleven cytosine H5-H6 and six thymine CH_3-H6 correlations were expected for the 17mer canonical GC DNA and all were observed in the DQF-COSY and TOCSY spectra with the corresponding C5 and C6 resonances being assigned in the ^{1}H-^{13}C HSQC spectrum.

Assignment of aromatic H6/8 – sugar H1' sequential connectivity

The H6/8-H1' sequential connectivity was identified for both strands in the NOESY spectrum measured in $^{2}H_2O$ at 25°C, confirming the sequence of the DNA. The sequential connectivity for A18-T34 is shown in Figure 3.18; all H1', H6 and H8 resonances were assigned. C6, C8 and C1' resonances were assigned by correlation with the ^{1}H-^{13}C HSQC spectrum.

Assignment of H2' and H2'' resonances

All H2' and H2'' protons were assigned in the DQF-COSY and TOCSY spectra. H6/8 – H2'/2'' sequential connectivities were identified in the NOESY spectrum measured at 25°C which provided important distance constraints for structure calculation. All C2' resonances were assigned despite the C2' region being overlapped in the ^{1}H-^{13}C HSQC.

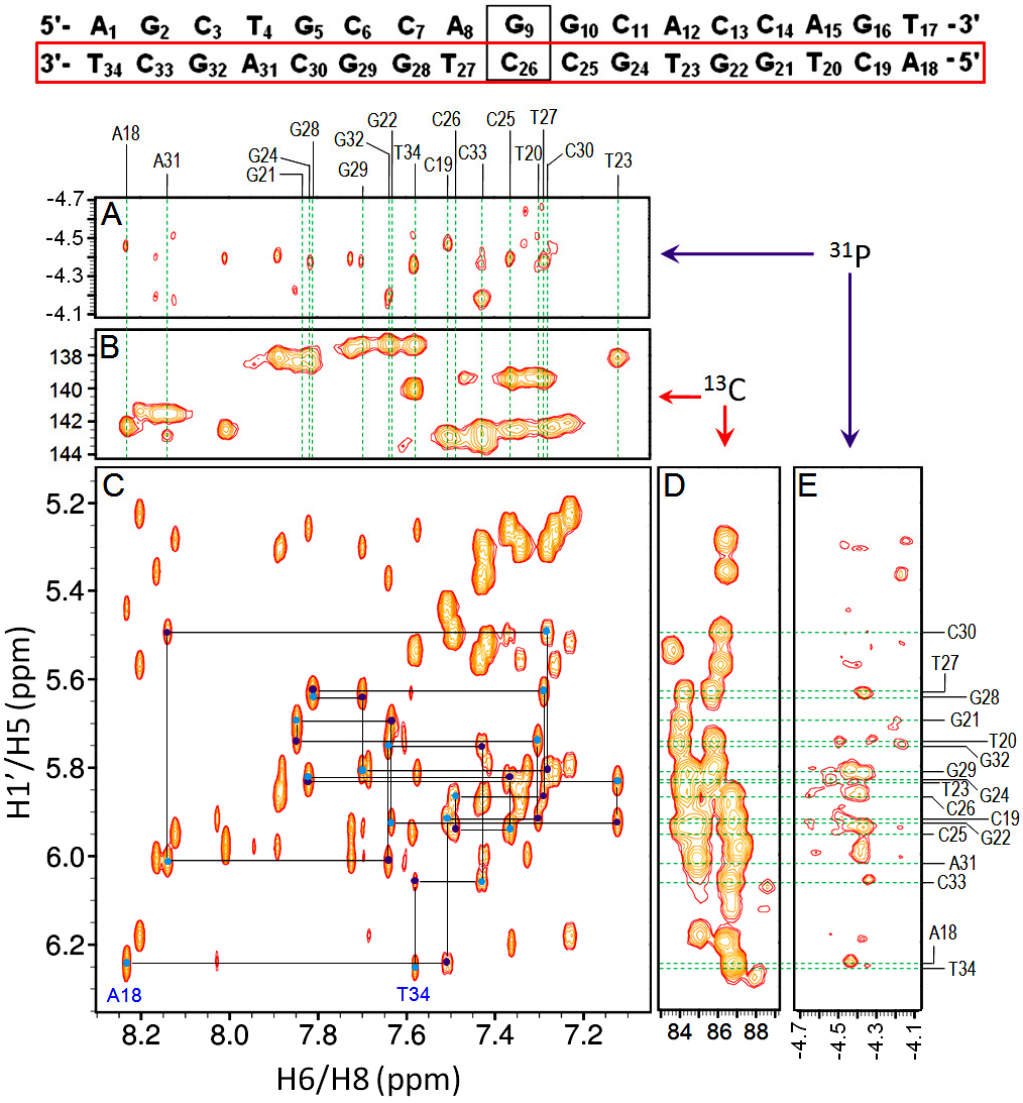


Figure 3.18. 600 MHz NOESY (τ_m = 250ms) spectrum (C) showing H6/8-H1' sequential NOE connectivities for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide connectivities are shown by blue circles and internucleotide connectivities by purple circles. 600 MHz ^1H-^{13}C HSQC spectra show the corresponding assignment for C6/C8 (B) and C1' (D) resonances. 600 MHz ^{31}P CPMG-HSQC-NOESY (τ_m = 500ms) spectra show the corresponding assignment for backbone ^{31}P resonances through H6/8-^{31}P (A) and H1'-^{31}P (E) NOE connectivities. The unassigned NOEs correspond to the H6/8-H1' sequential connectivity stretching between A1 to T17, which has been assigned and is shown in the Appendix.

Assignment of H3', H4' and H5'/5'' resonances

All H3' and H4' protons were assigned using H1'-H3'/H4' TOCSY peaks and the H6/8-H3'/H4' sequential connectivity for each strand in the NOESY spectrum. C3' (70-90 ppm) and C4' (85-95 ppm) were identified by correlation to the ^{1}H-^{13}C HSQC spectrum.

The above methodology was applied for H5'/5'' protons which were distinguished from each other on the basis of the intensity of their NOEs to H1'; H1'-H5'' cross peaks appeared more intense due to a shorter NOE distance. Not all H5'/5'' protons were assigned resonances and these are highlighted in section 3.2.4.

The C5' region in the ^{1}H-^{13}C HSQC spectrum were heavily overlapped with the exception of the 5' terminal residues (A1 and A18) which appeared separate from the other resonances. Consequently, with the exception of A1 and A18, all other C5' resonances were recorded as having a chemical shift of 66.2-69.0 ppm.

3.2.3 Assignment of phosphorus resonances

The ^{31}P resonances were assigned based on cross peaks observed in the ^{1}H-^{31}P CPMG-HSQC-NOESY spectrum based on H3'-^{31}P cross peaks. Using this approach all ^{31}P resonances were assigned with the exception of those for C6, C7, G10 and C11 residues. ^{31}P chemical shifts for residues C7 and G10 were assigned by identifying the NOEs to aromatic H6/8 and sugar H1' protons. Cross peaks were not observed for C6 in all regions of the ^{1}H-^{31}P CPMG-HSQC-NOESY spectrum

thus its ^{31}P resonance could not be assigned. The assignment of ^{31}P resonances using H3'-^{31}P cross peaks for residues A1-T17 is shown in Figure 3.19.

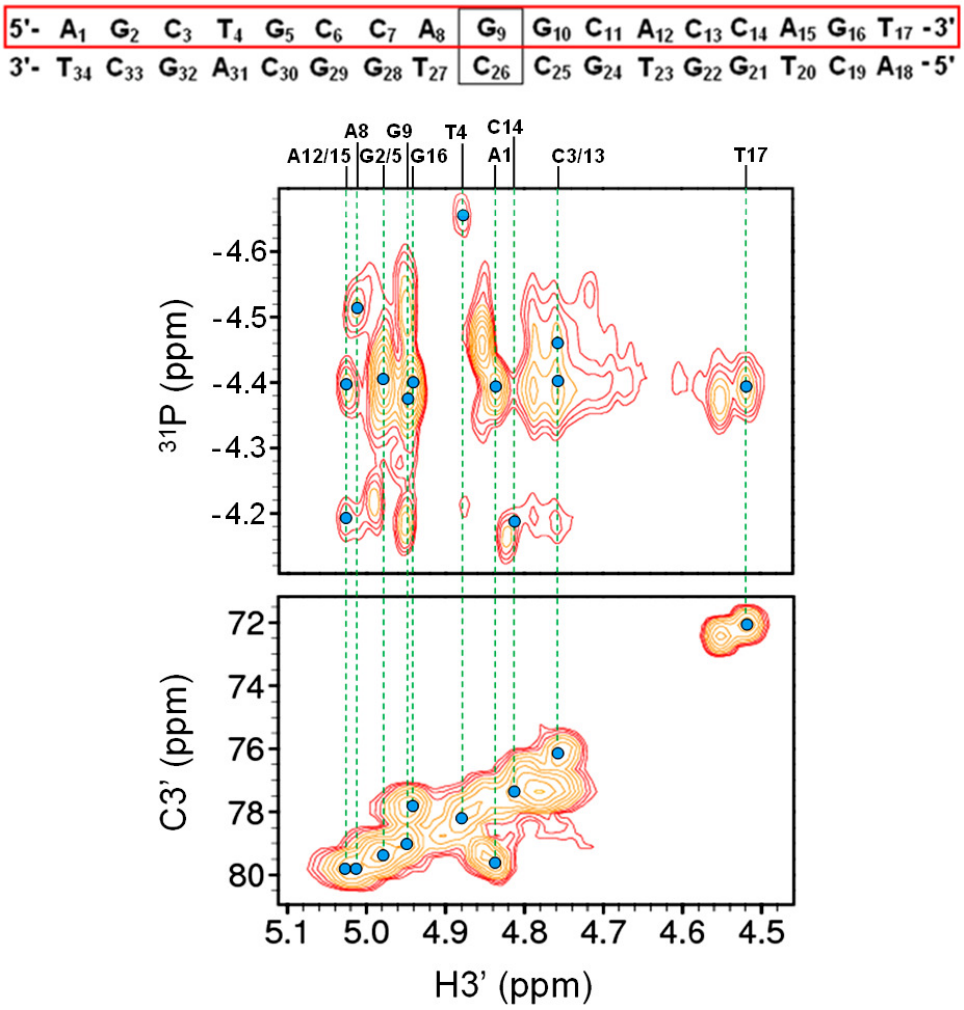


Figure 3.19. (Top panel) 600 MHz ^{1}H-^{31}P CPMG-HSQC-NOESY (τ_{mix} = 500ms) spectrum and (bottom panel) 600 MHz ^{1}H-^{13}C HSQC spectrum. The assignment of ^{31}P resonances is shown by means of H3'-^{31}P HSQC correlations for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) measured at 25°C. Unassigned peaks correspond to NOESY and HSQC peaks for residues A18-T34. These have been assigned and is shown in the Appendix.

3.2.4 Table of assignments

Based on the identification and assignment procedures described previously in the chapter, the chemical shifts of exchangeable and non-exchangeable proton, carbon and phosphorus resonances following assignment are summarised in Table 3.15 and 3.16.

Nucleotide Number	Imino H1	Amino $NH_2(1)$	Amino $NH_2(2)$	H2	$CH_3(5)$	H5	H6	H8	H1'	H2'	H2"	H3'	H4'	H5'	H5"
A1		8.14	6.24	8.03				8.01	5.98	2.46	2.63	4.84	4.21	3.69	3.69
G2	12.83	6.84						7.89	5.87	2.70	2.70	4.98	4.41	4.26	4.11
C3		8.14	5.99			5.31	7.43		6.00	2.06	2.52	4.75	4.27	4.10	4.18
T4	13.97				1.60		7.33		5.81	2.11	2.49	4.88	4.15		4.09
G5	12.83	6.40						7.88	5.85	2.64	2.68	4.98	4.38	4.08	4.13
C6		8.14	5.83			5.30	7.34		5.89	2.02	2.39	4.77	4.19	4.01	4.12
C7		8.53				5.56	7.43		5.29	1.99	2.27	4.79	4.03	3.97	
A8		7.91	6.38	7.58				8.13	5.95	2.69	2.86	5.01	4.35	4.14	3.96
G9	12.85	6.39						7.59	5.54	2.50	2.63	4.95	4.33		4.15
G10	12.84	6.39						7.58	5.81	2.46	2.66	4.91	4.34		4.15
C11		8.24	5.79			5.27	7.28		5.57	1.97	2.36	4.80	4.12		4.04
A12		7.91	6.07	7.66				8.21	6.19	2.67	2.87	5.00	4.40	4.25	4.13
C13		8.07	5.80			5.23	7.23		5.79	1.96	2.36	4.75	4.14	4.26	4.04
C14		8.51	5.89			5.52	7.42		5.35	1.98	2.29	4.81	4.06	3.99	
A15		7.92	6.35	7.66				8.17	6.01	2.71	2.85	5.03	4.38	3.98	4.11
G16	12.91	6.67	6.24					7.73	5.94	2.53	2.66	4.94	4.37	4.07	4.22
T17	13.97				1.55		7.37		6.20	2.21	2.21	4.52	4.07		4.23
A18				7.92				8.23	6.25	2.71	2.81	4.85	4.27	3.75	3.75
C19		8.25	5.98			5.44	7.51		5.92	2.12	2.46	4.78	4.24	4.19	4.10
T20	13.96				1.51		7.31		5.74	2.06	2.43	4.86	4.15	4.07	4.09
G21	12.84	6.39						7.85	5.69	2.70	2.74	4.99	4.36	4.20	4.13
G22	12.77	6.51	6.80					7.65	5.93	2.50	2.73	4.89	4.40	4.10	4.22
T23	13.62				1.31		7.13		5.83	2.02	2.47	4.85	4.16	4.22	4.11
G24	12.65	6.41						7.82	5.82	2.59	2.66	4.95	4.34	4.10	4.15
C25		8.07	5.82			5.26	7.37		5.94	2.16	2.47	4.78	4.23		4.13
C26		8.37	5.83			5.50	7.49		5.87	2.03	2.43	4.76	4.12	4.23	4.07
T27	13.97				1.59		7.29		5.63	2.01	2.35	4.82	4.09	4.01	
G28	12.94	6.84						7.82	5.64	2.66	2.74	4.98	4.32	4.15	4.09
G29	12.86	6.90						7.71	5.80	2.51	2.66	4.95	4.36		4.15
C30		8.29	5.85			5.30	7.29		5.50	1.95	2.31	4.81	4.10	3.99	
A31		7.90	6.25	7.58				8.14	6.02	2.71	2.87	5.02	4.37	4.19	4.11
G32	12.83	6.56						7.64	5.75	2.48	2.62	4.95	4.35		4.19
C33		8.25	5.69			5.38	7.43		6.06	2.20	2.46	4.75	4.21	4.06	4.13
T34	13.36				1.71		7.59		6.26	2.29	2.29	4.55	4.08		

Table 3.15. Summary of chemical shifts for ^{1}H proton resonances in 17mer canonical GC DNA. Boxes shaded in grey indicate that the absence of the proton in the specified residue and those shaded in blue indicate unassigned protons.

Nucleotide Number	C2	C5	C6	C8	C1'	C2'	C3'	C4'	C5'	^{31}P
A1	154.7			142.4	84.9	39.6	79.7	89.6	64.4	-4.41
G2				138.0	84.4	40.2	79.3	87.4		-4.43
C3		98.5	143.3		87.2	40.0	76.1	85.8		-4.43
T4		14.7	139.2		85.5	39.2	78.1	85.8		-4.68
G5				138.0	84.4	40.4	79.3	87.3		-4.43
C6		98.3	142.4		86.5	40.0	77.4	85.7		
C7		98.5	142.5		86.5	39.4	76.3	85.80		-4.18
A8	154.2			141.5	84.3	40.3	79.8	86.5		-4.54
G9				137.3	83.7	40.8	79.1	86.9		-4.38
G10				137.3	84.4	39.6	78.7	86.9		-4.38
C11		98.1	142.4		86.1	39.8	76.7	85.5		
A12	154.4			141.4	85.1	40.4	79.8	87.4		-4.42
C13		98.2	142.1		85.8	39.8	76.1	85.8		-4.48
C14		98.5	142.5		86.5	39.8	77.4	85.80		-4.21
A15	154.8			141.5	84.9	40.3	79.8	87.3		-4.22
G16				137.7	84.3	40.8	77.7	87.2		-4.42
T17		14.0	139.4		86.5	41.3	72.1	87.2		-4.41
A18	154.8			142.4	86.8	40.2	79.5	89.3	64.3	-4.48
C19		98.5	142.8		86.9	39.2	77.4	85.8		-4.49
T20		14.7	139.3		85.8	39.0	77.9	85.8		-4.36
G21				138.4	84.0	40.2	79.3	86.9		-4.24
G22				137.2	85.1	41.2	78.4	87.4		-4.50
T23		14.3	138.2		85.4	39.7	77.5	85.8		-4.58
G24				138.2	84.4	40.8	79.1	86.9		-4.38
C25		98.1	142.5		86.7	40.2	77.4	85.8		-4.41
C26		98.7	142.8		86.7	39.9	77.4	85.6		-4.48
T27		14.7	139.3		85.8	39.7	77.5	85.8		-4.41
G28				138.2	84.2	40.1	79.3	86.9		-4.43
G29				137.2	84.4	40.8	79.1	87.0		-4.40
C30		98.2	142.4		86.2	39.7	77.4	85.5		-4.40
A31	154.2			141.5	84.9	40.4	79.8	87.2		-4.42
G32				137.2	84.0	40.8	77.7	86.9		-4.21
C33		98.4	142.5		86.7	40.5	76.1	85.5		-4.38
T34		14.4	140.0		86.8	41.5	72.5	87.2		-4.40

Table 3.16. Summary of chemical shifts for ^{13}C and ^{31}P resonances in 17mer canonical GC DNA. Boxes shaded in grey indicate that the absence of the resonance in the specified residue and those shaded in blue indicate unassigned resonances. Boxes shaded in red indicate that the chemical shifts of C5' resonances are highly overlapped. The indicated C5' resonances all appear between 66.2-69.0 ppm.

3.2.5 Structure calculation

NOE distance constraints were determined from the NOESY spectra carried out in $^{1}H_2O$ and $^{2}H_2O$ whilst dihedral angles were given standard values based on the sugar pucker calculated from $^{3}J_{H1'\text{-}H2'}$ coupling constants in the DQF-COSY. Experimental constraints used for structure calculations using Xplor-NIH are given in Table 3.17.

Constraint Type	No. of constraints
Total NOEs	486
Strong (1.5 - 2.5 Å)	48
Medium (2.6 - 4.0 Å)	382
Weak (4.1 - 6.0 Å)	56
Hydrogen bond	102
Total dihedral angle	306
Helix (α, β, γ, δ, ε, ζ)	204
Glycosidic (χ)	34
Ribose pucker (υ_1, υ_2)	68
NOE constraints per residue	14.3
NOE and dihedral constraints per residue	23.3

Table 3.17. Summary of experimental constraints used in the structure calculation of 17mer canonical GC DNA using Xplor-NIH program.

From the 100 structures generated by Xplor-NIH, ten were selected based on their lowest energy and lowest RMSD [Figure 3.22, A, (RMSD = 0.50)] from which the average structure was generated (Figure 3.22, B). Good planarity of base pairs can be clearly identified and a canonical B-DNA type conformation is adopted.

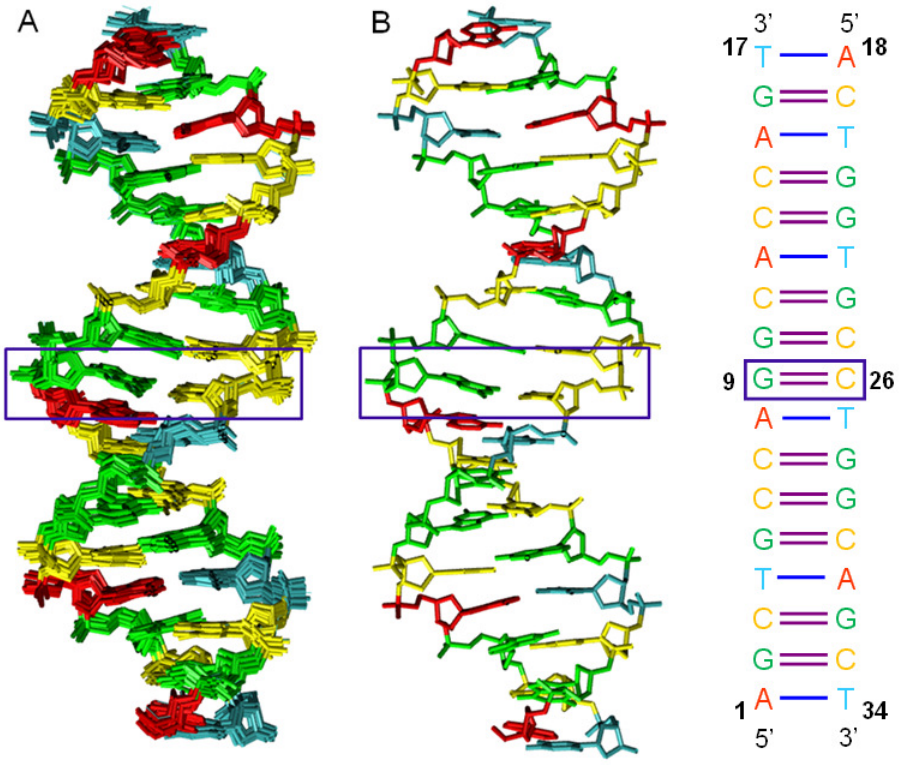


Figure 3.20. Illustration showing an overlay of the ten lowest energy and lowest RMSD NMR structures (A) and the final average structure (B) for 17mer canonical GC DNA calculated from Xplor-NIH. The sequence is shown on the right and the position of the G9-C26 base pair is outlined by the purple box.

3.2.6 Conformational analysis

To investigate the quality of the NMR structure of 17mer canonical GC DNA, the helical parameters and dihedral angle values obtained for the NMR structure were compared with those for the Dickerson DNA dodecamer crystal structure. Both structures were run through 3DNA and the average values are compared in the tables shown below.

Values for the local base pair and base pair step parameters (Tables 3.18 and 3.19) show good similarity to the Dickerson 12mer DNA. Large changes to the buckle, opening and

roll can be attributed to the 17mer canonical GC DNA being a solution state structure. In principle, the structure will be more susceptible to dynamic movement in solution and thus greater flexibility will occur. The large differences for the x-displacement and inclination values (Table 3.20) could be brought about by changes in the backbone dihedral angles, which are shown in Tables 3.7 and 3.8.

Helical Parameter	Average Value	
	Dickerson 12mer DNA	17mer canonical GC DNA
Shear (S_x)	-0.04	-0.03
Stretch (S_y)	-0.23	-0.28
Stagger (S_z)	0.14	0.00
Buckle (κ)	0.24	1.27
Propeller twist (ω)	-12.32	0.41
Opening (σ)	-0.55	-1.30

Table 3.18. Comparison of local base pair parameters between Dickerson DNA (PDB: 1BNA) and 17mer canonical GC DNA. Values were calculated using 3DNA and for 17mer canonical GC DNA are based on the average structure shown in Figure 3.20.

Helical Parameter	Average Value	
	Dickerson 12mer DNA	17mer canonical GC DNA
Shift (D_x)	-0.02	-0.02
Slide (D_y)	0.12	-0.63
Rise (D_z)	3.36	3.58
Tilt ($T\tau$)	-0.19	0.03
Roll ($P\rho$)	0.02	5.72
Helical twist (Ω)	35.58	35.04

Table 3.19. Comparison of local base pair step parameters between Dickerson DNA (PDB: 1BNA) and 17mer canonical GC DNA. Values were calculated using 3DNA and for 17mer canonical GC DNA are based on the average structure shown in Figure 3.20.

Helical Parameter	Average Value	
	Dickerson 12mer DNA	17mer canonical GC DNA
x-displacement (d_x)	0.13	-2.04
y-displacement (d_y)	0.02	0.03
Inclination (η)	0.70	9.53
Tip (θ)	-0.06	-0.06

Table 3.20. Comparison of local base pair helical parameters between Dickerson DNA (PDB: 1BNA) and 17mer canonical GC DNA. Values were calculated using 3DNA and for 17mer canonical GC DNA are based on the average structure shown in Figure 3.20.

3.3 Effect of oligonucleotide length on mismatch stability

The stability of the DNA double helix arises from the strength and number of internucleotide interactions throughout the structure; so in principle the longer the DNA molecule, the more stable it should be. It was shown in section 3.1 that the inclusion of a GT mismatch causes helical distortion leading to some degree of instability. This difference in stability cannot be easily quantified by NMR, so determination of the duplex melting temperature using UV absorbance melting was carried out. The change in absorbance to several length analogues (9mer to 17mer) was monitored between 20-90°C and the melting data was used to calculate thermodynamic parameters for each DNA oligonucleotide using the van't Hoff relationship (Chapter 2.5.2).

For each DNA sample, only data points from the primary UV absorbance melting data were used. Values for the experimentally calculated thermodynamic parameters are shown in Table 3.23. Inspection of the derivative plots shows, and keeping in view the method for estimating any errors, a narrowed dispersion around the T_m for the measured data with a symmetrical distribution.

Duplex	A (260)	Conc (μg/ml)	T_m (K)	ΔH (kJ mol^{-1})	ΔS (J K^{-1} mol^{-1})	ΔG (310) (kJ mol^{-1})
9mer GT	0.42	20.80	*	-62.59	227.13	-70.47
11mer GT	0.51	25.43	*	-78.94	271.84	-84.35
13mer GT	0.62	31.24	304.0	-103.85	342.81	-106.37
15mer GT	0.76	38.25	312.0	-105.00	349.55	-108.47
17mer GT	0.49	24.66	318.0	-106.60	381.55	-118.39

Table 3.23. Experimentally calculated thermodynamic parameters for UV absorbance melting experiments investigating the effect of oligonucleotide length on mismatch stability. The T_m values of 9mer and 11mer GT DNAs were unable to be recorded.

As expected, the stability of the mismatch GT DNA increased with increasing oligonucleotide length since more energy was required to disrupt the duplex into single strand oligonucleotides and the calculated ΔG values confirm this. An interesting observation was the small difference in ΔG values between the 13mer to a 15mer sequences compared to other values [Δ(ΔG) = -2.1 kJ mol^{-1}], which was accompanied by the loss of two GC base pairs. This was not expected as GC base pairs tend to increase helix stability compared to AT base pairs. The similarity in ΔG values may be caused by certain sequence orientations which give extra stability associated with factors such as better base stacking, for example, d(GGCC).

The base paired imino protons in the 9-15mer mismatch GT DNAs were identified based on the assignment of the NOESY spectra of 17mer mismatch GT DNA (Figure 3.1). As analogous NOESY experiments were not carried out on the 9-15mer mismatch GT DNAs, the imino protons were assigned on the basis of conserved chemical shifts between the 17mer and its smaller analogues (Figure 3.21).

The most important result was that the chemical shifts of the GT mispair imino protons were conserved irrespective of the length. These resonate at approximately 10.7 ppm and 11.8 ppm for G and T respectively. Figure 3.24 also highlights that the chemical shift of the T residue, which is adjacent to the GT mispair, also remains consistent ($\delta \approx$ 14.30 ppm). The chemical shifts of the imino protons are shown in Table 3.24, which remain consistent for each imino proton with each decrease in oligonucleotide length.

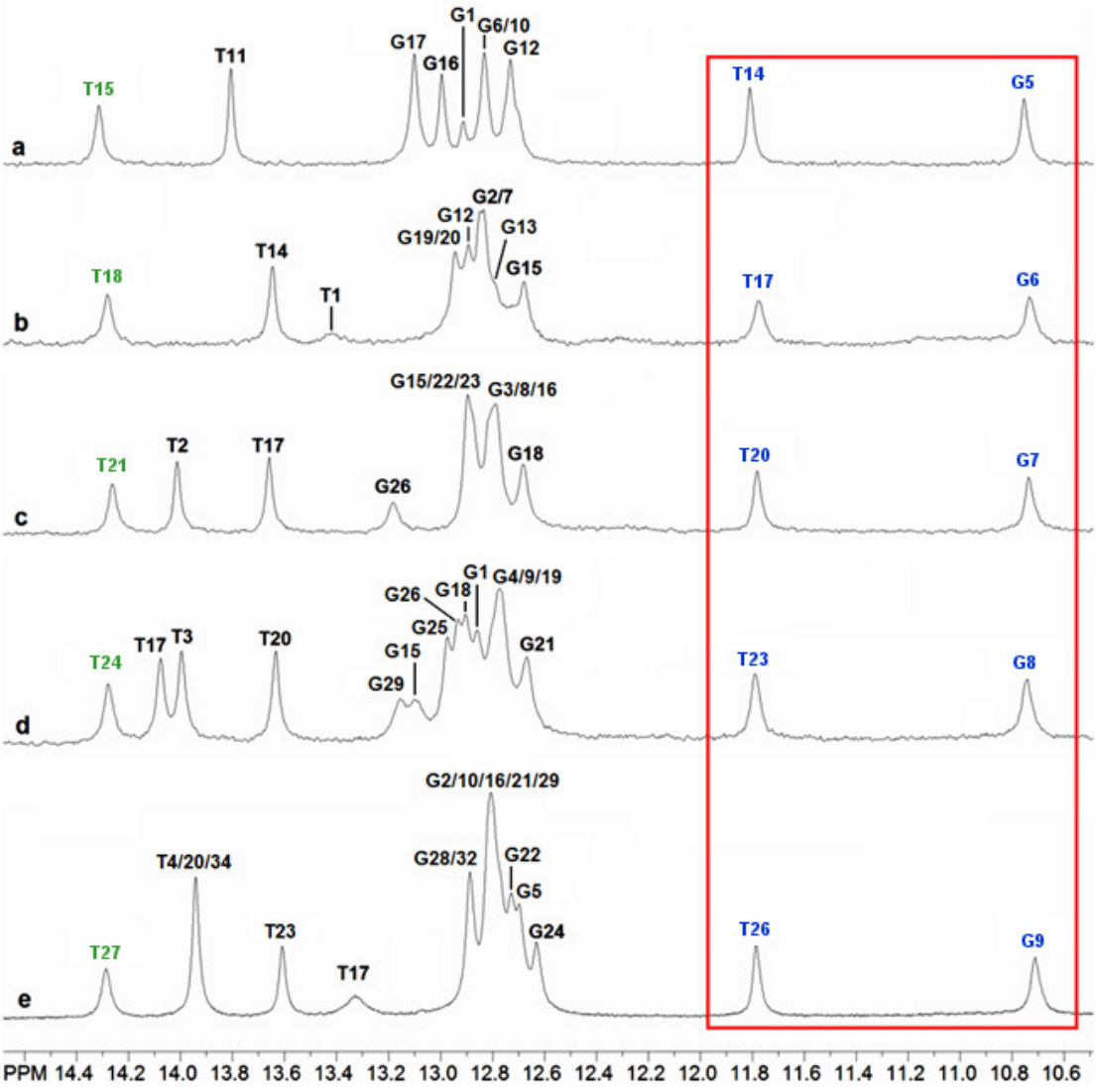


Figure 3.21. 400 MHz, 5°C (a-d) and 800 MHz, 2°C (e) 1D ^{1}H-NMR stack showing imino region of 17mer mismatch GT DNA and sequence length analogues, (a) 9mer GT, (b) 11mer GT, (c) 13mer GT, (d) 15mer GT and (e) 17mer GT. The conserved chemical shifts of the mispair GT imino protons and its immediately flanking T residue are highlighted by the red box and in green text respectively.

Base Pair No.	Base Pair	Imino Proton	Chemical Shift, δ (ppm)				
			17mer GT	15mer GT	13mer GT	11mer GT	9mer GT
1	A-T	T	13.95				
2	G-C	G	12.81	12.86			
3	C-G	G	12.89	13.16	13.21		
4	T-A	T	13.95	14.00	14.03	13.46	
5	G-T	G	12.70	12.78	12.81	12.88	12.94
6	C-G	G	12.81	12.94	12.92	12.98	13.13
7	C-G	G	12.89	12.98	12.92	12.98	13.02
8	A-T	T	14.29	14.29	14.28	14.32	14.34
9	G-T	G / T	10.72 / 11.79	10.75 / 11.80	10.78 / 11.76	10.77 / 11.88	10.78 / 11.84
10	G-T	G	12.81	12.78	12.81	12.88	12.86
11	C-G	G	12.64	12.67	12.72	12.72	12.76
12	A-T	T	13.61	13.64	13.65	13.69	13.84
13	C-G	G	12.73	12.78	12.81	12.83	12.86
14	C-G	G	12.81	12.91	12.92	12.93	
15	A-T	T	13.95	14.08	14.03		
16	G-C	G	12.81	13.10			
17	T-A	T	13.33				

Table 3.24. Imino proton chemical shifts in 9mer to 17mer mismatch GT DNA length analogues based on Figure 3.23.

^{1}H decoupled 1D ^{31}P-NMR experiments were also carried out to study how backbone ^{31}P resonances changed with increasing length (shown in appendix). Although assignments were made for ^{31}P resonances of the 17mer mismatch GT DNA, it proved difficult to transfer these assignments to the 1D ^{31}P-NMR spectra of the length analogues as peaks appeared very broad. All the ^{31}P resonances appeared in the expected range between -3.8 to 5.0 ppm.

3.4 Effect of mismatch orientation and sequence context on mismatch GT stability

The effect of oligonucleotide length has been noted in section 3.4, but it is not known how changing the mismatch orientation or sequence context affects mismatch DNA stability. Base stacking interactions contribute greatly to the stability of a DNA molecule but introduction of a mismatch base pair disrupts the base stacking in the local area, this section will investigate the extent of this disruption by UV absorbance melting (Figure 3.22). The calculated thermodynamic parameter values derived from the UV absorbance melting data are listed in Table 3.25.

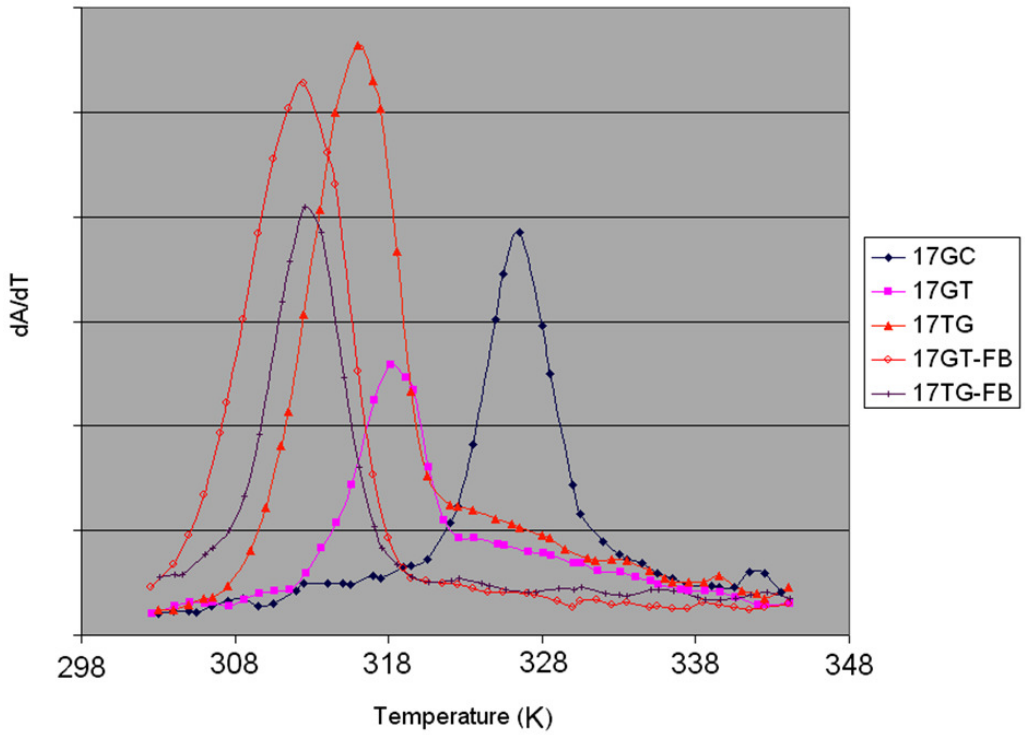


Figure 3.22. UV absorbance first derivative plot of 17mer mismatch GT DNA and its mismatch orientation and sequence context analogues. dA/dT values were calculated from primary UV absorbance melting data.

Inspection of the derivative plots shows, and keeping in view the method for estimating any errors, a narrowed dispersion around the T_m for the measured data with a symmetrical distribution.

Duplex	A (260)	Conc. (µg/ml)	T_m (K)	ΔH (kJ mol^{-1})	ΔS (J K^{-1} mol^{-1})	ΔG (310) (kJ mol^{-1})
17mer GC	0.51	25.42	326.5	-114.59	380.74	-118.14
17mer GT	0.49	24.66	317.1	-99.18	338.22	-104.95
17mer TG	0.76	38.25	316.0	-95.03	335.16	-103.99
17mer GT-FB	0.79	39.47	312.5	-46.01	188.44	-58.46
17mer TG-FB	0.83	41.35	312.6	-57.65	219.93	-68.24

Table 3.25. Experimentally calculated thermodynamic parameters for UV absorbance melting experiments investigating the effect of mismatch orientation and sequence context on stability of mismatched base pairs.

Figure 3.22 shows that the 17mer mismatch GT and TG DNAs were the most stable of the mismatch sequences with a T_m value approximately 10K lower than 17mer canonical GC DNA. Furthermore, there was a clear difference upon changing the sequence context. The change in sequence context results in a large number of AT base pairs being positioned towards the 3'end of the top strand. An increase in the number of AT base pairs reduces the stability of the molecule since they are less stable than GC base pairs.

Changing mismatch orientation was thought to have a greater effect on duplex stability since the base stacking between the two thymine residues at the mismatch site is broken. This was thought to have a contributing factor to mismatch stability from the NMR structure of 17mer mismatch GT DNA but as the data shows the difference in free energy (ΔG) between the GT and TG mismatch DNAs is approximately -1 kJ mol^{-1}.

This shows that quantitatively that base stacking does not contribute as great a difference as first thought.

The difference in free energy between GT-FB and TG-FB mismatch DNAs (-10 kJ mol^{-1}) gives further evidence that the increase in AT base pairs is the major contributing factor in decreasing mismatch stability.

In section 3.4, it was found that the chemical shifts of the GT mispair were consistent regardless of the sequence length. With a change in mismatch orientation, the imino proton resonances of the mispair undergo a change in chemical shift with the TG mispair [δ (G) = 10.23 ppm and δ (T) = 11.30 ppm] appearing further highfield in relation to the GT mispair [δ (G) = 10.72 ppm and δ (T) = 11.79 ppm] (Figure 3.23). This observation showed that there was better base stacking with a GT mispair resulting in lowfield shifting of the imino proton resonances. The change in chemical shift is quite dramatic with both G and T residues differing by approximately 0.5 ppm.

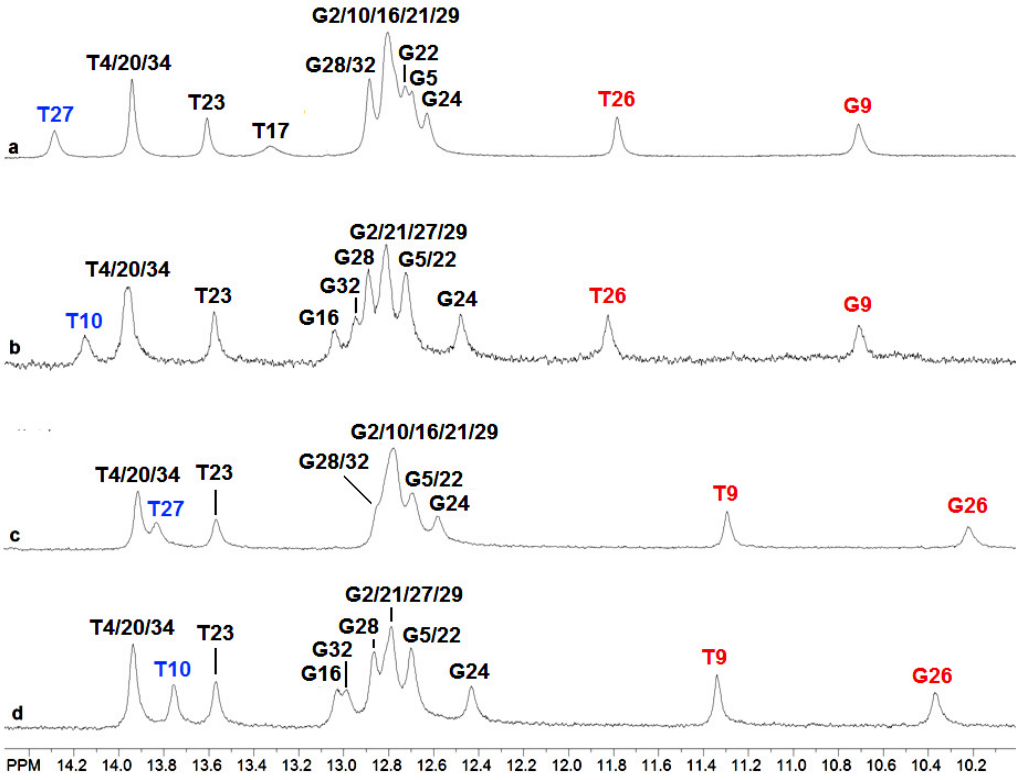


Figure 3.23. 400 MHz (b-d) and 800 MHz (a) 1D ^{1}H-NMR stack at 25°C showing imino region of 17mer mismatch GT DNA and its mismatch orientation and sequence context analogues, (a) 17mer GT, (b) 17mer GT-FB, (c) 17mer TG and (d) 17mer TG-FB. FB denotes that the base pairs immediately flanking the GT/TG mismatch are switched. The imino proton resonances of the mispair are highlighted in red with a clear difference in chemical shift observed upon a change in the orientation of the mismatch base pair. The T residue which immediately flanks the central mispair is highlighted in blue.

When the sequence context is changed, one would expect the mispair imino proton resonances to change. As shown in Figure 3.23, this was not true for 17mer mismatch GT-FB DNA where the mispair G and T imino protons appeared at similar chemical shifts to 17mer mismatch GT DNA. The chemical shift for T27 has also shifted further highfield in 17mer mismatch GT-FB DNA, indicating that the mismatch site is less stable when compared to 17mer mismatch GT DNA.

With 17mer mismatch TG-FB DNA, G26 undergoes a shift compared to 17mer mismatch TG DNA ($\Delta\delta$ = 0.16 ppm) and also appears sharper and more intense,

indicating less exchange with solvent. T10/27 also indicates that 17mer mismatch TG-FB DNA was more stable than 17mer mismatch TG DNA. Although T27 appears more lowfield, T10 is equally intense with no increase in line broadening indicating slow exchange with solvent. Imino proton chemical shifts for mismatch orientation and sequence context analogues are given in Table 3.26.

Base Pair No.	Base Pair	Imino Proton	Chemical Shift, δ (ppm)		Base Pair	Imino Proton	Chemical Shift, δ (ppm)	
			17mer GT	17mer TG			17mer GT-FB	17mer TG-FB
1	A-T	T	13.95	13.93	A-T	T	13.96	13.96
2	G-C	G	12.81	12.79	G-C	G	12.81	12.82
3	C-G	G	12.89	12.87	C-G	G	12.95	13.01
4	T-A	T	13.95	13.93	T-A	T	13.96	13.96
5	G-T	G	12.70	12.71	G-T	G	12.72	12.72
6	C-G	G	12.81	12.79	C-G	G	12.81	12.82
7	C-G	G	12.89	12.87	C-G	G	12.89	12.89
8	A-T	T	14.29	13.84	C-G	G	12.81	12.82
9	G-T / T-G	G / T	10.72 / 11.79	11.30 / 10.23	G-T / T-G	G / T	10.61 / 11.82	11.36 / 10.39
10	G-T	G	12.81	12.79	T-A	T	14.10	13.78
11	C-G	G	12.64	12.59	C-G	G	12.48	12.46
12	A-T	T	13.61	13.58	A-T	T	13.58	13.59
13	C-G	G	12.73	12.71	C-G	G	12.72	12.72
14	C-G	G	12.81	12.79	C-G	G	12.81	12.82
15	A-T	T	13.95	13.93	A-T	T	13.96	13.96
16	G-C	G	12.81	12.79	G-C	G	13.04	13.05
17	T-A	T	13.33		T-A	T		

Table 3.26. Imino proton chemical shifts for 17mer mismatch GT DNA and its mismatch orientation analogues and sequence context analogues. Values are recorded to two decimal places and are based on Figure 3.25.

In conclusion, the results show that irrespective of oligonucleotide length, mismatch orientation and sequence orientation, the imino proton resonances act like a fingerprint for each type of mismatched DNA, allowing quick identification of mismatch type.

The ^{1}H decoupled 1D ^{31}P-NMR spectra of the mismatch orientation and sequence context analogues shows all four were within the expected chemical shift range (-3.8 to -5.0 ppm). However, as with the length analogues, it was difficult to assign the spectra of sequences other than 17mer mismatch GT DNA, as the peaks were broad.

3.5 Interaction of mismatch GT DNAs with *Escherichia coli* MutS protein

The X-ray crystal structure of various mismatch and modified DNAs bound to prokaryotic and eukaryotic MutS protein have shown that the primary point of interaction is the mismatch/modified base pair of the DNA molecule.

To study the mismatch DNA-MutS interaction, the 1D ^{1}H-NMR spectra of mismatch DNAs (length, mismatch orientation and sequence context analogues) were recorded before and after the addition of sub-stoichiometric amounts of *E. coli* MutS protein. Similar titrations were carried out on modified 6-thioguanine and 6-sulphonateguanine DNAs which are discussed in Chapters 4.3 and 4.4.

E. coli MutS protein is considerably large in size (~ 200 kDa, dimer) hence it is difficult to characterise its NMR spectrum. Due to this limitation, monitoring changes to the protein resonances in standard NMR ligand binding experiments will be accompanied by a number of complications. These are greatly reduced when monitoring the changes to the NMR spectrum of mismatch DNA under sub-stoichiometric conditions of the addition of MutS, as there are fewer resonances which are also less overlapped. A significant advantage of monitoring DNA imino proton resonances is that they typically appear lowfield in the NMR spectrum so they can be monitored without any interference from the protein resonances. They are highly sensitive to changes in dynamics and can serve as a probe to monitor and observe the formation of mismatch DNA-MutS and modified DNA-MutS complexes.

Upon complex formation, changes to chemical shift and line broadening will be observed to the imino protons of residues involved in the interaction with MutS when compared to their unbound state. This approach provides both structural evidence for binding in solution state and also allows rapid determination of MutS affinity towards mismatch/modified DNA by NMR.

It is known that the protein binds to the mismatch GT base pair so changes in chemical shift and line broadening to the imino proton resonances of these bases were expected. However, the binding interaction between mismatch/modified DNA and MutS protein may not be the same for every sequence. Changes to specific imino proton resonances may indicate how the protein binds elsewhere to the DNA and this binding interaction may vary depending on the orientation of the base pairs and length of sequence.

3.5.1 Interaction of 17mer mismatch GT DNA and oligonucleotide length analogues

The assignment of imino proton resonances in the sequence length mismatch GT DNAs has already been shown previously (Chapter 3.4) and this provided a suitable starting point for assignment after MutS addition (Figure 3.21).

The titration of 9mer mismatch GT DNA with MutS protein (Figure 3.24) shows that the sequence was very stable even after the titration. This was evident as all imino proton resonances were still observed following the addition of MutS protein. A significant reduction in line intensity and increase in line broadening is observed for the mismatch

base pair, G5-T14, which confirms an interaction with the protein. Changes in line intensity occur when the exchange rate between exchangeable protons and the solvent increases, whereas a change in linewidth is commonly due to molecular binding. Evidence of an interaction is further strengthened by the chemical shift change to T14 following MutS addition ($\Delta\delta$ = 0.07 ppm). G1 is a terminal base and the chemical shift change is due to a greater likelihood of solvent exchange and fraying of the duplex. There is also a clear reduction in line intensity observed for T11 and T15 and the position of G1 has also shifted lowfield. T15 is adjacent to the GT mispair so the observed result is due to local conformational changes of the helix at the mismatch site during complex formation. The reduced intensity for T11 suggests it may be a secondary binding site for the protein.

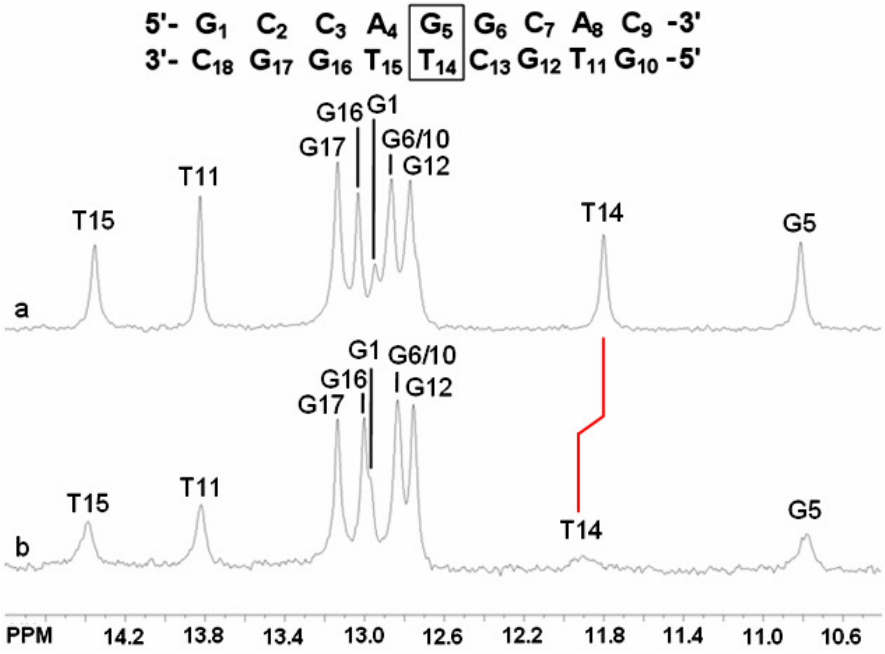


Figure 3.24. 400MHz 1D ^{1}H-NMR titration stack showing the effect of *E. coli* MutS addition to 9mer mismatch GT DNA (0.2mM, 90% $^{1}H_2O$+ 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 5°C where a. 9mer GT and b. 9mer GT + *E. coli* MutS (0.11 x 10^{-3} mM).

The results obtained for the 11mer mismatch GT DNA were similar to the 9mer sequence and gave evidence of an interaction between the protein and the GT mispair. However, the 1D ^{1}H-NMR spectrum (Figure 3.25) shows a different set of secondary changes to the imino proton resonances. Residue T17 exhibited the greatest change in chemical shift and this is important as T17 is part of the G6-T17 mispair but highfield shifts were observed for residues G19/20. G19/20 are not terminal bases so the chemical shift change may be promoted by a secondary phase of binding by MutS. Following the addition of *E. coli* MutS G12/19/20 resonated at the same chemical shift (δ = 12.95 ppm) and G13 was resolved from G2/7. Line broadening was observed to T18 once again highlighting that the effect of MutS binding is not only felt by the mispair.

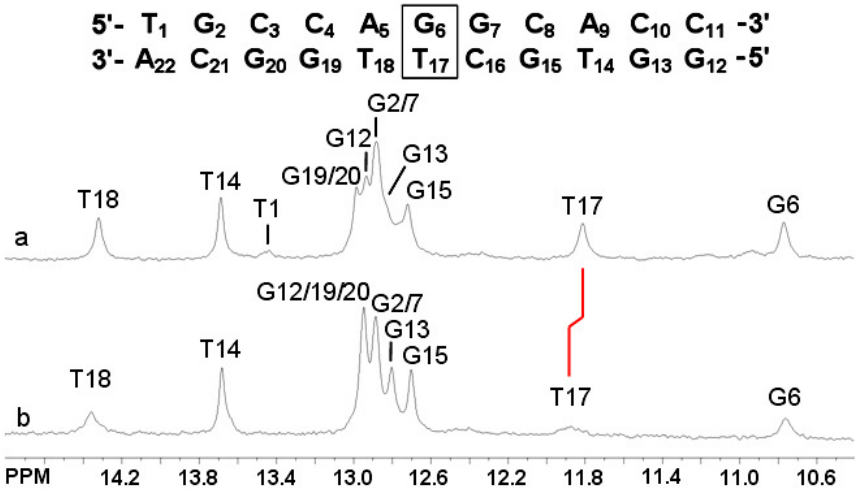


Figure 3.25. 400MHz 1D ^{1}H-NMR titration stack showing the effect of *E. coli* MutS addition to 11mer mismatch GT DNA (0.2mM, 90% $^{1}H_2O$+ 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 5°C where a. 11mer GT and b. 11mer GT + *E. coli* MutS (0.11 x 10^{-3} mM).

The NMR titration of 13mer mismatch GT DNA with MutS protein showed a large change in a number of imino proton resonances (Figure 3.26). The most significant was the disappearance of the T20 imino proton of the GT mispair. In fact, both the imino

proton resonances of the GT mispair were almost absent from the 1D ^{1}H-NMR spectrum. This observation clearly shows a strong interaction between *E. coli* MutS and 13mer mismatch GT DNA.

The imino proton resonance of G26 was not observed after the addition of MutS but this effect could be due to rapid exchange with solvent, as it is a terminal base. The same phenomenon was responsible for reduced line intensity for T14. T2 is in a similar position to G20 of the 11mer mismatch GT DNA and so its change in linewidth and intensity may be caused by secondary binding of the protein.

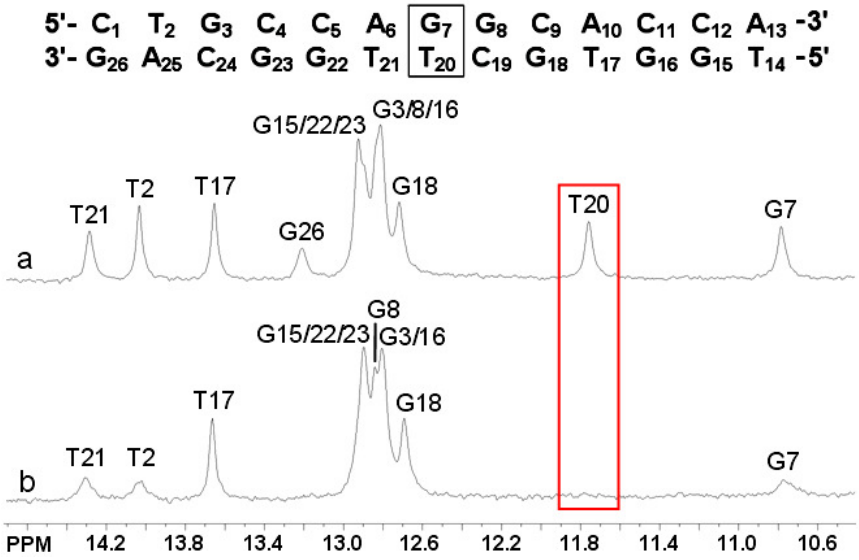


Figure 3.26. 400MHz 1D ^{1}H-NMR titration stack showing the effect of *E. coli* MutS addition to 13mer mismatch GT DNA (0.2mM, 90% $^{1}H_2O$+ 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 5°C where a. 13mer GT and b. 13mer GT + *E. coli* MutS (0.11 x 10^{-3} mM).

The titration of 15mer mismatch GT DNA (Figure 3.27) gives more evidence of a secondary binding site. Similar to the other mismatch DNA length analogues, a change in line intensity and linewidth was observed to the imino protons of the GT mispair (G8

and T23) and also the immediately flanking T residue (T24). The line intensity of the mispair imino proton resonances were approximately equal to each other before the titration. T23 was absent following the addition of *E. coli* MutS protein indicating stronger binding to T23. This supports the X-ray crystal structure which notes that the repair process removes the T residue when a GT mispair is present. T24 has been affected to the same extent as the mispair residues showing confirming a change in the local structure around the GT mispair. The imino proton for G15 was also absent, caused by rapid exchange with the solvent as it is a terminal residue.

The titration of the 15mer mismatch GT DNA has also yielded new information not seen with the other mismatch DNAs discussed so far. The absence of G29 and a change in chemical shift to G25 indicates a secondary stage of interaction towards the 5' end of the top strand, as observed with the 9 to 13mer mismatch GT DNAs. However, a perturbation in the imino proton resonance of G18 shows a possible interaction with the other half of the DNA molecule which was not observed with the 9 to 13mer mismatch GT DNAs. The implication presented here may be that as length increases, a series of secondary interactions may occur between *E. coli* MutS and mismatch GT DNA which help to anchor the protein in place whilst the repair process takes place at the mismatch site.

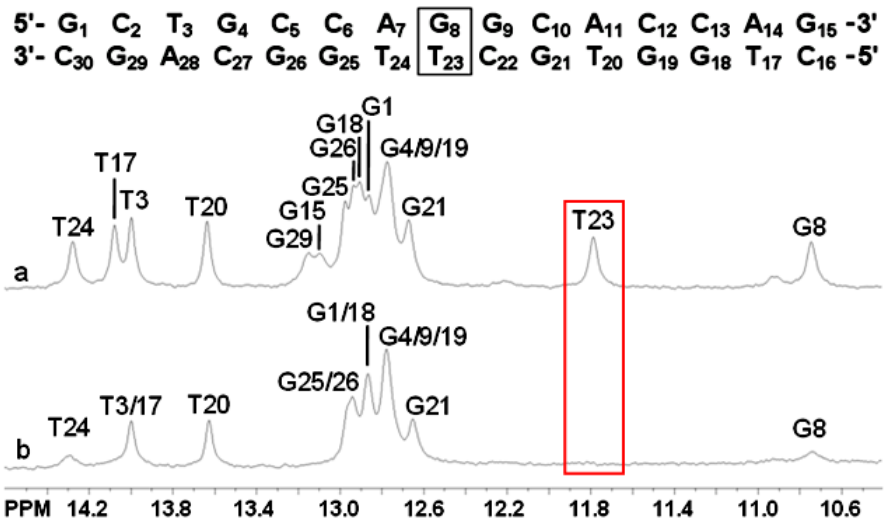


Figure 3.27. 400MHz 1D ^{1}H-NMR titration stack showing the effect of *E. coli* MutS addition to 15mer mismatch GT DNA (0.2mM, 90% $^{1}H_2O$+ 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 5°C where a. 15mer GT and b. 15mer GT + *E. coli* MutS (0.11 x 10^{-3} mM).

The titration of 17mer mismatch GT DNA indicates that it was still very stable even after the addition of *E. coli* MutS protein (Figure 3.28). However, this can be justified by the fact that the concentration of the 17mer mismatch GT DNA was significantly higher than the other mismatch GT DNAs. Subsequently, the degree of binding may be the same as the other mismatch DNAs although the final result does not support this.

The imino proton of T17 was absent (Figure 3.28, b) due to rapid exchange with solvent. G9 and T26 are in the GT mispair and are consequently directly affected by the addition of MutS protein. Although small, T26 experienced the biggest change in chemical shift ($\Delta\delta$ = -0.04 ppm).

The imino protons resonances of the two directly flanking base pairs to the mispair, G10 and T27, have also shown pertubations with G10 showing a very small chemical shift change ($\Delta\delta$ = -0.02 ppm) and T27 being less intense following MutS addition. G5 and

G22 which are both an equivalent distance from the mispair, three base pairs away, have also become overlapped again maybe indicative of secondary binding.

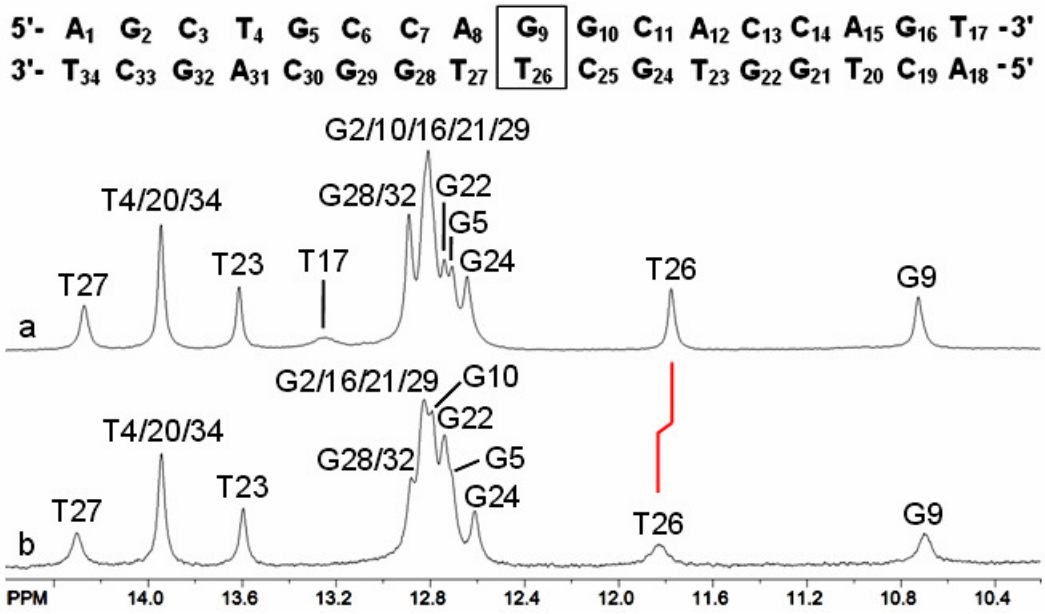


Figure 3.28. 800MHz 1D ^{1}H-NMR titration stack showing the effect of *E. coli* MutS addition to 17mer mismatch GT DNA (0.8mM, 90% $^{1}H_2O$+ 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 5°C where a. 17mer mismatch GT DNA and b. 17mer mismatch GT DNA + *E. coli* MutS (0.15 x 10^{-3} mM).

3.5.2 Interaction of 17mer mismatch GT DNA and mismatch orientation and sequence context analogues

The assignment of imino proton resonances in the sequence orientation mismatch GT DNAs has already been shown previously (section 3.5) and this provided a suitable starting point for assignment after MutS addition (Figure 3.23).

In 17mer mismatch GT-FB DNA, the mismatch G9 and T26 imino proton resonances could not been observed in the 1D ^{1}H-NMR spectrum following MutS addition (Figure 3.29). This suggests that the interaction between mismatch GT-FB DNA was better

when compared to 17mer mismatch GT DNA. It was discussed in section 3.2 that the NMR structure suggested additional stability at the mismatch site caused by better base stacking between G9-G10 and T26-T27. The change in the flanking bases breaks this base stacking and so the mismatch site may be less stable allowing a more accessible binding site for MutS protein. This was supported by the increase in linewidth in the imino proton for T10 and a decrease in line intensity of the G27 peak.

Beyond the mismatch site, G27 has shifted highfield overlapping with G28 indicating a weak interaction with the protein and G16 has shifted highfield due to faster exchange with the solvent. With the exception of the above changes, the other imino protons have remained practically unchanged.

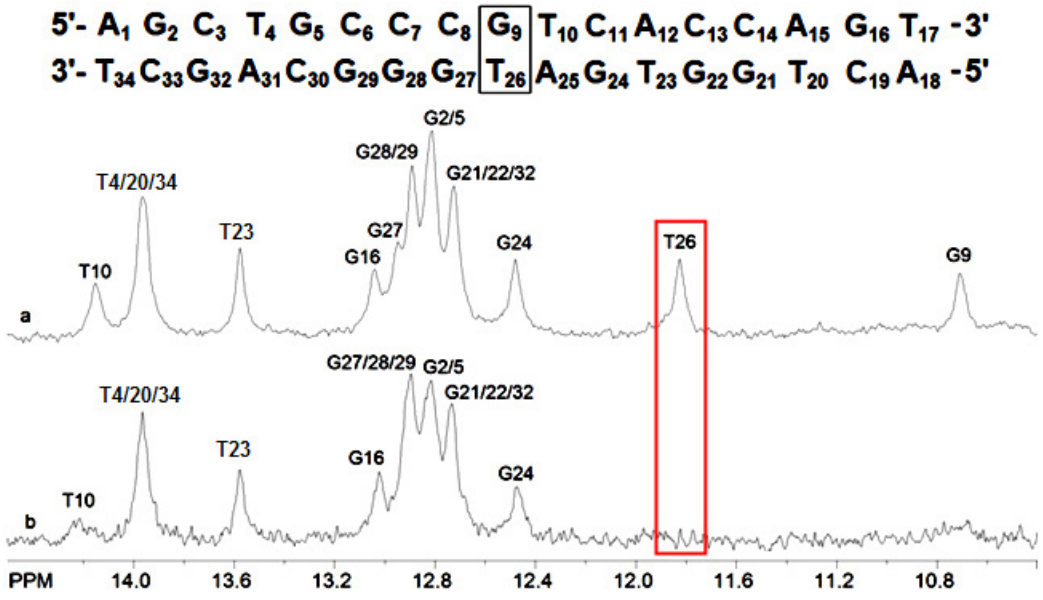


Figure 3.29. 400MHz 1D ^{1}H-NMR titration stack showing the effect of *E. coli* MutS addition to 17mer mismatch GT-FB DNA (0.2mM, 90% $^{1}H_2O$+ 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 5°C where a. 17mer GT-FB and b. 17mer GT-FB + *E. coli* MutS (0.11 x 10^{-3} mM).

As mentioned in Chapter 3.5, a change in the context of the mismatch was accompanied by a shift in the position of the mispair imino protons resonances. The region between 12.5-13.2 pm was typically where G imino protons appear. In 17mer mismatch TG DNA (Figure 3.32), this region was heavily overlapped so pertubations following MutS addition were not as clear to observe.

T9 and G26 which are in the mispair exhibit the greatest change in chemical shift [$\Delta\delta$ (T9) = 0.05 ppm and $\Delta\delta$ (G26) = 0.09 ppm] showing specific interaction with the protein; the interaction also resolved G10 from G2/16/21/29.

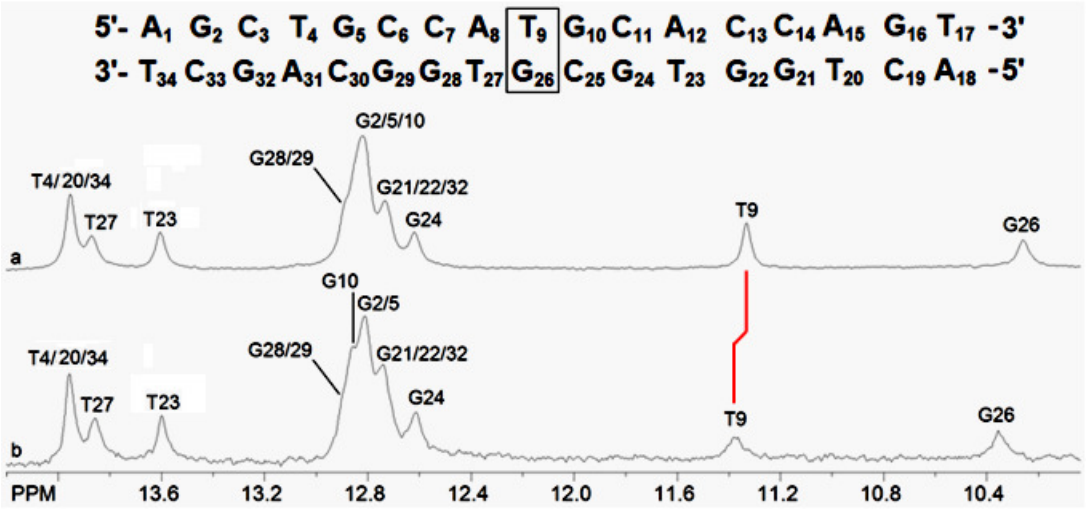


Figure 3.30. 400MHz 1D ^{1}H-NMR titration stack showing the effect of *E. coli* MutS addition to 17mer mismatch TG DNA (0.2mM, 90% $^{1}H_2O$+ 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 5°C where a. 17mer TG and b. 17mer TG + *E. coli* MutS (0.11 x 10^{-3} mM).

In 17mer mismatch TG-FB DNA, the change in flanking bases restored the G-G and T-T base stacking which resulted in the increased stability of 17mer mismatch GT DNA. This was confirmed in the 1D ^{1}H-NMR spectrum following addition of *E. coli* MutS

(Figure 3.31) where a large number of imino protons did not exhibit a change in chemical shift or linewidth.

Both imino proton resonances shifted lowfield [$\Delta\delta$ (T9) = 0.07 ppm and $\Delta\delta$ (G26) = 0.11 ppm] but despite this, T9 was more affected by the addition of MutS as its line intensity was significantly more reduced compared to G26. Surprisingly, the imino proton of T10 did not exhibit a chemical shift change.

Changes in chemical shift were also observed for T23, G24 and G27 showing that the effects of the protein are felt beyond the mispair. G16 is located towards the end of the DNA molecule and so its change in chemical shift is due to solvent exchange effects.

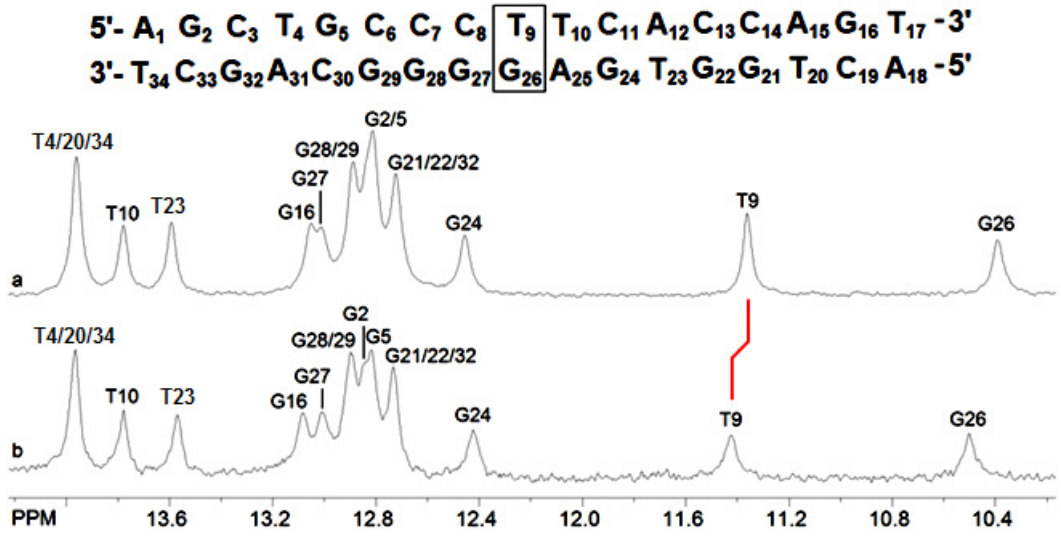


Figure 3.31. 400MHz 1D ^{1}H-NMR titration stack showing the effect of *E. coli* MutS addition to 17mer mismatch TG-FB DNA (0.2mM, 90% $^{1}H_2O$+ 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 5°C where a. 17mer TG-FB and b. 17mer TG-FB + *E. coli* MutS (0.11 x 10^{-3} mM).

Based on the crystal structure of the MutS-mismatch GT complex, the lowfield shifting of the mismatch T imino proton resonance could be explained by its interaction with Phe36 of MutS. The Phe36 residue intercalates into the DNA during complex formation as it stacks to the mismatch T base. Changes observed for other residues could result from conformational changes to the DNA, since it undergoes a strong kink of approximately 60° at the mismatch site in the DNA-MutS complex. The variation in line intensities and broadening could be an indicator of these conformational changes. Residues where no change was observed to the imino proton resonances are not involved in interactions with MutS.

To summarise, the NMR structure determination of 17mer mismatch GT DNA and 17mer canonical GC DNA have been successfully carried out to a high resolution with both giving low RMSD values for the ten best structures. The effect of mismatch orientation, sequence context and oligonucleotide length was investigated using UV absorbance melting and NMR spectroscopy with sequences studied analogous to the 17mer mismatch GT DNA. The results showed that the GT mispairs were more stable than TG mispairs but that irrespective of the sequence context and oligonucleotide length, the imino protons of the bases for each mismatch appear at distinct chemical shifts. These results correlated well to T_m values obtained using UV absorbance melting. MutS showed affinity for all of the above DNA sequences but the level of interaction seem to depend on the stability of the mispair and its base stacking interactions with neighbouring base pairs.

Chapter 4 – NMR structure determination of modified 6-thioguanine and 6-sulphonateguanine GC DNA

6-Thioguanine (6-TG) is well-known in medical research as an anti-leukaemia agent but prolonged exposure to UVA light causes oxidation of the sulphur atom forming a sulphonate group.[40] Further oxidation of 6-TG converts it to 6-sulphonateguanine or 2-aminopurine-6-sulfonic acid (G^{SO3})[40] and in doing so the anticancer property of 6-TG is abolished and the base becomes carcinogenic (Figure 4.1).

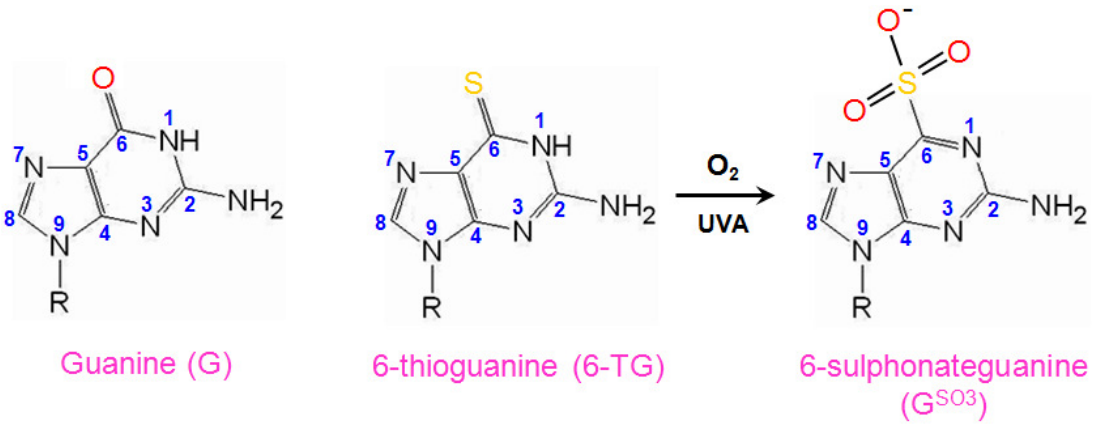


Figure 4.1. Comparison between guanine (G), 6-thioguanine (6-TG) and 6-sulphonateguanine (G^{SO3}) base structures.

In this chapter, the structure determination of 6-TG modified 13mer GC DNA by NMR spectroscopy is presented. The assignment of ^{1}H, ^{13}C and ^{31}P resonances and a summary of constraints used in the structure calculation is shown. Further to this, a study comparing the effect of 6-TG or G^{SO3} substitution into canonical GC DNA to the NMR spectrum is also presented. The effect of *E. coli* MutS addition on the various thio

modified DNAs studied in this chapter was also investigated by NMR titration and the results are presented here.

4.1 NMR structure determination of 6-thioguanine modified 13mer GC DNA

The assignment of proton, carbon and phosphorus resonances in 6-TG modified 13mer GC DNA and its structure calculation is presented in this chapter. NMR assignment was carried out with the same methodology used in Chapter 3. Hydrogen bonding in a 6-TG modified GC base pair is similar to a canonical GC base pair (Figure 1.18) as mentioned in Chapter 1, so a B-form DNA structure is still expected to be observed.

4.1.1 Identification and assignment of exchangeable proton resonances

The chemical shift of the 6-TG imino proton was expected to shift due to a change in the base pairing and this difference is clearly observed in its 1D ^{1}H-NMR spectrum when compared to that of canonical GC DNA (Figure 4.2). The change in the chemical shift of G7 was less than observed for the equivalent G residue (G9) in 17mer mismatch GT DNA. This indicates that the base stacking was better in the 6-TG base modified GC base pair compared to a GT mispair.

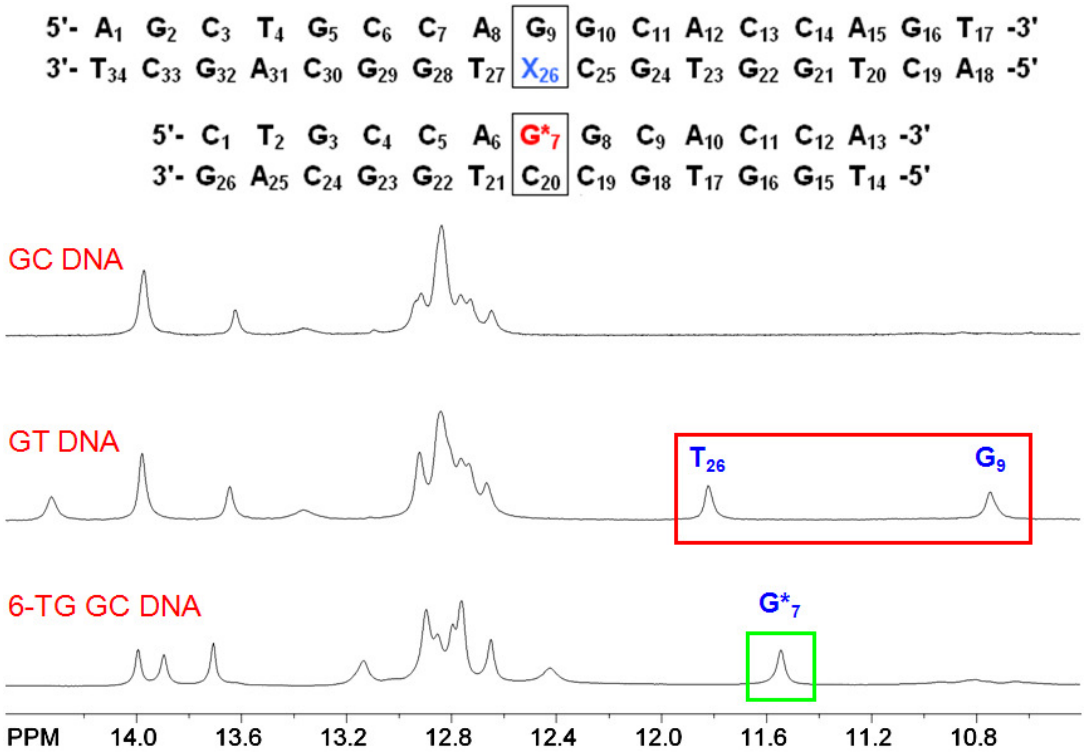


Figure 4.2. 800 MHz stack of 1D ^{1}H-NMR spectra showing comparison between canonical, non-canonical and modified base pairs measured at 2°C. (Top) canonical 17mer GC DNA (1mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) where $X_{26} = C_{26}$, (middle) mismatched 17mer GT DNA (1mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) where $X_{26} = T_{26}$ and (bottom) 6-TG modified 13mer GC DNA (0.8mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) where G^*_7 = 6-TG as shown in the sequences above.

Although, the imino proton corresponding to the G7 base is clearly shifted, its pattern of NOE cross peaks with NOESY spectrum was similar to that observed for a canonical GC base pair (Figure 4.3, panels left and right).

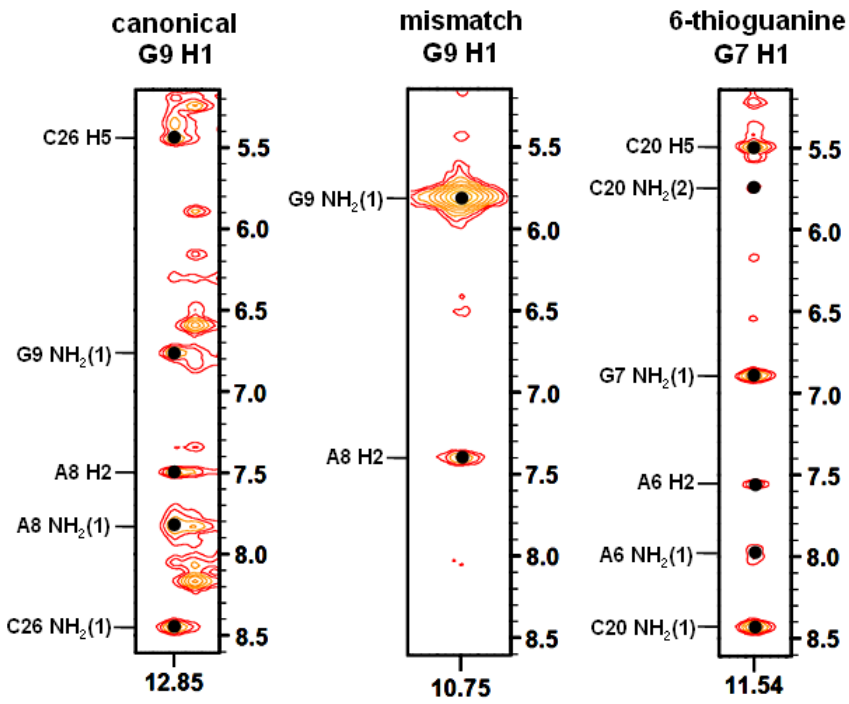


Figure 4.3. (Left panel) 700 MHz NOESY (τ_{mix} = 250ms) panel for GC base pair of 17mer canonical GC DNAat 2ºC, (centre panel) 800 MHz NOESY (τ_{mix} = 250ms) panel for GT base pair for 17mer mismatch GT DNA at 1ºC, (right panel) 800 MHz NOESY (τ_{mix} = 250ms) panel for 6-TG modified GC base pair for 6-TG modified 13mer GC DNA at 2ºC. A comparison illustrating the difference in base pair NOE cross peaks observed for canonical GC, mismatch GT and 6-TG modified GC imino proton resonances.

Exchangeable imino and amino protons were assigned in the 1D ^{1}H-NMR (Figure 4.4) and NOESY spectra (Figure 4.5) measured in $^{1}H_2O$. All imino proton resonances were observed in the typical chemical shift ranges for pyrimidines and purines with the exception of the G7 base. The imino proton resonances were assigned following the imino-imino proton sequential connectivity starting from G7. G7 appeared clearly separate from other resonances (δ = 11.54 ppm) and more importantly, sequential connectivity to G8 and T21 were clearly resolved and observed. Twelve imino proton resonances were assigned in the NOESY spectrum; only T14 remained unassigned. The sequential connectivity from G7 to G26 is shown in Figure 4.5 Assignment of the

sequential connectivity stretching from G7 to G15 was also carried but is shown in the appendix.

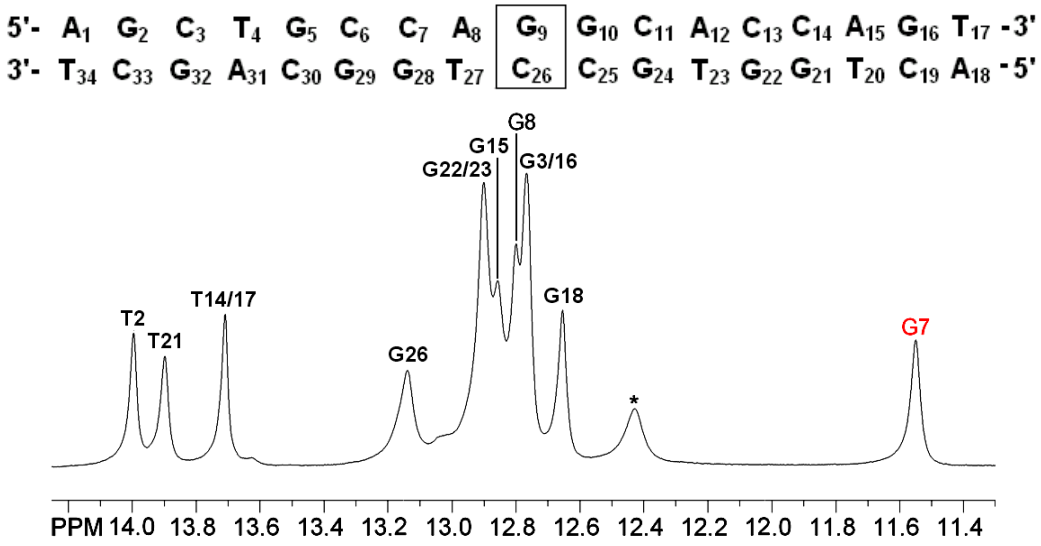


Figure 4.4. 800 MHz 1D ^{1}H-NMR spectrum showing imino proton region in 6-TG modified 13mer GC DNA (0.8mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C. The assignment of imino proton resonances is shown based on their identification in the NOESY spectrum measured in $^{1}H_2O$.

A key observation in the imino proton assignment was the peak at 12.43 ppm, marked by a * in Figure 4.4, which could not be identified through imino-imino NOE connectivities. Attributing this peak as an impurity was highly unlikely as the chemical shift is uncharacteristically lowfield. Impurities which are commonly observed for nucleic acids appear much further highfield between 0-5 ppm. The peak was most probably due to a minor secondary conformation of the 6-TG modified base (G7).

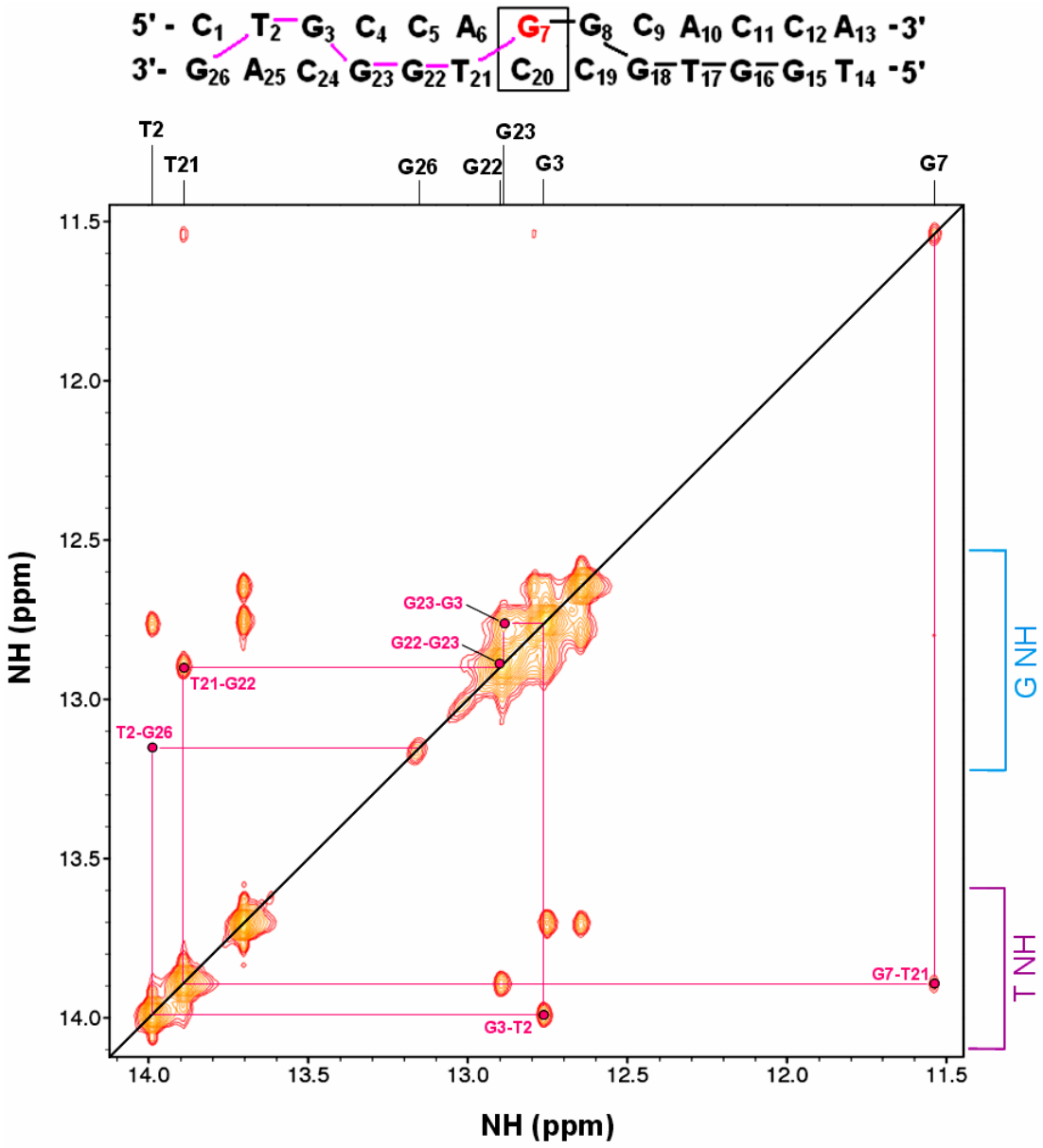


Figure 4.5. 800 MHz NOESY (τ_m = 250ms) spectrum of 6-TG modified 13mer GC DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C. The respective imino protons were identified and assigned based on the sequential connectivity stretching from G7 to G26 and this is indicated by the pink trace. The unassigned resonances in thie region correspond to the sequential connectivity stretching from G7 to G15 and this is shown in the Appendix)

Assignment of amino NH_2 protons

Imino proton assignments based on the above sequential connectivities were confirmed by assigning NOEs to proximal base paired amino protons. NOEs from imino to hydrogen bonded [NH_2(1)] (Figure 4.6) and non-hydrogen bonded [NH_2(2)] (shown in Appendix) appeared separate from one another allowing unambiguous identification and assignment. Assignment was assisted by identifying intrastrand and cross-strand NOEs for example, the cross strand between the A6 H2 to the G22 imino proton. The previously unassigned T14 was identified using strong NOEs to its base paired A H2 and NH_2(1) protons. One distinct feature of this region is the strong NOE between the G7 imino proton and the hydrogen bonded cytosine NH_2 of C20 to which it is base paired.

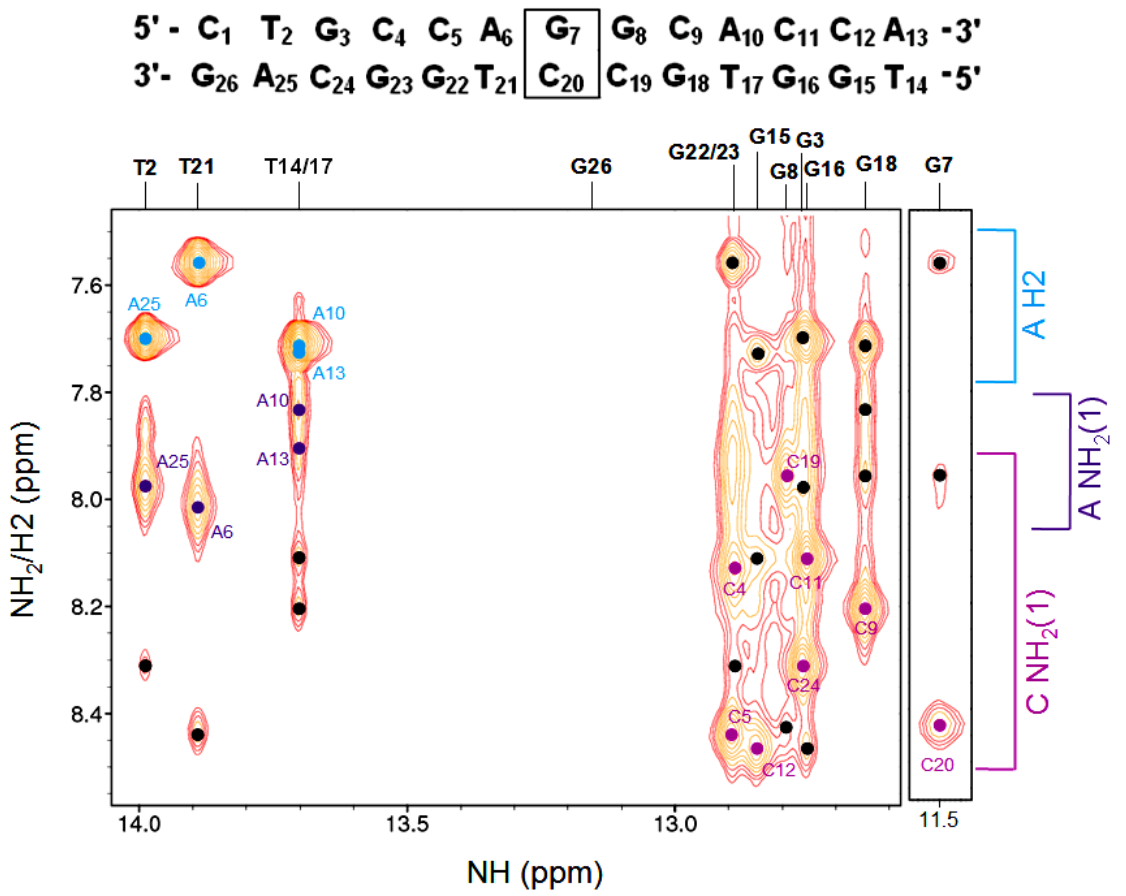


Figure 4.6. 800 MHz NOESY (τ_m = 250ms) spectrum of 6-TG modified 13mer GC DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C. Base pair imino-amino NOE connectivities to adenine H2 and hydrogen bonded cytosine/adenine NH_2 protons are indicated by blue, light purple and dark purple circles respectively. Important intrastrand and cross-strand NOE connectivities which confirmed the assignment are also highlighted by black circles in this region.

4.1.2 Identification and assignment of non-exchangeable proton and carbon resonances

A combination of DQF-COSY, TOCSY and NOESY spectra were used to identify the non-exchangeable protons in 6-TG modified 13mer GC DNA. The ^{1}H-^{13}C HSQC spectrum assisted the analysis and confirmed assignment.

Assignment of H2 resonances

All four adenine H2 resonances were assigned from the intense NOE cross peaks to their base-paired thymine imino protons observed in the $^{1}H_2O$ NOESY spectrum. Although A10, A13 and A25 have similar H2 chemical shifts they could be easily distinguished in the NOESY spectrum based on the dispersion of their base paired thymine imino protons. In contrast, in the HSQC spectrum, the corresponding C2 resonances were all overlapped regardless of them appearing separate from other resonances lowfield between 150-160 ppm.

With the exception of the below adenine H2 resonances, all other non-exchangeable resonances were assigned using homonuclear and heteronuclear NMR spectra recorded in 100% $^{2}H_2O$ solvent and at 25°C.

Assignment of cytosine H5-H6 and thymine CH_3-H6 connectivities

All nine cytosine H5-H6 correlations were identified using DQF-COSY and TOCSY spectra. The assignment was confirmed by observation of the same cross peaks in the NOESY spectrum, as the NOE distance is 2.4Å between the two protons and hence, the correlation appears as an intense NOE (Figure 4.7). Each H5-H6 correlation gave a coupling constant of 7-8 Hz as measured from the DQF-COSY spectrum, which is the expected value.

The pattern of cross peaks is similar to 17mer canonical GC DNA with the exception of C1 and C20. The position of the C1 H5-H6 cross peak was significantly different from all others as it is a terminal base and thus making its assignment unambiguous. The H6 resonance of C20 shifted further lowfield (7.52 ppm) due to its base pairing with G7. The identification of the H5-H6 correlation for C20 is significant as it is well resolved and therefore, it can be monitored . The assignment of H5-H6 correlations is based on the H6/8-H1' sequential connectivity which is presented later in this chapter (*Assignment of aromatic H6/8 – sugar H1' sequential connectivity*).

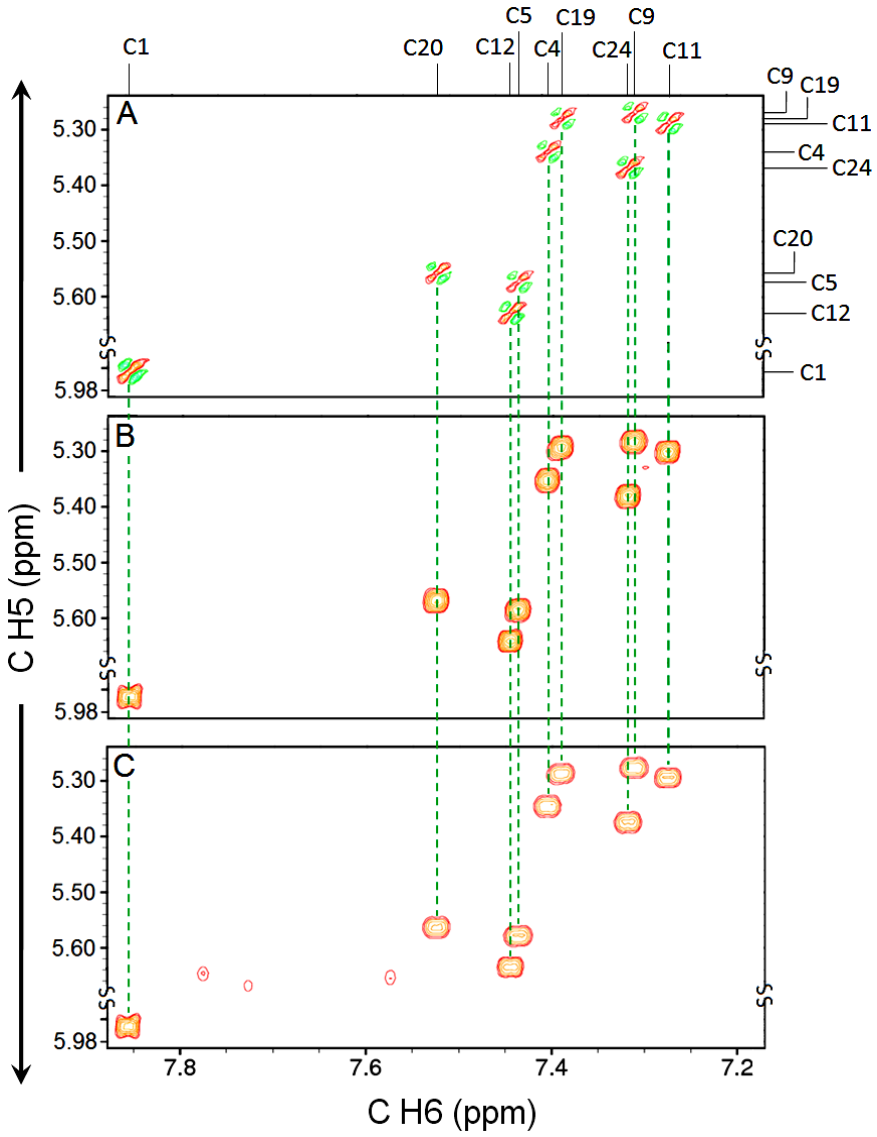


Figure 4.7. 800 MHz stack plots showing the assignment of cytosine H5-H6 connectivities in 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. DQF-COSY, B. TOCSY (SL_{mix} = 75ms) and C. NOESY (τ_m = 250ms).

Successful identification of H5 and H6 resonances allowed C5 and C6 resonances to be identified using the ^{1}H-^{13}C HSQC spectrum (Figure 4.8). C5 resonances appear at 95-100 ppm whereas the C6 resonances appear at 140-145 ppm.

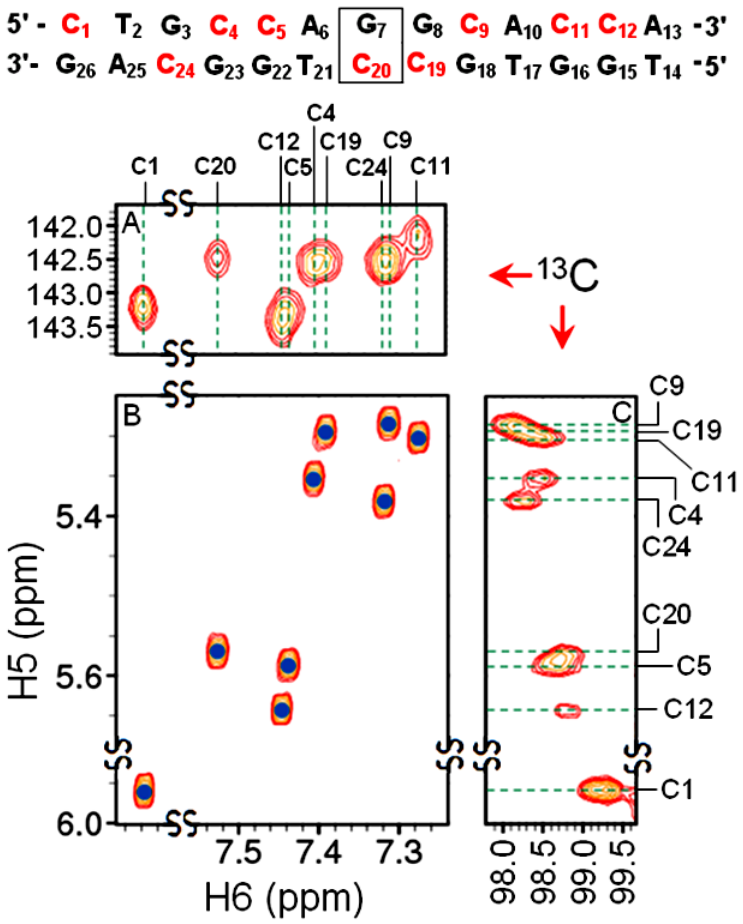


Figure 4.8. 800 MHz NOESY (τ_m = 250ms) spectrum (B) highlighting cytosine H5-H6 correlations in 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. 800 MHz ^1H-^{13}C HSQC panels show the assignment of the cytosine C6 (A) and C5 (C) chemical shifts.

Thymine CH_3-H6 correlations were assigned using the same methodology above using both DQF-COSY and TOCSY spectra. However, the cross peaks appear weaker as the protons are four bonds apart as opposed to cytosine H5-H6 correlations which are separated by three bonds. Assignment was confirmed as each CH_3-H6 correlation was also observed in the NOESY spectrum since the NOE distance is 2.9 Å. All four thymine CH_3-H6 correlations were identified and assigned. Assignment was on the basis of sequential connectivity of H6/8-H1' NOEs which is presented later (*Assignment of aromatic H6/8 – sugar H1' sequential connectivity*).

Similar to cytosine residues, identification of the CH_3 and H6 resonances in thymine residues allowed C5 (CH_3) and C6 resonances to be identified using the ^{1}H-^{13}C HSQC spectrum through the $^{1}J_{HC}$ correlation. C5 (CH_3) resonances appear at 0-20 ppm whereas the C6 resonances appear at 135-140 ppm.

Assignment of aromatic H5/CH_3 proton – aromatic H6/8 proton connectivities

Assignment of cytosine H5 and thymine CH_3 proton resonances were checked by the independent assignment of H5/CH_3 interresidue NOE connectivities to aromatic protons in the NOESY spectrum. Cross peaks can be observed between cytosine H5/thymine CH_3 protons on a given nucleotide (i) to the aromatic H6/H8 protons on the previous nucleotide (i-1) as the distance between them is within 5Å, thus can be seen in the NOESY spectrum. Carrying out this assignment enabled the key H6/H8-H1' sequential NOE connectivities to be identified.

Assignment of aromatic H6/8 – sugar H1' sequential connectivity

Sequential assignment of non-exchangeable protons was achieved by identifying NOE connectivities between aromatic and sugar protons in the $^{2}H_2O$ NOESY spectrum (Figure 4.9). The most important NOEs are those between aromatic H6/8 and sugar H1' protons. In helical B-form DNA the distance is 3.6-3.9Å so these will appear as intense cross peaks in the NOESY spectrum. Successful assignment at this step allows the assignment of all other resonances. H6 and H8 resonances are distinct from each other; cytosine and thymine H6 resonances are observed between 7.0-7.5 ppm, adenine H8

resonances between 8.0-8.5 ppm and guanine H8 resonances between 7.5-8.0 ppm. In addition, H1' protons resonate between 5.0-6.5 ppm and this broad range in chemical shifts allows cross peaks to be clearly observed since overlapping of resonances is reduced.

H6/8-H1' NOEs due to 5' and 3' end residues were identified first since the 5' ends have both intraresidue and interresidue connectivities whereas the 3' ends only show an intraresidue correlation. These peaks were also more intense than typically expected making their assignment unambiguous. Sequential connectivity was carried out once the terminal residues were assigned. The chemical shifts of the assigned H6/8 resonances was checked by assigning the NOESY region showing H6/8-H6/8 cross peaks. Sequential connectivity could also be carried in this region, which confirmed what was observed in the previous H6/8-H1' region.

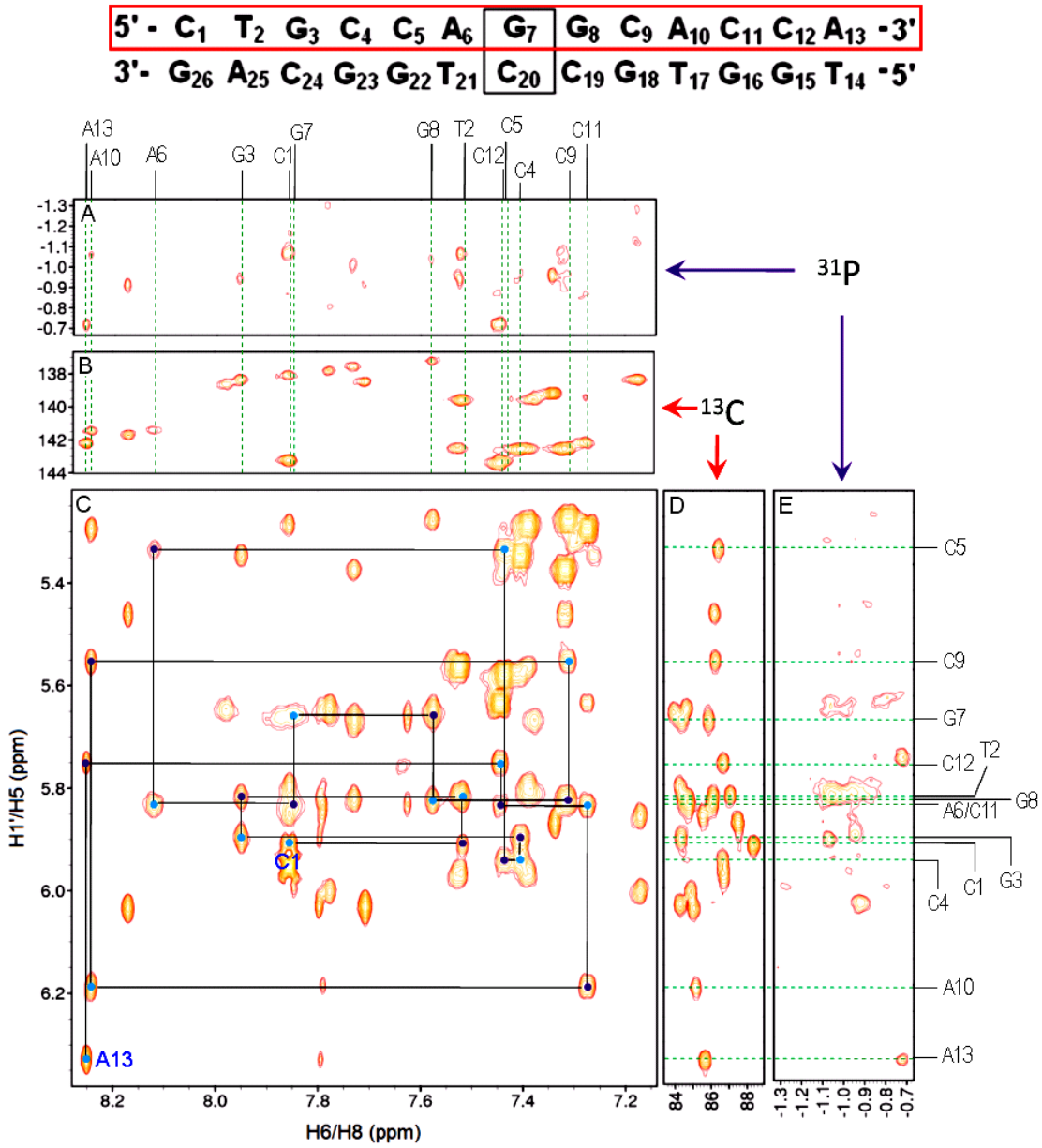


Figure 4.9. 800 MHz NOESY (τ_m = 250ms) spectrum (C) showing H6/8-H1' sequential NOE connectivities for residues C1-A13 in 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intraresidue connectivities are shown by blue circles and interresidue connectivities by purple circles. 800 MHz ^{1}H-^{13}C HSQC spectra show the corresponding assignment for C6/C8 (B) and C1' (D) resonances. 600 MHz ^{1}H-^{31}P CPMG-HSQC-NOESY (τ_m = 500ms) spectra show the corresponding assignment for backbone ^{31}P resonances through H6/8-^{31}P (A) and H1'-^{31}P (E) NOE connectivities. The unassigned NOEs seen in thie region correspond to the sequential connectivity stretching between T14 and G26, which has been assigned and is shown in the Appendix.

Assignment of H2' and H2'' resonances

The H2'/2'' protons resonate highfield between 1.8-3.0 ppm and can be differentiated in the DQF-COSY spectrum as the coupling constant of $^{3}J_{H1'\text{-}H2'}$ correlations are larger than $^{3}J_{H1'\text{-}H2''}$ correlations.

All H1'-H2' correlations were identified in the DQF-COSY spectrum (Figure 4.10). As with the 17mer DNAs, it was found that H2' protons resonated more highfield than H2'' protons. Assignment of H2' resonances allowed the H2'' resonances to be identified in the NOESY spectrum as both exhibit NOE cross peaks to the same H1' resonance for each residue. In the TOCSY spectrum they could not be distinguished as all cross peaks had similar intensities but in the NOESY spectrum H1'-H2'' cross peaks are more intense since H1'-H2'' distances are shorter than H1'-H2' for all sugar conformations. All H2' and H2'' resonances were assigned as clearly resolved peaks with the exception of G18, in which the resonances were overlapped.

By distinguishing the H2' and H2'' resonances, H6/8-H2'/2'' NOE connectivities can be assigned (Figure 4.11). These are important as they provide key distance constraints in the structure calculation of 6-TG modified 13mer GC DNA.

NOEs between aromatic H6/8 and sugar H2' protons are more intense due to the shorter distance between them. The H6/8-H2' NOE distance in DNA is normally 2.0-3.6 Å whereas H6/8-H2'' distances are usually 3.4-4.5 Å.

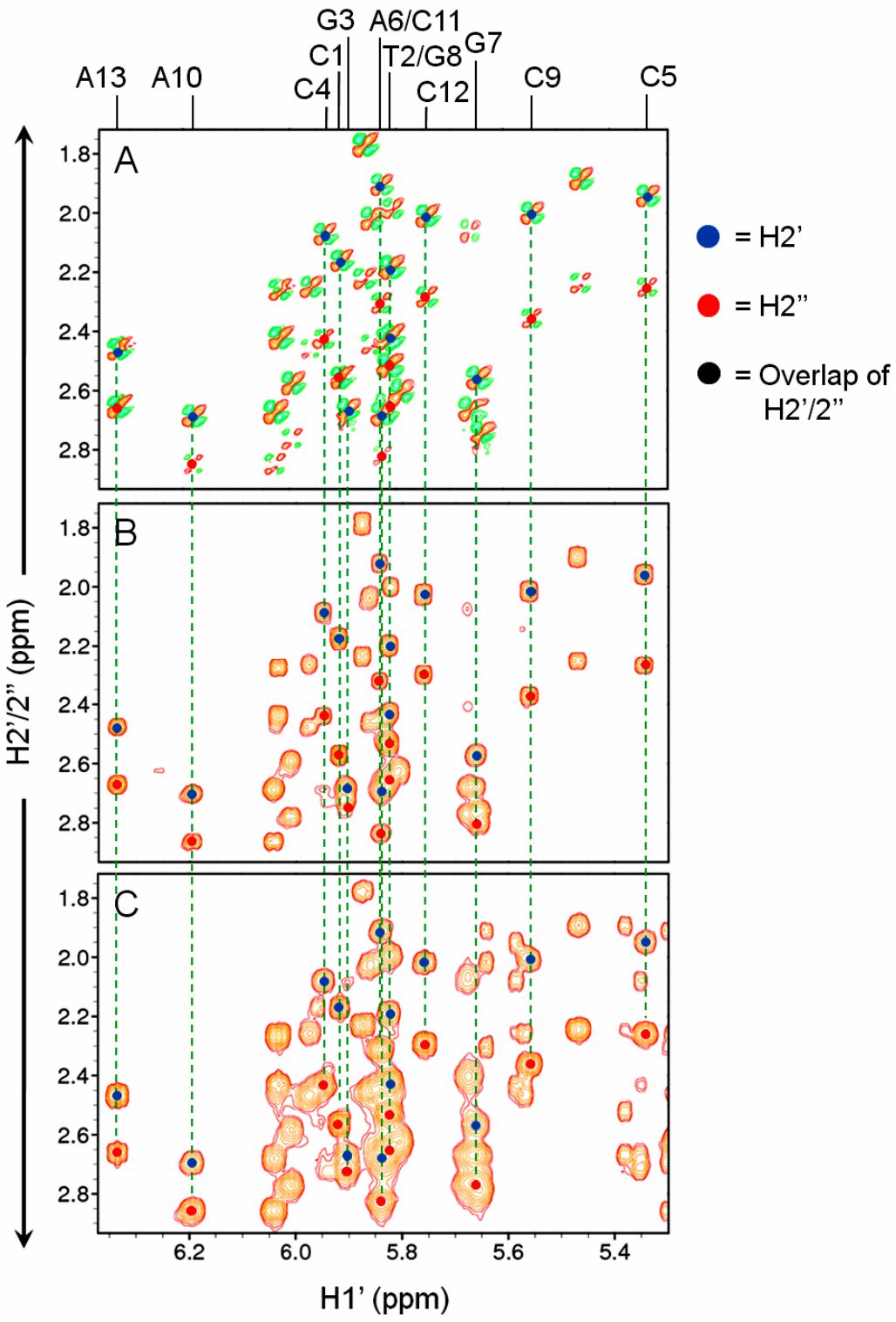


Figure 4.10. 800 MHz spectrum stack showing identification and assignment of H2'/H2'' resonances for C1-A13 in 6-thioguanine modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. DQF-COSY, B. TOCSY (SL_{mix} = 75ms) and C. NOESY (τ_m = 250ms). The unassigned peaks observed in this region correspond to H1'-H2'/2'' correlations for residues T14 to G26, which have been assigned and are shown in the Appendix.

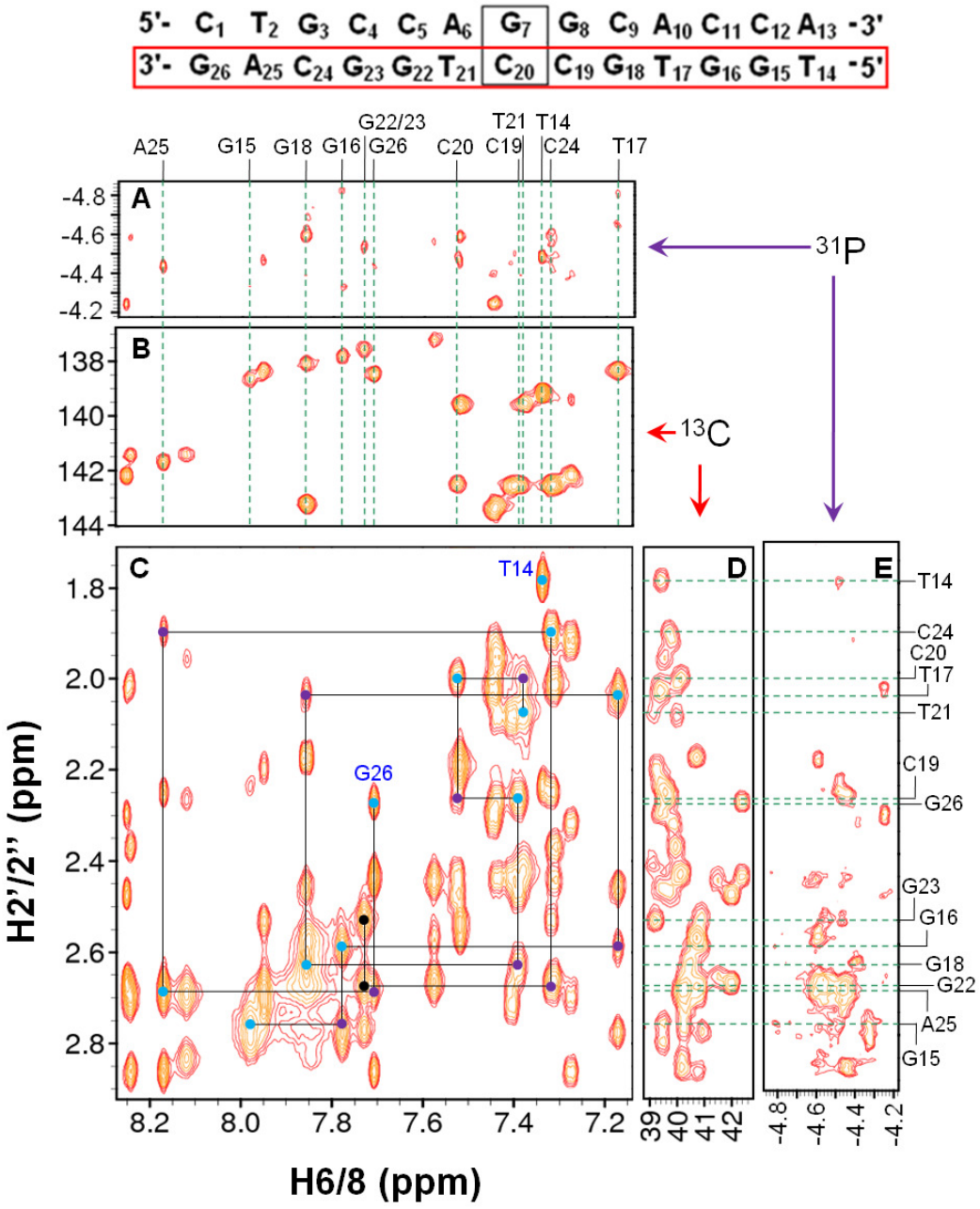


Figure 4.11. 800 MHz NOESY (τ_m = 250ms) spectrum (C) showing H6/8-H2' sequential NOE connectivities for residue T14-G26 in 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intraresidue connectivities are shown by blue circles and interresidue connectivities by purple circles. 800 MHz ^{1}H-^{13}C-HSQC spectra show the corresponding assignment for C6/8 (B) and C2' (D) resonances. 600 MHz ^{1}H-^{31}P CPMG-HSQC-NOESY (τ_m = 500ms) spectra show the corresponding assignment for backbone ^{31}P resonances through H6/8-^{31}P (A) and H2'-^{31}P (E) NOE connectivities respectively. The unassigned peaks shown above in the region correspond to the H6/8-H2' sequential connectivity between C1 and A1 and the H6/8-H2'' sequential connectivity for both strands. All the resonances have been assigned and have been shown in the Appendix.

As mentioned assigning the H3', H4' and H5'/5'' sugar resonances is much simpler by identifying of H2'-H3', H3'-H4' and H4'-H5' ^{3}J scalar coupled correlations in the DQF-COSY or TOCSY spectrum as these give intense cross-peaks. Although the DNA molecule was 13 nucleotides long, there was still a significant level of overlap for H3'-H4' and H4'-H5'/5'' connectivities. H2'-H3' connectivities were used to assign a number of H3' resonances and this is further explained below.

Assignment of H3' resonances

The H3' resonances were assigned using the H6/8-H3' sequential connectivity. H3' protons resonate between 4.5-5.2 ppm and the most immediate observation in the NOESY spectrum is the clear separation between the four terminal base H6/8-H3' cross peaks observed at approximately 4.6 ppm. These were assigned based on the chemical shift of the respective aromatic H6/8 protons. As seen with in Chapter 3, there is a clear divide between the chemical shift of purine and pyrimidine H3' protons. Seventeen H3' resonances were assigned in this region of the NOESY out of a possible thirty four. The other H3' resonances, which were all pyrimidine residues coincidentally, were assigned using H2'-H3' NOEs. In this region pyrimidine H2' resonances appear further highfield and are more resolved. Correlating the H3' resonances assignment with the ^{13}C HSQC spectrum identified the C3' resonances which appear between 70-85 ppm.

Assignment of H4' and H5'/5'' resonances

H4' resonances were assigned by first identifying H6/8-H4' NOEs. The majority of assignments made at this point would be those of purines as they give clear NOE cross peaks. The remaining H4' resonances were assigned through H1'-H4' cross peaks in the NOESY spectrum. The DQF-COSY and TOCSY spectra were less helpful for this task due to heavy overlapping between H1'-H4' and H1'-H5'/5'' correlations. Further to this, in the NOESY spectrum cross peaks to H5'/5'' protons appeared very weak or were not observed resulting in a large number of unassigned resonances. Using the ^{13}C-HSQC to assist in identifying H5'/5'' resonances was difficult as the C5' region was heavily overlapped. The spectra showing the assignment of H4' and H5'/5'' resonances are given in the appendix.

4.1.3 Assignment of phosphorus resonances

A combination of ^{1}H decoupled 1D ^{31}P-NMR (not shown) and 2D ^{1}H-^{31}P CPMG-HSQC-NOESY experiments were used to identify and assign ^{31}P resonances. In the CPMG-HSQC-NOESY, the most intense peaks are those from H3'-^{31}P and H5'/5''-^{31}P correlations but the resonances of H4' overlap heavily with the H5'/5'' and consequently, the assignment of ^{31}P resonances was based on H3'-^{31}P correlations (Figure 4.12).

H3'-^{31}P correlations were observed for all nucleotides in the DNA allowing complete assignment of ^{31}P resonances. The peak intensities of H3'-^{31}P cross peaks for C20 and

T21 were much weaker than all other nucleotides. These two bases are located near the G7 which may cause greater distortion in the phosphate backbone leading to the observed lower peak intensity. The conformational analysis carried out on the NMR structure, which is shown later in this chapter, should give indications as to whether this is the case.

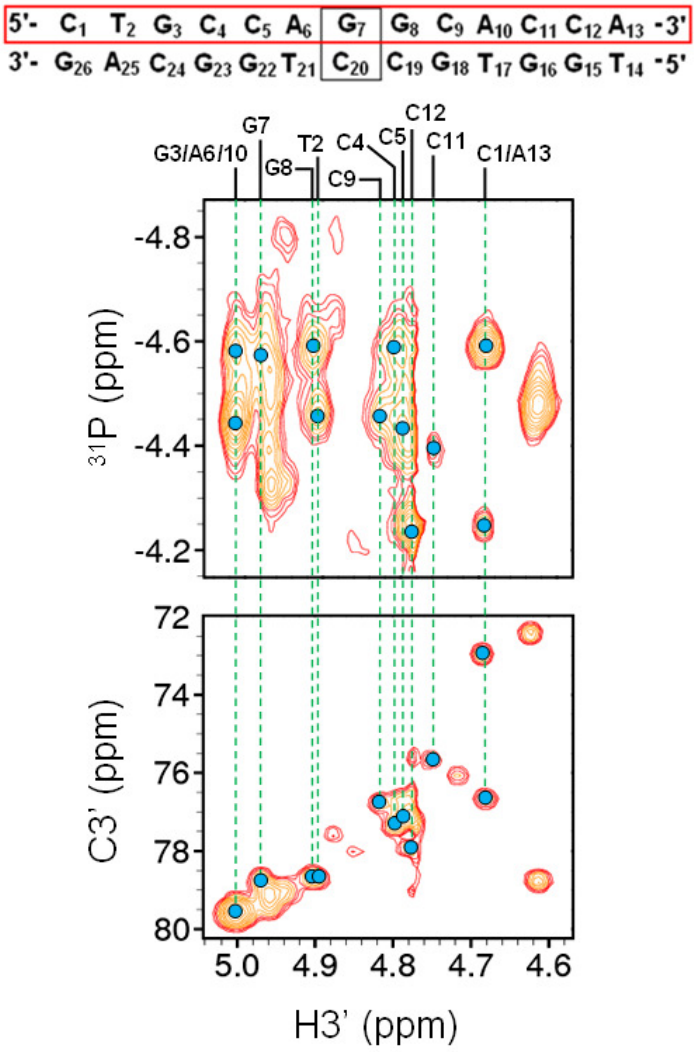


Figure 4.12. (Top panel) 600 MHz ^{1}H-^{31}P CPMG-HSQC-NOESY (τ_{mix} = 500ms) spectrum and (bottom panel) 600 MHz ^{1}H-^{13}C HSQC spectrum. The assignment of ^{31}P resonances is shown by means of H3'-^{31}P HSQC correlations for residues C1-A13 in 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) measured at 25°C. The unassigned peaks correspond to NOESY and HSQC correlations for residues T14 and G26. The assignment for the other ^{31}P resonances is shown in the Appendix.

4.1.4 Table of assignments

Based on the identification and assignment procedures outlined in the methodology, the chemical shifts of exchangeable and non-exchangeable proton, carbon and phosphorus resonances are summarised in Tables 4.1 and 4.2.

Nucleotide Number	Imino H1	Amino $NH_2(1)$	Amino $NH_2(2)$	H2	CH_3 (5)	H5	H6	H8	H1'	H2'	H2"	H3'	H4'	H5'	H5"
C1		7.86	7.12			5.96	7.86		5.92	2.18	2.57	4.68	4.12	3.78	3.82
T2	13.70				1.71		7.52		5.82	2.20	2.53	4.90	4.20	4.07	
G3	12.85	7.06	7.39					7.95	5.90	2.69	2.71	5.00	4.40	4.09	4.16
C4						5.35	7.41		5.95	2.08	2.44	4.80	4.23		
C5		8.44	6.35			5.58	7.44		5.34	1.96	2.27	4.80	4.05		
A6		8.01	6.18	7.56				8.12	5.83	2.69	2.83	5.00	4.35	4.04	
G7	11.54							7.85	5.66	2.57	2.78	4.97	4.37	4.16	
G8	12.79	6.89	7.33					7.58	5.83	2.44	2.66	4.91	4.36	4.17	
C9		8.20	6.31			5.28	7.31		5.56	2.01	2.37	4.82	4.14		
A10		7.84	6.09	7.71				8.24	6.19	2.70	2.86	5.01	4.41	4.12	4.26
C11		8.11	6.60			5.30	7.27		5.84	1.92	2.32	4.75	4.12		
C12		8.31	6.44			5.64	7.45		5.76	2.02	2.30	4.79	4.07		
A13		7.97	6.41	7.70				8.25	6.34	2.48	2.67	4.69	4.19	4.08	
T14	13.99				1.60		7.34		5.87	1.79	2.24	4.62	4.03	3.65	3.65
G15	12.76	7.12	7.26					7.98	5.65	2.76	2.79	4.96	4.34	3.97	4.05
G16	12.76	7.10	7.06					7.78	6.01	2.59	2.78	4.94	4.42	4.19	
T17	13.70				1.37		7.17		5.86	2.03	2.46	4.88	4.19	4.15	4.22
G18	12.65	7.25						7.86	5.80	2.63	2.63	4.97	4.35		
C19		7.96	6.54			5.29	7.39		5.97	2.26	2.47	4.79	4.26		
C20		8.43	6.89			5.57	7.52		5.82	2.00	2.43	4.72	4.10		
T21	13.89				1.62		7.38		5.68	2.08	2.41	4.85	4.14	4.06	4.04
G22	12.89	6.82	7.36					7.73	5.67	2.68	2.77	4.99	4.36	4.04	4.11
G23	12.77							7.73	5.82	2.53	2.67	4.96	4.37		
C24		8.47				5.38	7.32		5.47	1.90	2.25	4.79	4.09		
A25			6.59	7.70				8.17	6.04	2.69	2.86	5.01	4.37	3.99	4.10
G26	13.13							7.71	6.03	2.27	2.44	4.63	4.17	4.12	4.24

Table 4.1. Summary of chemical shifts for ^{1}H proton resonances in 6-TG modified 13mer GC DNA. Boxes shaded in grey indicate that the absence of the proton in the specified residue and those shaded in blue indicate unassigned protons.

Nucleotide Number	C2	C5	C6	C8	C1'	C2'	C3'	C4'	C5'	^{31}P
C1		99.3	143.3		88.4	40.7	76.7	88.2	63.0	-4.59
T2		14.6	139.6		87.0	39.3	78.7	86.4		-4.46
G3				138.4	84.3	40.3	79.6	87.2		-4.45
C4		98.5	142.6		86.6	40.0	77.3	85.8		-4.58
C5		98.7	143.3		86.4	39.6	77.2	85.7		-4.45
A6	154.4			141.4	84.8	40.2	79.6	87.6		-4.45
G7				138.0	83.9	40.8	78.8	87.6		-4.57
G8				137.2	84.8	41.5	78.7	87.2		-4.59
C9		98.1	142.5		86.2	39.9	76.8	85.6		-4.46
A10	154.9			141.5	85.2	40.7	79.6	87.4		-4.59
C11		98.5	142.2		86.0	39.8	75.7	85.1		-4.39
C12		98.8	143.5		88.7	39.4	77.9	85.8		-4.24
A13	154.8			142.2	85.7	42.0	73.0	88.0		-4.25
T14		14.1	139.2		87.5	39.4	78.8	88.8	64.0	-4.52
G15				138.6	84.7	39.6	79.1	87.6		-4.59
G16				137.8	84.9	40.9	79.0	87.2		-4.33
T17		14.7	138.4		85.5	39.4	77.6	85.7		-4.64
G18				138.1	84.3	40.5	79.1	87.6		-4.57
C19		98.2	142.5		86.7	39.6	76.8	86.0		-4.39
C20		98.7	142.5		86.1	40.2	76.1	85.4		-4.49
T21		14.1	139.6		85.9	39.2	78.0	86.0		-4.22
G22				137.6	84.3	40.3	79.6	87.4		-4.54
G23				137.6	84.4	40.9	79.2	87.0		-4.32
C24		98.3	142.6		86.2	39.8	77.2	85.4		-4.42
A25	154.8			141.7	85.1	40.2	79.6	87.0		-4.45
G26				138.5	84.3	42.4	72.4	87.4		-4.47

Table 4.2. Summary of chemical shifts for ^{13}C and ^{31}P resonances in 6-TG modified 13mer GC DNA. Boxes shaded in grey indicate that the absence of the resonance in the specified residue. Boxes shaded in red indicate that the chemical shifts of C5' resonances are highly overlapped. The indicated C5' resonances all appear between 65.9-69.1 ppm.

4.1.5 Charge calculations

The substitution of the O6 atom with a sulphur in the 6-TG base meant that the electron density across the base changed and thus, the atom charges would differ compared to a normal guanine base.

To solve this problem charge calculations were carried out on the 6-TG base using the program Gaussian03W. This carries out *ab initio* calculations to the molecule of interest to generate its properties from first principles. This is in contrast to the empirical techniques used in molecular mechanics. For the 6-TG base two types of calculations were used; Møller-Plesset, and B3LYP.

Møller-Plesset (MP) methods are based on Hartree-Fock methods but with an additional correction, known as the Møller-Plesset correlation energy correction. This takes electron correlation into account, using perturbation theory. If the correction is curtailed at the second order the calculation is known as MP2 and at higher orders MP3, MP4 and so on.

The B3LYP method is a Hybrid functional method. It is a combination taking components of both Hartree-Fock and Density functional methods. The Hartree-Fock method is an approximation method for solving the Schrödinger equation and Density functional methods simplify the Hartree-Fock method by considering electron density as opposed to electron wave functions. Combining aspects of two methods increase the accuracy and speed of calculations carried out. B3LYP methods generate correct

conformations for energy minimised structures well, but are not as good at obtaining correct energies. This drawback was not serious for 6-TG calculations as the resulting properties generated were incorporated into the Xplor-NIH structure determination method. The calculated charges for the 6-TG base are shown in Figure 4.13.

Atom No.	Atom Type	Charge
1	H	0.446
2	O	-0.608
3	C	-0.032
4	H	0.124
5	H	0.057
6	C	0.104
7	H	0.111
8	O	-0.467
9	C	0.736
10	H	0.034
11	N	-0.384
12	C	0.345
13	H	0.080
14	N	-0.657
15	C	0.384
16	C	0.051
17	N	-0.476
18	H	0.316
19	C	0.889
20	N	-1.106
21	H	0.480
22	H	0.461
23	N	-0.695
24	C	0.276
25	C	0.421
26	H	0.024
27	C	-0.622
28	H	0.196
29	H	0.116
30	O	-0.745
31	H	0.459
32	S	-0.318

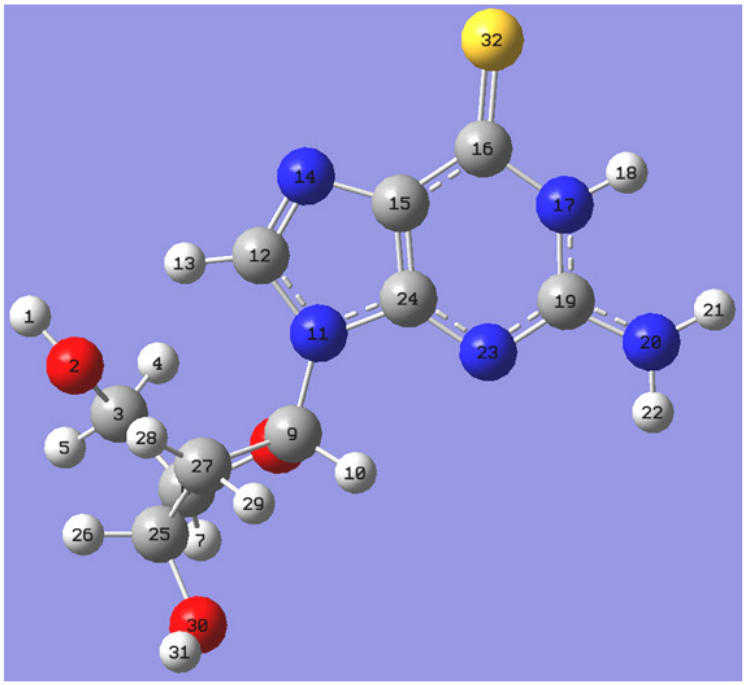


Figure 4.13. Summary of atom charges calculated from MP2 and B3LYP calculations for 2'-deoxy-6-thioguanine carried using the Gaussian03W program.

As the summary of atomic charges shows, the results correlate well to those published by Guerra *et al.*[42] and Villani[43], with the sulphur being less electronegative than most of the oxygen atoms in the structure. From this, it can be hypothesised that during the

structure calculations, the base pairing involving the 6-TG base should orientate itself with an elongated S6•••H41-N hydrogen bond and a possible compression of the O2•••H21 hydrogen bond and that the base pair may appear more opened or even buckled.

4.1.6 Structure calculation

Experimental distance and dihedral angle constraints were obtained using CcpNMR analysis as outlined in Chapter 3.1.5. The number of NMR experimental constraints used is summarised in Table 4.3 for the structure calculation of 6-TG modified 13mer GC DNA carried out using the Xplor-NIH program.

Constraint Type	No. of constraints
Total NOEs	368
Strong (1.5 - 2.5 Å)	30
Medium (2.6 – 4.0 Å)	311
Weak (4.1 - 6.0 Å)	27
Hydrogen bond	78
Total dihedral angle	252
Helix (α, β, γ, δ, ε, ζ)	174
Glycosidic (χ)	26
Ribose pucker (υ_1, υ_2)	52
NOE constraints per residue	14.1
NOE and dihedral constraints per residue	23.8

Table 4.3. Summary of NMR experimental constraints used in the structure calculation of 6-TG modified 13mer GC DNA carried out with the Xplor-NIH program.

As with the structure determination of the 17mer DNAs in Chapter 3, no constraints were included for H4', H5' and H5'' NOEs as they were too overlapped in the spectrum.

For the same reason, NOE correlations between sugar protons were not included in the structure calculation with the exception of H1'-H2'/2'', H1'-H3' and H2'/2''-H3' whose NOEs were well-resolved.

From the 100 structures generated from Xplor-NIH, ten were selected based on the criteria of lowest energy and lowest RMSD. The average structure was then generated from these ten structures and the results of these two stages are shown in Figure 4.14. The best ten structures show a good overlay (RMSD = 0.46) of B-form conformational DNA duplexes (Figure 4.14, A).

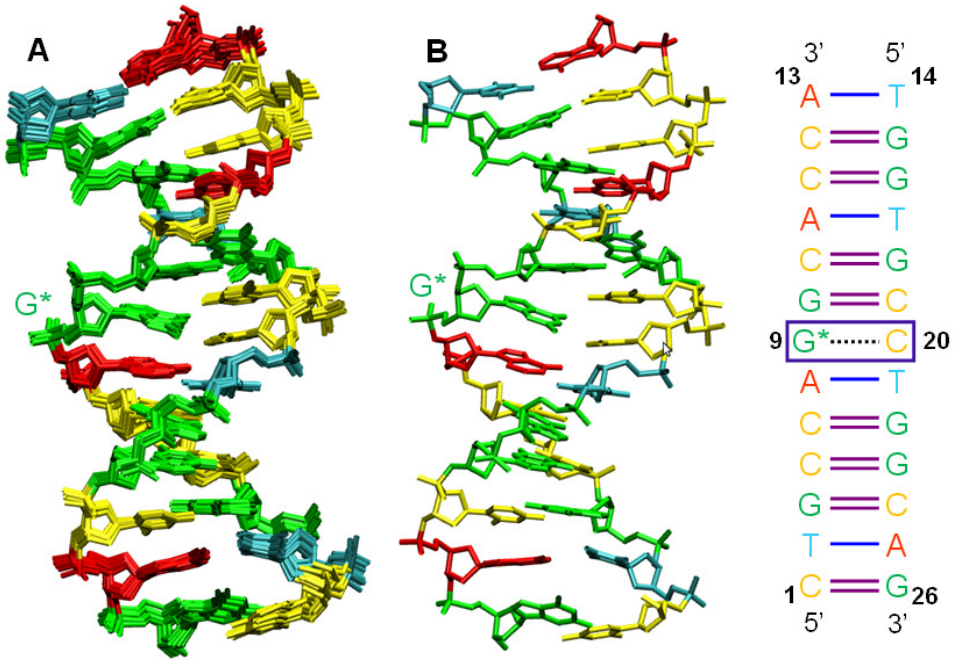


Figure 4.14. Illustration showing an overlay of the ten lowest energy and lowest RMSD and energy structures (A) and the final structure (B) for 6-TG modified 13mer GC DNA calculated from Xplor-NIH. Sequence is shown on the right and the position of the 6-TG (G*) base is highlighted on the structures.

4.1.7 Conformational analysis

Conformational analysis was carried out using 3DNA on the final structure obtained from Xplor-NIH (Figure 4.14, B). Similar to the 17mer DNA structures (Chapter 3.1.6), helical parameters were analysed to monitor any major changes in the structure (Tables 4.4 - 4.6). The dihedral angles were also given by 3DNA and these were analysed also (Tables 4.7 - 4.8).

Since the inclusion of a modified 6-TG base was the only selective change in the structure compared to canonical GC DNA, the helical parameters around this base were highlighted. The values obtained for the buckle and opening parameters were higher than other base pairs confirming that the modified base pair is wider, as reported in previous literature.[40] Despite the wider base pair, the propeller twist value did not increase by a significant amount so good planarity was maintained. The combination of these changes could be crucial when considering the interaction of 6-TG modified GC DNA with *E. coli* MutS. A wider opening may allow an access route for the protein-DNA complex to form.

The backbone dihedral angles given by 3DNA showed good consistency with little deviation from the input values. The values for χ varied the most but this was expected as greater error bounds were given for this angle. For the sugar dihedral angles, the majority of nucleotides were $C_{2'}$-*endo* in conformation with nine having a $C_{1'}$-*exo* pucker. $C_{1'}$-*exo* is close to $C_{2'}$-*endo* so the overall structure does not deviate much from B-form DNA.

Base Pair	Nucleotides	Shear (S_x)	Stretch (S_y)	Stagger (S_z)	Buckle (κ)	Propeller Twist (ω)	Opening (σ)
1	C-G	-0.83	-1.65	-2.25	5.81	21.23	-35.89
2	T-A	-0.06	-0.10	1.19	-21.78	-1.43	0.80
3	G-C	0.18	-0.45	1.34	12.74	15.87	-7.26
4	C-G	-0.09	-0.06	0.35	-16.11	3.54	3.81
5	C-G	0.12	-0.22	0.25	-14.47	-4.70	-9.10
6	A-T	-0.12	-0.23	1.26	19.92	-13.39	-2.85
7	G-C	0.20	-0.55	1.55	23.53	-13.09	-19.98
8	G-C	0.55	-0.75	1.58	-5.51	-2.34	-7.01
9	C-G	0.03	-0.29	1.02	-30.14	5.31	-3.84
10	A-T	-0.20	-0.16	0.83	3.92	4.16	3.62
11	C-G	-0.19	-0.47	1.00	4.57	16.72	0.55
12	C-G	-0.10	-0.30	1.17	-12.91	-4.47	0.35
13	A-T	-0.05	-0.28	0.62	9.57	1.44	-4.48
Average		-0.04	-0.42	0.76	-1.60	2.22	-6.25

Table 4.4. Helical parameters describing local base pair relationships in the 6-TG modified 13mer GC DNA NMR structure. Parameters were calculated using 3DNA software from the average structure shown in Figure 4.14, B.

Step No.	Step	Shift (D_x)	Slide (D_y)	Rise (D_z)	Tilt ($T\tau$)	Roll ($P\rho$)	Helical Twist (Ω)
1	CT / AG	2.71	-0.68	4.90	-10.78	11.02	40.54
2	TG / CA	-0.56	-0.78	2.05	-2.17	-1.93	35.81
3	GC / GC	1.06	-1.61	4.29	3.21	13.00	33.87
4	CC / GG	-1.94	-1.27	2.68	2.34	3.15	33.58
5	CA / TG	0.01	-0.26	2.40	-7.48	-9.75	36.56
6	AG / CT	-0.67	-0.73	3.48	1.24	-2.11	34.10
7	GG / CC	1.26	-1.33	3.97	-3.20	15.88	39.56
8	GC / GC	0.11	-1.70	4.05	0.39	2.00	34.76
9	CA / TG	0.82	-0.77	1.90	3.05	-3.37	35.96
10	AC / GT	-0.35	-1.35	3.18	-5.22	7.57	33.42
11	CC / GG	0.03	-1.42	3.60	5.05	17.13	33.00
12	CA / TG	-0.64	-0.73	2.45	-0.02	3.26	35.26
Average		0.15	-1.05	3.25	-1.13	4.65	35.54

Table 4.5. Helical parameters describing local base pair step parameters in the 6-TG modified 13mer GC DNA NMR structure. Parameters were calculated using 3DNA software from the average structure shown in Figure 4.14, B.

Step No.	Step	X-displacement (d_x)	Y-displacement (d_y)	Inclination (η)	Tip (θ)
1	CT / AG	-2.41	-5.13	15.26	14.92
2	TG / CA	-1.09	0.72	-3.13	3.51
3	GC / GC	-4.83	-1.14	21.31	-5.26
4	CC / GG	-2.59	3.63	5.42	-4.03
5	CA / TG	0.50	-0.71	-15.03	11.53
6	AG / CT	-0.89	1.36	-3.59	-2.12
7	GG / CC	-3.64	-2.10	22.36	4.50
8	GC / GC	-3.22	0.25	3.35	-0.65
9	CA / TG	-0.95	-1.06	-5.44	-4.92
10	AC / GT	-3.36	-0.17	12.85	8.87
11	CC / GG	-4.48	0.63	27.77	-8.18
12	CA / TG	-1.55	1.05	5.37	0.04
Average		-2.38	-0.22	7.21	1.52

Table 4.6. Helical parameters describing local base pair helical parameters in the 6-TG modified 13mer GC DNA NMR structure. Parameters were calculated using 3DNA software from the average structure shown in Figure 4.14, B.

Dihedral Angle	Average Value		
	Dickerson 12mer DNA	**17mer canonical GC DNA**	**6-TG modified 13mer GC DNA**
α	-62.8	-64.2	-75.7
β	-169.3	-169.1	-154.1
γ	59.5	48.3	56.7
δ	122.7	137.6	135.9
ε	164.2	178.0	172.1
ζ	-105.0	-99.8	97.2
χ	-117.3	-110.2	-117.4

Table 4.7. Comparison of backbone dihedral angles between Dickerson DNA (PDB: 1BNA), 17mer canonical GC DNA and 6-TG modified 13mer GC DNA. Values were calculated using 3DNA and for 17mer canonical GC DNA are based on the average structure shown in Chapter 3.3.5. For 6-TG modified 13mer GC DNA, the values are based on the average structure given in Figure 4.14.

Dihedral Angle	Average Value		
	Dickerson 12mer DNA	17mer canonical GC DNA	6-TG modified 13mer GC DNA
ν_0	-38.7	-22.4	-28.0
ν_1	40.8	32.3	36.4
ν_2	-30.5	-29.6	-30.5
ν_3	17.3	17.6	15.2
ν_4	21.7	2.8	7.8
Amplitude	46.4	32.3	36.1
Phase	127.6	156.5	149.0

Table 4.8. Comparison of sugar ring dihedral angles between Dickerson DNA (PDB: 1BNA), 17mer canonical GC DNA and 6-TG modified 13mer GC DNA. Values were calculated using 3DNA and for 17mer canonical GC DNA are based on the average structure shown in Chapter 3.3.5. For 6-TG modified 13mer GC DNA, the values are based on the average structure given in Figure 4.14.

4.2 NMR study of 6-sulphonateguanine modified 13mer DNAs

As mentioned at the start of this chapter, 6-sulphonateguanine or 2-aminopurine-6-sulfonic acid is the oxidised form of 6-thioguanine which is obtained due to prolonged exposure to UVA light. The conversion is shown in Figure 4.15. In this section, the effect of the G^{SO3} group is presented primarily through the change in chemical shift of the base paired imino proton. The 1D spectrum of the imino region for each DNA molecule is shown and a table giving the chemical shifts of the imino proton resonances is presented at the end (Table 4.9).

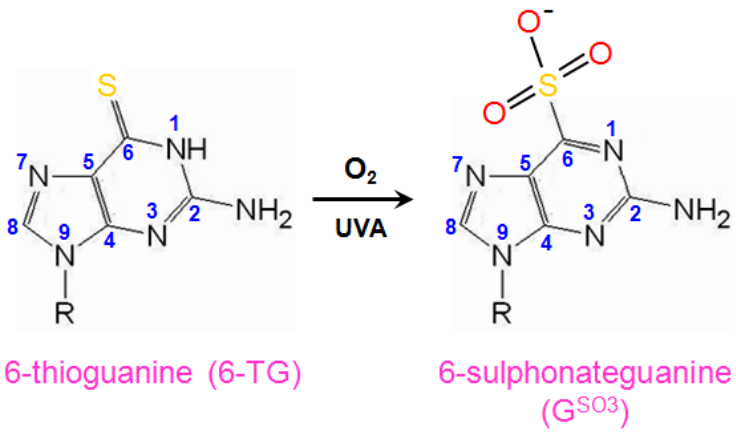


Figure 4.15. Illustration showing the chemical structures of 6-thioguanine and 6-sulphonateguanine bases and the oxidation conversion process between the two.

As the chemical structures show there is no imino proton in G^{SO3} and this absence was clearly observed in the imino region of the 1D ^{1}H-NMR spectrum of G^{SO3} modified 13mer GC DNA (Figure 4.16).

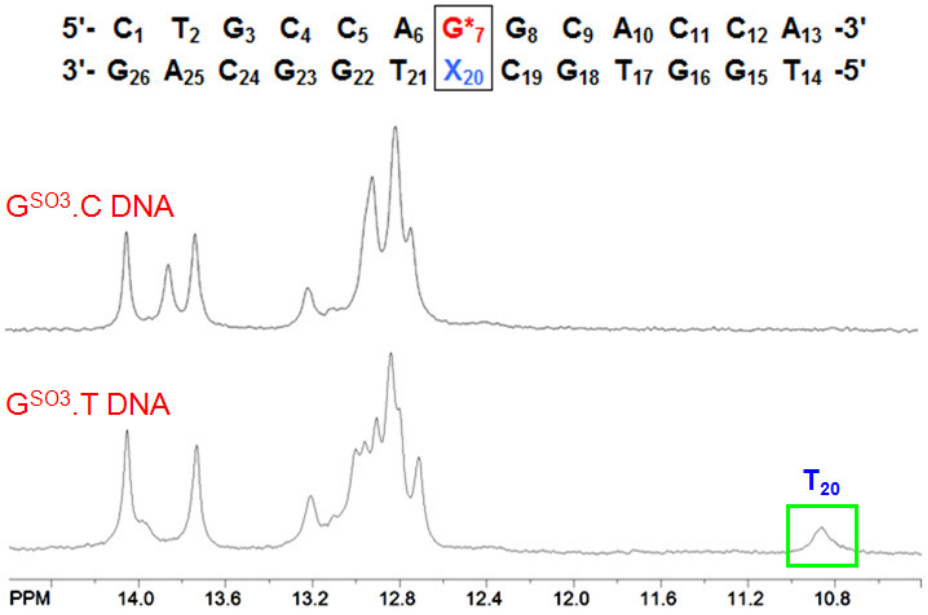


Figure 4.16. 400 MHz stack of 1D ^{1}H-NMR spectra showing comparison between G^{SO3} modified GC and GT base pairs measured at 5°C. (Top) G^{SO3} modified 13mer GC DNA (0.2mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) and (bottom) G^{SO3} modified 13mer GT DNA (0.2mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2). The highfield shifted imino proton resonances of the mispaired T is highlighted by the green box.

It was evident that with the exception of the modified G^{SO3} base, the rest of the imino region remains relatively unchanged. There is broadening of some peaks but the chemical shifts remain the same with the exception of T21 which has shifted lowfield upon introduction of a mispaired T base. The low peak intensity also suggests that it is exchanging very rapidly with the solvent. Further analysis using 2D spectra such as TOCSY and NOESY will provide more accurate indications of changes in the structure conformation through the pattern of cross peaks compared to 6-TG. The assignment of imno protons for G^{SO3} modified 13mer GC DNA is shown in Figure 4.17.

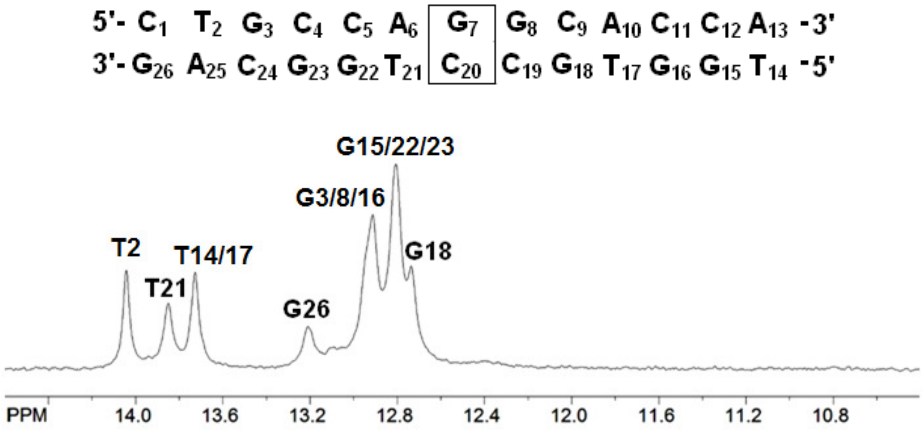


Figure 4.17. 400 MHz 1D ^{1}H-NMR spectrum showing imino proton region in G^{SO3} modified 13mer GC DNA (0.2mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C. The assignment of imino proton resonances is shown and is based on the same region of 6-TG modified 13mer GC DNA.

A sample of G^{SO3} modified mismatch GT DNA was also studied. It has already been shown in Chapter 3.1 that a GT mismatch will appear further highfield in the imino region due to the disruption in canonical base pairing and stacking and the same was observed for G^{SO3} modified GT DNA.

The 1D ^{1}H-NMR spectrum of G^{SO3} modified GT DNA showed the highfield shifted imino proton corresponding to the mismatched T base (T20, δ = 10.77 ppm) and the absence of the imino proton of G7. Similarly to G^{SO3} modified GC DNA, chemical shifts of the other imino protons remained relatively unchanged showing good stability in the helical structure (Figure 4.18).

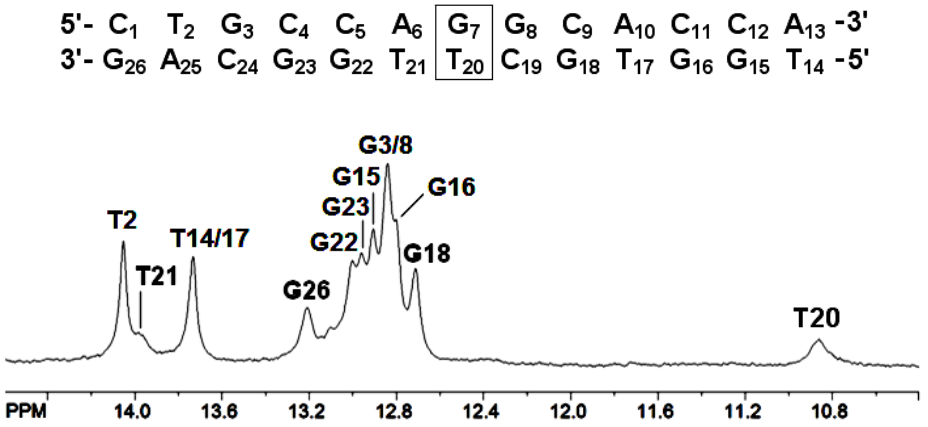


Figure 4.18. 400 MHz 1D ^{1}H-NMR spectrum showing imino proton region in G^{SO3} modified 13mer GT DNA (0.2mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C. The assignment of imino proton resonances is shown and is based on the same region of 6-TG modified 13mer GC DNA.

T20 appears broad indicating faster exchange between the imino proton with the solvent. A significant finding was that the T20 resonates at 10.83 ppm, which is close in chemical shift to the guanine residue when a GT mismatch is incorporated as shown in Chapter 3. To further compare, the thymine residue in the 17mer GT mismatch had a chemical shift of 11.80 ppm, a difference of 1 ppm. This gives strong evidence that the base stacking at the G7-T20 modified base pair is extremely disrupted. However, the lowfield shifting of T21 suggests that there is still some degree of base stacking to T20, an effect similar to that seen for the 17mer mismatched GT DNA. The broad

appearance indicates a much faster rate of exchange with the solvent than observed in GSO3 modified 13mer GC DNA.

Nucleotide Number	δ Imino Proton (ppm)		
	6-TG.C	G^{SO3}.C	G^{SO3}.T
T2	13.70	14.02	13.99
G3	12.85	12.78	12.78
G7	11.54		
G8	12.79	12.78	12.78
T14	13.99	13.70	13.67
G15	12.76	12.89	12.84
G16	12.76	12.78	12.74
T17	13.70	13.70	13.67
G18	12.65	12.71	12.65
T20			10.80
T21	13.89	13.82	13.91
G22	12.89	12.89	12.94
G23	12.77	12.89	12.90
G26	13.13	13.18	13.15

Table 4.9. Comparison of imino proton chemical shifts in 6-TG and G^{SO3} modified 13mer DNAs. Boxes shaded in grey indicate that the imino proton is absent from the DNA molecule.

4.3 Interaction of 6-thioguanine modified 13mer GC DNA with *E. coli* MutS

The affinity of *E. coli* MutS towards mismatch DNAs has already been documented in literature[46,51,59] and demonstrated in Chapter 3.6. Affinity of MutS towards modified DNAs is less known and it is not clear if the mode of binding is the same for modified base pairs as it is for mismatch base pairs.

The NMR titration of 6-TG modified 13mer GC DNA is shown in Figure 4.19. The results show that there was an interaction with *E. coli* MutS as the peak intensity has been reduced for the G7 imino proton peak. The peak has also become more broad

which further gives supporting evidence of an interaction. An increased linewidth suggests a more rapid exchange between the G7 imino proton and the solvent.

A binding interaction between the protein will induce some conformational change in the local area at the modified base pair site and this will promote the faster exchange. With 17mer mismatch GT DNA, a reduction in the line intensity and increased linewidth was clearly observed for the T residue (T27) immediately flanking the GT mispair; the same effect is not observed for 6-TG modified 13mer GC DNA. The peak for the equivalent flanking T imino proton (T21) does not show a reduction in line intensity but remains relatively unchanged. This leads to the implication that the 6-TG modified GC base pair was more stable that GT mispair, which is expected as the 6-TG modified GC base pair behaves more like a canonical base pair.

The peak at approximately 12.50 ppm, which was identified as a possible secondary minor conformation of the modified G7 base, did not appear in the 1D 1H-NMR spectrum upon addition of MutS protein. This result is important as it demonstrates that the protein binds to the DNA and locks the DNA into a single conformation.

Whereas the imino proton of T21 remained unaffected by the addition of *E. coli* MutS, there was a dramatic effect of the imino protons of T2 and T14. The peak had reduced in intensity by almost half and was less intense than the peak for T21, but had also increased in linewidth slightly. Since both are towards the end of the molecule the result observed is most likely caused by rapid exchange with solvent, particularly with T14 as it is a terminal residue.

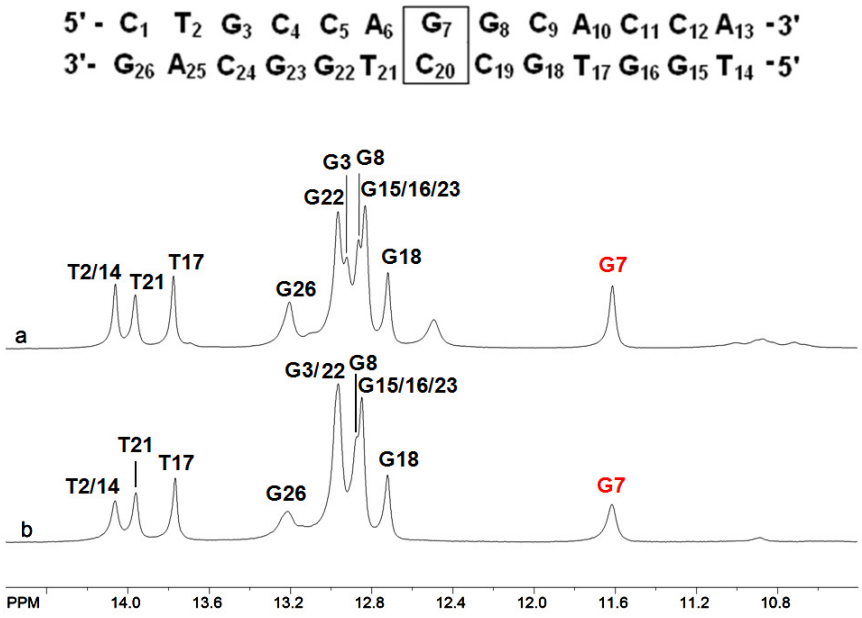


Figure 4.19. 800MHz 1D ^{1}H-NMR titration stack showing the effect of *E. coli* MutS addition to 13mer 6-thioguanine modified GC DNA (0.8mM, 90% $^{1}H_2O$+ 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 5°C where a. 13mer 6-thioguanine modified GC DNA and b. 13mer 6-thioguanine modified GC DNA + *E. coli* MutS (0.11 x 10^{-3} mM).

Although the change to the imino proton chemical shifts following titration was minor, there was a change in linewidth to a few peaks suggesting that the interaction between 6-TG modified 13mer GC DNA and *E. coli* MutS is weaker when compared to 17mer mismatch GT DNA.

4.4 Interaction of 13mer 6-sulphonateguanine modified GC and GT DNA with *E. coli* MutS

The oxidation of 6-TG to G^{S03} forms a base with a very bulky group at the 6-position. The result is that a base pair between G^{SO3} and C residues is very unlikely to form. This leaves potentially a much larger binding site than expected at the modified base pair for MutS protein to bind to.[137]

As with the 6-TG modified 13mer GC DNA, large changes to the chemical shift of imino proton resonances were not seen although a slight drop in line intensity was observed for residues T2/14. The imino proton resonance of G26 did appear broader but once again this was most likely due to rapid exchange with solvent (Figure 4.20).

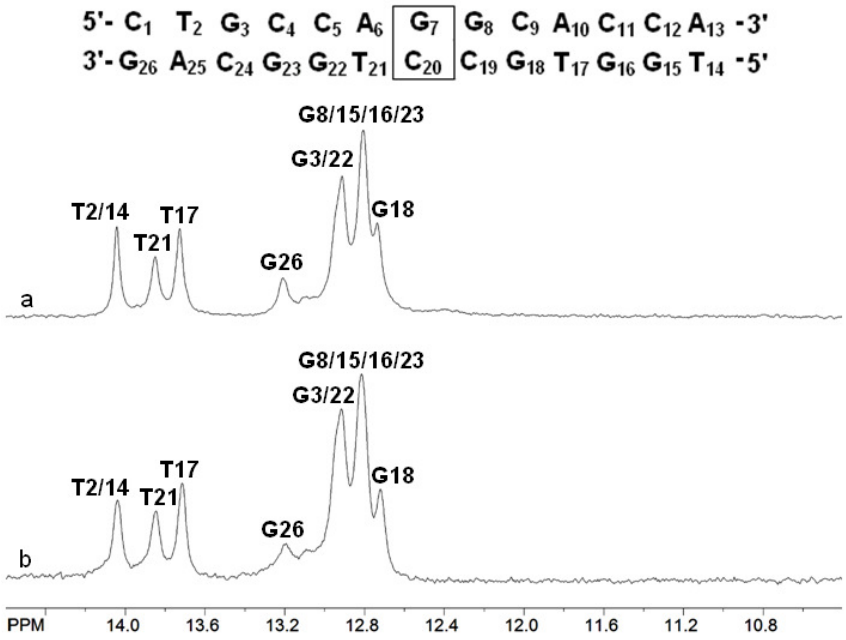


Figure 4.20. 400MHz 1D ^{1}H-NMR stack showing the effect of *E. coli* MutS addition to G^{SO3} modified 13mer GC DNA (0.2mM, 90% $^{1}H_2O$+ 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 5°C where a. 13mer 6-sulphonate guanine modified GC DNA and b. 13mer 6- sulphonateguanine modified GC DNA + *E. coli* MutS (0.11 x 10^{-3} mM).

Substitution of the cytosine residue base paired to the G^{SO3} base with a thymine residue means that the base pairing and stacking changes in a similar manner to a change from canonical GC to mismatch GT. It has already been shown in section 4.2 that the substitution of cytosine for a thymine residue results in an additional imino proton resonance which was observed in the 1D ^{1}H-NMR spectrum (Figure 4.18).

As there were no highfield shifted imino protons in G^{SO3} modified 13mer GC DNA, comparing the magnitude of binding to *E. coli* MutS using the imino region would not be useful. Figure 4.21 illustrates the differences in the dispersion of the imino protons between the two following the addition of MutS. When a modified G^{SO3}.T base pair is present, despite the non-canonical character of the base pair the imino protons were still well resolved. Even after the addition of MutS protein, the chemical shifts of the imino protons remain similar and the disappearance of the T20 imino proton shows a degree of selective binding to the modified base pair.

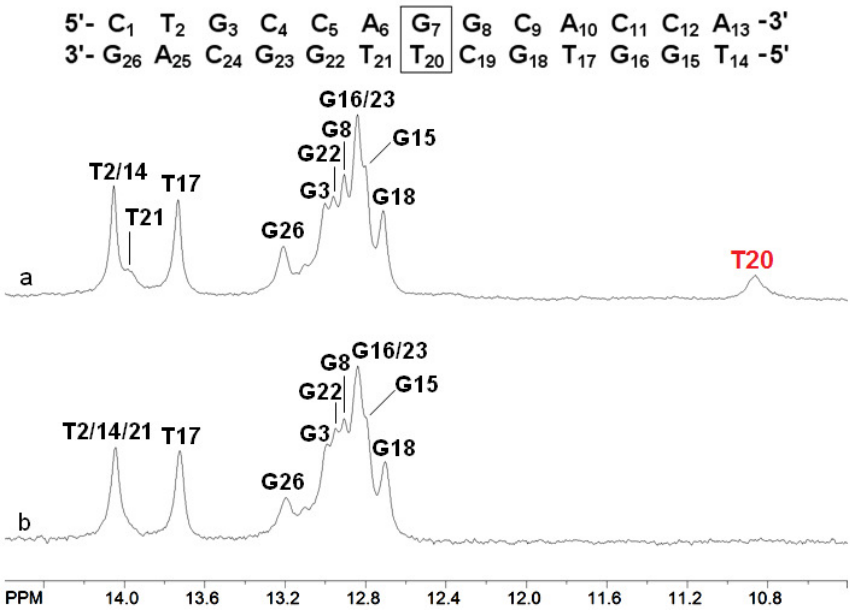


Figure 4.21. 400MHz 1D ^{1}H-NMR stack showing the effect of *E. coli* MutS addition to G^{SO3} modified 13mer GT DNA (0.2mM, 90% $^{1}H_2O$+ 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 5°C where a. G^{SO3} modified 13mer GT DNA and b. G^{SO3} modified 13mer GT DNA + *E. coli* MutS (0.11 x 10^{-3} mM).

The performed titrations provided evidence that the formation of mismatched DNA-MutS and modified DNA-MutS complexes could be observed by NMR upon addition of MutS. Selective changes to assigned G and T base pair imino proton resonances of the mismatch were observed proving that the interaction of *E. coli* MutS and the mismatch occurred in a specific manner.

Chapter 5 – Conformational analysis of free and MutS bound mismatch DNA and NMR characterisation of NTD-MutS

In this chapter, a comparison of the structure of *apo* and MutS bound GT mismatch DNA is described. However, as shown in Chapter 3.6, discrete changes were caused to the imino protons of GT mispairs of various GT DNAs upon the addition of full length MutS. In view of this NMR observation and the need to further probe the above interaction, the preparation of the N-terminal domain (NTD) of *E. coli* MutS (15 kDa, monomer) was carried out.

To date, no structure exists of the *apo* NTD of MutS protein which is directly involved in mismatch DNA binding. Consequently, vital structural knowledge regarding the conformational change of MutS upon binding to mismatch DNA is currently unknown. The molecular size of WT-*E. coli* MutS protein is significantly large (97 kDa, monomer) and its structure determination by NMR is beyond the application of current methodology and modern techniques. To simplify the spectrum of WT-*E. coli* MutS, the isolation and NMR characterisation of the 117 residue NTD of *E. coli* MutS (15 kDa, monomer) was carried out. Determination of the NMR solution structure of NTD-MutS should assist in the elucidation of the conformational change accompanying the binding to mismatch DNA by *E. coli* MutS protein.

5.1 Conformational analysis of the *apo* NMR structure of 17mer mismatch GT DNA and bound DNA in the WT-MutS complex

The NMR structure determination of the *apo* 17mer mismatch GT DNA has been determined in Chapter 3.1. The sequence characterised mirrors that which was used in the X-ray crystallography study of the 30mer mismatch GT DNA – *E. coli* MutS protein complex (~ 220 kDa) and this presents an opportunity to compare the *apo* NMR structure with the bound DNA crystal structure. It is important to point out that the DNA was structurally ordered only for the first sixteen base pairs and the remaining fourteen base pairs were invisible.

Helical parameters were calculated using the 3DNA software for each structure and the following section will give a comparison of the parameter values with major changes being highlighted. The two structures are shown in Figure 5.1 along with the corresponding sequence. As the degree of structural distortion is greatest at the site of mismatch only the values for the GT mispair and its immediately flanking four base pairs were considered during the comparison.

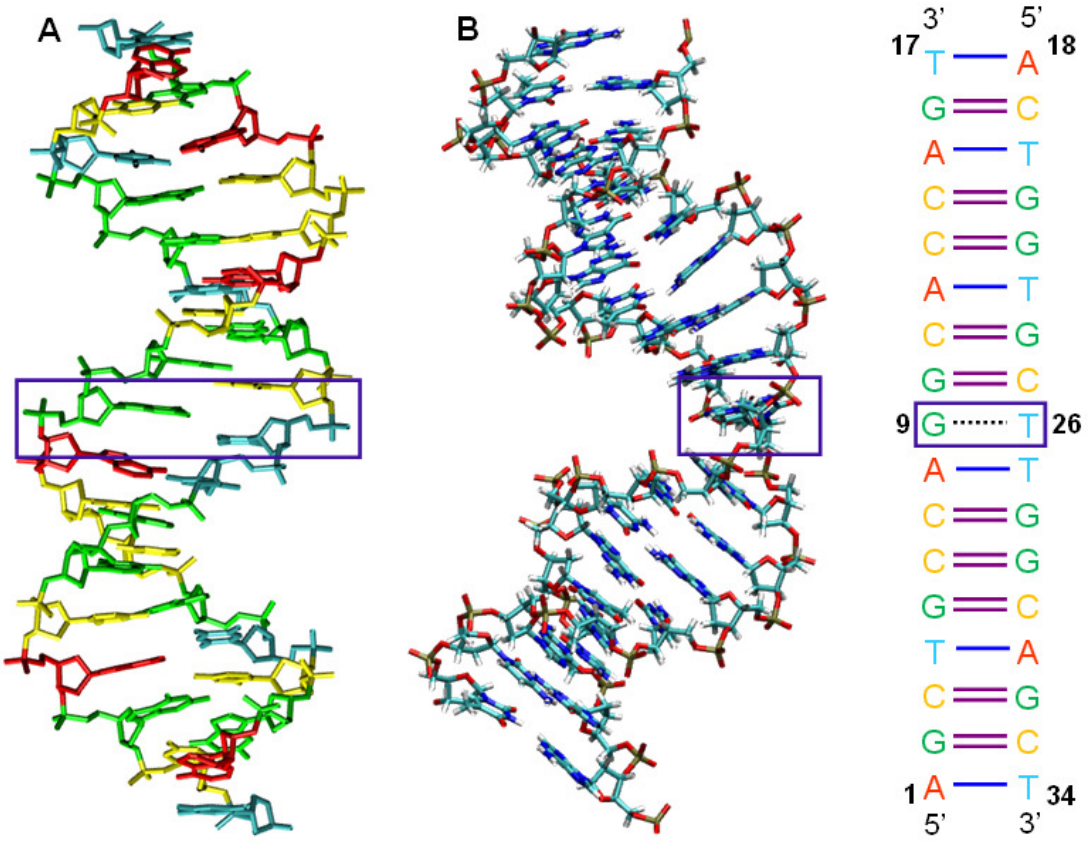


Figure 5.1. Comparison of the *apo* NMR structure of 17mer mismatch GT DNA (A) and bound DNA in the WT-MutS complex crystal structure (B) (PDB: 1E3M). The corresponding sequence of the mismatch GT DNA is shown on the right. The position of the G9-T26 mispair is highlighted by the purple boxes.

Comparison of the local base pair parameters (Table 5.1) shows large changes to the buckle (κ = -31.99) and opening (σ = -55.37) parameters to the GT mispair when bound to the protein. The combination of these changes allows the NTD of MutS to bind to the DNA and for the key Phe36 residue to intercalate with the GT mispair. The change of shear value from -1.35 to 4.65 also shows that the mispair had reorientated itself in order to accommodate the stacking between the mispaired T residue and Phe36. Surprisingly, although the shear value has increased, the propeller twist has reduced to -2.58, signifying that the two bases in the GT mispair are almost planar to each other. There are also changes seen for the immediately flanking base pairs to the GT mispair

particularly the buckle parameter with values of 11.34 and -22.42 for A8-T27 and G10-C25 respectively. These changes are an indication of hydrogen bond interactions between A8-T27 and G10-C25 and the Phe-X-Glu binding motif involved in mismatch recognition.

A

Base Pair	Nucleotides	Shear (S_x)	Stretch (S_y)	Stagger (S_z)	Buckle (κ)	Propeller Twist (ω)	Opening (σ)
7	C-G	-0.04	-0.43	0.77	-2.92	-5.55	-4.97
8	A-T	0.27	-0.05	0.42	20.82	-13.49	3.28
9	G-T	-1.35	-0.52	1.91	5.47	-8.94	4.84
10	G-C	0.10	-0.31	1.20	3.35	-0.59	1.35
11	C-G	0.13	-0.22	0.92	-14.52	-3.86	1.33
Average		-0.18	-0.31	1.04	2.44	-6.49	1.17

B

Base Pair	Nucleotides	Shear (S_x)	Stretch (S_y)	Stagger (S_z)	Buckle (κ)	Propeller Twist (ω)	Opening (σ)
7	C-G	0.23	-0.01	-0.04	2.00	-10.65	1.10
8	A-T	0.32	-0.12	0.18	11.34	-5.40	0.35
9	G-T	4.65	0.17	-0.71	-31.99	-2.58	-55.37
10	G-C	-0.26	0.01	-0.31	-22.42	-10.59	-4.85
11	C-G	0.07	-0.14	-0.29	11.45	5.64	0.88
Average		1.00	-0.02	-0.23	-5.92	-4.72	-11.58

Table 5.1. Comparison of local base pair parameters for A. NMR structure of *apo* 17mer mismatch GT DNA and B. bound GT DNA structure in the WT-MutS complex. Changes in important parameters are highlighted in light blue and yellow for *apo* and bound GT DNAs respectively.

Comparison of the local helical parameters (Table 5.2) shows that in the bound DNA a large change was seen to the y-displacement and tip parameters for the GT mispair. The increase in y-displacement gives evidence that the DNA helix is getting slightly wider at the GT mispair which may be done to accommodate the protein during recognition binding. Significant changes were also observed to the inclination values for the steps

involving the base pairs immediately flanking the GT mispair (Table 5.2, steps 8 and 10). This effect may be caused by the change in y-displacement in the GT mispair since as similar changes were not observed for the flanking base pairs. Consequently, the inclination change compensates for the structural change at the GT mispair.

A

Step No.	Step	x-displacement (d_x)	y-displacement (d_y)	Inclination (η)	Tip (θ)
7	CA / TG	-1.24	-0.81	-2.52	-5.02
8	AG / TT	-3.20	-2.32	4.81	13.27
9	GG / CT	-1.15	0.54	-2.64	-2.90
10	GC / GC	-2.98	0.29	11.09	-2.71
11	CA / TG	-1.98	0.48	4.22	-7.44
Average		-2.11	-0.36	2.99	-0.96

B

Step No.	Step	x-displacement (d_x)	y-displacement (d_y)	Inclination (η)	Tip (θ)
7	CA / TG	0.62	-0.55	6.48	-3.65
8	AG / TT	-4.33	2.75	60.58	-10.73
9	GG / CT	0.71	-23.94	10.61	27.36
10	GC / GC	-3.30	-2.15	41.38	0.94
11	CA / TG	-2.40	-0.41	15.07	2.62
Average		-1.74	-4.86	26.82	3.31

Table 5.2. Comparison of local helical parameters for A. NMR structure of *apo* 17mer mismatch GT DNA and B. bound GT DNA structure in GT DNA-*E. coli* MutS complex. Changes in important parameters are highlighted in light blue and yellow for *apo* and bound GT DNAs respectively.

In the comparison of the local base pair step parameters (Table 5.3), the most striking change was the change in helical twist for the GT mispair. In the *apo* NMR structure the helical twist value was 38.92 which is expected for B-form DNA but in contrast, the helical twist in the bound DNA was 8.07, which was significantly lower than this and indicates that the helix was in an unwound state at the GT mispair. The helical twist

value of 23.34 observed for step 10 shows that the effect of this unwinding of the helix was greater for the G10-C25 base pair than for A8-T27. The change in roll values to the flanking base pairs to the GT mispair may be induced by intercalation of the Phe-X-Glu binding motif of WT- MutS protein

A

Step No.	Step	Shift (D_x)	Slide (D_y)	Rise (D_z)	Tilt ($T\tau$)	Roll ($P\rho$)	Helical Twist (Ω)
7	CA / TG	0.69	-0.85	2.34	3.04	-1.53	35.33
8	AG / TT	0.45	-1.46	3.82	-7.22	2.61	31.86
9	GG / CT	-0.19	-0.92	3.60	1.93	-1.76	38.92
10	GC / GC	-0.01	-1.16	3.91	1.68	6.86	36.28
11	CA / TG	0.00	-1.01	2.37	4.41	2.50	34.61
Average		0.19	-1.08	3.21	0.77	1.74	35.40

B

Step No.	Step	Shift (D_x)	Slide (D_y)	Rise (D_z)	Tilt ($T\tau$)	Roll ($P\rho$)	Helical Twist (Ω)
7	CA / TG	0.54	0.72	3.09	2.27	4.03	36.03
8	AG / TT	-1.85	1.48	6.72	9.71	54.86	33.32
9	GG / CT	2.65	0.58	3.24	-4.22	1.64	8.07
10	GC / GC	1.10	0.01	2.73	-0.46	20.21	23.34
11	CA / TG	0.09	-0.53	3.41	-1.48	8.48	31.96
Average		0.51	0.45	3.84	1.16	17.84	26.54

Table 5.3. Comparison of local base pair step parameters for A. NMR structure of *apo* 17mer mismatch GT DNA and B. bound GT DNA structure in GT DNA-*E. coli* MutS complex. Changes in important parameters are highlighted in light blue and yellow for *apo* and bound GT DNAs respectively.

The combination of a very small helical twist and the large y-displacement mentioned above shows that recognition of the mismatch is accompanied by quite a large change in conformation in the bound DNA. However, the more significant changes seem to occur at the GT mispair and its flanking base pairs suggesting the distortion of the DNA helix is very localised.

Ideally, the *apo* NMR structure would be compared to the MutS bound GT DNA whereby its structure has been determined using NMR spectroscopy. This would give a more accurate comparison as the structure determination would have been carried out under similar experimental conditions. However, as shown in Chapter 3.6, selective changes were observed to the imino proton resonances of 17mer mismatch GT DNA (Chapter 3.6.1, Figure 3.31). These changes were observed after the addition of WT-MutS protein. The clear evidence of a tertiary fold in the NTD of MutS protein provides a strong foundation for its structure determination in solution state.

5.2 Expression of normal and isotopically labelled NTD-MutS for NMR

The expression, isolation and purification of highly pure WT- MutS had previously been carried out in the group by Dr. F. Vendeix (Figure 5.2) and this protein was used in NMR titrations to study the interaction between WT-*E. coli* MutS and mismatch/modified DNAs in Chapters 3.4, 3.5, 4.3 and 4.4.

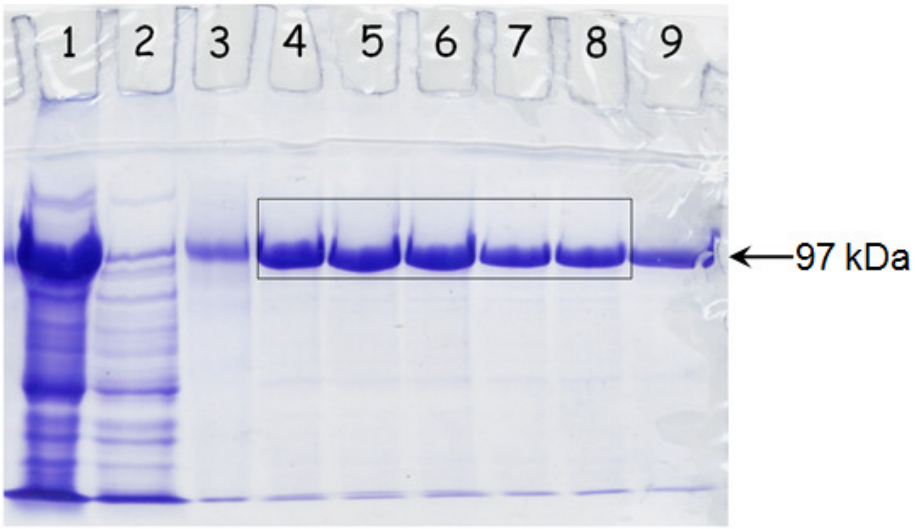


Figure 5.2. PAGE assay of WT-*E. coli* MutS where 1 = Cell lysate, 2 = Crude extract after column loading, 3 and 9 = *E. coli* MutS marker (97 kDa) and 4 - 8 = Pure WT-*E. coli* MutS. PAGE assay carried out by Dr. F. Vendeix.[88]

Despite the successful isolation of WT- MutS protein (97 kDa, monomer), no isolation, biochemical or NMR studies have ever been carried out involving the NTD-MutS protein (15 kDa, monomer) which is key for mismatch recognition. This study demonstrates the first successful attempt at expressing the isotopically normal NTD-MutS protein domain and its ^{15}N enriched and ^{13}C, ^{15}N enriched derivatives to a high yield and purity.

Following expression and isolation, SDS-PAGE gels assays were performed on the NTD-MutS following purification in order to check its purity (Figure 5.3). Lanes 6-8 of the figure provide unambiguous evidence for the high expression and purity of the isolated NTD-MutS and thus justify the NMR investigation undertaken. In addition, gels were run after NMR experiments to monitor the stability of the protein.

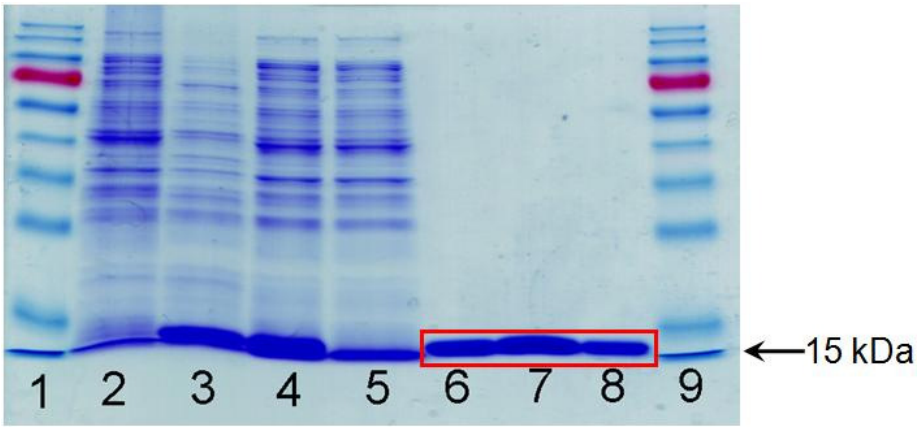


Figure 5.3. SDS-PAGE assay of NTD-MutS where 1 = marker, 2 = *E. coli* BL21 before induction, 3 = *E. coli* BL21 after induction, 4 = crude extract before column loading, 5 = crude extract after column loading, 6, 7, 8 = MutS aliquots (red box) and 9 = marker. PAGE assay carried out by Dr. M. Zampini.

5.3 Mass spectrometry of NTD-MutS protein

The NTD-MutS could be produced as isotopically enriched derivatives and this was important for its study by NMR as it provides access to more advanced NMR techniques which could be used to characterise its structure. However, before such experiments took place it was important to determine the level of enrichment as this will directly impact the NMR spectra being measured. The mass spectra of unlabelled NTD-MutS and ^{15}N isotopically labelled are shown in Figures 5.4 and 5.5 respectively. The difference in mass was calculated to be 181 corresponding to an enrichment level of over 95%.

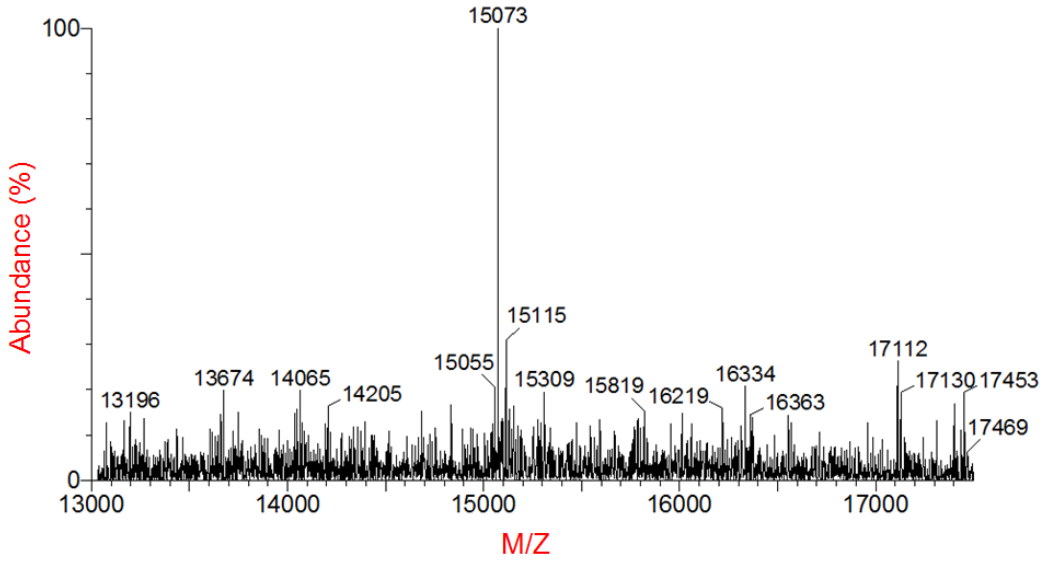


Figure 5.4. Mass spectrum of unlabelled NTD-MutS protein showing a mass of 15073 Da for the protein. Spectrum was recorded using ESI-MS.

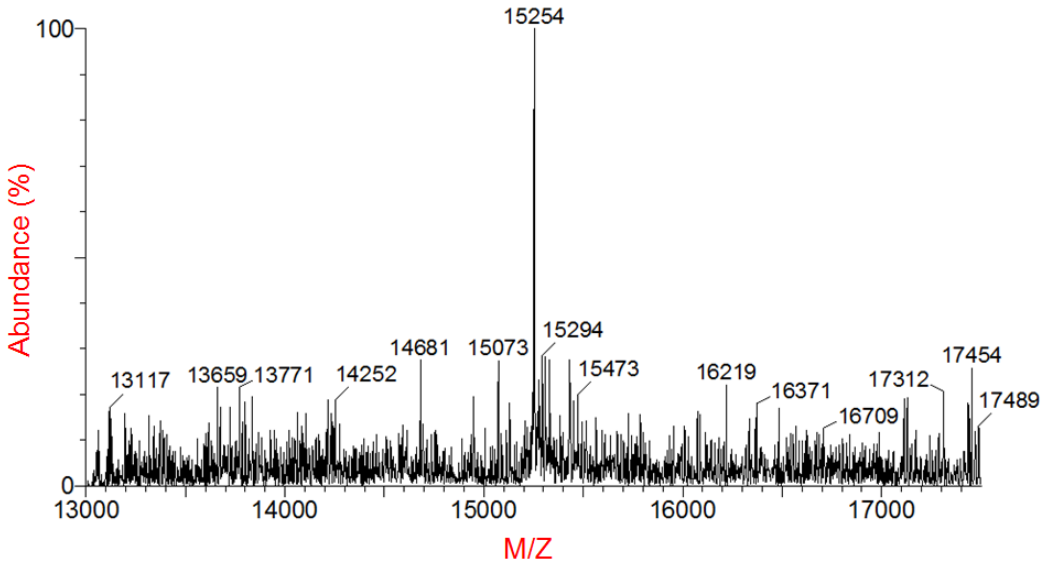


Figure 5.5. Mass spectrum of ^{15}N isotopically labelled NTD-MutS protein showing a mass of 15254 Da for the labelled protein. Spectrum was recorded using ESI-MS.

5.4 Identification and sequence specific assignment of amide backbone of NTD-MutS protein

The NMR assignment of NTD-MutS was carried out using 1D, 2D and 3D NMR techniques as noted in the methodology given in Figure 2.27. Every amino acid spin system is different in terms of its structure and the proton environments it represents. Spin systems can be broadly classed into two main groups; aliphatic and aromatic spin systems. Each amino acid spin system can be identified from cross peak patterns in TOCSY and NOESY spectra, which are characteristic to each one.

5.2.1 Characterisation of protein fold

A comparison of the 1D ^{1}H-NMR spectra of unlabelled (Figure 5.6, A) and ^{13}C, ^{15}N isotopically labelled NTD-MutS (Figure 5.6, B) with that of full length WT-*E. coli* MutS (Figure 5.6, C) gives strong evidence that both the WT-protein and NTD-MutS are well were folded based on the dispersion of various proton resonances and their expected linewidths. It is striking that despite being significantly smaller than the WT-protein (97 kDa, monomer) the peaks observed for the NTD-MutS domain (15 kDa, monomer) appear to be well conserved throughout in the spectrum of WT-protein.

Detailed inspection of the 1D ^{1}H-NMR spectrum of unlabelled NTD-MutS in $^{1}H_2O$ at 25°C revealed a clearly resolved highfield shifted peak resonating at (-) 0.2 ppm (Figure 5.6, A). The chemical shift of this peak is probably due to an aromatic ring current effect on methyl protons such as Val, Leu or Ile. Upon reducing the sample temperature

to 2°C, the same peak also showed a chemical shift change ($\Delta\delta$ = 0.10 ppm). More significantly, this peak is also observed in the ^{13}C, ^{15}N labelled NTD-MutS 1D ^{1}H-NMR spectrum suggesting an important role for the proton group in the structure of the NTD-MutS domain.

In the 1D ^{1}H-NMR spectrum of WT-*E. coli* MutS, this highfield shifted peak was no longer observed as the increase in molecular size would be accompanied by a large increases in linewidth and rendering the peak invisible. The exchange retarded amide protons, observed lowfield of 9.0 ppm, are another important feature of NTD-MutS and which is conserved in the spectrum of the full length protein.

Identification and assignment of the resonances described above may provide important information on tertiary interactions which occur in NTD-MutS protein. Further to the shifted resonances, the observed resonance linewidths of the spectrum correlated well with those expected from a monomeric 15 kDa protein in Figure 5.6, A and B.

From an assignment viewpoint, the line intensity observed for the 1D ^{1}H-NMR spectrum of ^{13}C, ^{15}N isotopically labelled NTD-MutS protein was much higher and with fewer impurities in comparison to the unlabelled protein. This would be advantageous when NMR experiments were carried and will be discussed further in this chapter.

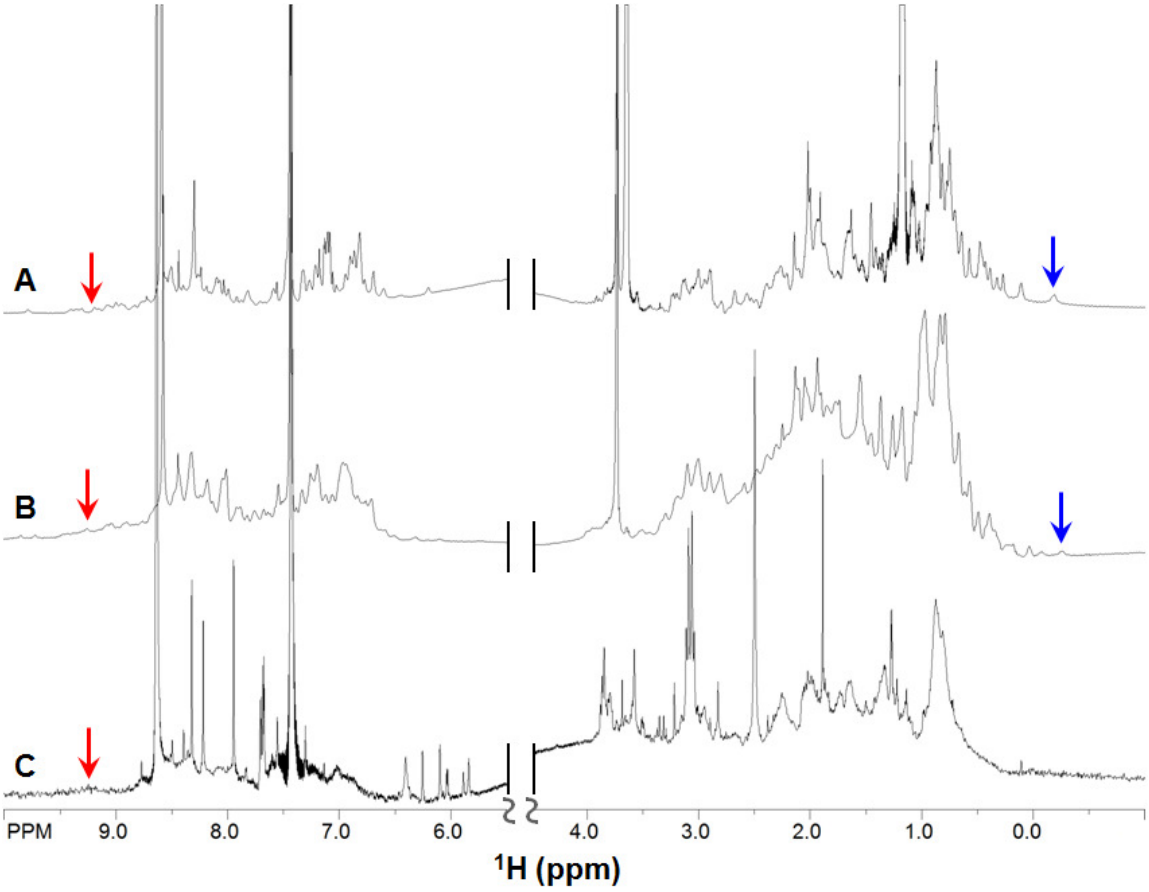


Figure 5.6. Spectra stack comparing 1D ^{1}H-NMR spectra A. unlabelled NTD-MutS (0.6mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) measured at 800 MHz, B. ^{13}C, ^{15}N isotopically labelled NTD-MutS (0.4mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) and C. unlabelled full length WT-MutS (0.2mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2). A and C were measured at 800 MHz and B was measured at 700 MHz; all spectra were recorded at 25°C. The red arrow indicates the observation of lowfield shifted amide proton whilst the blue arrow indicates a possible highfield shifted methyl group. The observation of sharp and intense peaks in the spectrum for WT-*E. coli* MutS correspond to buffer impurities.

To seek unambiguous evidence for the endowment of a good amide backbone and tertiary structure of NTD-MutS, a ^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N isotopically labelled NTD-MutS was measured in $^{1}H_2O$ at 25°C (Figure 5.7). The spectrum shows very good sensitivity and spectral simplification achieved with the observation of well resolved cross peaks.

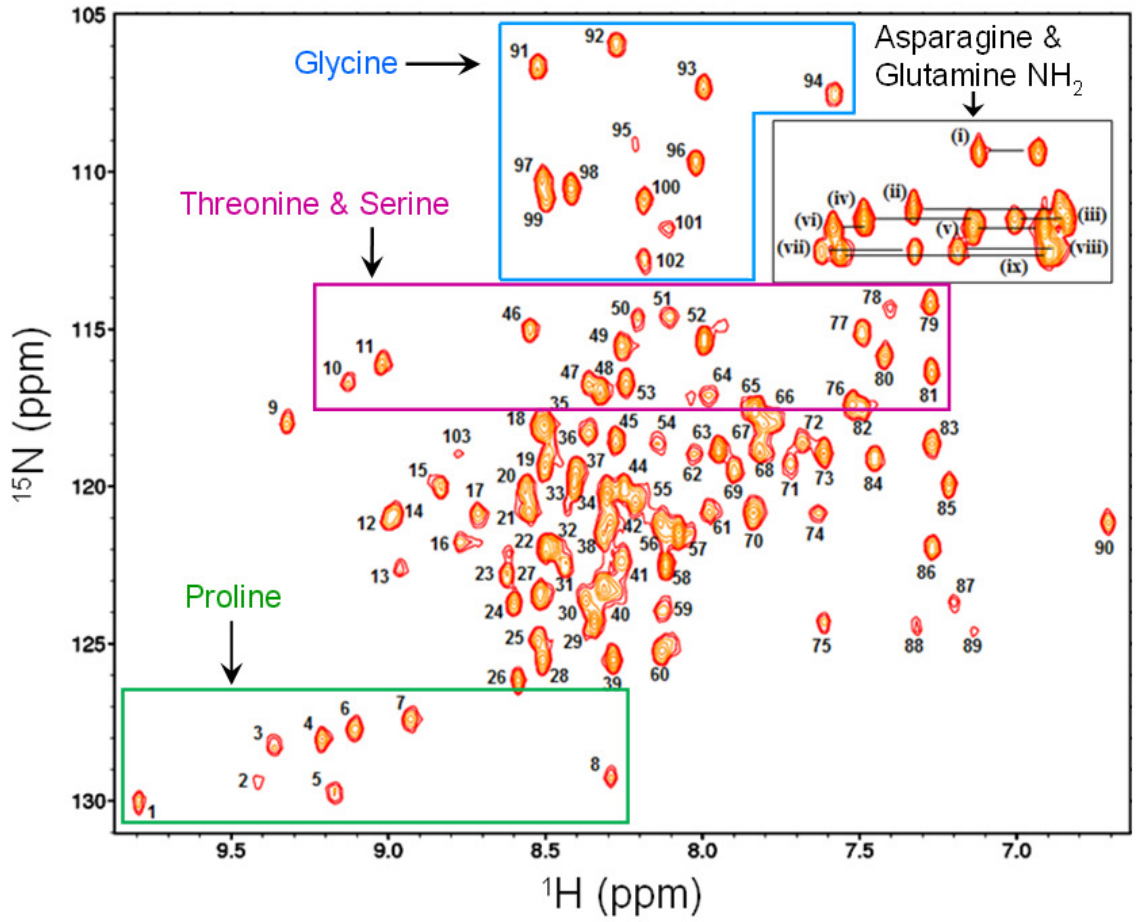


Figure 5.7. 800 MHz ^{1}H-^{15}N HSQC spectrum of NTD-MutS protein (0.6mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. The different coloured boxes indicate the characteristic backbone NH resonances of certain amino acids where blue, purple and green correspond to glycine, threonine and serine, and proline respectively. Asparagine and glutamine side chain NH_2 groups are highlighted by the black box.

In the ^{1}H-^{15}N HSQC spectrum, a single resonance should be observed for every amide NH group in the protein sequence, the NH group can include those in the amide backbone or those in the side chain NH_2 groups. The HSQC spectrum shows a good dispersion of cross peaks both in the ^{1}H and ^{15}N dimensions which demonstrates that the protein is endowed with both α-helical and β-sheet secondary structural elements. Successful measurement of the well-dispersed HSQC spectrum below laid a strong foundation for further NMR studies on a uniformly doubly labelled ^{13}C, ^{15}N NTD-MutS with the aim to assign the amide backbone of the protein, sequence specifically.

There were 102 visible cross peaks in the ^{1}H-^{15}N HSQC spectrum, excluding those for asparagine and glutamine NH_2 side chain groups. This illustrates that a majority of the peaks are well dispersed with just a few being overlapped which could only be separated by going into the third dimension when collecting NMR data. A second observation was the nine sets of cross peaks (i-ix) which corresponds to asparagine and glutamine NH2 groups which are highlighted in Figure 5.7. There are only nine asparagine and glutamine residues in the protein sequence and this provides unambiguous evidence for their identification.

The ^{1}H-^{15}N HSQC spectrum was also used to make initial identification of amino acid residues based on the amide ^{15}N chemical shift. The amide ^{15}N resonances of most amino acids typically cover a chemical shift range of 115-120 ppm, but for a few amino acids they appear characteristically further highfield or lowfield than this. This provided a useful tool for assignment as they provide a starting point for establishing the sequential connectivity. An example was that for glycine residues, the amide ^{15}N resonances typically appear the most highfield of all the amino acids between 105-115 ppm and this is highlighted in Figure 5.7, along with the position of amide ^{15}N resonances for other selected amino acids.

5.2.2 Identification of $d_{\alpha N}$ sequential connectivity

Preparation of a ^{13}C, ^{15}N isotopically labelled sample enabled 3D NMR experiments to be carried out on the NTD-MutS. This simplified the spectrum by dispersing the resonances into a third dimension thus resolving any overlapping cross peaks. A suite of 3D experiments were implemented which dispersed peaks into different third frequency dimensions (^{13}C) and thereby resolving any overlapped amide cross peaks.

Assignment started with the HNCA experiment. This correlates the amide proton of an amino acid residue to its own C_α resonance. The same experiment also gives the sequential connectivity from the C_α resonance of a given amino acid with the amide proton of the next amino acid in the sequence.

Ninety-eight frequency planes were clearly observed in the 3D HNCA spectrum on the basis of ^{15}N chemical shifts. The amide ^{15}N spins of glycine residues appear in the highfield region of the HSQC spectrum and identification of sequential connectivity started from the four resolved glycine peaks (Figure 5.7, peaks 91-94).

Similar to DNA, sequential connectivities can be made between adjacent amino acid residues in a protein. For each panel that was observed, one would expect to see two peaks, an intraresidue peak and a sequential peak and those corresponding to intraresidue connectivities typically appear more intense.

A number of sequential connectivities could be identified in the HNCA spectrum. An example is the selected sequential walk stretching between glutamic acid 84 (E84) and tyrosine 99 (Y99) which is shown in Figure 5.9. A clear sequence specific assignments is shown by the residue panel G90 which can be identified in the HSQC spectrum as peak 91 (Figure 5.8). In this manner, in principle, one should be able to sequentially connect the first and last amino acid of the protein, however, this was not entirely possible in the case of NTD-MutS protein due to a number of breaks where the sequential peak could not be observed in the HNCA panel.

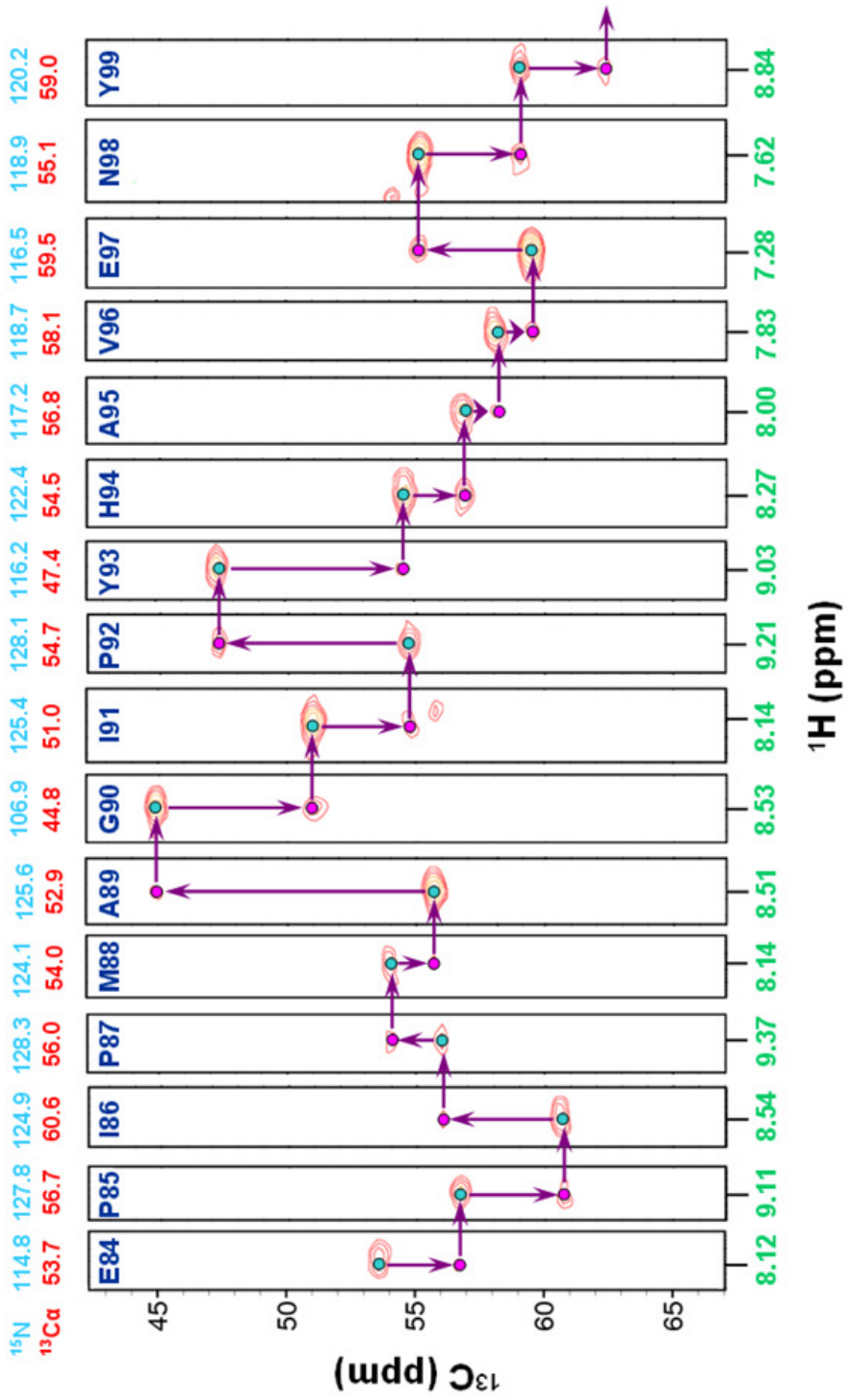


Figure 5.8. 600 MHz 3D HNCA spectrum of ^{13}C, ^{15}N isotopically labelled NTD-MutS protein (0.4mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, pH 6.2, 50mM PO_4^{3-} and 50mM NaCl) measured at 25°C. The figure shows the selected amide ^{1}H – Cα sequential connectivity between E84 and Y99. Light blue circles correspond to the intraresidue connectivities and pink circles indicate the sequential connectivity to the next residue. The chemical shift of backbone amide protons are indicated in green, those for the amide ^{15}N resonances are indicated in light blue and red indicates the chemical shift of $^{13}C\alpha$ resonances

Panels in the HNCA spectrum could be correlated to the HNCACB spectrum as they are both similar experiments, with the exception that the amide proton is correlated to both C_α and C_β resonances of the amino acid side chain, in the HNCAB experiment. This is important as identifying the C_α and C_β chemical shifts means that 3D NMR experiments can be performed which are ^{13}C edited where more assignments can be made, e.g. ^{1}H-^{13}C HSQC-TOCSY.

Seventy-seven panels were observed in the HNCACB spectrum which was not much lower when compared to the HNCA spectrum. The appearance of peaks corresponding to C_α resonances at the same ^{13}C chemical shift as observed in the HNCA spectrum helped confirm their assignment. In addition to these peaks, C_β peaks were also observed and these were distinguished from C_α peaks as they appeared with negative spin in the spectrum.

Under optimum conditions, one should be able to trace the same sequential connectivity between C_α and NH resonances as observed in the HNCA spectrum. In addition, the sequential connectivity should be also observed between C_β and NH resonances along the amide backbone.

The sequential connectivity between arginine 78 (R78) and serine 81 (S81) is shown in Figure 5.9, which shows the difference between the HNCA and HNCACB experiments. The comparison also confirms the identity of peak 97 in the HSQC spectrum (Figure 5.8) arising from G79 in the sequence.

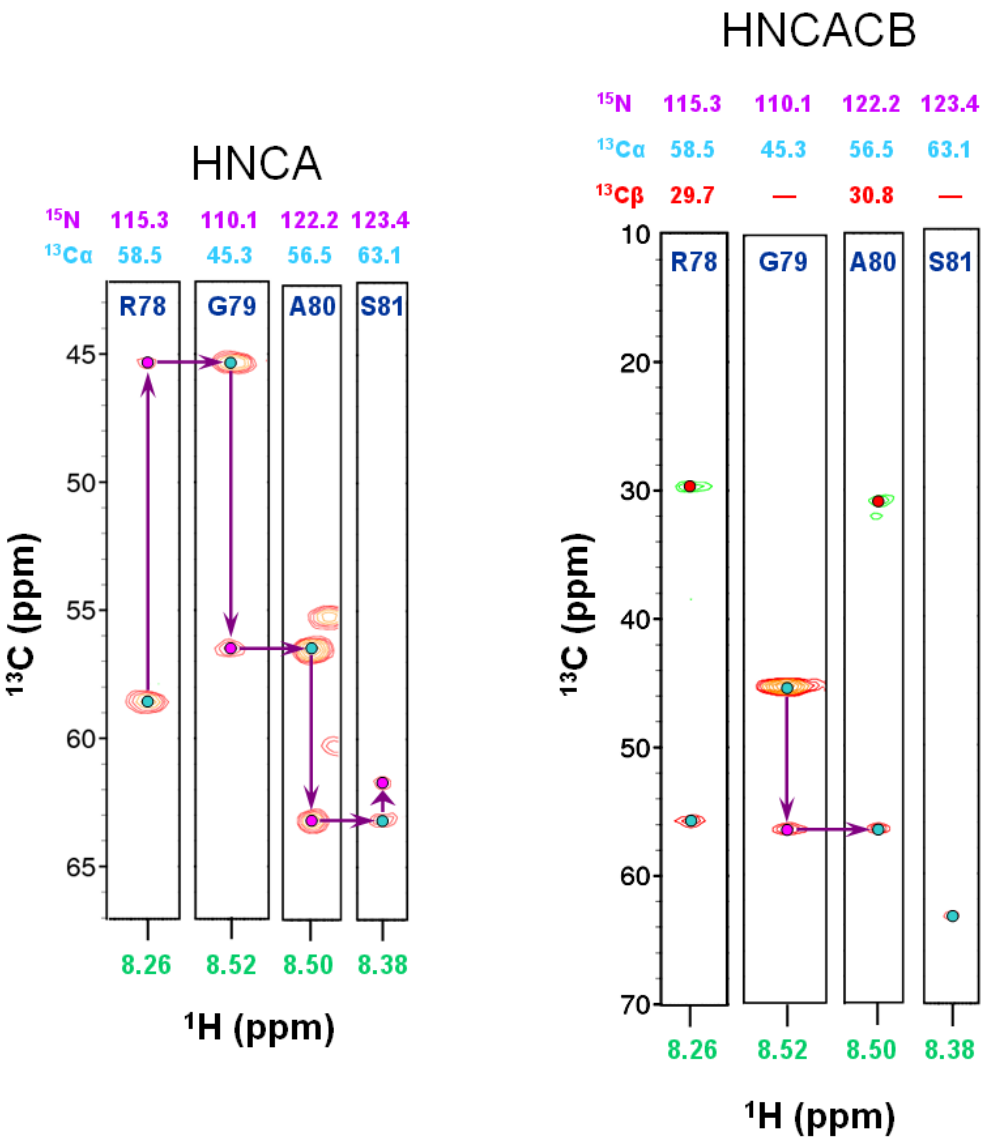


Figure 5.9. Comparison between 3D HNCA (600 MHz) and HNCACB (700 MHz) experiments measured for ^{13}C, ^{15}N isotopically labelled NTD-MutS (0.4mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, pH 6.2, 50mM PO_4^{3-} and 50mM NaCl) at 25°C. The selected sequential connectivity is between Arg78 (R78) and Ser81 (S81). Light blue circles correspond to the intraresidue connectivities, pink circles indicate the sequential connectivity to the next residue and red circles indicate the intraresidue C_β-NH correlation. The chemical shift of backbone amide protons are indicated in green, those for the amide ^{15}N resonances are indicated in purple, $^{13}C_\alpha$ chemical shifts are given in light blue and red indicates the chemical shift of $^{13}C_\beta$ resonances.

In order to confirm the amino acid type using the chemical shift data obtained from the HNCA and HNCACB experiments, a 3D HCCH-TOCSY was carried out. This experiment is similar to a 3D ^{13}C edited TOCSY spectrum with the advantage exception that it is superior in sensitivity. The HCCH-TOCSY allows all the proton groups in the side chain to be assigned, with the exception of aromatic ring groups. Although the experiment was measured in $^{1}H_2O$, the sensitivity of the spectrum was good and a number of spin systems (Figure 5.10) were identified based on the previous identification of both C_α and C_β chemical shifts.

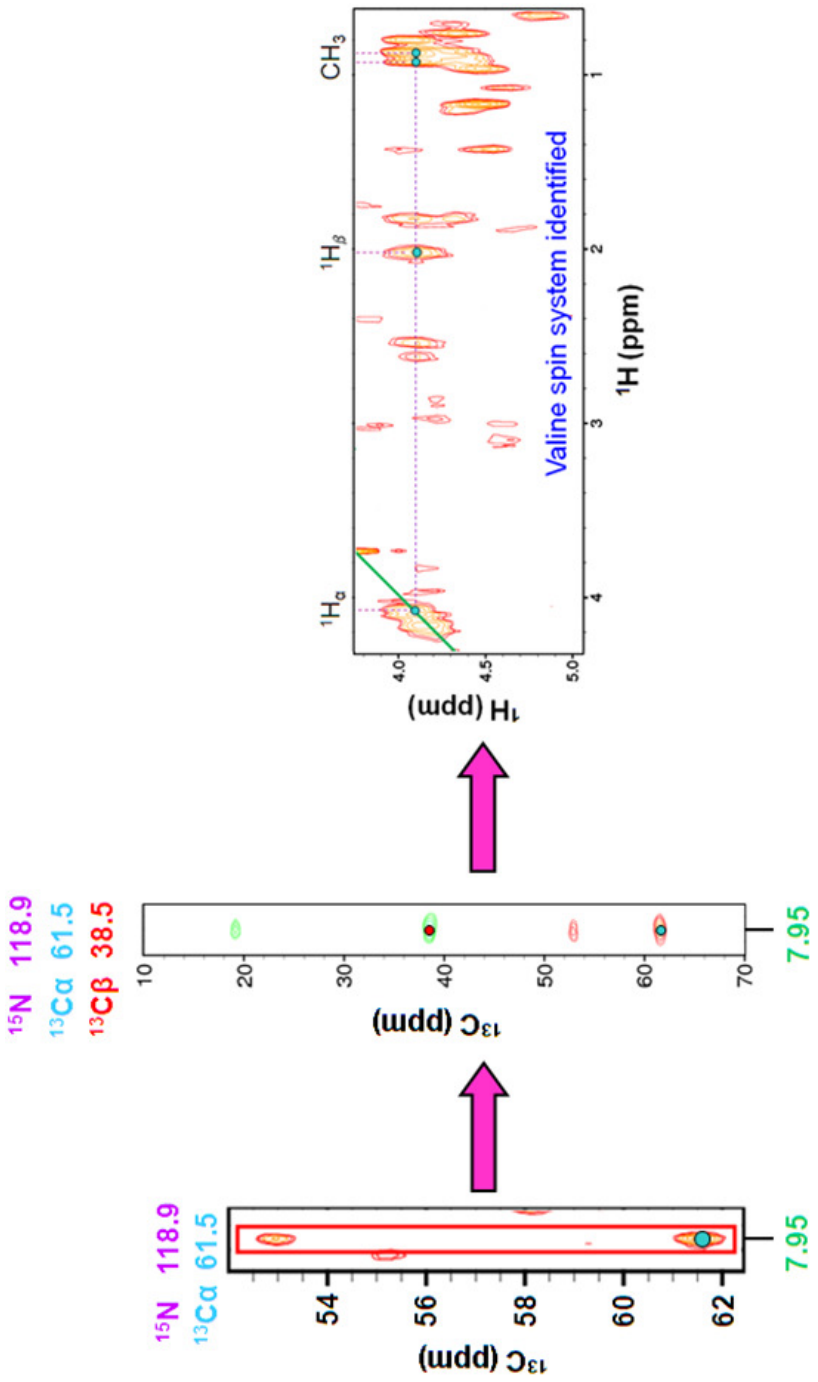


Figure 5.10. Identification of valine spin system in ^{13}C, ^{15}N isotopically labelled NTD-MutS (0.4mM, 90% $^{1}H_2O$ + 10% $^{2}H_2O$, pH 6.2, 50mM PO_4^{3-} and 50mM NaCl) at 25°C using 3D experiments; (bottom) HNCA panel, (centre) HNCACB panel and (top) HCCH-TOCSY panel. Light blue circles correspond to the intraresidue connectivities, pink circles indicate the sequential connectivity to the next residue and red circles indicate the intraresidue C_β-NH correlation. The chemical shift of backbone amide protons are indicated in green, those for the amide ^{15}N resonances are indicated in purple, $^{13}C_\alpha$ chemical shifts are given in light blue and red indicates the chemical shift of $^{13}C_\beta$ resonances. The position of the side chain proton groups are indicated in the HCCH-TOCSY panel.

To gain insight into the nature of tertiary interactions exhibited by NTD-MutS, a ^{1}H-^{1}H correlated NOESY spectrum was measured in $^{2}H_2O$ at 25°C. The spectrum shows a large number of both intra and interresidue NOE cross peaks which are required for structure determination of the domain. Among the various correlations, a most striking region was the one corresponding to methyl group protons correlating with aromatic ring protons (Figure 5.11). Identification of this structurally significant long range NOEs provides additional vital evidence for a well disposed tertiary structure of NTD-MutS.

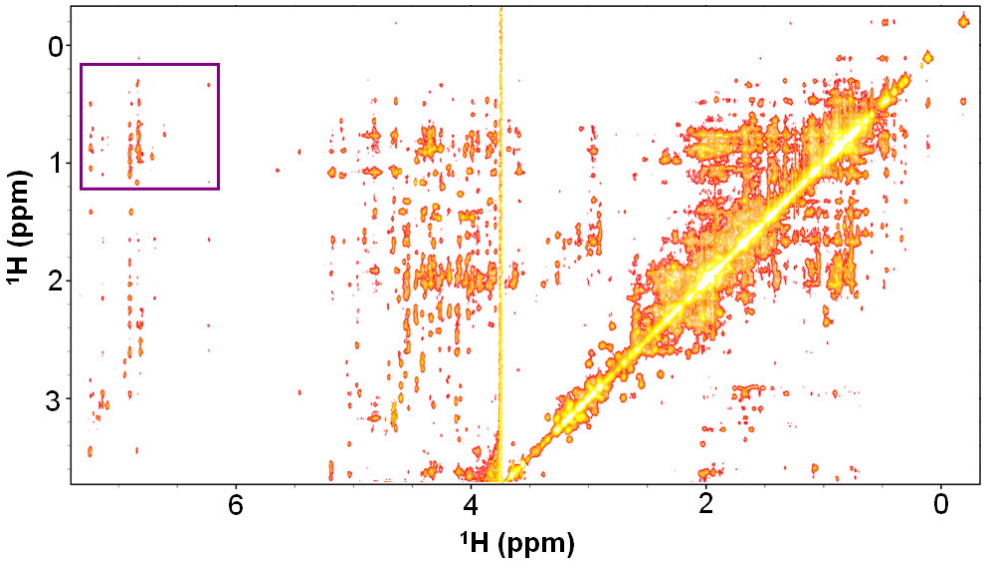


Figure 5.11. 800 MHz NOESY spectrum of unlabelled NTD-MutS (0.6mM, 100% $^{2}H_2O$, pH 6.2, 50mM PO_4^{3-} and 50mM NaCl) measured at 25°C. The box highlights NOEs between aromatic and methyl group protons.

A number of NOEs were observed to the highfield shifted residue in the NOESY spectrum with the most noticeable between two methyl groups (δ = -0.20 ppm and δ = 0.50 ppm). As mentioned earlier (section 5.2.1), these methyl groups may arise from an aliphatic spin system such as a Leucine, Isoleucine or Valine whose resonances have been shifted highfield because of an aromatic ring current effect. Cross peaks in the

HCCH-TOCSY identified the spin system as an isoleucine residue (Figure 5.12) although no cross peaks were observed to the Hγ protons. The Hγ protons were however identified through cross peaks from the high shifted methyl group in the NOESY spectrum.

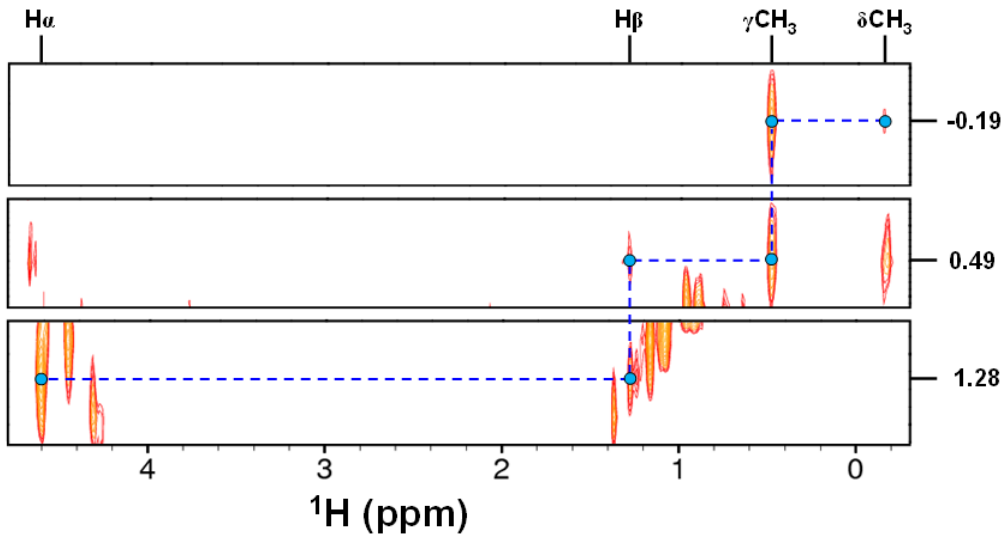


Figure 5.12. 600 MHz HCCH-TOCSY spectrum showing the assignment of an isoleucine residue in ^{13}C, ^{15}N isotopically labelled NTD-MutS (0.6mM, 100% $^{2}H_2O$, pH 6.2, 50mM PO_4^{3-} and 50mM NaCl) measured at 25°C. Cross peaks to the Hγ protons are not observed and were assigned in the NOESY spectrum.

Another clear observation from the NOESY spectrum was the region in which the aromatic spin systems were identified. This allowed a number of aromatic spin systems to be assigned. For a phenylalanine, NOEs should be observed between all its aromatic protons (H2/6, H3/5 and H4). This NOE pattern between aromatic protons in phenylalanine residues are shown in Figure 5.13. In the NTD-MutS protein, there are four phenylalanine residues and three were identified in the NOESY spectrum. The highfield shifted phenylalanine system (purple line) could be induced by a similar aromatic ring current effect as seen for the methyl group protons in Figure 5.6.

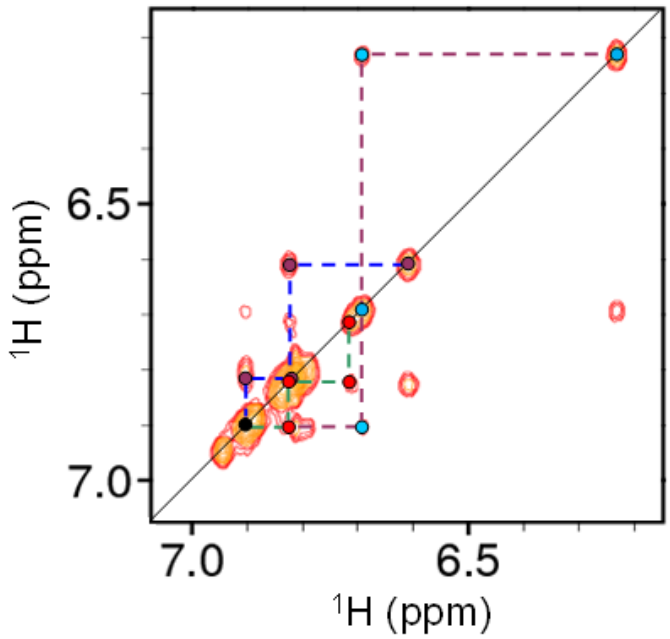


Figure 5.13. 800 MHz NOESY spectrum of unlabelled NTD-MutS (0.6mM, 100% ^2H_2O, pH 6.2, 50mM PO_4^{3-} and 50mM NaCl) measured at 25°C. The region highlights NOEs between aromatic protons for phenylalanine residues, with each one being identified by different coloured lines.

Similar to phenylalanine residues, the NOE patterns for tyrosine aromatic spin systems were also observed in the same region. As H4 has been replaced with a hydroxyl group in tyrosine, NOEs will only be observed to H2/6 and H3/5 protons. There are six tyrosine residues in the sequence of NTD-MutS and four of these were observed in the NOESY spectrum (Figure 5.14).

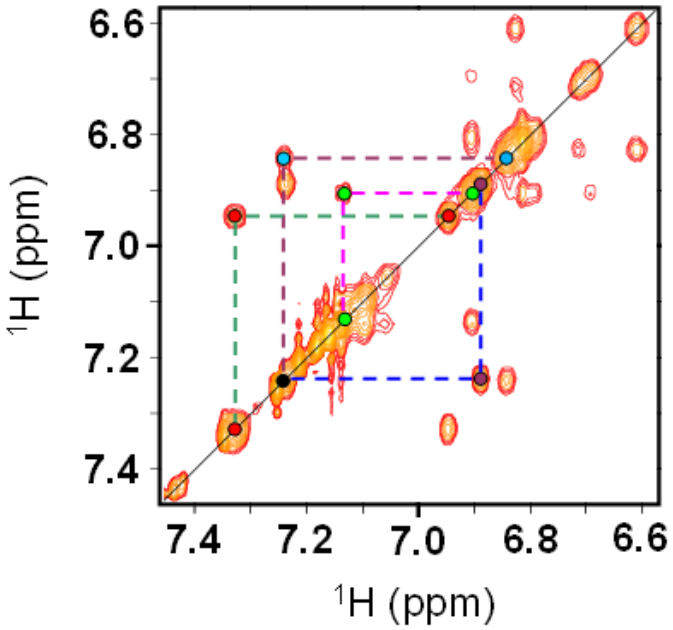


Figure 5.14. 800 MHz NOESY spectrum of unlabelled NTD-MutS (0.6mM, 100% ^2H_2O, pH 6.2, 50mM PO_4^{3-} and 50mM NaCl) measured at 25°C. The region highlights NOEs between aromatic protons for tyrosine residues, with each one being identified by different coloured lines.

Despite the good quality NOESY spectrum, cross peaks were either too overlapped to provide accurate distance constraints, or were missing from the spectrum due to the solvent used. Therefore, in order to generate distance constraints 3D NOESY type experiments would have to be carried out on the ^{13}C, ^{15}N isotopically labelled sample.

To conclude, a comparison of the helical parameters between the *apo* NMR structure of 17mer mismatch GT DNA and the bound DNA of the GT DNA-MutS protein complex show that upon mismatch recognition the change in helical conformation primarily occurs at the GT mispair and its flanking base pairs. The isolation of isotopically normal and enriched NTD-MutS has been accomplished to a high purity and yield and NMR spectroscopy has been used to confirm that the domain has a tertiary fold.

Progress has been made to gain information on the structure of *apo* NTD-MutS by carrying out the assignment of the amide backbone. The assignments made in this chapter provides a good foundation towards the structure determination of the *apo* structure of NTD-MutS in solution state. The isotope enrichment level obtained during protein preparation and the level of sensitivity in NMR data collected so far fully justifies the NMR investigation carried on the NTD-MutS protein and warrants further investigation towards its structure determination in solution state.

A recent paper by Sachadyn[60] investigating the conservation and diversity of MutS proteins has shown that a critical GXFXE binding motif, which contains the Phe-X-Glu motif mentioned in Chapter 1.4.2, is conserved in all MutS and MSH6 homologues. This chapter has shown that the amide backbone ^{15}N resonances of glycine residues appear well resolved from other residues in the ^{1}H-^{15}N HSQC spectrum (Figure 5.7) and that a glycine residue is part of the highly conserved GXFXE motif. Consequently, it is possible given the quality of data presented in this chapter to identify and assign not only the glycine residue, but the entire GXFXE motif in the *apo* NTD-MutS. This could prove significant as it is still not clear how MutS is recruited to newly replicated DNA and how it is able to recognise mismatches. Hence, understanding how this motif behaves in the *apo* NTD-MutS structure may lead to better understanding of the mode of recognition by MutS.

Chapter 6 – Conclusions and future work

I have successfully carried out the structure determination of mismatched GT DNA and its canonical GC derivative to a high resolution, using NMR spectroscopy, to investigate structural differences caused by non-canonical base pairs. Both structures adopt a B-form DNA conformation, although a comparison of the helical parameters revealed that the substitution of a GT mismatch induces a change in the local conformation, with the GT mispair adopting a sheared orientation. Large changes were also seen to particular conformational parameters for certain base pairs giving strong evidence that the effect induced by the GT mispair is experienced throughout the helix. Further to its structure determination, I have demonstrated the mismatch DNA's affinity for *E. coli* MutS repair protein in solution state for the first time using NMR titration by monitoring selective changes to the imino protons of its NMR spectrum.

A combination of UV absorbance melting and NMR studies were carried out on a number of analogues to the 17mer mismatch GT DNA in which the mismatch orientation, sequence context and oligonucleotide length were changed. Results obtained using UV absorbance melting show that the stability of a GT mismatch DNA is influenced by the mismatch orientation, sequence context and oligonucleotide length and produced the stability series GC > GT > TG > GT-FB ≈ TG-FB. Affinity for WT-MutS protein through monitoring of imino proton chemical shifts by NMR was also investigated on these sequences and the results supported the UV absorbance melting data. The results showed that the 17mer mismatch DNA chosen for NMR investigation

was optimum in terms of length, sequence context and mismatch orientation and hence, fully justifies the NMR structure determination undertaken in this project. The NMR structures given in this thesis are to a high standard and are comparable to those given in literature.

Further to the mismatched GT DNA, I have shown that thio modified DNAs, such as thioguanine (6-TG) DNA, adopt a stable structure in solution with retention of affinity for WT-MutS protein. The NMR structure determination of a 6-TG modified 13mer GC DNA enabled a comparison to the 17mer mismatch GT DNA and gave a perspective into how the degree of helical distortion varies upon introducing a different DNA lesion. Despite a single, unique substitution of an oxygen to a sulphur atom in 6-thioguanine, a dramatic highfield shift was observed to its imino proton ($\Delta\delta \approx 1.0$ ppm). Changes were also observed to a number of non-exchangeable protons in adjacent residues which were induced by the substitution of sulphur. The NMR data presented in this thesis for 6-TG modified DNA correlates well with that given in the literature and the NMR structure produced is similar to ones which have been deposited in the RCSB PDB database

The substitution of 6-sulphonateguanine or 2-aminopurine-6-sulfonic acid (G^{SO3}), an oxidised derivative of 6-TG, showed an even greater distortion of the DNA helical structure. This was clearly evident when monitoring the imino protons of a GT mispair where the guanine residue was substituted for a G^{SO3} modified base. The imino proton of the mispaired T resonance was shifted highfield by an additional 1.0 ppm when compared to a mismatch GT base pair showing clearly that the base stacking has been interrupted resulting in a less stable helix. Affinity for WT-MutS was demonstrated

during NMR titration with the G^{SO3}.T modified duplex whereby the imino proton of the mispaired T residue (T20) was not observed following addition of sub-stoichoimetric amounts of WT-MutS.

A striking result of the project is the successful isolation of the N-terminal domain (NTD) of *E. coli* MutS protein which adopts a stable, tertiary fold in solution and thereby provides an opportunity to compare its structure with the domain when it is bound to the full length WT-protein. The assignment of the amide backbone of NTD-MutS showed that a number of sequential walks could be identified from which numerous assignments were made. This included the sequence specific assignment of a number of glycine residues. The NMR analysis also identified the chemical shifts of highfield shifted isoleucine and phenylalanine residues, which may have an important role in the structure of NTD-MutS.

It has been shown that the stability of the GT mispair in DNA is influenced by the mismatch orientation and sequence context. However, it is not known how the differences in stability relate to structural change and hence, the NMR structure determination of these analogues is justified in order to determine how mismatch stability relates to helical distortion.

The effect of other modified bases on the base pairing properties in DNA and helical conformation could be explored. Recently, it was discovered that guanine sulphinate (G^{SO2})[138] was a stable product from the oxidation of 6-TG. Investigation of this

modified base would be a good continuation on what has already been established in this thesis.

The isolation of the NTD-MutS provides a good opportunity to study its binding to various DNA mismatches and previous biochemical studies on *Taq* MutS have shown that mismatched DNA can be successfully photo cross-linked to Phe39 in the N-terminal region of the protein.[139] This method can be adopted to characterise mismatch GT DNA cross-linked to *E. coli* NTD-MutS using NMR. Another excellent opportunity presented by the NTD-MutS is for its use as a segmental labelling probe of the domain architecture of full-length WT-MutS.

New methods are being continually developed to probe the mismatch repair pathway such as the recent DNA scissors developed by Yang *et al.*[140] which have been used to measure the degree of duplex bending when MutS interacts with mismatched DNA. Other examples are the use of deuterium exchange mass spectrometry to probe structural change during binding to mismatch DNA[141] and the use of FRET TACKLE[142] to study the kinetics during binding.

These recent studies give further evidence that there is a lot to be learned about this DNA repair pathway and the work carried out in this project has established solid laboratory protocols for resonance assignment and structure determination of DNAs and proteins. The results provide a strong foundation for further study of the interaction of mismatched and modified DNAs with *E. coli* MutS with the aim of better understanding the mode of recognition during mismatch binding in DNA mismatch repair.

Dessemination of results

The results of this work have been presented at two RSC NMR meetings and these are listed below.

1. RSC NMR meeting, 2008, University of Cambridge – Awarded for bursary

2. RSC Young Researchers NMR meeting, 2009, University of Manchester – Abstract selected for oral presentation and was shortlisted for a prize at the meeting.

Papers to be published

The results of the project described in Chapters 3 to 5 will be published in peer reviewed journals and they are listed below.

1. Tony Cheung, Franck Vendeix, Yao-Zhong Xu and Vasudevan Ramesh (2010), NMR studies of mismatch DNA and its interaction with *E. coli* MutS protein, *Nucleic Acids Research*, (To be submitted).

2. Tony Cheung, Franck Vendeix, Yao-Zhong Xu and Vasudevan Ramesh (2010), NMR studies of thioguanine modified DNA and its interaction with *E. coli* MutS protein, *DNA Repair*, (To be submitted).

References

1. V. A. Bloomfield, D. M. Crothers and I. Tinoco Jr., Nucleic Acids: Structures, Properties and Functions, 2000, University Science Books, Sausalito, California.

2. W. Saenger, Principles of nucleic acid structure, 1984, Springer-Verlag, USA.

3. G. M. Blackburn and M. J. Gait, Nucleic Acids in Chemistry and Biology, 1996, Oxford University Press, U.K.

4. J. Watson and F. Crick, *Nature*, 1953, **171**, 737-738.

5. R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fatini and M. L. Kopka, *Science*, 1982, **216**, 475-485.

6. B. N. Conner, T. Takano, S. Tanaka, K. Itaknra and R. E. Dickerson, *Nature (London)*, 1982, **295**, 294.

7. A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel and A. Rich, *Nature (London)*, 1979, **282**, 680-686.

8. F. M. Pohl and T. M. Jovin, *J. Mol. Biol.*, 1972, **67**, 375-396.

9. Y. Mitsui, R. Langridge, R. C. Grant, M. Kodama, R. D. Wells, B. E. Shortle and C. R. Cantor, *Nature (London)*, 1970, **228**, 1166-1169.

10. A. T. Phan, V. Kuryavyi and D. J. Patel, *Curr. Opin. Struct. Biol.*, 2006, **16**, 288–298.

11. A. Rich and S. Zhang, *NATURE*, 2003, **4**, 566-572.

12. E. C. Friedberg, DNA Repair, 1995, W. H. Freeman and Company, New York.

13. E. C. Friedberg, *Cell Research*, 2008, **18**, 3-7.

14. T. Brown and W. N. Hunter, Non-Watson-Crick Base Associations in DNA and RNA revealed by X-ray diffraction methods: mismatches, modified bases and non-duplex DNA, 1997, John Wiley and Sons Ltd., England.

15. B. Singer and M. K. Dosanjh, *Mut. Res.*, 1990, **233**, 45-51.

16. T. Brown and O. Kennard, *Curr. Opin. Struct. Biol.*, 1992, **2**, 354-360.

17. K. Ezaz-Nikpay and G. L. Verdine, *Chemistry and Biology*, 1994, **1**, 235-240.

18. V. Thiviyanathan, A. Somasunderam, D. E. Volk, T. K. Hazra, S. Mitra and D. G. Gorenstein, *Biochem. Biophys. Res. Commun.*, 2008, **366**, 752–757.

19. Y. Osakada, K. Kawai, M. Fujitsuka and T. Majima, *Nuc. Acids Res.*, 2008, **36**, 5562–5570.

20. T. A. Kunkel and D. A. Erie, *Ann. Rev. Biochem.*, 2005, **74**, 681-710.

21. K-P. Hopfner and J. A. Tainer, *Structure*, 2000, **8**, R237-R241.

22. P. Modrich, *J. Biol. Chem.*, 1997, **272**, 24727-24730.

23. P. Hsieh and K. Yamane, *Mechanisms of Ageing and Development*, 2008, **129**, 391-407.

24. R. D. Kolodner and G. T. Marsischky, *Curr. Opin. Genet.*, 1999, **9**, 89-96.

25. M. Cho, S. Chung, S. Heo, J. Ku and C. Ban, *Biosens. Bioelectron.*, 2007, **22**, 1376-1381.

26. P. Sachadyn, A. Stanislawska-Sachadyn and J. Kur, *J. Biotech.*, 2007, **127**, 229-234.

27. P. Abad-Valle, M. T. Fernández-Abedul and A. Costa-García, *Biosens. Bioelectron.*, 2007, **22**, 1642-1650.

28. H. Chen, X-J. Liu, Y-L. Liu, J-H. Jiang, G-L. Shen and R-Q. Yu, *Biosens. Bioelectron.*, 2009, **24**, 1955-1961.

29. Z. Z. Zdraveski, J. A. Mellow, C. K. Farinelli, J. M. Essignmann and M. G. Marinus, *J. Biol. Chem.*, 2002, **277**, 1255-1260.

30. Y. Sedletska, L. Fourrier and J-M. Malinge, *J. Mol. Biol.*, 2007, **369**, 27-40.

31. F. R. Salsbury Jr., J. E. Clodfelter, M. B. Gentry, T. Hollis and K. Drotschmann Scarpinato, *Nuc. Acids Res.*, 2006, **34**, 2173–2185.

32. L. Stojic, R. Brun and J. Jiricny, *DNA Repair*, 2004, **3**, 1091–1101.

33. P. D. Hoffman, H. Wang, C. W. Lawrence, S. Iwai, F. Hanaoka and J. B. Hays, *DNA Repair*, 2005, **4**, 983-993.

34. H. Wang, P. D. Hoffman, C. Lawrence and J. B. Hays, *Environ. Mol. Mutagen.*, 2006, **47**, 296-306.

35. K. Taira, S. Nakamura, K. Nakano, D. Maehara, K. Okamoto, S. Arimoto, D. Loakes, L. Worth Jr., R. M. Schaaper, K. Seio, M. Sekine, K. Negishi and T. Negishi, *Mut. Res.*, 2008, **640**, 107–112.

36. P. Hsieh, *Mutation Research/DNA Repair*, 2001, **486**, 71-87.

37. A. Mazurek, C. N. Johnson, M. W. Germann and R. Fishel, *PNAS*, 2009, **106**, 4177-4182.

38. H. Chung, C. G. Lopez, D. J. Young, J. F. Lai, J. Holmstrom, D. Ream-Robinson, B. L. Cabrera and J. M. Carethers, *Oncogene*, 2010, **29**, 2172-2180

39. H. Chung, C. G. Lopez, J. Holmstrom, D. J. Young, J. F. Lai, D. Ream-Robinson and J. M. Carethers, *Human Molecular Genetics*, 2010, 1-10.

40. L. Somerville, E. Y. Krynetski, N. F. Krynetskaia, R. D. Beger, W. Zhang, C. A. Marhefka, W. E. Evans and R. W. Kriwacki, *J. Biol. Chem.*, 2003, **278**, 1005-1011.

41. X. Zhang, G. Jeffs, X. Ren, P. O'Donovan, B. Montaner, C. M. Perrett, P. Karran and Y-Z. Xu, *DNA Repair*, 2007, **6**, 344-354.

42. C. F. Guerra, E. J. Baerends and F. M. Bickelhaupt, *Cent. Eur. J. Chem.*, 2008, **6**, 15-21.

43. G. Vilani, *J. Phys. Chem. B*, 2009, **113**, 2128-2134.

44. J. Bohon and C. R. de los Santos, *Nuc. Acids Res.*, 2003, **31**, 1331-1338.

45. P. F. Swann, T. R Waters, D. C. Moulton, Y-Z. Xu, Q. Zheng, M. Edwards and R. Mace, *Science*, 1996, **273**, 1109-1111.

46. R. Wagner, P. Debbie and M. Radman, *Nuc. Acids Res.*, 1995, **23**, 3944-3948.

47. M. Jones, R. Wagner and M. Radman, *Genetics*, 1987, **115**, 605-610.

48. G. Obmolova, C. Ban, P. Hsieh and W. Yang, *Nature*, 2000, **407**, 703-710.

49. M. H. Lamers, A. Perrakis, J. H. Enzlin, H. H. K. Winterwerp N. de Wind and T. K. Sixma, *Nature*, 2000, **407**, 711-716.

50. E. Jacobs-Palmer and M. M. Hingorani, *J. Mol. Biol.*, 2007, **366**, 1087-1098.

51. M. S. Junop, G. Obmolova, K. Rausch, P. Hsieh and W. Yang, *Mol. Cell*, 2001, **7**, 1-12.

52. C. Spampinato and P. Modrich, *J. Biol. Chem.*, 2000, **275**, 9863-9869.

53. T. Selmane, M. J. Schofield, S. Nayak, C. Du and P. Hsieh, *J. Mol. Biol.*, 2003, **334**, 949-965.

54. W. Yang, *DNA Repair*, 2006, **5**, 654-666.

55. G. Natrajan, M. H. Lamers, J. H. Enzlin, H. H. K. Winterwerp, A. Perrakis and T. K. Sixma, *Nuc. Acids Res.*, 2003, **31**, 4814-4821.

56. M. J. Schofield, F. E. Brownewell, S. Nayak, C. Du, E. T. Kool and P. Hsieh, *J. Biol. Chem.*, 2001, **276**, 45505-45508.

57. A. Yanamoto, M. J. Schofield, I. Biswas and P. Hsieh, *Nuc. Acids Res.*, 2000, **28**, 3564-3569.

58. J. H. G. Lebbink, D. Georgijevic, G. Natrajan, a. Fish, H. H. K. Winterwerp, T. K. Sixma and N. de Wind, *EMBO J*, 2006, **25**, 409-419.

59. J. J. Warren, T. J. Pohlhaus, A. Changela, R. R. Iyer, P. L. Modich and L. S. Beese, *Mol. Cell*, 2007, **26**, 579-592.

60. P. Sachadyn, *Mutation Research*, 2010, **694**, 20-30.

61. K. J. Breslauer, *Methods Enzymol.*, 1995, **259**, 221-242.

62. D. A. Skoog, D. M. West and F. J. Holler, Fundamentals of Analytical Chemistry 7th Ed., 1995, Brooks/Cole Publishing Co.

63. J. R. Yates, C. I. Ruse and A. Nakorchevsky, *Annu. Rev. Biomed. Eng.*, 2009, **11**, 49–79.

64. S. Shah and S. H Friedman, *NATURE Protocols*, 2008, **3**, 351-356.

65. R. K. Harris, E. D. Becker, S. M. Cabral de Menezes, R. Goodfellow and P. Granger, *Pure Appl. Chem.*, 2001, **73**, 1795-1818.

66. K. Wüthrich, NMR of proteins and nucleic acids, 1986, John Wiley & Sons Inc., USA.

67. H. Günther, NMR Spectroscopy – Basic principles, concepts and applications in chemistry 2nd Ed., 1995, John Wiley & Sons Ltd., England.

68. R. J. Abraham, J. Fisher and P. Loftus, Introduction to NMR Spectroscopy, 1988, John Wiley & Sons Ltd., England.

69. J. Keeler, Understanding NMR Spectroscopy, 2005, John Wiley & Sons Ltd., England.

70. J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, M. Rance and N. J. Skelton, Protein NMR Spectroscopy: Principles and Practice 2nd Ed., 2007, Elsevier Academic Press, San Diego, U.S.A.

71. T. D. W. Claridge, High-Resolution NMR techniques in Organic Chemistry 2nd Ed., 2009, Elsevier Academic Press, London, U.K.

72. A. Kouranov, L. Xie, J. de la Cruz, L. Chen, J. Westbrook, P. E. Bourne and H. M. Berman, *Nuc. Acids Res.*, 2006, **34**, D302–D305.

73. F. H. T. Nelissen, A. J. van Gammeren, M. Tessari, F. C. Girard, H. A. Heus and S. S. Wijmenga, *Nuc. Acids Res.*, 2008, **36**, e89.

74. L. G. Scott and M. Hennig, *Meth. .Mol. Biol.*, 2008, **452**, 29-61.

75. W. Liu, H. M. Vu and D. R. Kearns, *Biochimica et Biophysica Acta*, 2002, **1574**, 93-99.

76. V. Tugarinov, R. Muhandiram, A. Ayed and L. E. Kay, *J. Am. Chem. Soc.*, 2002, **124**, 10025-10035.

77. W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T. E. Malliavin and M. Nilges, *Bioinformatics*, 2007, **23**, 381–382.

78. G. Cornilescu, F. Delaglio and A. Bax, *J. Biomol. NMR*, 1999, **13**, 289–302.

79. D. MacDonald and P. Lu, *Current Opinion in Structural Biology*, 2002, **12**, 337-343.

80. R. S. Lipsitz and N. Tjandra, *Annu. Rev. Biophys. Biomol. Struct.*, 2004, **33**, 387-413.

81. A. Roberts, J. G. Pelton and D. E. Wemmer, *Magn. Reson. Chem.*, 2006, **44**, S71-S82.

82. Z. Zhou, H. Feng, D. Flemming Hansen, H. Kato, E. Luk, D. I. Freedberg, L. E Kay, C. Wu and Y. Bai, *Nat. Struct. Mol. Biol.*, 2008, **15**, 868-869.

83. D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman, *Comput. Phys. Commun.*, 1995, **91**, 1-41.

84. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**,187-217.

85. W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger and W. F. van Gunsteren, *J. Phys. Chem. A*, 1999, **103**, 3596-3607.

86. S. Halouska, Y. Zhou, D. F. Becker and R. Powers, *Proteins*, 2009, **75**, 12–27.

87. T. Cierpicki, L. E. Risner, J. Grembecka, S. M. Lukasik, R. Popovic, M. Omonkowska, D. D Shultis, N. J. Zeleznik-Le and J. H. Bushweller, *Nat. Struct. Mol. Biol.*, 2010, **17**, 62-69.

88. F. Vendeix, 2003, PhD thesis, University of Manchester.

89. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse, *Science*, 1989, **246**, 64–71.

90. KaleidaGraph, http://www.synergy.com/kg.htm

91. R. Reed, D. Holmes, J. Weyers and A. Jones, Practical skills in biomolecular sciences 2nd Ed., 2003, Chpater 66 - Descriptive statistics, 475-477, Prentice Hall.

92. F. H. Martin, O. C. Uhlenbeck and P. Doty, *J. Mol. Biol.*, 1971, **57**, 201-215.

93. P. N. Borer, B. Dengler and I. Tinoco Jr., *J. Mol. Biol.*, 1974, **86**, 843-853.

94. R. I. Boysen, Y. Wang, H. Hong Keah and M. T. W. Hearn, *Biophys. Chem.*, 1999, **77**, 79-97.

95. D. I. Hoult, *J. Magn. Reson.*, 1976, **21**, 337-347.

96. M. Piotto, V. Saudek and V. Sklenář, *J. Biomol. NMR*, 1992, **2**, 661-665.

97. W.P. Aue, E. Bartholdi and R. R. Ernst, *J. Chem. Phys.*, 1976, **64**, 2229-2246.

98. A. Bax and R. Freeman, *J. Magn. Reson.*, 1981, **44**, 542-561.

99. D. Neuhaus, G. Wagner, M. Vašák, J. H. R. Kägi and K. Wüthrich, *Eur. J. Biochem*, 1985, **151**, 257–273.

100. M. Rance and P. E. Wright, *J. Magn. Reson.*, 1986, **66**, 372-378.

101. L. Braunschweiler and R. R. Ernst, *J. Magn. Reson.*, 1983, **53**, 521-528.

102. A. Bax and D. G. Davis, *J. Magn. Reson.*, 1985, **65**, 355-360.

103. S. Macura and R. R. Ernst, *Mol. Phys.*, 1980, **41**, 95.

104. S. Macura, Y. Huang, D. Suter and R. R. Ernst, *J. Magn. Reson.*, 1981, **43**, 259.

105. D. Brühwiler and G. Wagner, *J. Magn. Reson.*, 1986, **69**, 546-551.

106. G. Bodenhausen and D. J. Ruben, *Chem. Phys. Lett.*, 1980, **69**, 185-189.

107. G. A. Morris and R. Freeman, *J. Am. Chem. Soc.*, 1979, **101**, 760-762.

108. G. A. Morris, *J. Am. Chem. Soc.*, 1980, **102**, 428–429.

109. B. Luy and J. P. Marino, *J. Am. Chem. Soc.*, 2001, **123**, 11306–11307.

110. L.E. Kay, M. Ikura, R. Tschudin, and A. Bax, *J. Magn. Reson.*, 1990, **89**, 496-514.

111. M. Wittekind and L. Mueller, *J. Magn. Reson. B*, 1993, **101**, 201-205.

112. A. Bax, G. M. Clore and A. M. Gronenborn, *J. Magn. Reson.*, 1990, **88**, 425-431.

113. Spinworks, http://www.umanitoba.ca/chemistry/nmr/spinworks/

114. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax, *J. Biomol. NMR*, 1995, **6**, 277-293.

115. T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco.

116. W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, *PROTEINS: Structure, Function, and Bioinformatics*, 2005, **59**, 687–696.

117. L. Žídek, R. Štefl and V. Sklenář, *Curr. Opin. Struct. Biol.*, 2001, **11**, 275-281.

118. D. R. Hare, D. E. Wemmer, S-H. Chou, G. Drobny and B. R. Reid, *J. Mol. Biol.*, 1983, **171**, 319-336.

119. Y-Y. Tseng and S-H. Chou, *J. Chin. Chem. Soc.*, 1999, **46**, 699-708.

120. M. R. Conte, C. J. Bauer and A. N. Lane, *J. Biomol. NMR*, 1996, **7**, 190-206.

121. P. J. Lukavsky and J. D. Puglisi, *Method. Enzymol.*, 2005, **394**, 399-416.

122. L. E. Kay, *J. Magn. Reson.*, 2005, **173**, 193–207.

123. V. Tugarinov, P. M. Hwang and L. E. Kay, *Annu. Rev. Biochem.*, 2004, **73**, 107–146.

124. N. Tjandra and A. Bax, *Science*, 1997, **278**, 1111–1114.

125. A. G. Tzakos, C. R. R. Grace, P. J. Lukavsky and R. Riek, *Annu. Rev. Biophys. Biomol. Struct.*, 2006, **35**, 319–42.

126. L. E. Kay and K. H. Gardner, *Curr. Opin. Struct. Biol.*, 1997, **7**, 722-731.

127. A. Leach, Molecular modeling: Principles and practice 2nd Ed., 2001, Prentice Hall.

128. C. D. Schwieters, J. J. Kuszewski, N. Tjandra and G. M. Clore, *J. Magn. Reson.*, 2003, **160**, 65–73.

129. C. D. Schwieters, J. J. Kuszewski and G. M. Clore, *Prog. Nuc. Mag. Res. Sp.*, 2006, **48**, 47–62.

130. J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein. B. D. Sykes, P. E. Wright and K. Wüthrich, *Pure and Appl. Chem.*, 1998, **70**, 117-142.

131. C. D. Schwieters and G. M. Clore, *J. Magn. Reson.*, 2001, **149,** 239–244.

132. G. Zheng, X-J. Lu and W. K. Olson, *Nuc. Acids Res.*, 2009, **37,** W240-W246.

133. 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, H. Suzuki, C. S. Tung, E. Westhof, C. Wolberger and H. M. Berman, *J. Mol. Biol.*, 2001, **313**, 229-237.

134. H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider, *Biophys. J.*, 1992, **63**, 751-759.

135. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *Nuc. Acids Res.*, 2000, **28**, 235-242.

136. E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller; E. Nakatani, C. F. Schulte, D. E. Tolmie, R. K. Wenger, H. Yao, J. L. Markley, *Nuc. Acids Res.,* 2007, **36**, D402-D408.

137. C. Gu and Y. Wang, *Nuc. Acids Res.*, 2007, **35**, 3693-3704.

138. X. Ren, F. Li, G. Jeffs, X. Zhang, Y-Z. Xu and P. Karran, *Nuc. Acids Res.*, 2010, **38**, 1832-1840.

139. V. A. Malkov, I. Biswas, R. D. Camerini-Otero and P. Hsieh, *J. Biol. Chem.*, 1997, **272**, 23811-23817.

140. H. Gu, W. Yang and N. C. Seeman, *J. Am. Chem. Soc.*, 2010, **132**, 4352-4357.

141. M. L. Mendillo, C. D. Putnam, A. O. Mo, J. W. Jamison, S. Li, V. L. Woods Jr. and R. D. Kolodner, *J. Biol. Chem.*, 2010, **285**, 13170-13182.

142. L. E. Sass, C. Lanyi, K. Weninger and D. A. Erie, *Biochemistry*, 2010, **49**, 3174-3190.

Appendix I – NMR data processing scripts

i. Bruker NMR data conversion script

```
bruk2pipe -in /Desktop/TCdata_22_05_08/10/ser  \
  -bad 0.0 -aswap -DMX -decim 2040 -dspfvs 20 -grpdly 67.9862060546875  \
  -xN              2048   -yN                80   -zN               160  \
  -xT              1024   -yT                40   -zT                80  \
  -xMODE            DQD   -yMODE    States-TPPI   -zMODE    States-TPPI  \
  -xSW         9803.922   -ySW          4539.265   -zSW          13192.612  \
  -xOBS         699.948   -yOBS          176.008   -zOBS            70.933  \
  -xCAR           4.774   -yCAR           41.678   -zCAR           118.023  \
  -xLAB              1H   -yLAB              13C   -zLAB               15N  \
  -ndim               3   -aq2D           States                          \
  -out /Desktop/TCdata_22_05_08/10/fid/test%03d.fid -verb -ov
```

ii. Varian NMR data conversion script

```
var2pipe -in /home/john/Desktop/TCdata_020508/protein_hnca_25C.fid/fid \
 -noaswap  \
  -xN              1728   -yN                96   -zN                96  \
  -xT               864   -yT                48   -zT                48  \
  -xMODE        Complex   -yMODE        Complex   -zMODE        Complex  \
  -xSW          7200.072   -ySW          4525.399   -zSW          1919.985  \
  -xOBS          599.893   -yOBS          150.851   -zOBS            60.794  \
  -xCAR            4.774   -yCAR           56.005   -zCAR           118.023  \
  -xLAB                H   -yLAB              C13   -zLAB               N15  \
  -ndim                3   -aq2D           States                          \
  -out /Desktop/TCdata_020508/protein_hnca_25C.fid/fid1/test%03d.fid -verb -ov
```

iii. NMRpipe data processing and phasing script

For 2D data, the script below was used in NMRpipe to process and phase the data.

```
xyz2pipe -in /Desktop/TCdata_22_05_08/10/fid1/test%03d.fid -x -verb \
|nmrPipe -fn SOL -fl 50\
|nmrPipe -fn POLY -time\
|nmrPipe -fn GMB -lb -2.0 -gb 0.08 -c 0.5\
|nmrPipe -fn ZF -auto\
|nmrPipe -fn FT -auto\
|nmrPipe -fn PS -p0 39.2 -p1 143.2 -di\
|nmrPipe -fn TP\
|nmrPipe -fn SP -off 0.5 -end 0.98 -c 0.5\
|nmrPipe -fn ZF -auto\
|nmrPipe -fn FT -auto\
|nmrPipe -fn PS -p0 17.6 -p1 -7.8 -di\
|nmrPipe -fn TP\
|pipe2xyz -out /Desktop/TCdata_22_05_08/10/ft/test%03d.ft2 -y
```

For 3D data, the script below is added following the 2D script shown previously to account for the third frequency dimension.

```
xyz2pipe -in /Desktop/TCdata_22_05_08/10/ft/test%03d.ft2 -z -verb\
|nmrPipe -fn SP -off 0.5 -end 0.98 -c 0.5\
|nmrPipe -fn ZF -auto\
|nmrPipe -fn FT -auto\
|nmrPipe -fn PS -p0 -5.0 -p1 22.0 -di\
|pipe2xyz -out /Desktop/TCdata_22_05_08/10/ft/test%03d.ft3 –z
```

iv. NMRpipe data convert script

For 2D data**:**

```
/sparky/bin/./pipe2ucsf /Desktop/TCdata_22_05_08/15/post_hsqc.ft2
/Desktop/TCdata_22_05_08/15/post_hsqc.ucsf
```

For 3D data:

```
xyz2pipe -verb -in /Desktop/TCdata_22_05_08/10/ft/test%03d.ft3 -x\
| nmrPipe -out /Desktop/TCdata_22_05_08/10/ft/test%03d.spec -ov

/sparky/bin/./pipe2ucsf -123 /Desktop/TCdata_22_05_08/10/ft/test%03d.spec
/Desktop/TCdata_22_05_08/10/ft/HNCACB.ucsf
```

Appendix II – DNA structure calculation scripts

i. Xplor-NIH simulated annealing script

```
from pdbTool import PDBTool
from xplorPot import XplorPot
from rdcPotTools import create_RDCPot
from varTensorTools import create_VarTensor
import varTensorTools
from ivm import IVM
from potList import PotList
import protocol
from protocol import initMinimize
from ivm import IVM
from xplor import command
import random
from atomAction import SetProperty
from simulationTools import StructureLoop
from vec3 import Vec3
from psfGen import seqToPSF
from xplorPot import XplorPot
from varTensorTools import create_VarTensor
import varTensorTools
from ivm import IVM
from potList import PotList
import protocol
from avePot import AvePot
from simulationTools import MultRamp, StaticRamp, InitialParams, StructureLoop,
AnnealIVM, FinalParams
from simulationTools import AnnealIVM
from monteCarlo import randomizeTorsions

xplor.parseArguments()

command = xplor.command
from noePotTools import create_NOEPot

protocol.initParams("nucleic")
protocol.initTopology("nucleic")

rampedParams=[]
highTempParams=[]

init_t  = 3500.
final_t=25
bathTemp=2000
```

```
seqToPSF(open('Gstrand.seq').read(), seqType='dna')
seqToPSF(open('Tstrand.seq').read(), seqType='dna', startResid=18)

for atom in AtomSel("all"):
   atom.setPos( Vec3(float(atom.index())/10,
                     random.uniform(-0.5,0.5),
                     random.uniform(-0.5,0.5)) )
   pass

protocol.fixupCovalentGeom(useVDW=1,maxIters=100)

pots = PotList()

noex = create_NOEPot("noex",
                     "17GTH2O100327.tbl")
noex.setPotType("soft")
rampedParams.append( MultRamp(2,30.,"noex.setScale( VALUE )") )

noen = create_NOEPot("noen",
                     "17GTD2O100327.tbl")
noen.setPotType("soft")
rampedParams.append( MultRamp(2,30.,"noen.setScale( VALUE )") )

hbon = create_NOEPot("hbon",
                      "GThbond.tbl")
hbon.setPotType("hard")
rampedParams.append( MultRamp(2,30.,"hbon.setScale( VALUE )") )

protocol.initDihedrals("GTtor.tbl",
                  scale=5,
                  useDefaults=0)
highTempParams.append( StaticRamp("pots['CDIH'].setScale(10)") )
rampedParams.append( StaticRamp("pots['CDIH'].setScale(200)") )


pots.add( XplorPot('CDIH') )
pots['CDIH'].setThreshold( 5 )

xplor.command("@GTplanc.tbl")

pots.add( XplorPot("BOND") )
pots.add( XplorPot("DIHE") )
pots.add( XplorPot("ANGL") )
pots.add( XplorPot("IMPR") )

rampedParams.append( MultRamp(0.4,1.0,"pots['ANGL'].setScale(VALUE)"))
rampedParams.append( MultRamp(0.1,1.0,"pots['IMPR'].setScale(VALUE)"))

pots.add( XplorPot("VDW") )
rampedParams.append( MultRamp(0.9,0.78,
                     "xplor.command('param nbonds repel VALUE end end')") )
```

```
rampedParams.append( MultRamp(.004,4,
                    "xplor.command('param nbonds rcon VALUE end end')") )

pots.add(noex)
pots.add(noen)
pots.add(hbon)

pots.append(AvePot(XplorPot("plan",xplor.simulation)) )


from selectTools import IVM_groupRigidSidechain
from selectTools import IVM_breakRiboses

dyn = IVM()
protocol.initDynamics(dyn,potList=pots)
protocol.torsionTopology(dyn)


minc = IVM()
protocol.initMinimize(minc,potList=pots)
IVM_groupRigidSidechain(minc)
protocol.cartesianTopology(minc,"not resname ANI")


from simulationTools import AnnealIVM
cool = AnnealIVM(initTemp =init_t,
                    finalTemp=final_t,
                    tempStep =12.5,
                    ivm=dyn,
                    rampedParams = rampedParams)

cart_cool = AnnealIVM(initTemp =init_t,
                    finalTemp=25,
                        tempStep =12.5,
                    ivm=minc,
                    rampedParams = rampedParams)

def calcOneStructure( structData ):
                    randomizeTorsions(dyn)
                    InitialParams( rampedParams )
                    InitialParams( highTempParams )

protocol.initNBond(repel=1.2,
                    cutnb=100,
                    tolerance=45,
                    selStr="name P")

protocol.initDynamics(dyn,
                    potList=pots,
                    bathTemp=init_t,
                    initVelocities=1,
```

```
                    finalTime=800,
                    numSteps=8000,
                    printInterval=100)
dyn.setETolerance( init_t/100 )
dyn.run()

protocol.initNBond()
InitialParams( rampedParams )

protocol.initDynamics(dyn,
                    potList=pots,
                    numSteps=100,
                    finalTime=.2 ,
                    printInterval=100)

cool.run()


protocol.initMinimize(dyn,
                    printInterval=50)
dyn.run()
protocol.initDynamics(minc,
                    potList=pots,
                    numSteps=100,
                    finalTime=.4 ,
                    printInterval=100)
cart_cool.run()

protocol.initMinimize(minc,
                    potList=pots,
                    dEPred=10)
minc.run()
structData.writeStructure(pots)

simWorld.setRandomSeed( 785 )
outPDBFilename = 'SCRIPT_STRUCTURE.pdb'

StructureLoop(numStructures=100,
                    pdbTemplate=outPDBFilename,
                    structLoopAction=calcOneStructure,
                    genViolationStats=1,
                    averageTopFraction=0.3,
                    averageContext=FinalParams(rampedParams),
                    averageSortPots=[pots['BOND'],pots['ANGL'],pots['IMPR'],
                    noex,noen,pots['CDIH'],hbon],
                    averagePotList=pots).run()
```

ii. Xplor-NIH refinement script

```
from pdbTool import PDBTool
from xplorPot import XplorPot
from rdcPotTools import create_RDCPot
from varTensorTools import create_VarTensor
import varTensorTools
from ivm import IVM
from potList import PotList
import protocol
from protocol import initMinimize
from ivm import IVM
from xplor import command
import random
from atomAction import SetProperty
from simulationTools import StructureLoop
from vec3 import Vec3
from psfGen import seqToPSF
from xplorPot import XplorPot
from ivm import IVM
from potList import PotList
import protocol
from avePot import AvePot
from simulationTools import MultRamp, StaticRamp, InitialParams, StructureLoop, AnnealIVM
from simulationTools import AnnealIVM
xplor.parseArguments()

command = xplor.command
from noePotTools import create_NOEPot

protocol.initParams("nucleic")
protocol.initTopology("nucleic")

seed=10
numberOfStructures=100
startStructure=0
outFilename = "SCRIPT_STRUCTURE.pdb"

rampedParams=[]

init_t=2000
final_t=25
bathTemp=2000

startFile="annealGC13meru_24.pdb"

simWorld.setRandomSeed(seed)
seqToPSF(open('Gstrand.seq').read(), seqType='dna')
seqToPSF(open('Cstrand.seq').read(), seqType='dna', startResid=14)
```

```
protocol.initCoords(startFile)
protocol.covalentMinimize()

pots = PotList()
crossTerms=PotList('cross terms')

noex = create_NOEPot("noex",
                     "13GCH2O060809.tbl")
noex.setPotType("soft")

nonex = create_NOEPot("nonex",
                     "13GCD2O060809.tbl")
nonex.setPotType("hard")

hbon = create_NOEPot("hbon",
                      "GChbond.tbl")
hbon.setPotType("hard")
hbon.setScale(1000)
rampedParams.append( MultRamp(0.2,30.,"hbon.setScale( VALUE )") )

protocol.initDihedrals("GCtor.tbl",
                scale=5)
pots.append(AvePot(XplorPot,"cdih") )
rampedParams.append( StaticRamp("pots['CDIH'].setScale(200)") )

xplor.command("@GCplan.tbl")

pots.add( XplorPot("BOND") )
pots.add( XplorPot("DIHE") )
pots.add( XplorPot("ANGL") )
pots.add( XplorPot("IMPR") )

rampedParams.append( MultRamp(0.4,1.0,"pots['ANGL'].setScale(VALUE)"))
rampedParams.append( MultRamp(0.1,1.0,"pots['IMPR'].setScale(VALUE)"))

protocol.initNBond(cutnb=4.5)
pots.add( XplorPot("VDW") )
rampedParams.append( MultRamp(0.9,0.78,
                     "xplor.command('param nbonds repel VALUE end end')") )
rampedParams.append( MultRamp(.004,4,
                     "xplor.command('param nbonds rcon VALUE end end')") )

pots.add(noex)
pots.add(nonex)
pots.add(hbon)
pots.append(AvePot(XplorPot("plan",xplor.simulation)) )

mini = IVM()
from selectTools import IVM_groupRigidSidechain
from selectTools import IVM_breakRiboses
IVM_groupRigidSidechain(mini)
```

```
IVM_breakRiboses(mini, sel=0, breakSelStr="name O4' or name C1'")
protocol.cartesianTopology(mini,"not resname ANI")

protocol.initMinimize(mini,
                numSteps=20)
mini.fix("not resname ANI")
mini.run()

dyn = IVM()
protocol.initDynamics(dyn,potList=pots)
IVM_groupRigidSidechain(dyn)
IVM_breakRiboses(dyn, sel=0, breakSelStr="name O4' or name C1'")
protocol.torsionTopology(dyn)

from selectTools import IVM_groupRigidSidechain
minc = IVM()
protocol.initMinimize(minc,potList=pots)
IVM_groupRigidSidechain(minc)
IVM_breakRiboses(minc, sel=0, breakSelStr="name O4' or name C1'")
protocol.cartesianTopology(minc,"not resname ANI")
anneal= AnnealIVM(initTemp =init_t,
          finalTemp=25,
          tempStep =25,
          ivm=dyn,
          rampedParams = rampedParams)

InitialParams( rampedParams )

protocol.initMinimize(dyn,
                numSteps=1000)
dyn.run()

from simulationTools import testGradient

def calcOneStructure( structData ):

InitialParams( rampedParams )

protocol.initNBond(repel=1.2,
          cutnb=100,
          tolerance=45,
          selStr="name P")

protocol.initDynamics(dyn,
          initVelocities=1,
          bathTemp=init_t,
          potList=pots,
          finalTime=10)
          dyn.run()
protocol.initNBond()
```

```
protocol.initDynamics(dyn,
                      finalTime=0.2,
                      numSteps=0,
                      eTol_minimum=0.001
                      )
anneal.run()

protocol.initMinimize(dyn)

dyn.run()

minc.run()

structData.writeStructure(pots,crossTerms)


def accept(potList):
        """
        return True if current structure meets acceptance criteria
        """
if pots['noex'].violations()>2:
        return False
if pots['nonex'].violations()>2:
        return False
if pots['CDIH'].violations()>2:
        return False
if pots['BOND'].violations()>2:
        return False
if pots['IMPR'].violations()>2:
        return False

        return True

from simulationTools import StructureLoop, FinalParams
StructureLoop(numStructures=numberOfStructures,
        startStructure=startStructure,
        pdbTemplate=outFilename,
        structLoopAction=calcOneStructure,
        genViolationStats=1,
        averagePotList=pots,
        averageSortPots=[pots['BOND'],pots['ANGL'],pots['IMPR'],
        noex,nonex,pots['CDIH'],hbon],
        averageTopFraction=0.3,
        averageContext=FinalParams(rampedParams),
        averageFilename="SCRIPT_ave.pdb",
        averageFitSel="name P",
        averageCompSel="not resname ANI and not name H*"     ).run()
```

iii. Xplor-NIH average structure script

```
from pdbTool import PDBTool
from xplorPot import XplorPot
from rdcPotTools import create_RDCPot
from varTensorTools import create_VarTensor
import varTensorTools
from ivm import IVM
from potList import PotList
import protocol
from protocol import initMinimize
from ivm import IVM
from xplor import command
import random
from atomAction import SetProperty
from simulationTools import StructureLoop
from vec3 import Vec3
from psfGen import seqToPSF
from xplorPot import XplorPot
from ivm import IVM
from potList import PotList
import protocol
from avePot import AvePot
from simulationTools import MultRamp, StaticRamp, InitialParams, StructureLoop, AnnealIVM
from simulationTools import AnnealIVM
xplor.parseArguments()

command = xplor.command
from noePotTools import create_NOEPot

protocol.initParams("nucleic")
protocol.initTopology("nucleic")

seed=56789
numberOfStructures=1
startStructure=0
outFilename = "SCRIPT_STRUCTURE.pdb"

rampedParams=[]

init_t=2000
final_t=25
bathTemp=2000

startFile="averageb.pdb"

simWorld.setRandomSeed(seed)

seqToPSF(open('Gstrand.seq').read(), seqType='dna')
seqToPSF(open('Cstrand.seq').read(), seqType='dna', startResid=14)
```

```
protocol.initCoords(startFile)
protocol.covalentMinimize()

pots = PotList()
crossTerms=PotList('cross terms')

noex = create_NOEPot("noex",
                     "13GCH2O060809.tbl")
noex.setPotType("soft")

nonex = create_NOEPot("nonex",
                     "13GCD2O060809.tbl")
nonex.setPotType("hard")

hbon = create_NOEPot("hbon",
                      "GChbond.tbl")
hbon.setPotType("hard")
hbon.setScale(1000)
rampedParams.append( MultRamp(0.2,30.,"hbon.setScale( VALUE )") )


protocol.initDihedrals("GCtor.tbl",
                scale=5)
pots.append(AvePot(XplorPot,"cdih") )
rampedParams.append( StaticRamp("pots['CDIH'].setScale(200)") )

pots.add( XplorPot("BOND") )
pots.add( XplorPot("DIHE") )
pots.add( XplorPot("ANGL") )
pots.add( XplorPot("IMPR") )

rampedParams.append( MultRamp(0.4,1.0,"pots['ANGL'].setScale(VALUE)"))
rampedParams.append( MultRamp(0.1,1.0,"pots['IMPR'].setScale(VALUE)"))

protocol.initNBond(cutnb=4.5)
pots.add( XplorPot("VDW") )
rampedParams.append( MultRamp(0.9,0.78,
                     "xplor.command('param nbonds repel VALUE end end')") )
rampedParams.append( MultRamp(.004,4,
                     "xplor.command('param nbonds rcon VALUE end end')") )

pots.add(noex)
pots.add(nonex)
pots.add(hbon)

mini = IVM()
from selectTools import IVM_groupRigidSidechain
from selectTools import IVM_breakRiboses

dyn = IVM()
protocol.initDynamics(dyn,potList=pots)
```

```
IVM_groupRigidSidechain(dyn)
IVM_breakRiboses(dyn, sel=0, breakSelStr="name O4' or name C1'")
protocol.torsionTopology(dyn)

from selectTools import IVM_groupRigidSidechain
minc = IVM()
protocol.initMinimize(minc,potList=pots,numSteps=100)
protocol.cartesianTopology(minc,"resid 6 and resid 28 and resid 29")

minc.run()

protocol.initRamaDatabase('nucleic')
pots.append(AvePot(XplorPot,"rama") )
rampedParams.append( MultRamp(1,1,"xplor.command('rama scale VALUE end')"))

minc2 = IVM()
protocol.initMinimize(minc2,potList=pots,numSteps=100)
IVM_groupRigidSidechain(minc2)
protocol.cartesianTopology(minc2,"resid 6 and resid 28 and resid 29")

minc2.run()

minc3 = IVM()
protocol.initMinimize(minc3,potList=pots,numSteps=200)
IVM_groupRigidSidechain(minc3)
protocol.cartesianTopology(minc3,"all")

minc3.run()

command("write coordinates output=averageem2.pdb end")
from simulationTools import StructureLoop, FinalParams
StructureLoop(numStructures=numberOfStructures,
        startStructure=startStructure,
        pdbTemplate=outFilename,
        genViolationStats=1,
        averagePotList=pots,
        averageSortPots=[pots['BOND'],pots['ANGL'],pots['IMPR'],
        ],
        averageTopFraction=0.25,
        averageFitSel="name P",
        averageCompSel="not resname ANI and not name H*"     ).run()
```

Appendix III – Data tables

i. Conformational analysis - 17mer canonical GC DNA

Base Pair	Nucleotides	Shear (S_x)	Stretch (S_y)	Stagger (S_z)	Buckle (κ)	Propeller Twist (ω)	Opening (σ)
1	A-T	-0.29	-0.25	1.08	21.18	-1.29	-4.00
2	G-C	0.08	-0.39	-0.85	15.04	-4.21	2.36
3	C-G	-0.25	-0.52	-0.98	-10.29	-1.12	-0.17
4	T-A	0.12	-0.30	0.67	-15.35	-3.92	-4.18
5	G-C	0.19	-0.23	0.67	8.60	3.61	0.37
6	C-G	-0.12	-0.27	-0.35	-6.29	7.00	-0.42
7	C-G	-0.05	-0.29	0.06	-8.05	3.22	-0.61
8	A-T	-0.18	-0.30	0.51	5.76	-2.43	-4.54
9	G-C	0.07	-0.22	-0.07	13.34	-0.53	-2.83
10	G-C	0.02	-0.27	-0.33	10.56	1.30	0.95
11	C-G	0.11	-0.37	-0.31	-14.98	4.13	-2.90
12	A-T	0.11	0.04	0.16	6.40	3.93	4.06
13	C-G	-0.35	-0.24	-0.79	0.76	11.01	-4.10
14	C-G	-0.07	-0.36	0.33	-5.90	-0.67	-4.85
15	A-T	-0.01	-0.23	0.54	14.87	-7.08	-0.29
16	G-C	-0.04	-0.47	-0.65	11.00	-4.56	1.40
17	T-A	0.23	-0.08	0.24	-25.14	-1.49	-2.37
Average		-0.03	-0.28	0.00	1.27	0.41	-1.30

Local base pair parameters in the NMR structure of 17mer canonical GC DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 3.20.

Step No.	Step	Shift (D_x)	Slide (D_y)	Rise (D_z)	Tilt (Tτ)	Roll (Pρ)	Helical Twist (Ω)
1	AG / CT	0.35	-0.44	3.77	8.42	6.15	36.72
2	GC / GC	-0.20	-0.67	4.39	-0.56	11.24	35.74
3	CT / AG	-0.21	-0.50	3.72	-6.62	5.21	36.73
4	TG / CA	0.39	-0.53	2.31	-0.70	1.20	34.33
5	GC / GC	-0.01	-0.77	4.05	3.73	8.94	34.99
6	CC / GG	-0.19	-0.68	3.68	-2.93	4.72	34.01
7	CA / TG	-0.34	-0.59	2.91	-1.48	1.90	33.85
8	AG / TT	0.01	-0.62	3.06	2.58	4.76	34.88
9	GG / CC	0.35	-0.64	3.54	1.74	7.49	34.34
10	GC / GC	-0.27	-0.75	4.70	-0.97	9.37	35.91
11	CA / TG	0.73	-0.53	2.59	-1.11	1.68	33.63
12	AC / GT	-0.97	-0.89	3.59	1.01	7.48	32.83
13	CC / GG	-0.28	-0.75	3.91	-3.88	4.44	35.03
14	CA / TG	0.43	-0.47	2.37	0.33	1.91	35.11
15	AG / CT	-0.04	-0.53	3.46	4.34	6.30	35.53
16	GT / AC	-0.11	-0.67	5.22	-3.43	8.66	37.02
Average		-0.02	-0.63	3.58	0.03	5.72	35.04

Local base pair step parameters in the NMR structure of 17mer canonical GC DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 3.20.

Step No.	Step	X-displacement (d_x)	Y-displacement (d_y)	Inclination (η)	Tip (θ)
1	AG / CT	-1.66	0.77	9.89	-13.55
2	GC / GC	-3.22	0.22	18.61	0.93
3	CT / AG	-1.61	-0.70	8.33	10.60
4	TG / CA	-1.03	-0.73	2.03	1.19
5	GC / GC	-2.91	0.69	15.06	-6.29
6	CC / GG	-2.01	-0.20	8.11	5.03
7	CA / TG	-1.28	0.38	3.27	2.54
8	AG / TT	-1.70	0.35	7.97	-4.32
9	GG / CC	-2.33	-0.30	12.79	-2.98
10	GC / GC	-3.10	0.25	15.37	1.59
11	CA / TG	-1.12	-1.41	2.90	1.92
12	AC / GT	-2.94	1.89	13.34	-1.80
13	CC / GG	-2.04	-0.23	7.40	6.47
14	CA / TG	-0.98	-0.68	3.16	-0.55
15	AG / CT	-1.83	0.72	10.41	-7.16
16	GT / AC	-2.90	-0.55	13.78	5.46
Average		-2.04	0.03	9.53	-0.06

Local base pair helical parameters in the NMR structure of 17mer canonical GC DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 3.20.

Base	Nucleotides	C1'-N (χ)	P-O5' (α)	O5'-C5' (β)	C5'-C4' (γ)	C4'-C3' (δ)	C3'-O3' (ε)	O3'-P (ζ)
1	A	-107.0			47.4	140.1	-179.4	-96.2
2	G	-101.1	-63.2	-175.0	44.0	137.1	176.9	-101.2
3	C	-113.6	-65.0	-166.6	48.8	137.9	177.4	-95.8
4	T	-114.2	-61.6	-165.2	48.4	137.1	176.4	-98.3
5	G	-104.7	-65.5	-172.8	47.8	137.1	175.4	-102.9
6	C	-114.1	-65.3	-169.4	49.1	137.1	176.3	-100.4
7	C	-117.9	-63.5	-167.5	50.0	136.7	177.4	-101.3
8	A	-108.1	-65.0	-168.7	49.7	138.1	177.5	-100.0
9	G	-104.1	-66.5	-172.6	49.1	137.7	178.4	-98.8
10	G	-105.0	-66.0	-170.0	48.1	137.2	177.3	-100.8
11	C	-117.8	-64.8	-165.3	49.5	137.6	177.5	-96.1
12	A	-106.1	-63.9	-172.3	48.6	138.2	-179.0	-117.0
13	C	-118.8	-66.5	-174.8	49.9	137.5	179.6	-99.2
14	C	-117.7	-62.7	-166.1	50.2	137.9	177.3	-96.4
15	A	-104.4	-64.1	-173.5	48.5	136.9	176.9	-101.7
16	G	-105.0	-65.8	-173.1	49.2	138.7	-180.0	-95.4
17	T	-114.1	-59.5	-158.9	43.1	136.8		
Average		-110.2	-64.3	-169.5	48.3	137.6	110.4	-100.1

Backbone dihedral angles for residues A1-T17 in the NMR structure of 17mer canonical GC DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 3.20.

Base	Nucleotides	C1'-N (χ)	P-O5' (α)	O5'-C5' (β)	C5'-C4' (γ)	C4'-C3' (δ)	C3'-O3' (ε)	O3'-P (ζ)
18	A	-105.4			46.5	139.5	179.9	-95.7
19	C	-112.5	-63.2	-166.8	47.8	137.4	176.5	-95.7
20	T	-113.3	-61.4	-167.0	47.6	135.7	176.3	-104.1
21	G	-105.5	-66.4	-173.1	49.4	138.5	176.0	-98.0
22	G	-109.1	-65.2	-170.0	47.5	137.8	175.4	-97.1
23	T	-114.4	-61.1	-167.3	48.5	135.6	177.9	-108.6
24	G	-105.5	-67.9	-172.5	50.9	139.4	179.2	-98.3
25	C	-113.9	-63.7	-166.9	49.0	137.5	176.1	-99.6
26	C	-113.1	-64.9	-166.2	48.3	138.0	175.1	-96.8
27	T	-111.9	-62.6	-168.8	47.9	135.8	176.5	-99.8
28	G	-105.3	-65.2	-171.1	48.4	137.1	176.9	-100.4
29	G	-105.7	-65.1	-168.7	48.8	137.4	177.0	-103.7
30	T	-118.5	-63.5	-167.4	51.1	137.8	176.8	-101.7
31	A	-105.2	-65.8	-172.3	49.8	138.2	177.0	-98.8
32	G	-104.6	-65.3	-173.4	48.8	137.5	176.9	-99.1
33	C	-114.2	-63.2	-167.2	49.1	137.9	178.3	-94.1
34	T	-113.1	-59.9	-160.1	42.1	137.0		
Average		-110.1	-64.0	-168.7	48.3	137.5	177.0	-99.5

Backbone dihedral angles for residues A18-T34 in the NMR structure of 17mer canonical GC DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 3.20.

Base	Nucleotides	ν_0	ν_1	ν_2	ν_3	ν_4	Amplitude	Phase	Pucker
1	A	-18.4	30.0	-29.5	19.7	-1.0	30.8	163.0	$C_{2'}$-endo
2	G	-22.1	32.1	-29.4	17.4	2.8	32.1	156.3	$C_{2'}$-endo
3	C	-27.1	32.1	-29.7	18.0	2.2	32.1	157.5	$C_{2'}$-endo
4	T	-23.1	32.9	-29.6	17.1	3.6	32.7	155.0	$C_{2'}$-endo
5	G	-23.3	33.1	-29.8	17.1	3.7	32.9	154.8	$C_{2'}$-endo
6	C	-22.1	32.2	-29.5	17.5	2.7	32.2	156.5	$C_{2'}$-endo
7	C	-23.3	32.6	-29.2	16.6	4.0	32.4	154.3	$C_{2'}$-endo
8	A	-21.5	31.8	-29.5	17.9	2.2	31.9	157.5	$C_{2'}$-endo
9	G	-22.0	32.0	-29.5	17.6	2.6	32.1	156.7	$C_{2'}$-endo
10	G	-22.4	32.3	-29.4	17.3	3.1	32.2	155.9	$C_{2'}$-endo
11	C	-23.5	33.0	-29.5	16.8	4.1	32.8	154.3	$C_{2'}$-endo
12	A	-23.3	33.2	-29.8	17.2	3.6	32.9	154.9	$C_{2'}$-endo
13	C	-22.1	32.0	-29.4	17.4	2.8	32.1	156.4	$C_{2'}$-endo
14	C	-22.7	32.4	-29.4	17.1	3.4	32.3	155.4	$C_{2'}$-endo
15	A	-22.9	32.6	-29.4	16.9	3.6	32.5	155.0	$C_{2'}$-endo
16	G	-21.9	32.3	-29.9	18.1	2.3	32.3	157.4	$C_{2'}$-endo
17	T	-20.4	31.4	-29.6	18.8	0.8	31.6	159.7	$C_{2'}$-endo
Average		-22.5	32.2	-29.5	17.6	2.7	32.2	156.5	

Deoxyribose sugar dihedral angles for residues A1-T17 in the NMR structure of 17mer canonical GC DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 3.20.

Base	Nucleotides	ν_0	ν_1	ν_2	ν_3	ν_4	Amplitude	Phase	Pucker
18	A	-21.3	32.1	-29.8	18.4	1.6	32.1	158.2	$C_{2'}$-endo
19	C	-22.1	32.2	-29.5	17.6	2.6	32.2	156.6	$C_{2'}$-endo
20	T	-22.2	32.3	-29.6	17.5	2.9	32.3	156.3	$C_{2'}$-endo
21	G	-21.7	32.0	-29.6	17.9	2.3	32.1	157.3	$C_{2'}$-endo
22	G	-21.9	31.9	-29.4	17.5	2.7	32.0	156.6	$C_{2'}$-endo
23	T	-22.0	32.1	-29.4	17.5	2.6	32.1	156.6	$C_{2'}$-endo
24	G	-22.7	32.5	-29.5	17.2	3.3	32.4	155.6	$C_{2'}$-endo
25	C	-24.6	33.7	-29.4	16.0	5.2	33.2	152.3	$C_{2'}$-endo
26	C	-22.6	32.9	-30.0	17.9	2.8	32.8	156.4	$C_{2'}$-endo
27	T	-22.2	32.3	-29.6	17.6	2.7	32.3	156.5	$C_{2'}$-endo
28	G	-20.2	31.2	-29.7	18.9	0.7	31.6	160.1	$C_{2'}$-endo
29	G	-23.6	32.8	-29.2	16.4	4.3	32.5	153.7	$C_{2'}$-endo
30	T	-23.5	33.7	-30.3	17.5	3.6	33.4	155.1	$C_{2'}$-endo
31	A	-21.9	32.3	-29.8	18.0	2.2	32.3	157.3	$C_{2'}$-endo
32	G	-23.9	32.2	-29.2	16.3	4.6	32.8	153.2	$C_{2'}$-endo
33	C	-22.5	32.8	-29.8	17.6	3.0	32.6	156.0	$C_{2'}$-endo
34	T	-19.4	30.8	-29.7	19.3	-0.1	31.4	161.4	$C_{2'}$-endo
Average		-22.3	32.3	-29.6	17.6	2.8	32.4	156.4	

Deoxyribose sugar dihedral angles for residues A18-T34 in the NMR structure of 17mer canonical GC DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 3.20.

Base Pair	Hydrogen Bond	Distance (Å)
A1-T34	H61-O4	2.86
	N1-H3	2.86
G2-C33	O6-H41	2.81
	H1-N3	2.79
	H21-O2	2.70
C3-G32	O6-H41	2.68
	H1-N3	2.71
	H21-O2	2.64
T4-A31	H61-O4	2.72
	N1-H3	2.74
G5-C30	O6-H41	2.83
	H1-N3	2.81
	H21-O2	2.67
C6-G29	O6-H41	2.80
	H1-N3	2.75
	H21-O2	2.60
C7-G28	O6-H41	2.71
	H1-N3	2.71
	H21-O2	2.59
A8-T27	H61-O4	2.72
	N1-H3	2.78
G9-C26	O6-H41	2.71
	H1-N3	2.81
	H21-O2	2.79
G10-C25	O6-H41	2.81
	H1-N3	2.75
	H21-O2	2.58
C11-G24	O6-H41	2.63
	H1-N3	2.71
	H21-O2	2.67
A12-T23	H61-O4	3.19
	N1-H3	2.94
C13-G22	O6-H41	2.87
	H1-N3	2.87
	H21-O2	2.82
C14-G21	O6-H41	2.54
	H1-N3	2.70
	H21-O2	2.74
A15-T20	H61-O4	2.88
	N1-H3	2.73
G16-C19	O6-H41	2.67
	H1-N3	2.66
	H21-O2	2.57
T17-A18	H61-O4	2.88
	N1-H3	2.88

Base pair hydrogen bond distances in 17mer canonical GC DNA. Values are recorded to two decimal places and are based on structures shown in Figure 3.20.

ii. Conformational analysis - 17mer mismatch GT DNA (800 MHz)

Base	Nucleotides	C1'-N (χ)	P-O5' (α)	O5'-C5' (β)	C5'-C4' (γ)	C4'-C3' (δ)	C3'-O3' (ε)	O3'-P (ζ)
1	A	-106.2			60.1	136.6	176.2	-106.3
2	G	-105.0	-72.7	-176.5	58.4	135.3	178.2	-97.3
3	C	-123.5	-72.3	-165.5	56.4	133.8	175.0	-98.0
4	T	-120.8	-70.2	-173.6	58.6	133.1	178.0	-95.8
5	G	-104.3	-78.3	-173.3	56.0	134.1	175.8	-97.1
6	C	-121.9	-76.0	-168.0	55.9	133.2	175.4	-106.0
7	C	-121.4	-71.1	-175.6	59.2	134.4	175.3	-93.9
8	A	-104.8	-72.8	-175.8	58.4	134.0	175.3	-100.1
9	G	-116.9	-71.6	-165.2	57.3	133.9	175.1	-95.2
10	G	-113.5	-73.9	-169.6	56.6	133.9	174.3	-101.6
11	C	-125.0	-69.2	-173.0	58.7	133.8	174.7	-99.6
12	A	-109.5	-75.9	-172.0	56.5	135.2	176.3	-106.8
13	C	-121.3	-75.1	-172.8	56.2	134.2	-180.0	-93.8
14	C	-121.5	-76.5	-164.8	56.0	133.5	176.3	-98.5
15	A	-104.9	-77.2	-174.1	57.8	134.4	177.2	-100.5
16	G	-113.6	-75.7	-172.2	57.2	134.3	174.3	-100.9
17	T	-121.4	-66.2	-171.8	59.9	135.7		
Average		-115.0	-73.4	-171.5	57.6	134.3	153.6	-99.5

Backbone dihedral angles for residues A1-T17 in the NMR structure of 17mer mismatch GT DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 3.11.

Base	Nucleotides	C1'-N (χ)	P-O5' (α)	O5'-C5' (β)	C5'-C4' (γ)	C4'-C3' (δ)	C3'-O3' (ε)	O3'-P (ζ)
18	A	-105.6			59.9	135.8	176.8	-102.2
19	C	-121.6	-74.8	-170.5	58.0	134.3	177.4	-96.1
20	T	-121.9	-71.4	-165.7	57.8	134.8	180.0	-95.1
21	G	-106.6	-80.6	-171.7	56.0	133.8	175.4	-103.0
22	G	-116.7	-76.1	-173.9	57.3	134.2	173.2	-98.4
23	T	-121.4	-67.5	-173.9	59.1	133.4	-180.0	-99.1
24	G	-108.2	-80.0	-168.2	56.6	135.2	176.2	-99.9
25	C	-121.4	-73.5	-170.4	58.5	134.2	174.6	-93.4
26	T	-121.2	-67.4	-173.2	60.4	133.5	177.3	-110.3
27	T	-121.5	-69.2	-175.4	60.0	134.7	175.9	-101.4
28	G	-109.4	-75.7	-169.9	56.9	134.2	175.2	-96.1
29	G	-108.5	-74.4	-171.0	56.6	134.2	175.4	-102.3
30	T	-121.9	-74.0	-170.3	57.1	134.5	175.5	-97.0
31	A	-108.4	-73.4	-170.5	56.8	134.6	179.2	-101.6
32	G	-106.9	-77.8	-174.4	57.0	135.2	175.1	-99.3
33	C	-121.4	-75.9	-170.8	56.8	133.5	176.9	-93.4
34	T	-121.3	-72.7	-163.3	57.6	134.6		
Average		-115.5	-74.0	-170.8	57.8	134.4	154.0	-99.3

Backbone dihedral angles for residues A18-T34 in the NMR structure of 17mer mismatch GT DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 3.11.

Base	Nucleotides	ν_0	ν_1	ν_2	ν_3	ν_4	Amplitude	Phase	Pucker
1	A	-26.3	34.8	-29.7	15.4	6.6	34.2	150.2	C_2-endo
2	G	-28.2	36.0	-29.7	14.2	8.5	35.3	147.3	C_2-endo
3	C	-30.5	37.3	-29.7	12.9	10.7	36.6	144.3	C_2-endo
4	T	-31.3	37.8	-29.6	12.3	11.7	37.0	143.1	C_1-exo
5	G	-30.1	37.1	-29.6	13.1	10.4	36.3	144.8	C_2-endo
6	C	-31.2	37.7	-29.6	12.4	11.5	37.0	143.3	C_1-exo
7	C	-29.7	36.8	-29.7	13.4	9.9	36.1	145.4	C_2-endo
8	A	-30.3	37.2	-29.7	13.0	10.5	36.4	144.6	C_2-endo
9	G	-30.4	37.3	-29.7	12.9	10.7	36.5	144.4	C_2-endo
10	G	-30.3	37.3	-29.7	13.0	10.5	36.5	144.6	C_2-endo
11	C	-30.6	37.4	-29.7	12.8	10.8	36.6	144.2	C_2-endo
12	A	-28.2	36.0	-29.8	14.3	8.4	35.3	147.5	C_2-endo
13	C	-30.2	37.2	-29.7	13.1	10.4	36.4	144.7	C_2-endo
14	C	-31.0	37.6	-29.7	12.6	11.3	36.9	143.6	C_1-exo
15	A	-29.8	36.9	-29.7	13.4	10.0	36.1	145.3	C_2-endo
16	G	-29.6	36.9	-29.8	13.5	9.8	36.1	145.6	C_2-endo
17	T	-27.9	37.1	-31.5	16.2	7.1	36.3	150.0	C_2-endo
Average		-29.7	37.0	-29.8	13.4	9.9	36.2	145.5	

Deoxyribose sugar dihedral angles for residues A1-T17 in the NMR structure of 17mer mismatch GT DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 3.11.

Base	Nucleotides	ν_0	ν_1	ν_2	ν_3	ν_4	Amplitude	Phase	Pucker
18	A	-29.0	37.6	-31.2	15.3	8.3	36.8	148.2	C_2-endo
19	C	-31.1	37.7	-29.6	12.5	11.4	36.9	143.4	C_1-exo
20	T	-28.8	36.4	-29.8	14.0	8.9	35.6	146.8	C_2-endo
21	G	-29.2	36.5	-29.7	13.7	9.4	35.8	146.1	C_2-endo
22	G	-29.4	36.7	-29.7	13.5	9.7	35.9	145.7	C_2-endo
23	T	-30.0	37.0	-29.7	13.2	10.2	36.2	145.0	C_2-endo
24	G	-29.8	36.9	-29.7	13.4	10.0	36.2	145.3	C_2-endo
25	C	-29.2	36.6	-29.7	13.7	9.5	35.8	146.0	C_2-endo
26	T	-20.8	37.5	-29.6	12.6	11.1	36.7	143.8	C_1-exo
27	T	-30.0	37.1	-29.7	13.1	10.3	36.3	144.9	C_2-endo
28	G	-28.4	36.1	-29.8	14.2	8.7	35.4	147.2	C_2-endo
29	G	-30.7	37.4	-29.6	12.6	11.1	36.7	143.9	C_1-exo
30	T	-29.9	37.1	-29.7	13.3	10.1	36.3	145.1	C_2-endo
31	A	-30.5	37.3	-29.6	12.8	10.8	36.6	144.1	C_2-endo
32	G	-29.1	36.5	-29.7	13.8	9.3	35.8	146.3	C_2-endo
33	C	-29.6	36.8	-29.7	13.4	9.9	36.1	145.4	C_2-endo
34	T	-27.6	35.6	-29.8	14.7	7.8	34.9	148.4	C_2-endo
Average		-29.0	36.9	-29.8	13.5	9.8	36.1	145.6	

Deoxyribose sugar dihedral angles for residues A18-T34 in the NMR structure of 17mer mismatch GT DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 3.11.

Base Pair	Hydrogen Bond	Distance (Å)
A1-T34	H61-O4	3.02
	N1-H3	2.88
G2-C33	O6-H41	2.56
	H1-N3	2.80
	H21-O2	2.94
C3-G32	O6-H41	2.61
	H1-N3	2.73
	H21-O2	2.75
T4-A31	H61-O4	3.01
	N1-H3	2.85
G5-C30	O6-H41	2.89
	H1-N3	2.90
	H21-O2	2.81
C6-G29	O6-H41	3.01
	H1-N3	3.02
	H21-O2	2.91
C7-G28	O6-H41	2.61
	H1-N3	2.70
	H21-O2	2.71
A8-T27	H61-O4	3.09
	N1-H3	2.81
G9-T26	O6-H3	3.50
	H1-O2	3.07
G10-C25	O6-H41	2.99
	H1-N3	2.90
	H21-O2	2.71
C11-G24	O6-H41	2.98
	H1-N3	2.84
	H21-O2	2.61
A12-T23	H61-O4	3.10
	N1-H3	2.93
C13-G22	O6-H41	2.62
	H1-N3	2.81
	H21-O2	2.92
C14-G21	O6-H41	2.82
	H1-N3	2.87
	H21-O2	2.81
A15-T20	H61-O4	2.71
	N1-H3	2.72
G16-C19	O6-H41	2.60
	H1-N3	2.69
	H21-O2	2.68
T17-A18	H61-O4	3.09
	N1-H3	3.07

Base pair hydrogen bond distances in 17mer mismatch GT DNA. Values are recorded to two decimal places and are based on structures shown in Figure 3.11.

iii. ^{1}H chemical shift table - 17mer mismatch GT DNA (1000 MHz)

Number	Nucleotide Number	H2	CH_3(5)	H5	H6	H8	H1'	H2'	H2"	H3'	H4'	H5'	H5"
1	A1	7.94				8.01	5.98	2.46	2.63	4.83	4.21	3.69	3.69
2	G2					7.89	5.86	2.68	2.70	4.98	4.40	4.26	4.12
3	C3			5.30	7.42		6.00	2.05	2.51	4.74	4.27		4.17
4	T4		1.61		7.32		5.80	2.10	2.48	4.87	4.14		4.08
5	G5					7.88	5.84	2.62	2.68	4.97	4.37	4.09	4.14
6	C6			5.29	7.34		5.88	2.01	2.37	4.76	4.18	4.11	4.23
7	C7			5.56	7.42		5.30	2.01	2.29	4.78	4.05	4.01	4.10
8	A8	7.42				8.19	6.00	2.72	2.85	5.01	4.34	4.09	4.00
9	G9					7.50	5.65	2.46	2.65	4.91	4.34		4.16
10	G10					7.54	5.80	2.41	2.66	4.88	4.31		4.16
11	C11			5.28	7.27		5.56	1.98	2.35	4.81	4.11	4.04	4.08
12	A12	7.69				8.21	6.18	2.67	2.87	4.99	4.40	4.26	4.12
13	C13			5.22	7.23		5.79	1.94	2.35	4.75	4.14	4.03	4.26
14	C14			5.52	7.42		5.34	1.98	2.28	4.80	4.05	3.98	4.10
15	A15	7.70				8.17	6.00	2.71	2.84	5.02	4.38	3.98	4.10
16	G16					7.72	5.93	2.53	2.65	4.93	4.36	4.18	4.22
17	T17		1.55		7.36		6.19	2.20	2.22	4.51	4.07		4.22
18	A18	8.03				8.23	6.24	2.71	2.81	4.85	4.27	3.75	3.75
19	C19			5.43	7.51		5.92	2.11	2.46	4.78	4.24	4.10	4.18
20	T20		1.61		7.30		5.73	2.05	2.42	4.85	4.14	4.07	4.10
21	G21					7.85	5.69	2.69	2.75	4.99	4.36	4.05	4.12
22	G22					7.65	5.92	2.50	2.72	4.89	4.39	4.11	4.20
23	T23		1.32		7.12		5.82	2.00	2.45	4.84	4.16	4.19	4.11
24	G24					7.82	5.82	2.58	2.63	4.94	4.34	4.06	4.11
25	C25			5.20	7.30		6.00	2.05	2.51	4.73	4.25	4.10	4.19
26	T26		1.73		7.40		5.83	1.97	2.28	4.78	3.99		4.03
27	T27		1.60		7.39		5.56	2.18	2.39	4.84	4.14	3.96	4.18
28	G28					7.79	5.63	2.65	2.73	4.97	4.31	4.07	3.97
29	G29					7.71	5.80	2.51	2.65	4.95	4.36		4.15
30	C30			5.30	7.28		5.49	1.93	2.30	4.79	4.10	3.99	4.16
31	A31	7.61				8.14	6.01	2.71	2.88	5.02	4.37	4.18	3.99
32	G32					7.64	5.74	2.48	2.61	4.95	4.34	4.24	4.18
33	C33			5.37	7.43		6.05	2.20	2.46	4.74	4.20	4.04	4.12
34	T34		1.73		7.58		6.25	2.28	2.28	4.55	4.07	4.04	4.15

Summary of chemical shifts for ^{1}H resonances in 17mer mismatch GT DNA from NOESY spectra measured at 1000MHz.

iv. Conformational analysis - 17mer mismatch GT DNA (1000 MHz)

Base Pair	Nucleotides	Shear (S_x)	Stretch (S_y)	Stagger (S_z)	Buckle (κ)	Propeller Twist (ω)	Opening (σ)
1	A-T	0.27	0.00	0.54	5.22	-16.73	-2.86
2	G-C	0.02	-0.25	0.00	15.92	-12.75	-10.90
3	C-G	-0.17	-0.28	0.30	-2.96	.13.43	-9.48
4	T-A	0.00	-0.22	0.36	-2.41	-26.10	2.18
5	G-C	-0.19	-0.35	0.36	8.48	-10.56	0.52
6	C-G	0.12	-0.39	0.53	-10.21	-6.34	-3.32
7	C-G	0.06	-0.41	0.51	-8.22	-5.84	-3.57
8	A-T	0.53	-0.01	0.91	22.33	-7.67	1.68
9	G-T	-0.27	-1.11	2.49	11.68	-12.15	-1.00
10	G-C	-0.02	-0.51	0.76	0.62	-7.44	-4.10
11	C-G	0.11	-0.36	0.38	-11.66	-9.17	-2.44
12	A-T	0.46	-0.05	0.53	6.49	-12.57	0.47
13	C-G	0.04	-0.32	0.08	-6.69	3.55	-5.24
14	C-G	0.03	-0.37	0.39	-8.34	-6.51	-2.05
15	A-T	0.27	-0.19	0.50	11.17	-18.81	-1.60
16	G-C	0.01	-0.27	0.06	3.48	-2.48	-5.85
17	T-A	0.30	-0.10	0.22	-8.61	-12.19	-1.13
Average		0.09	-0.31	0.52	1.55	-10.24	-2.86

Local base pair parameters in the 1000 MHz NMR structure of 17mer mismatch GT DNA based on the average structure in Figure 3.15 and calculated using 3DNA.

Step No.	Step	x-displacement (d_x)	y-displacement (d_y)	Inclination (η)	Tip (θ)
1	AG / CT	-1.64	1.61	7.32	-3.01
2	GC / GC	-2.83	-0.01	21.71	0.20
3	CT / AG	-1.82	-1.27	15.39	-2.99
4	TG / CA	-1.58	-0.79	14.81	3.92
5	GC / GC	-3.21	0.16	21.76	2.13
6	CC / GG	-3.58	-0.32	26.76	1.84
7	CA / TG	-1.12	-0.47	1.59	0.93
8	AG / TT	-2.44	-1.06	5.62	14.12
9	GG / CT	-0.85	1.93	-1.10	-11.41
10	GC / GC	-3.00	-0.33	18.07	-3.03
11	CA / TG	-1.53	-0.19	13.48	-2.63
12	AC / GT	-4.31	0.71	24.52	2.59
13	CC / GG	-2.16	0.14	17.34	2.56
14	CA / TG	-1.62	0.01	12.57	-0.12
15	AG / CT	-3.08	0.77	18.76	0.03
16	GT / AC	-1.99	-1.01	6.10	-3.71
Average		-2.30	-0.01	14.04	0.09

Local base pair helical parameters in the 1000 MHz NMR structure of 17mer mismatch GT DNA based on the average structure in Figure 3.15 and calculated using 3DNA.

Step No.	Step	Shift (D_x)	Slide (D_y)	Rise (D_z)	Tilt ($T\tau$)	Roll ($P\rho$)	Helical Twist (Ω)
1	AG / CT	-0.78	-0.60	2.76	1.70	4.15	32.71
2	GC / GC	-0.01	-0.68	3.73	-0.13	14.56	37.43
3	CT / AG	0.94	-0.39	2.99	1.84	9.50	35.12
4	TG / CA	0.34	-0.37	2.68	-2.47	9.36	35.99
5	GC / GC	-0.24	-0.85	3.83	-1.39	14.24	36.46
6	CC / GG	0.12	-1.01	2.91	-1.08	15.67	31.66
7	CA / TG	0.25	-0.63	2.21	-0.57	0.97	35.58
8	AG / TT	-0.25	-1.11	3.83	-8.36	3.33	33.67
9	GG / CT	-0.61	-0.65	3.90	7.75	-0.75	39.23
10	GC / GC	0.39	-0.85	3.65	1.88	11.22	35.00
11	CA / TG	0.22	-0.51	2.29	1.71	8.76	37.23
12	AC / GT	-0.55	-1.06	3.68	-1.43	13.56	30.21
13	CC / GG	-0.23	-0.51	3.32	-1.67	11.30	36.90
14	CA / TG	0.00	-0.50	2.39	0.07	7.61	34.72
15	AG / CT	-0.46	-0.83	3.42	-0.02	11.08	33.20
16	GT / AC	0.87	-0.86	3.81	2.32	3.81	36.24
Average		0.00	-0.71	3.21	0.01	8.65	35.08

Local base pair step parameters in the 1000 MHz NMR structure of 17mer mismatch GT DNA based on the average structure in Figure 3.15 and calculated using 3DNA.

Base	Nucleotides	C1'-N (χ)	P-O5' (α)	O5'-C5' (β)	C5'-C4' (γ)	C4'-C3' (δ)	C3'-O3' (ε)	O3'-P (ζ)
1	A	-105.3			60.0	138.2	-179.9	-104.7
2	G	-99.6	-80.8	-173.5	58.3	136.2	176.0	-103.9
3	C	-115.2	-73.1	-173.4	60.5	136.0	178.0	-88.4
4	T	-111.8	-68.1	-179.6	55.2	130.4	-179.8	-93.2
5	G	-99.3	-86.2	-169.9	56.0	134.9	174.0	-103.1
6	C	-119.3	-69.3	-173.1	57.3	133.9	-179.8	-89.4
7	C	-115.8	-80.4	-173.8	51.9	130.1	-179.9	106.0
8	A	-101.6	-72.4	-178.5	60.7	135.8	178.3	-107.4
9	G	-113.7	-66.3	-170.3	56.8	138.6	180.0	-108.7
10	G	-107.5	-71.6	-173.5	57.4	140.9	173.9	-91.5
11	C	-117.3	-71.3	-173.6	54.6	132.3	179.0	-100.0
12	A	-102.6	-77.8	-174.1	58.6	136.0	179.5	-101.2
13	C	-118.5	-81.6	-173.8	57.4	132.6	-180.0	-109.3
14	C	-117.4	-71.5	-172.4	58.6	138.0	179.6	-93.3
15	A	-102.4	-75.8	-174.8	55.2	134.6	-179.9	-101.7
16	G	-108.1	-81.3	-170.5	55.4	134.9	172.9	-95.0
17	T	-120.5	-60.4	-178.8	58.0	130.3		
Average		-110.3	-74.2	-174.0	57.2	134.9	43.2	-86.6

Backbone dihedral angles for residues A1-T17 in the 1000 MHz NMR structure of 17mer mismatch GT DNA based on the average structure in Figure 3.15 and calculated using 3DNA.

Base	Nucleotides	C1'-N (χ)	P-O5' (α)	O5'-C5' (β)	C5'-C4' (γ)	C4'-C3' (δ)	C3'-O3' (ε)	O3'-P (ζ)
18	A	-114.1			59.5	137.1	-179.8	-104.1
19	C	-118.7	-79.5	-169.5	61.1	136.2	180.0	-91.9
20	T	-119.1	-69.2	-176.2	54.5	130.4	-179.9	-94.4
21	G	-98.5	-85.0	-168.1	55.2	134.8	176.3	-101.9
22	G	-105.4	-76.0	-171.5	55.8	136.7	175.2	-96.0
23	T	-116.3	-65.6	-173.5	54.8	132.3	180.0	-101.8
24	G	-105.7	-80.1	-172.9	55.7	136.1	-180.0	-101.9
25	C	-115.4	-75.0	-173.3	58.5	134.4	174.6	-92.0
26	T	-124.2	-62.4	-171.0	55.0	132.7	180.0	-97.8
27	T	-119.1	-71.4	-177.3	58.4	131.1	175.0	-102.0
28	G	-106.2	-71.6	-174.7	55.7	133.4	179.1	-97.7
29	G	-105.8	-79.2	-170.6	53.1	134.1	176.0	-99.3
30	T	-119.2	-71.1	-175.0	58.9	134.0	-179.9	-104.1
31	A	-100.5	-77.0	-172.1	61.3	137.1	-179.9	-102.6
32	G	-101.3	-82.5	-172.6	59.2	136.8	172.8	-101.7
33	C	-121.7	-71.1	-175.3	58.4	133.0	-179.9	-91.1
34	T	-114.3	-69.0	179.9	56.6	130.0		
Average		-112.1	-74.1	-150.9	57.2	134.1	43.1	-98.8

Backbone dihedral angles for residues A1-T17 in the 1000 MHz NMR structure of 17mer mismatch GT DNA based on the average structure in Figure 3.15 and calculated using 3DNA.

Base	Nucleotides	ν_0	ν_1	ν_2	ν_3	ν_4	Amplitude	Phase	Pucker
1	A	-23.4	33.2	-29.8	17.2	3.7	32.9	154.9	C_2-endo
2	G	-27.5	35.8	-30.0	14.9	7.7	35.1	148.7	C_2-endo
3	C	-25.6	34.4	-29.7	15.7	6.1	33.9	151.1	C_2-endo
4	T	-33.5	38.7	-29.0	10.4	14.2	38.1	139.6	C_1-exo
5	G	-29.8	37.2	-29.9	13.6	9.9	36.3	145.5	C_2-endo
6	C	-29.8	36.8	-29.5	13.1	10.2	36.1	145.0	C_2-endo
7	C	-35.5	40.1	-29.3	9.6	15.9	39.6	137.8	C_1-exo
8	A	-26.7	35.1	-29.7	15.1	7.1	34.5	149.5	C_2-endo
9	G	-25.5	35.5	-31.3	17.4	4.9	35	153.3	C_2-endo
10	G	-21.7	33.1	-31.1	19.5	1.2	33.3	159.2	C_2-endo
11	C	-32.8	38.8	-29.7	11.5	13.1	38.1	141.3	C_1-exo
12	A	-26.5	35.0	-29.6	15.2	6.9	34.3	149.8	C_2-endo
13	C	-32.8	38.9	-29.8	11.6	13.0	38.1	141.4	C_1-exo
14	C	-21.8	31.9	-29.4	17.7	2.4	32.0	157.0	C_2-endo
15	A	-28.6	36.2	-29.6	13.9	9.0	35.5	146.7	C_2-endo
16	G	-27.3	35.3	-29.5	14.6	7.7	34.6	148.5	C_2-endo
17	T	-32.9	39.0	-29.8	11.5	13.1	38.1	141.3	C_1-exo
Average		-28.3	36.2	-29.8	14.3	8.6	35.6	147.7	

Deoxyribose sugar dihedral angles for residues A1-T17 in the 1000 MHz NMR structure of 17mer mismatch GT DNA based on the average structure in Figure 3.15 and calculated using 3DNA.

Base	Nucleotides	ν_0	ν_1	ν_2	ν_3	ν_4	Amplitude	Phase	Pucker
18	A	-24.9	34.1	-29.7	16.1	5.3	33.6	152.2	C_2-endo
19	C	-26.5	34.9	-29.7	15.2	6.9	34.3	149.8	C_2-endo
20	T	-33.8	39.0	-29.1	10.3	14.5	38.4	139.4	C_1-exo
21	G	-28.8	36.4	-29.7	14.0	9.0	35.6	146.6	C_2-endo
22	G	-25.5	34.5	-29.9	15.9	5.8	34.0	151.5	C_2-endo
23	T	-32.5	38.5	-29.6	11.6	12.9	37.8	141.5	C_1-exo
24	G	-27.0	35.4	-29.8	15.0	7.3	34.7	149.2	C_2-endo
25	C	-29.4	36.8	-29.8	13.7	9.6	36.0	145.9	C_2-endo
26	T	-30.4	37.0	-29.3	12.4	11.1	36.3	143.7	C_1-exo
27	T	-32.9	38.4	-29.2	11.0	13.5	37.8	140.6	C_1-exo
28	G	-30.4	37.2	-29.6	12.9	10.7	36.5	144.3	C_2-endo
29	G	-30.2	37.3	-29.8	13.1	10.4	36.5	144.7	C_2-endo
30	C	-30.5	37.3	-29.7	12.9	10.8	36.5	144.3	C_2-endo
31	A	-24.7	33.9	-29.7	16.2	5.2	33.5	152.5	C_2-endo
32	G	-27.1	35.8	-30.3	15.5	7.1	35.1	149.7	C_2-endo
33	C	-28.7	35.9	-29.2	13.3	9.5	35.2	145.9	C_2-endo
34	T	-34.4	40.0	-29.8	10.8	14.5	39.2	139.6	C_1-exo
Average		-29.3	36.6	-29.6	13.5	9.7	35.9	146.0	

Deoxyribose sugar dihedral angles for residues A1-T17 in the 1000 MHz NMR structure of 17mer mismatch GT DNA based on the average structure in Figure 3.15 and calculated using 3DNA.

v. Titration data – 17mer mismatch GT DNA and length analogues

Imino proton	Chemical shift before MutS addition (δ_1)	Chemical shift after MutS addition (δ_2)	δ_2-δ_1
G1	12.94	12.97	0.03
G5	10.78	10.78	0.00
G6	12.86	12.84	-0.02
G10	12.86	12.84	-0.02
T11	13.84	13.82	-0.02
G12	12.76	12.76	0.00
T14	11.84	11.91	0.07
T15	14.34	14.39	0.05
G16	13.02	13.00	-0.02
G17	13.13	13.13	0.00

The chemical shifts (δ, ppm) of imino proton resonances in 9mer mismatch GT DNA before and after addition of *E. coli* MutS protein; data is based on Figure 3.24.

Imino proton	Chemical shift before MutS addition (δ_1)	Chemical shift after MutS addition (δ_2)	δ_2-δ_1
T1	13.46	-	-
G2	12.88	12.89	0.01
G6	10.77	10.76	-0.01
G7	12.88	12.89	0.01
G12	12.93	12.95	0.02
G13	12.83	12.80	-0.03
T14	13.69	13.68	-0.01
G15	12.72	12.70	-0.02
T17	11.82	11.88	0.06
T18	14.32	14.36	0.04
G19	12.98	12.95	-0.03
G20	12.98	12.95	-0.03

The chemical shifts (δ, ppm) of imino proton resonances in 11mer mismatch GT DNA before and after addition of *E. coli* MutS protein; data is based on Figure 3.25.

Imino proton	Chemical shift before MutS addition (δ_1)	Chemical shift after MutS addition (δ_2)	δ_2-δ_1
T2	14.03	14.03	0.00
G3	12.81	12.81	0.00
G7	10.78	10.78	0.00
G8	12.81	12.85	0.04
T14	14.03	14.03	0.00
G15	12.92	12.90	-0.02
G16	12.81	12.81	0.00
T17	13.65	13.67	0.02
G18	12.72	12.70	-0.02
T20	11.76	-	N/A
T21	14.28	14.31	0.03
G22	12.92	12.90	-0.02
G23	12.92	12.90	-0.02
G26	13.21	-	N/A

The chemical shifts (δ, ppm) of imino proton resonances in 13mer mismatch GT DNA before and after addition of *E. coli* MutS protein; data is based on Figure 3.26.

Imino proton	Chemical shift before MutS addition (δ_1)	Chemical shift after MutS addition (δ_2)	δ_2-δ_1
G1	12.86	12.87	0.01
T3	14.00	14.00	0.00
G4	12.78	12.78	0.00
G8	10.75	10.75	0.00
G9	12.78	12.78	0.00
G15	13.10	-	N/A
T17	14.08	14.00	-0.08
G18	12.91	12.87	-0.04
G19	12.78	12.78	0.00
T20	13.64	13.63	-0.01
G21	12.67	12.65	-0.02
T23	11.80	-	N/A
T24	14.29	14.30	0.01
G25	12.98	12.94	-0.04
G26	12.94	12.94	0.00
G29	13.16	-	N/A

The chemical shifts (δ, ppm) of imino proton resonances in 15mer mismatch GT DNA before and after addition of *E. coli* MutS protein; data is based on Figure 3.27.

Imino proton	Chemical shift before MutS addition (δ_1)	Chemical shift after MutS addition (δ_2)	δ_2-δ_1
G2	12.81	12.83	0.02
T4	13.95	13.94	-0.01
G5	12.70	12.71	0.01
G9	10.72	10.70	-0.02
G10	12.81	12.79	-0.02
G16	12.81	12.83	0.02
T17	13.33	-	-
T20	13.95	13.94	-0.01
G21	12.81	12.83	0.02
G22	12.73	12.74	0.01
T23	13.61	13.60	-0.01
G24	12.64	12.61	-0.03
T26	11.79	11.83	0.04
T27	14.29	14.30	0.01
G28	12.89	12.88	-0.01
G29	12.81	12.83	0.02
G32	12.89	12.88	-0.01
T34	13.95	13.94	-0.01

The chemical shifts (δ, ppm) of imino proton resonances in 17mer mismatch GT DNA before and after addition of *E. coli* MutS protein; data is based on Figure 3.28.

vi. Titration data – Mismatch orientation and sequence context analogues of 17mer mismatch GT DNA

Imino proton	Chemical shift before MutS addition (δ_1)	Chemical shift after MutS addition (δ_2)	δ_2-δ_1
G2	12.81	12.81	0.00
T4	13.96	13.96	0.00
G5	12.74	12.74	0.00
G9	10.61	-	N/A
T10	14.15	14.22	0.07
G16	13.04	13.02	-0.02
T20	13.96	13.96	0.00
G21	12.81	12.82	0.01
G22	12.74	12.74	0.00
T23	13.58	13.58	0.00
G24	12.48	12.48	0.00
T26	11.82	-	-
G27	12.81	12.82	0.01
G28	12.89	12.90	0.01
G29	12.81	12.81	0.00
G32	12.95	12.90	-0.05
T34	13.96	13.97	0.01

The chemical shifts (δ, ppm) of imino proton resonances in 17mer mismatch GT-FB DNA before and after addition of *E. coli* MutS protein; data is based on Figure 3.29. T17 is absent from the table as it was not observed in the 1D ^{1}H-NMR spectrum.

Imino proton	Chemical shift before MutS addition (δ_1)	Chemical shift after MutS addition (δ_2)	δ_2-δ_1
G2	12.79	12.82	0.03
T4	13.93	13.93	0.00
G5	12.71	12.71	0.00
T9	11.30	11.38	0.05
G10	12.79	12.83	0.04
G16	12.79	12.82	0.03
T20	13.93	13.93	0.00
G21	12.79	12.79	0.00
G22	12.71	12.72	0.01
T23	13.58	13.59	0.01
G24	12.59	12.59	0.00
G26	10.23	10.32	0.09
T27	13.84	13.84	0.00
G28	12.87	12.87	0.00
G29	12.79	12.80	0.01
G32	12.87	12.90	0.03
T34	13.93	13.96	0.03

The chemical shifts (δ, ppm) of imino proton resonances in 17mer mismatch TG DNA before and after addition of *E. coli* MutS protein; data is based on Figure 3.30. T17 is absent from the table as it was not observed in the 1D ^{1}H-NMR spectrum.

Imino proton	Chemical shift before MutS addition (δ_1)	Chemical shift after MutS addition (δ_2)	δ_2-δ_1
G2	12.82	12.82	0.00
T4	13.96	13.96	0.00
G5	12.72	12.72	0.00
T9	11.36	11.43	0.07
T10	13.78	13.78	0.00
G16	13.05	13.08	0.03
T20	13.96	13.96	0.00
G21	12.82	12.82	0.00
G22	12.72	12.72	0.00
T23	13.59	13.57	-0.02
G24	12.46	12.43	-0.02
G26	10.39	10.50	0.11
G27	12.82	12.84	0.02
G28	12.89	12.89	0.00
G29	12.82	12.82	0.00
G32	13.01	13.01	0.00
T34	13.96	13.96	0.00

The chemical shifts (δ, ppm) of imino proton resonances in 17mer mismatch TG-FB DNA before and after addition of *E. coli* MutS protein; data is based on Figure 3.31. T17 is absent from the table as it was not observed in the 1D ^{1}H-NMR spectrum.

vii. Conformational analysis - 6-thioguanine modified 13mer GC DNA

Base	Nucleotides	C1'-N (χ)	P-O5' (α)	O5'-C5' (β)	C5'-C4' (γ)	C4'-C3' (δ)	C3'-O3' (ε)	O3'-P (ζ)
1	C	-153.2			59.7	137.5	-159.7	-59.9
2	T	-121.9	-89.1	-170.5	56.8	135.4	177.9	-107.5
3	G	-108.8	-78.6	-170.1	57.7	139.7	167.7	-92.1
4	C	-133.8	-69.7	-170.1	56.7	132.2	-179.0	-112.6
5	C	-123.7	-75.5	-170.1	58.7	136.1	173.7	-106.1
6	A	-103.3	-68.5	-170.2	63.8	141.0	-176.2	-95.9
7	G	-92.8	-76.4	-170.1	54.6	132.0	162.1	-81.5
8	G	-128.8	-74.9	-170.7	56.6	132.0	174.9	-100.6
9	C	-136.5	-73.7	-167.7	57.7	133.9	175.5	-92.5
10	A	-109.1	-77.7	-170.0	55.5	134.5	174.5	-103.6
11	C	-124.4	-77.4	-170.1	57.7	134.4	-178.0	-91.4
12	C	-120.7	-79.9	-168.9	53.5	131.2	176.9	-109.7
13	A	-106.0	-75.3	-169.9	56.8	141.0		
Average		-120.2	-76.4	-169.9	57.4	135.5	57.5	-96.1

Backbone dihedral angles for residues C1-A13 in the NMR structure of 6-thioguanine modified 13mer GC DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 4.14.

Base	Nucleotides	C1'-N (χ)	P-O5' (α)	O5'-C5' (β)	C5'-C4' (γ)	C4'-C3' (δ)	C3'-O3' (ε)	O3'-P (ζ)
14	T	-121.7			59.8	136.1	178.4	-92.7
15	G	-104.1	-78.1	-170.0	57.2	134.5	176.1	93.3
16	G	-124.3	-85.4	-168.0	52.6	132.4	177.2	-98.4
17	T	-118.2	-76.1	-170.1	58.9	134.1	-177.1	-122.0
18	G	-111.7	-72.4	178.8	58.8	140.1	175.1	-93.9
19	C	-124.3	-70.6	-167.3	56.1	133.6	-179.2	-93.8
20	C	-124.2	-80.3	-166.8	56.2	131.7	169.2	-91.3
21	T	-117.3	-66.4	-170.1	58.1	134.7	167.9	-97.1
22	G	-73.0	-76.6	-152.2	52.3	148.6	145.5	-86.9
23	G	-121.0	-64.4	-170.6	62.6	130.6	-159.5	-110.1
24	C	-126.9	-74.4	-170.8	44.9	134.8	171.7	-98.3
25	A	-106.3	-71.0	-170.1	57.4	138.3	177.7	-100.4
26	G	-115.1	-83.5	-162.3	52.4	143.0		
Average		-114.5	-74.9	-138.3	55.9	136.3	85.3	-82.6

Backbone dihedral angles for residues T14-G26 in the NMR structure of 6-thioguanine modified 13mer GC DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 4.14.

Base	Nucleotides	ν_0	ν_1	ν_2	ν_3	ν_4	Amplitude	Phase	Pucker
1	C	-23.9	33.5	-29.8	16.7	4.2	33.1	153.9	C_2-endo
2	T	-29.1	36.6	-29.7	13.8	9.3	35.8	146.2	C_2-endo
3	G	-24.8	35.6	-32.2	18.8	3.6	35.4	155.4	C_2-endo
4	C	-31.4	37.5	-29.3	11.9	11.9	36.9	142.6	C_1-exo
5	C	-26.6	35.0	-29.7	15.2	6.9	34.4	149.7	C_2-endo
6	A	-18.0	29.8	-29.6	20.2	-1.6	30.8	164.1	C_2-endo
7	G	-32.6	39.2	-30.4	12.4	12.5	38.3	142.4	C_1-exo
8	G	-34.3	39.7	-29.5	10.6	14.5	38.9	139.4	C_1-exo
9	C	-30.6	37.5	-29.8	12.9	10.8	36.7	144.2	C_2-endo
10	A	-28.8	36.3	-29.6	13.8	9.2	35.6	146.4	C_2-endo
11	C	-29.6	36.8	-29.6	13.3	9.9	36.0	145.4	C_2-endo
12	C	-34.7	39.8	-29.6	10.3	14.9	39.2	139.1	C_1-exo
13	A	-22.3	34.9	-33.3	21.4	0.3	35.3	160.6	C_2-endo
Average		-28.2	36.3	-30.2	14.7	8.2	35.9	148.4	

Deoxyribose sugar dihedral angles for residues C1-A13 in the NMR structure of 6-thioguanine modified 13mer GC DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 4.14.

Base	Nucleotides	ν_0	ν_1	ν_2	ν_3	ν_4	Amplitude	Phase	Pucker
14	T	-27.3	35.5	-29.8	14.8	7.6	34.8	148.8	C_2-endo
15	G	-29.4	36.7	-29.7	13.5	9.7	35.9	145.7	C_2-endo
16	G	-31.8	38.0	-29.5	11.9	12.2	37.3	142.4	C_1-exo
17	T	-30.8	37.6	-29.7	12.7	11.1	36.8	143.8	C_1-exo
18	G	-19.9	31.2	-29.9	19.2	0.3	31.6	160.7	C_2-endo
19	C	-31.1	37.7	-29.6	12.4	11.5	36.9	143.3	C_1-exo
20	C	-33.1	39.0	-29.8	11.4	13.3	38.3	141.1	C_1-exo
21	T	-28.9	36.7	-30.3	14.4	8.8	36.0	147.2	C_2-endo
22	G	-20.6	38.1	-38.9	28.3	-5.5	39.8	168.2	C_2-endo
23	G	-33.6	38.9	-29.1	10.4	14.3	38.3	139.6	C_1-exo
24	C	-29.0	36.5	-29.7	13.7	9.3	35.7	146.3	C_2-endo
25	A	-22.8	32.8	-29.8	17.5	3.2	32.6	155.8	C_2-endo
26	G	-22.5	36.3	-35.2	23.2	-0.7	37.0	162.1	C_2-endo
Average		-27.8	36.5	-30.8	15.6	7.3	36.2	149.6	

Deoxyribose sugar dihedral angles for residues T14-G26 in the NMR structure of 6-thioguanine modified 13mer GC DNA. Values were calculated using the 3DNA program and are based on the average structure in Figure 4.14.

viii. Titration data – 6-thioguanine and 6-sulphonateguanine modified 13mer GC DNA

Imino proton	Chemical shift before MutS addition (δ_1)	Chemical shift after MutS addition (δ_2)	δ_2-δ_1
T2	14.06	14.05	-0.01
G3	12.85	12.89	0.04
G7	11.54	11.54	0.00
G8	12.79	12.80	0.01
T14	13.99	13.98	-0.01
G15	12.76	12.77	0.01
G16	12.76	12.77	0.01
T17	13.70	13.69	-0.01
G18	12.65	12.65	0.00
T21	13.89	13.89	0.00
G22	12.89	12.89	0.00
G23	12.77	12.77	0.00
G26	13.13	13.14	0.01

The chemical shifts (δ, ppm) of imino proton resonances in 13mer 6-thioguanine modified GC DNA before and after addition of *E. coli* MutS protein; data is based on Figure 4.19.

Imino proton	Chemical shift before MutS addition (δ_1)	Chemical shift after MutS addition (δ_2)	δ_2-δ_1
T2	14.02	14.01	-0.01
G3	12.89	12.89	0.00
G8	12.78	12.78	0.00
T14	14.02	14.01	-0.01
G15	12.78	12.78	0.00
G16	12.78	12.78	0.00
T17	13.70	13.68	-0.02
G18	12.71	12.69	-0.02
T21	13.83	13.82	-0.01
G22	12.89	12.89	0.00
G23	12.78	12.78	0.00
G26	13.19	13.16	-0.03

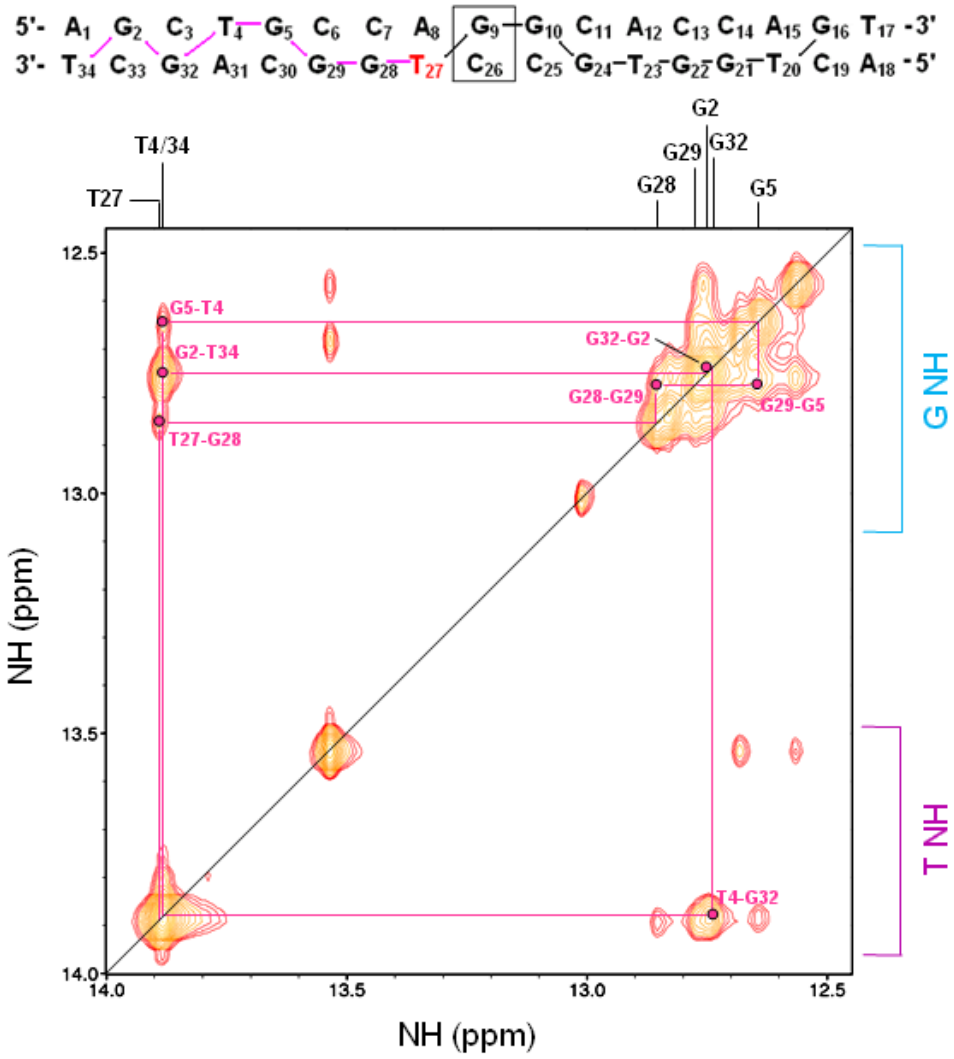
The chemical shifts (δ, ppm) of imino proton resonances in 13mer 6-sulphonateguanine modified GC DNA before and after addition of *E. coli* MutS protein; data is based on Figure 4.20.

Imino proton	Chemical shift before MutS addition (δ_1)	Chemical shift after MutS addition (δ_2)	δ_2-δ_1
T2	14.05	14.05	0.00
G3	13.00	12.99	-0.01
G8	12.91	12.91	0.00
T14	14.05	14.05	0.00
G15	12.80	12.80	0.00
G16	12.84	12.84	0.00
T17	13.73	13.73	0.00
G18	12.71	12.70	-0.01
T20	10.87	-	N/A
T21	13.99	14.05	0.06
G22	12.96	12.95	-0.01
G23	12.84	12.84	0.00
G26	13.21	13.20	-0.01

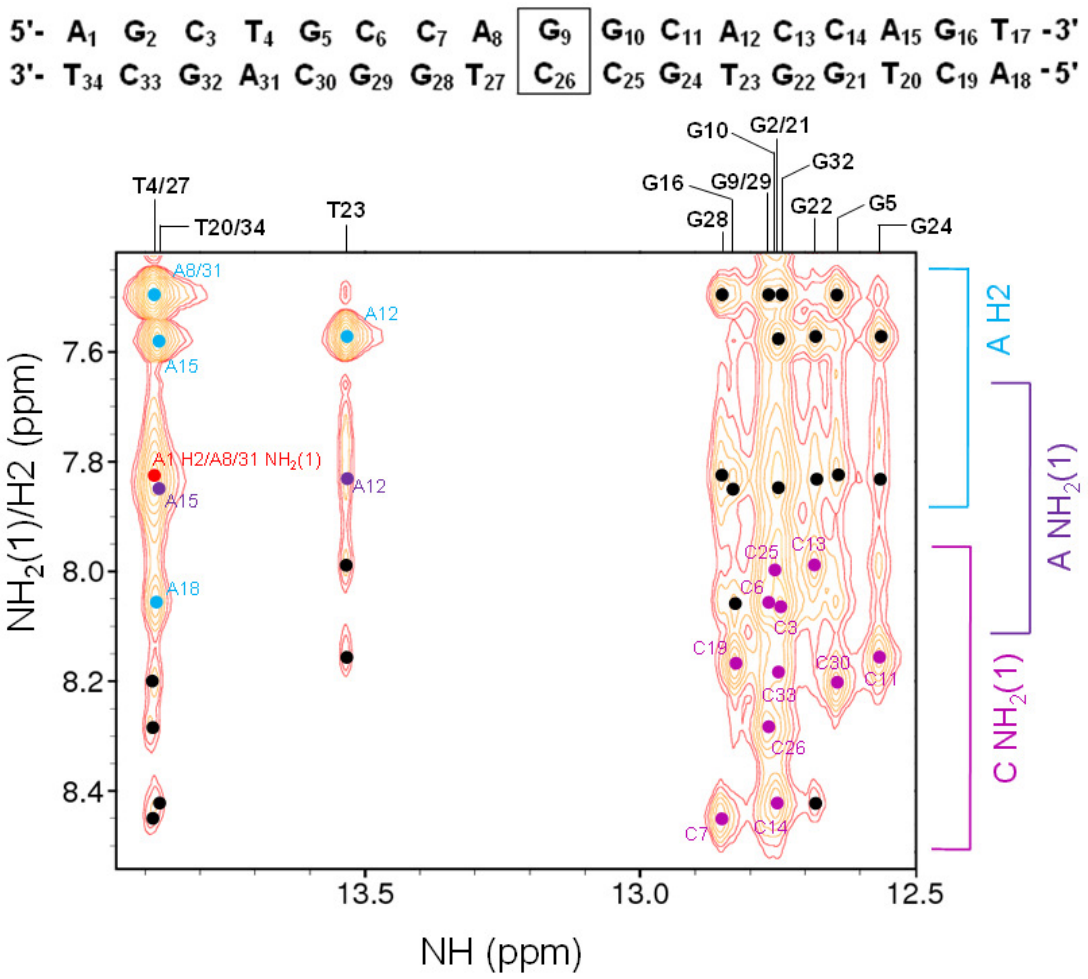
The chemical shifts (δ, ppm) of imino proton resonances in 13mer 6-sulphonateguanine modified GT DNA before and after addition of *E. coli* MutS protein; data is based on Figure 4.21.

Appendix IV – 2D NMR spectra for 17mer canonical GC, 17mer mismatch GT and 6-TG modified 13mer GC DNAs

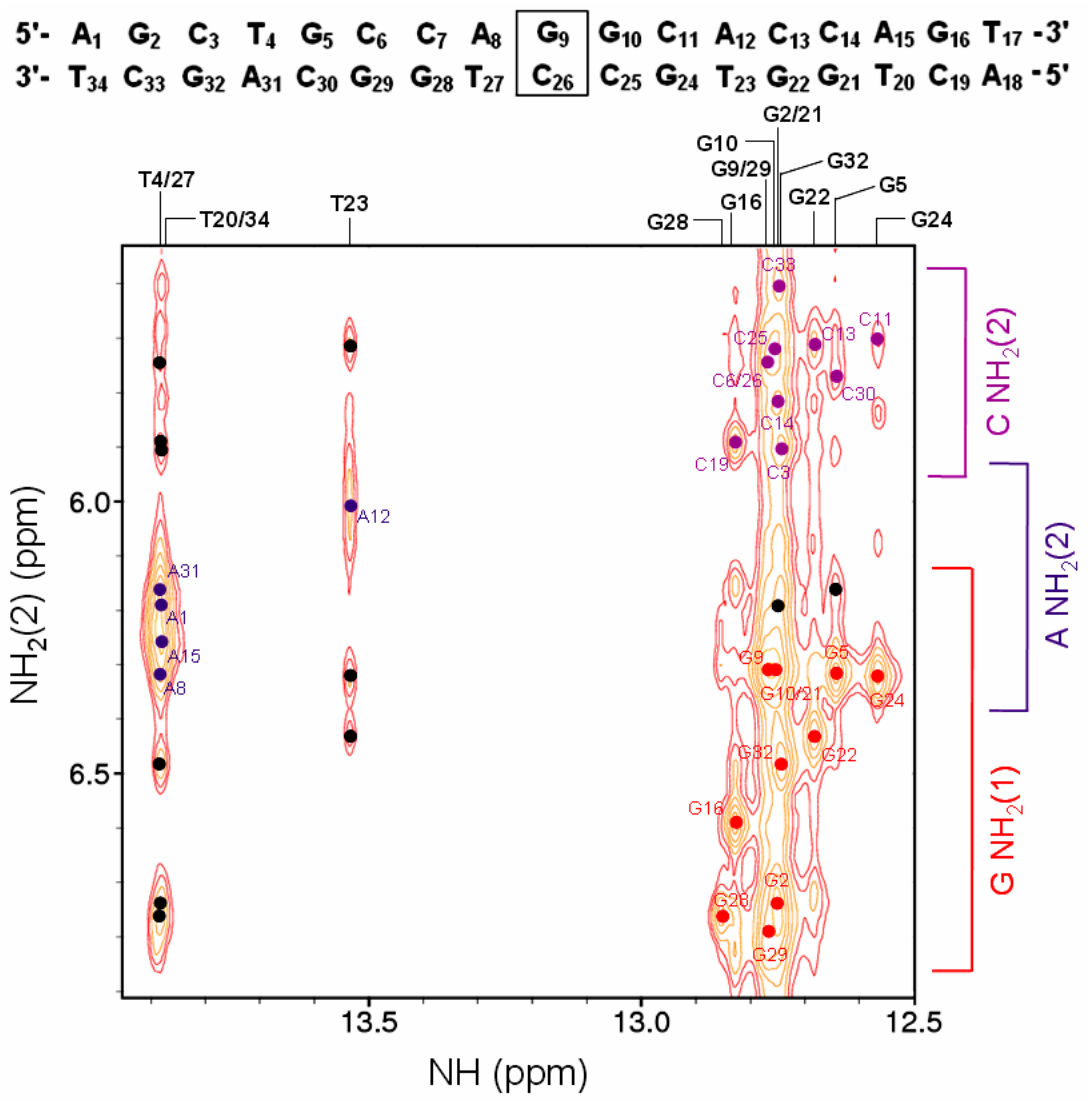
i. 17mer canonical GC DNA



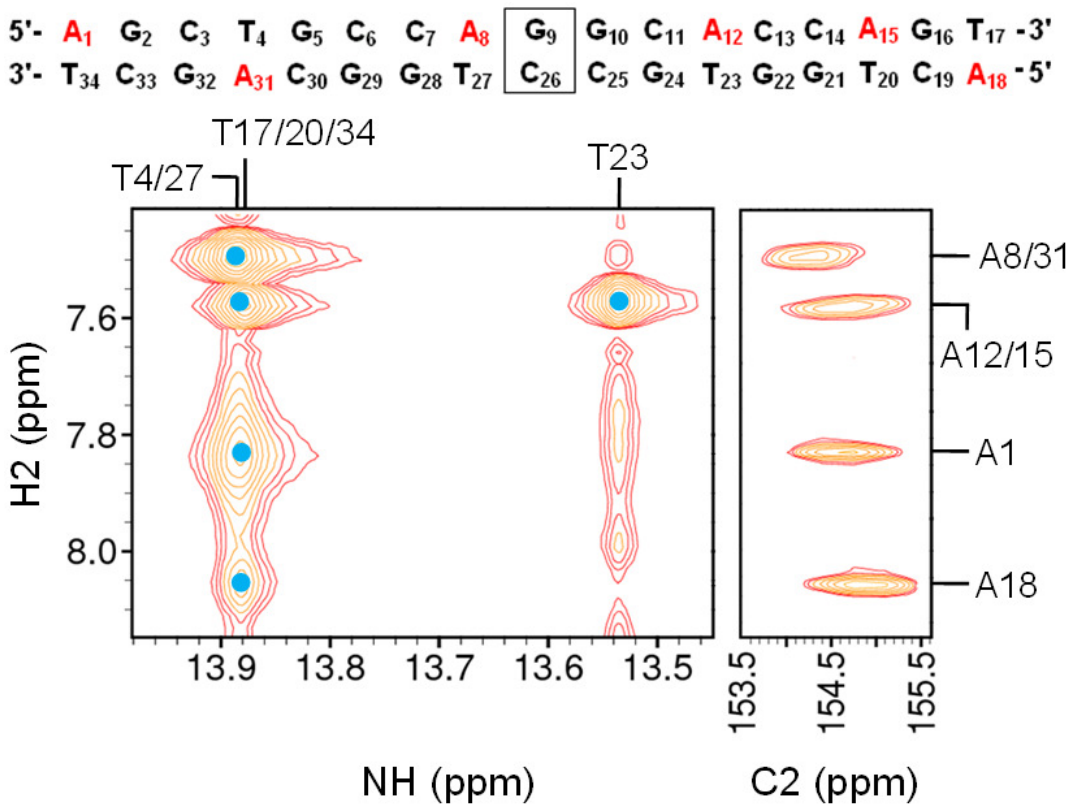
700 MHz NOESY (τ_m = 250ms) spectrum of 17mer canonical GC DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C. The respective imino protons were identified and assigned based on the sequential connectivity starting from T27 to T34 (pink line) as illustrated in the sequence above.



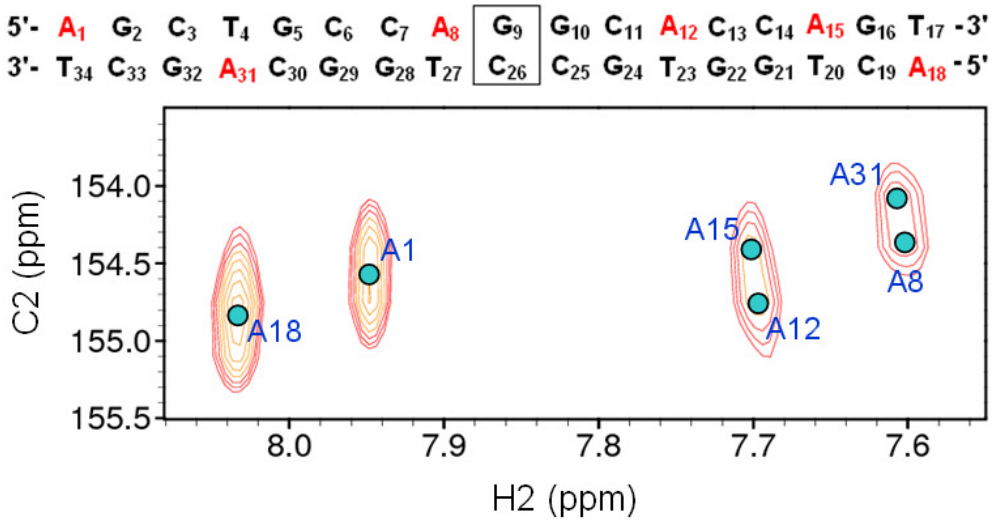
700 MHz NOESY (τ_m = 250ms) spectrum of 17mer canonical GC DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C. Base pair NOEs to adenine H2 and hydrogen bonded cytosine/adenine NH_2 protons are indicated by blue, light purple and dark purple circles respectively. Important intra/interstrand NOE connectivities which confirmed assignment are also highlighted by black circles in this region.



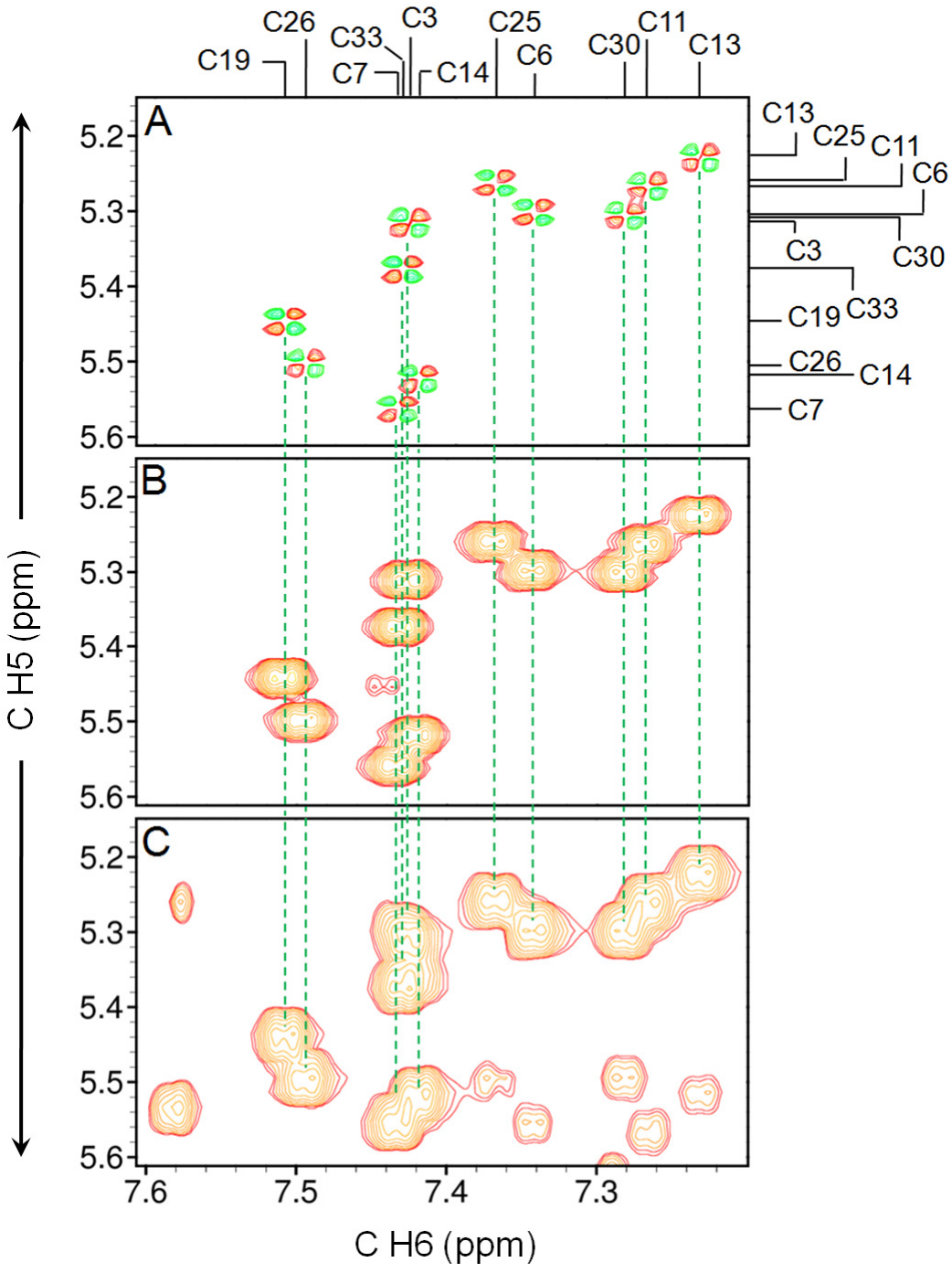
700 MHz NOESY (τ_m = 250ms) spectrum of 17mer canonical GC DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C. Base pair imino-amino NOE connectivities to hydrogen bonded guanine NH_2 and non-hydrogen bonded cytosine/adenine NH_2 protons are indicated by red, light purple and dark purple circles respectively. Important intra/interstrand NOE connectivities which confirmed assignment are also highlighted by black circles.



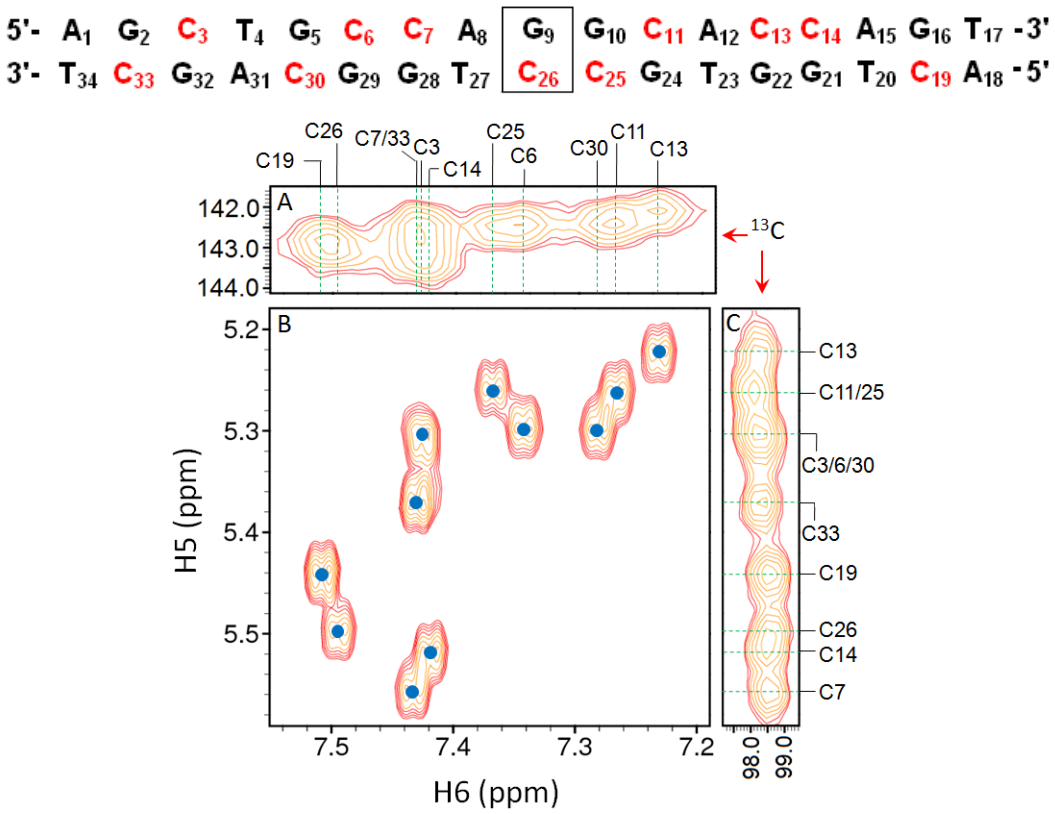
(Left panel) 700 MHz NOESY (τ_m = 250ms) spectrum of 17mer canonical GC DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C, (Right panel) 600 MHz ^1H-^{13}C HSQC spectrum of 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) measured at 25°C. The assignment of adenine H2 resonances is based on the NOE connectivities to thymine H3 imino resonances. Correlation to the ^1H-^{13}C HSQC spectrum identified the chemical shifts of C2 resonances.



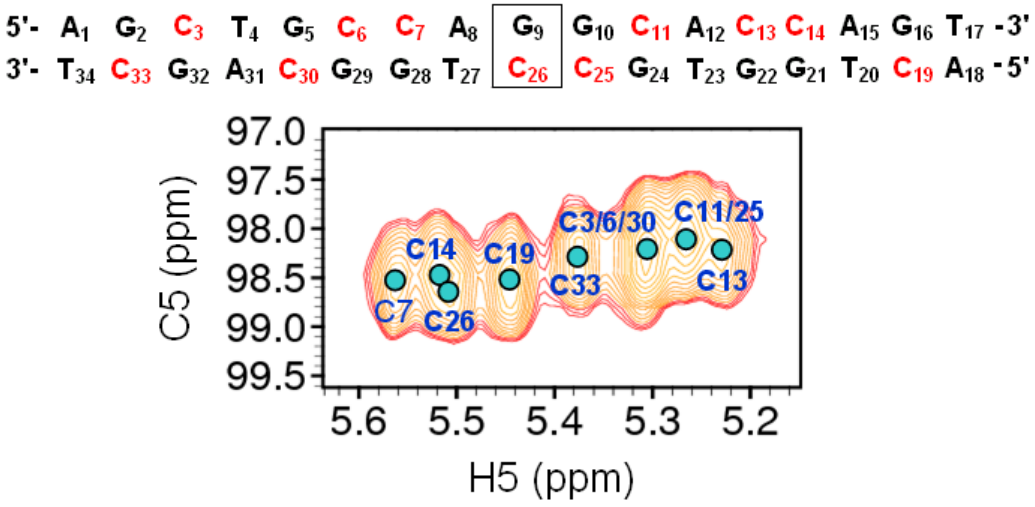
600 MHz ^1H-^{13}C HSQC spectrum of 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C showing the assignment of C2 resonances.



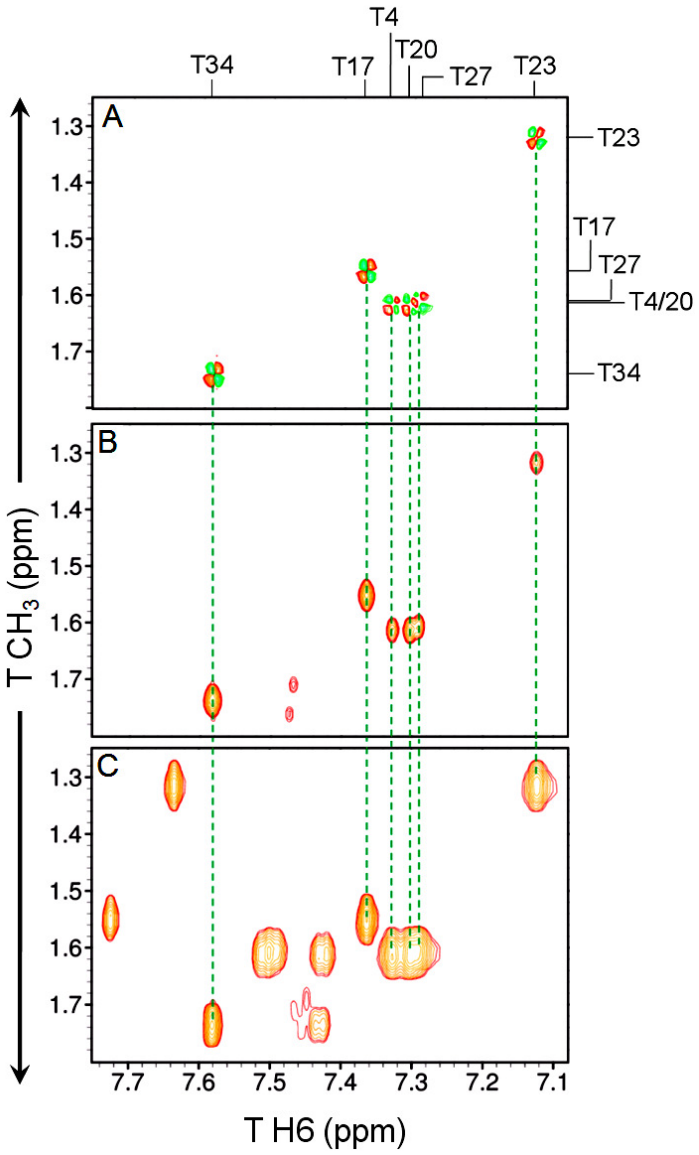
Spectrum stack showing the assignment of cytosine H5-H6 connectivities in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, pH 6.2, 50mM PO_4^{3-} and 50mM NaCl) at 25°C where A. DQF-COSY (800 MHz), B. TOCSY (600 MHz, SL_{mix} = 75ms) and C. NOESY (600 MHz, τ_m = 250ms).



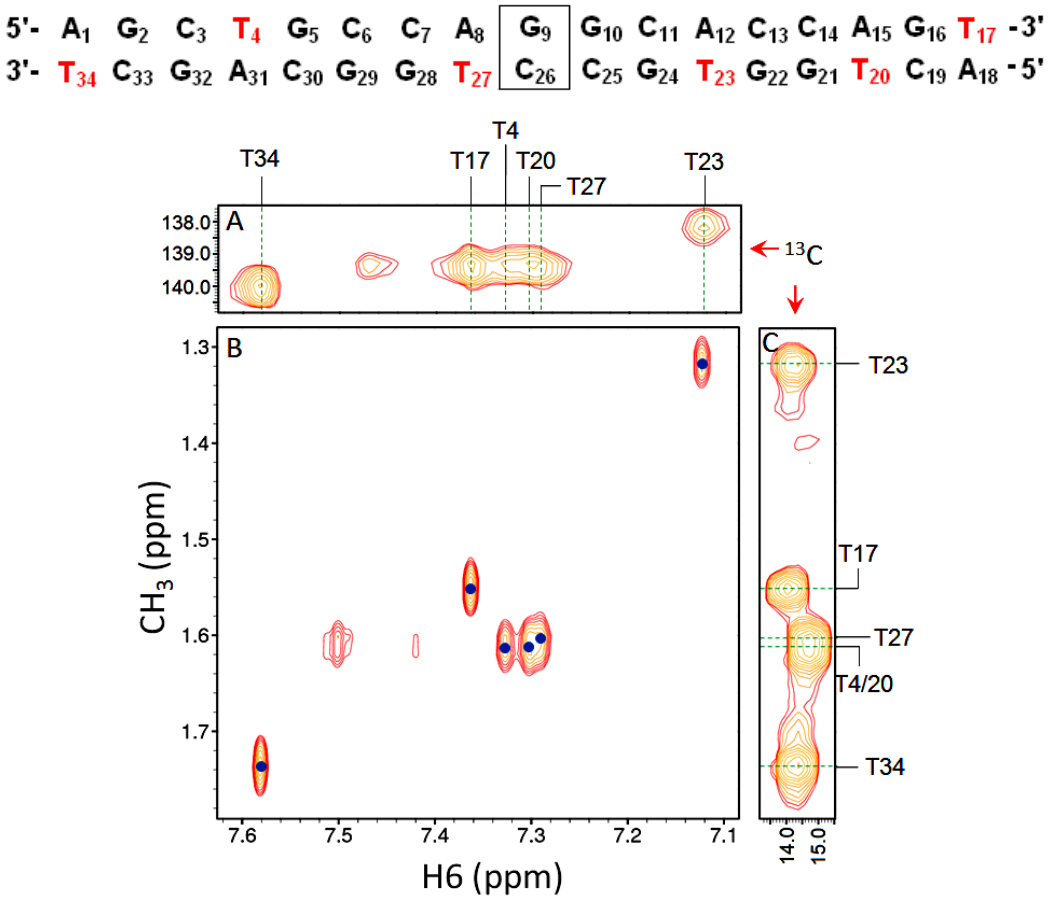
600 MHz NOESY (τ_m = 250ms) spectrum (B) highlighting cytosine H5-H6 correlations in 17mer canonical GC DNA (1mM, 100% ^2H_2O, pH 6.2, 50mM PO_4^{3-} and 50mM NaCl) at 25°C. 800 MHz ^1H-^{13}C HSQC panels show the assignment of the cytosine C6 (A) and C5 (C) chemical shifts.



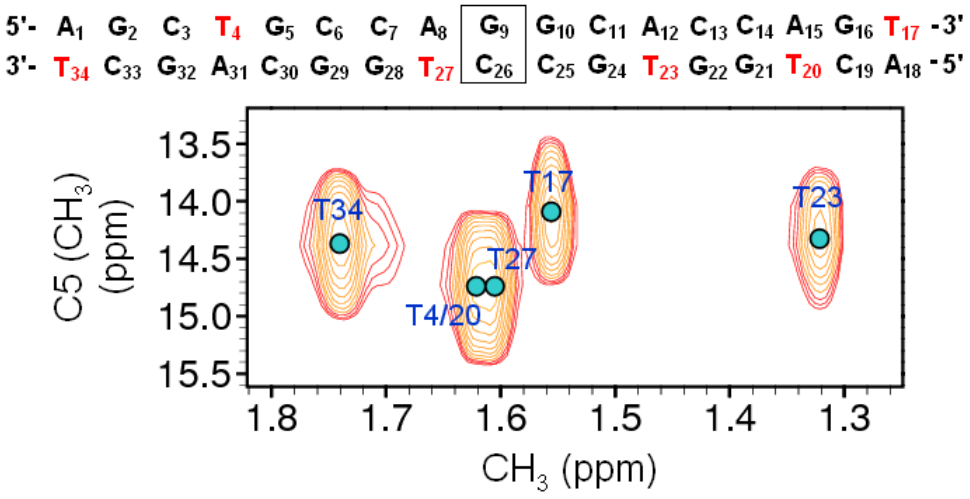
600 MHz ^1H-^{13}C HSQC spectrum showing the assignment of the cytosine C5 chemical shifts for 17mer canonical GC DNA (1mM, 100% ^2H_2O, pH 6.2, 50mM PO_4^{3-} and 50mM NaCl) at 25°C.



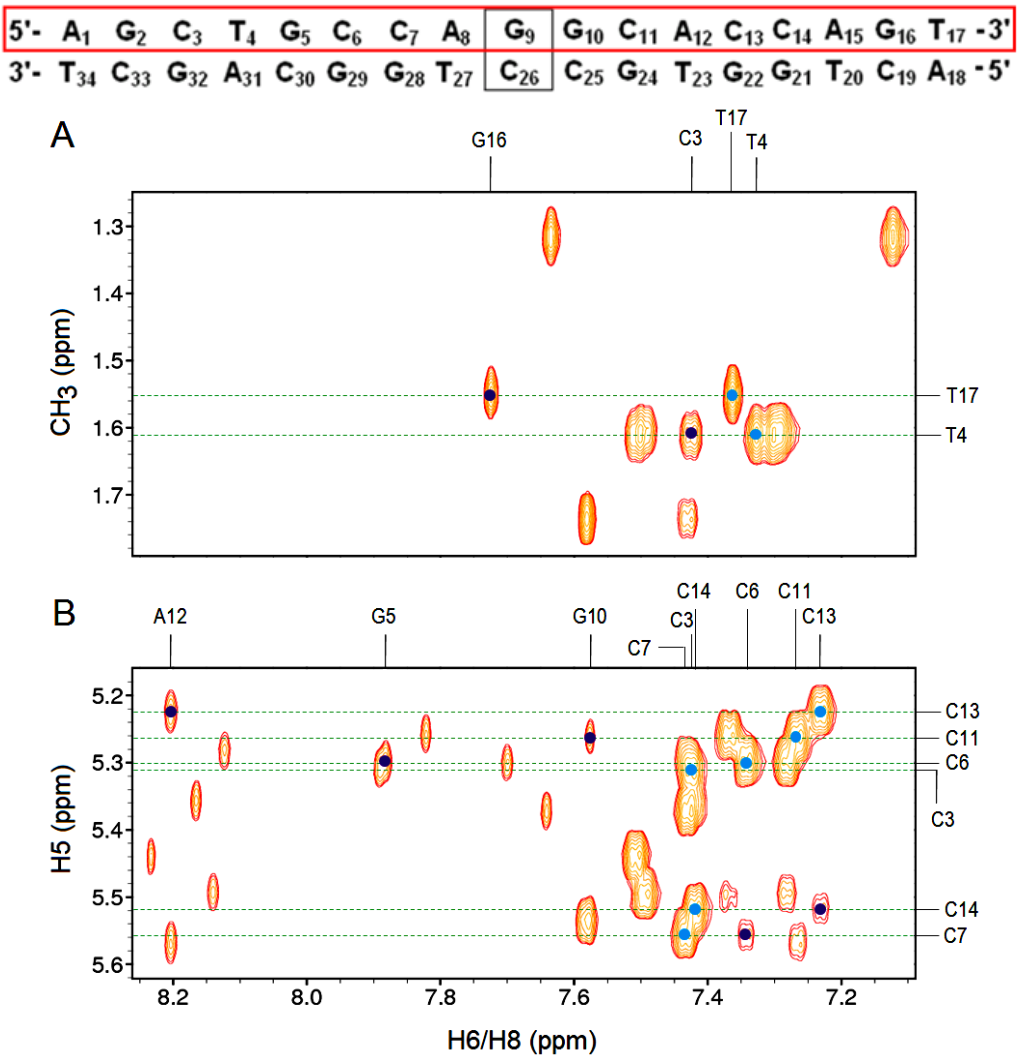
Stack plot of spectra showing the assignment of thymine CH_3-H6 NOE connectivities in 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where (A) DQF-COSY (800 MHz), (B) TOCSY (600 MHz, SL_{mix} = 75ms) and (C) NOESY (600 MHz, τ_m = 250ms).



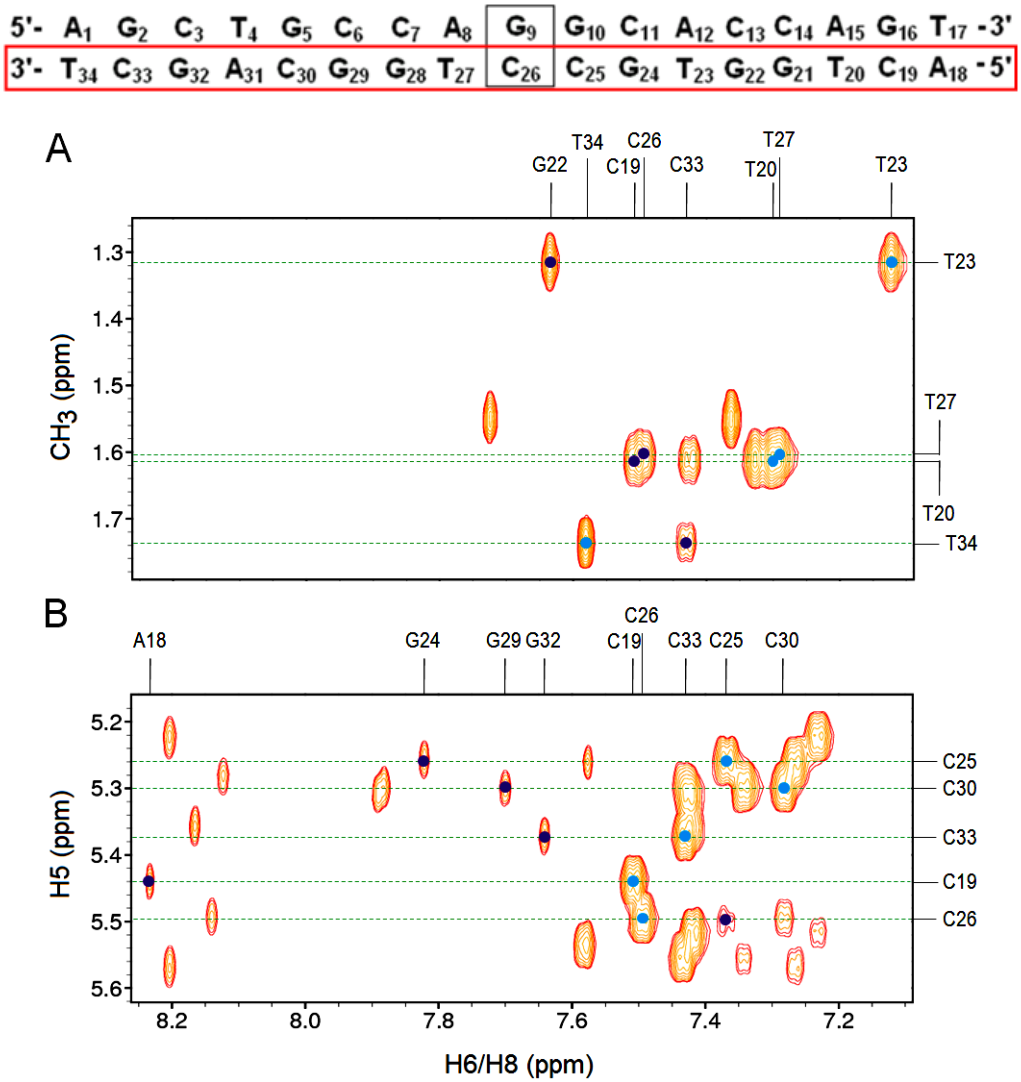
600 MHz NOESY (τ_m = 250ms) spectrum (B) highlighting thymine CH_3-H6 correlations in 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. 600 MHz ^1H-^{13}C HSQC panels show the assignment of the thymine C6 (A) and C5 (C) chemical shifts.



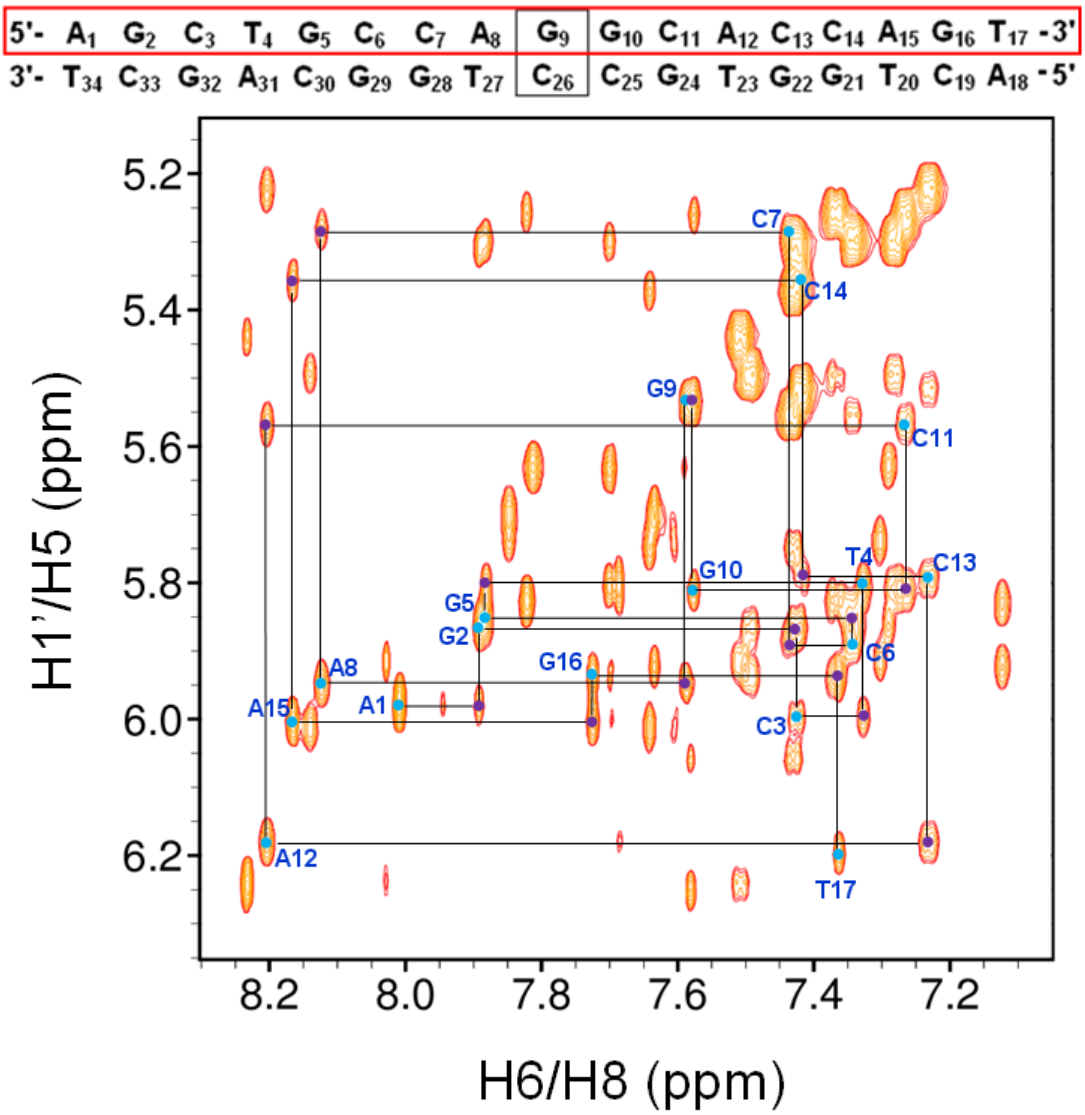
600 MHz ^1H-^{13}C HSQC spectrum showing the assignment of the thymine CH_3 (C5) chemical shifts in 17mer canonical GC DNA. (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



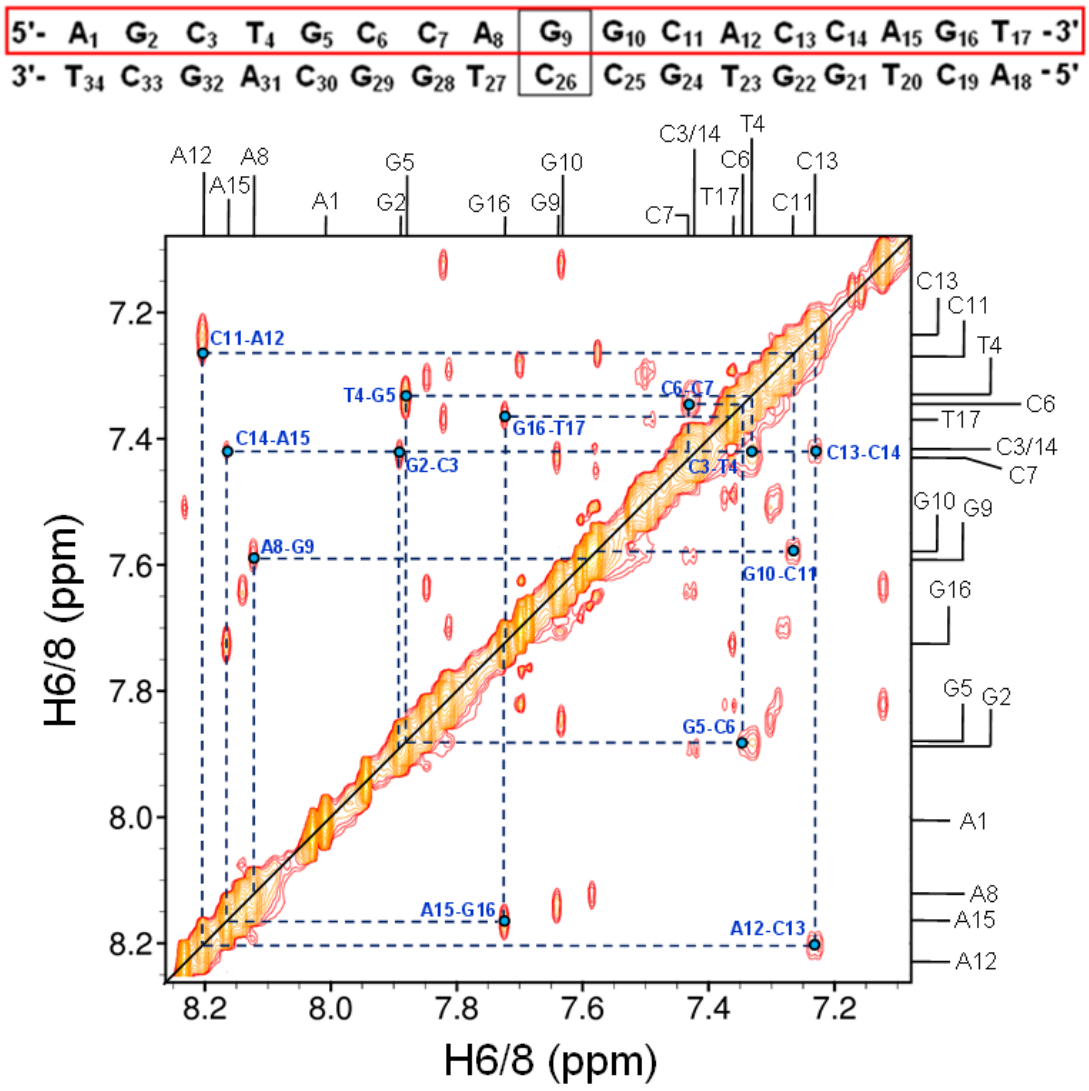
600 MHz NOESY (τ_m = 250ms) spectra showing internucleotide H5/CH_3 (i)-H6/8 (i-1) NOE connectivities for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where (A) Thymine and (B) Cytosine connectivities. Scalar coupled connectivities are indicated by blue circles and internucleotide connectivities by purple circles.



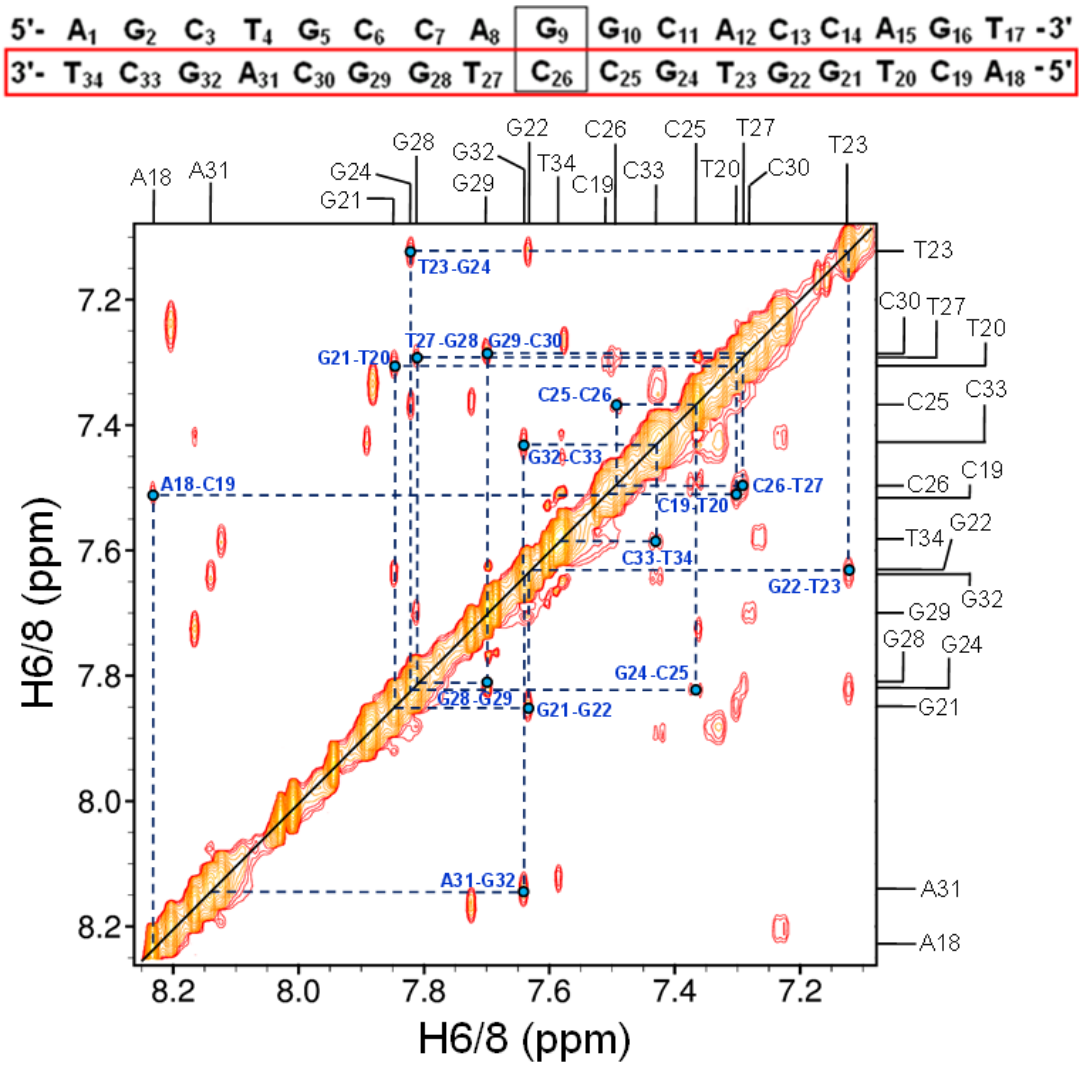
600 MHz NOESY (τ_m = 250ms) spectra showing internucleotide H6/8 (i-1)-H5/CH_3 (i) NOE connectivities for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where (A) Thymine and (B) Cytosine connectivities. Scalar coupled connectivities are indicated by blue circles and internucleotide connectivities by purple circles.



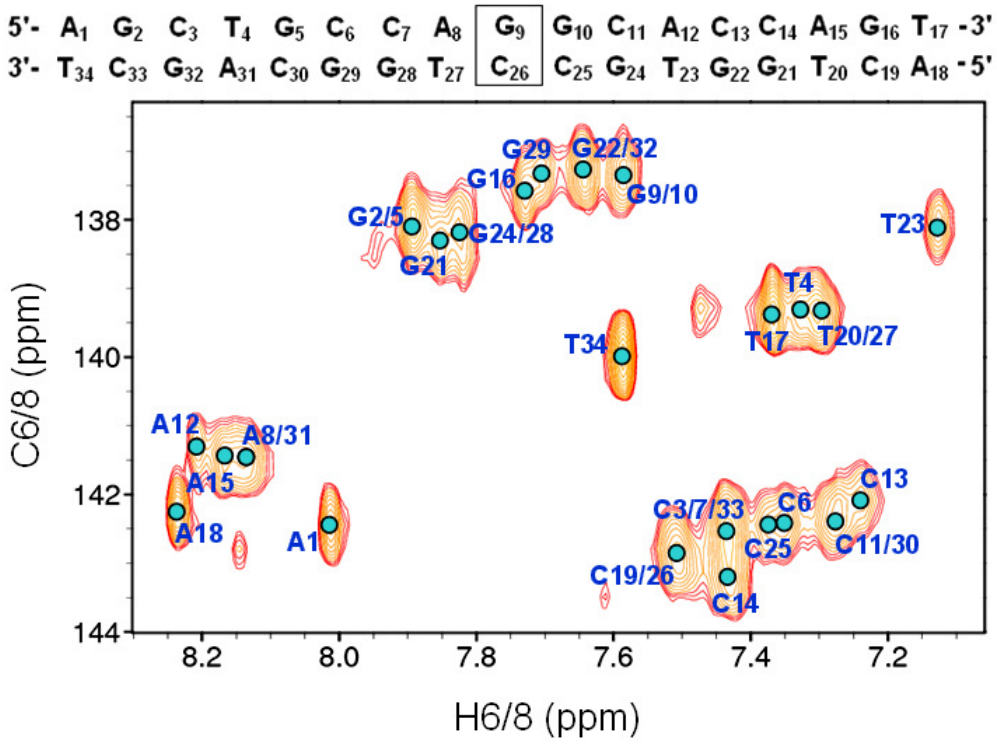
600 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H1' sequential NOE connectivities for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide and internucleotide connectivities are shown by blue and purple circles respectively.



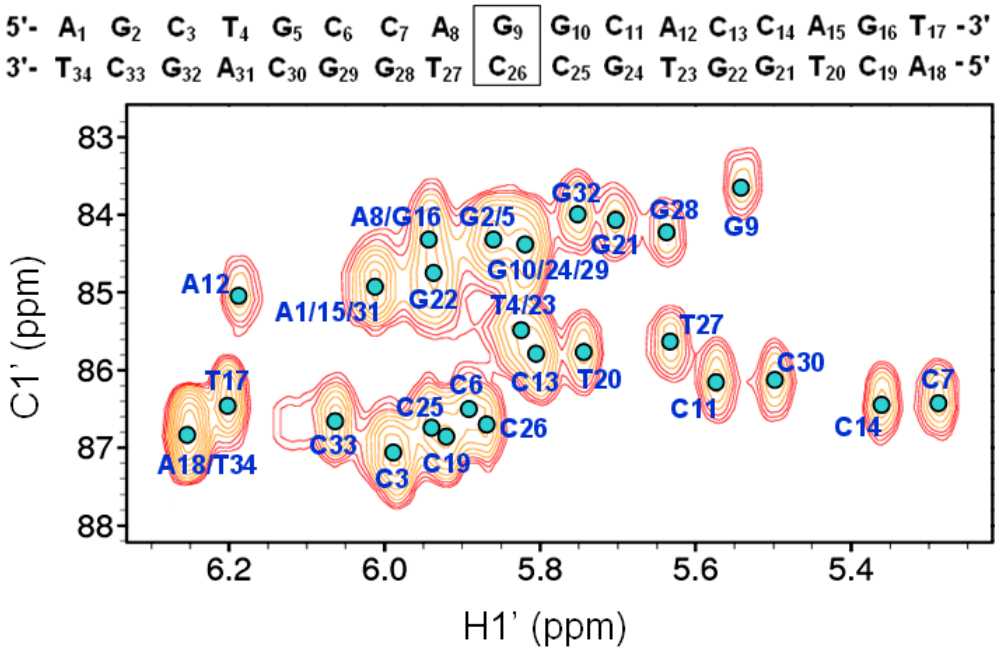
600 MHz NOESY (τ_m = 250ms) spectrum showing internucleotide H6/8-H6/8 NOE connectivities for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



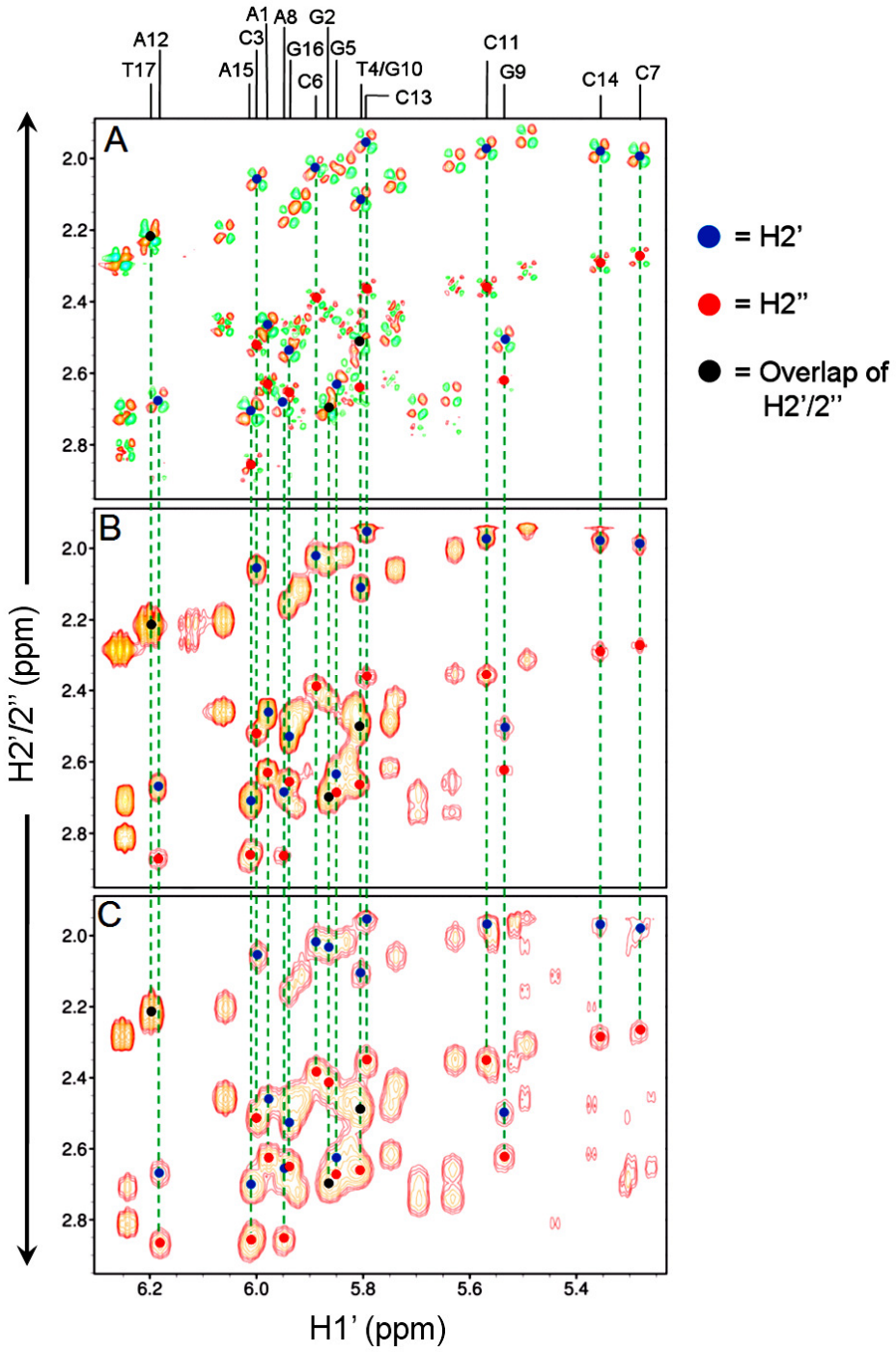
600 MHz NOESY (τ_m = 250ms) spectrum showing internucleotide H6/8-H6/8 NOE connectivities for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



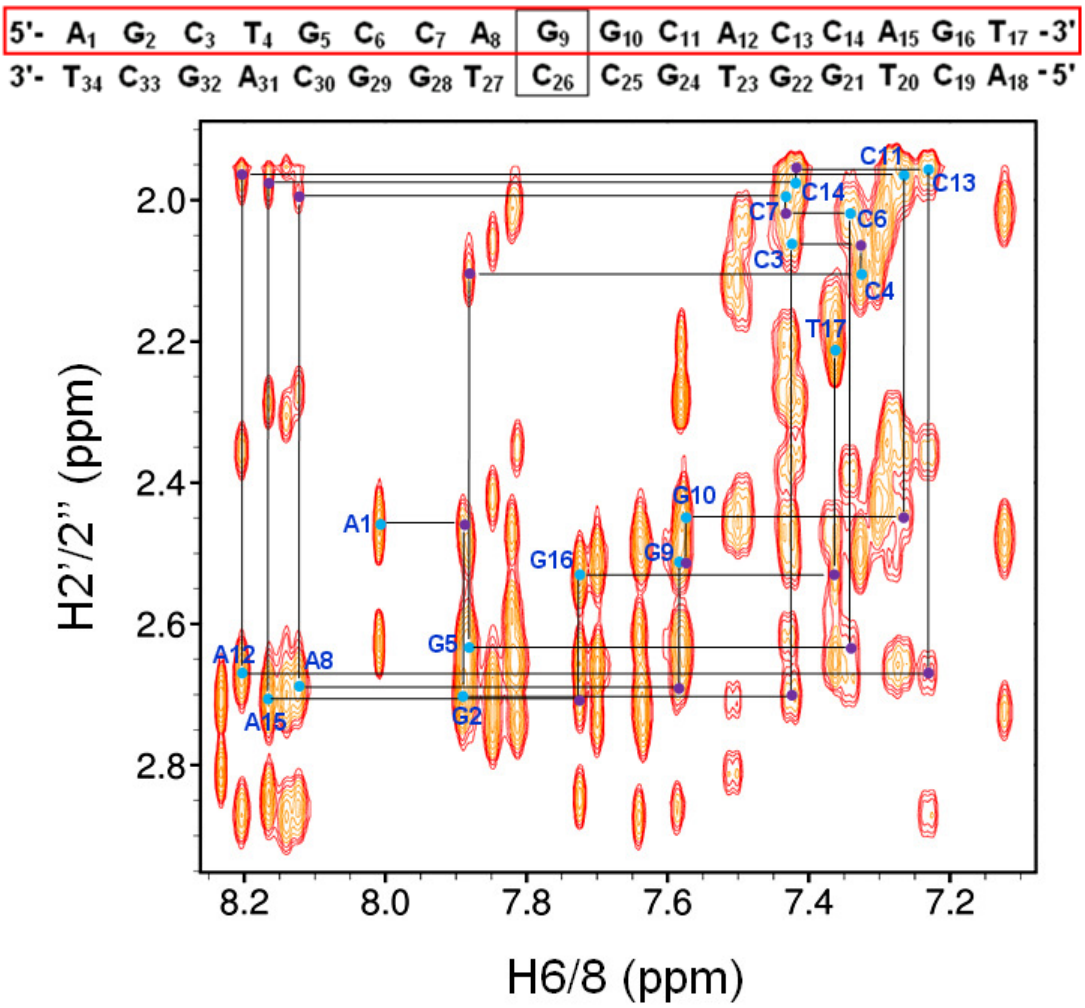
600 MHz ^{1}H-^{13}C HSQC spectrum showing the assignment of C6/C8 resonances in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



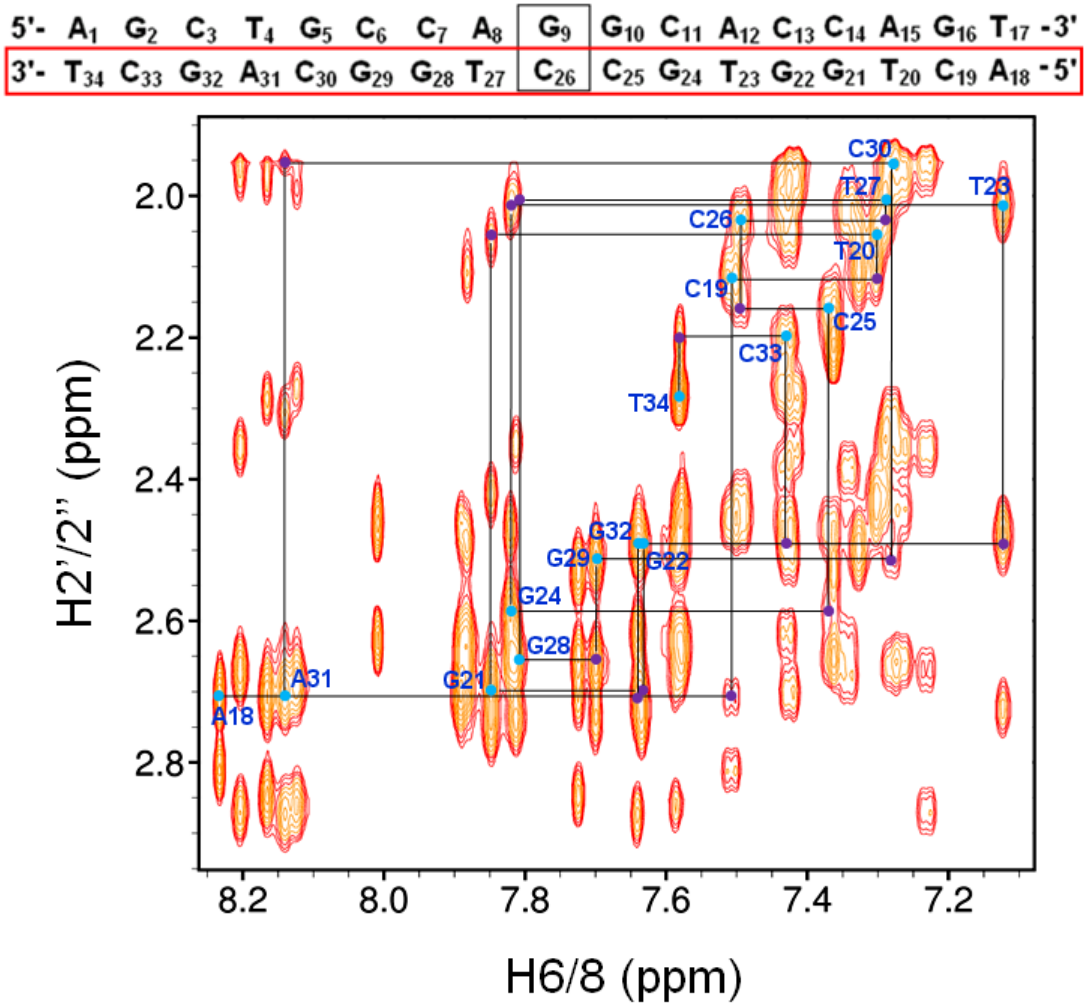
600 MHz ^{1}H-^{13}C HSQC spectrum show the assignment of C1' resonances in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



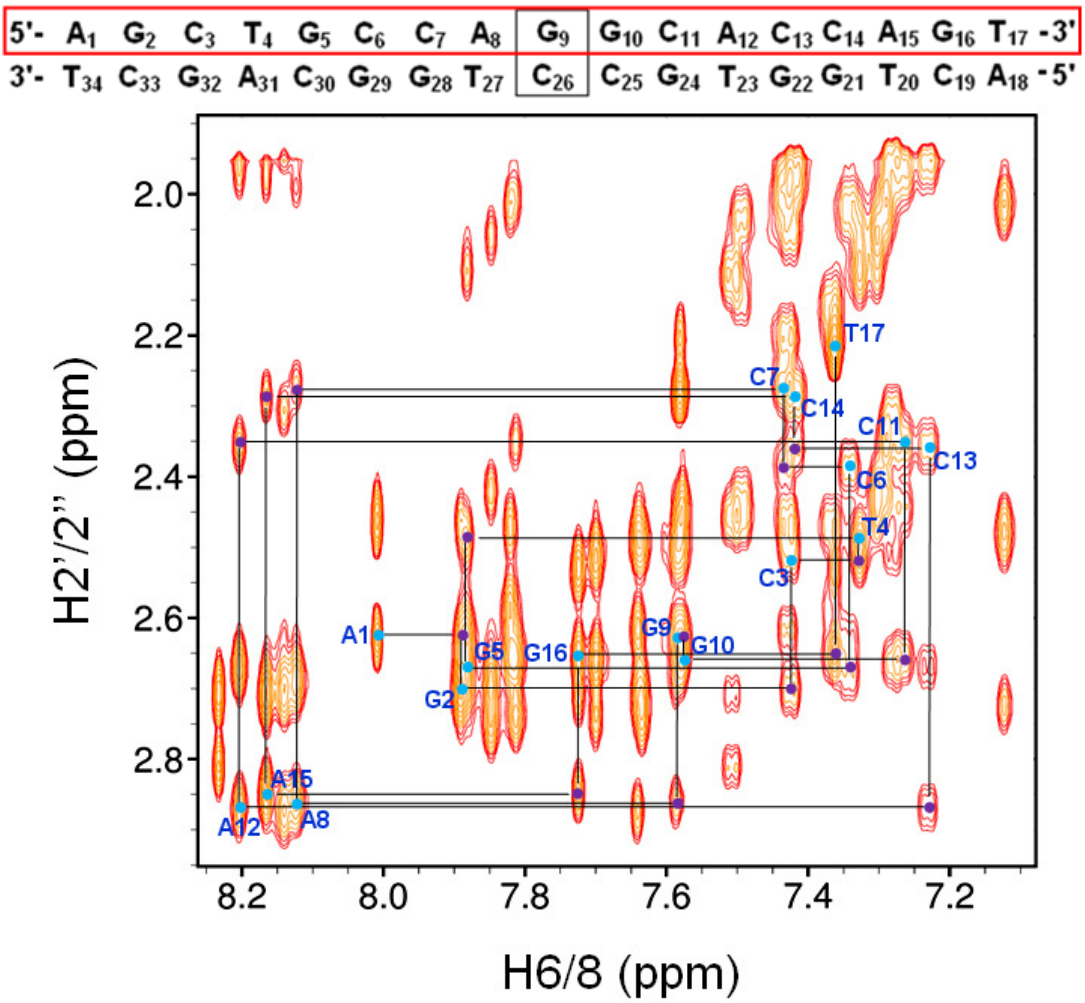
Spectrum stack showing identification and assignment of H2'/H2'' resonances for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. DQF-COSY (800 MHz), B. TOCSY (600 MHz, SL_{mix} = 75ms) and C. NOESY (600MHz, τ_m = 250ms).



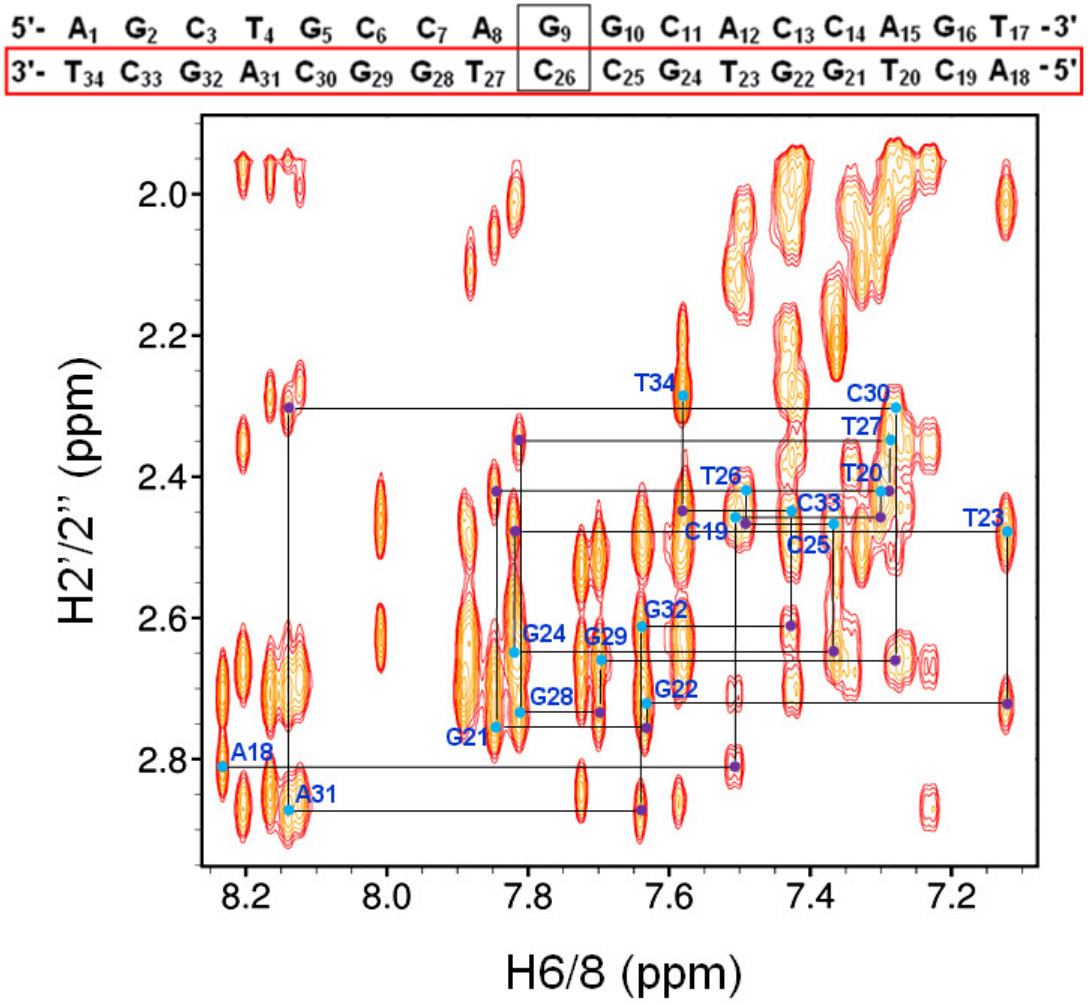
600 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2' sequential NOE connectivities for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide and internucleotide connectivities are shown by blue and purple circles respectively.



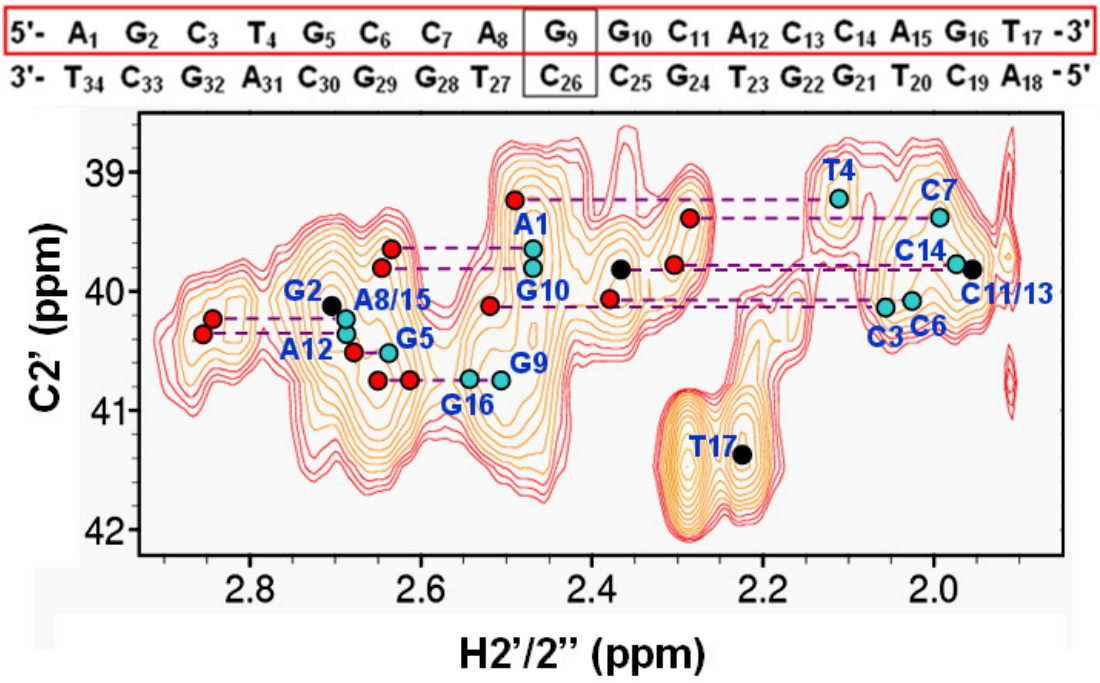
600 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2' sequential NOE connectivities for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide and internucleotide connectivities are shown by blue and purple circles respectively.



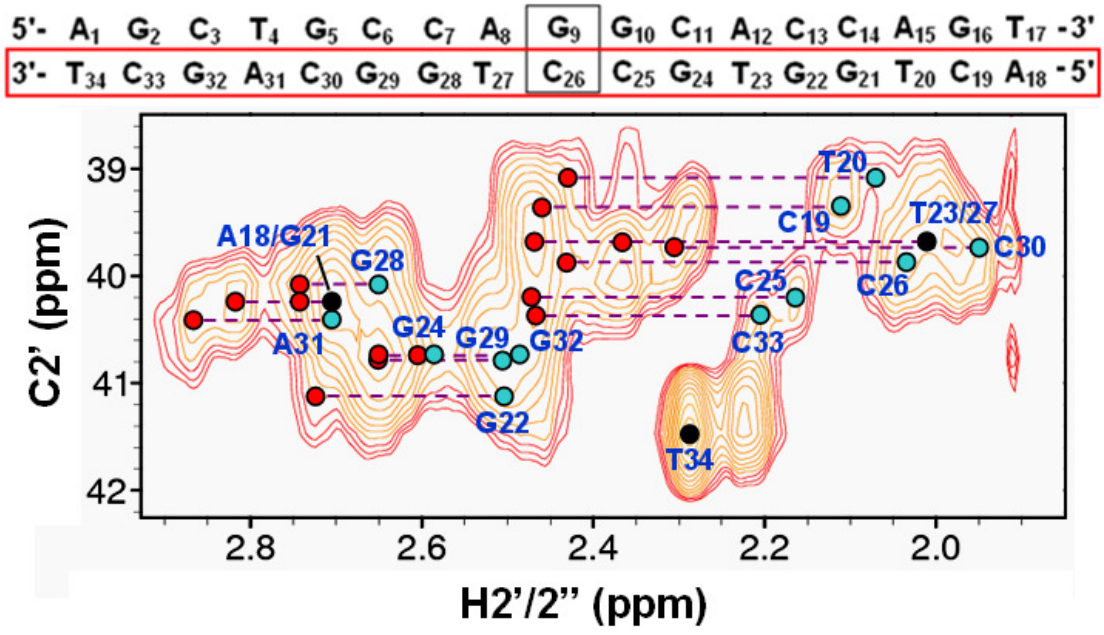
600 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2'' sequential NOE connectivities for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide and internucleotide connectivities are shown by blue and purple circles respectively.



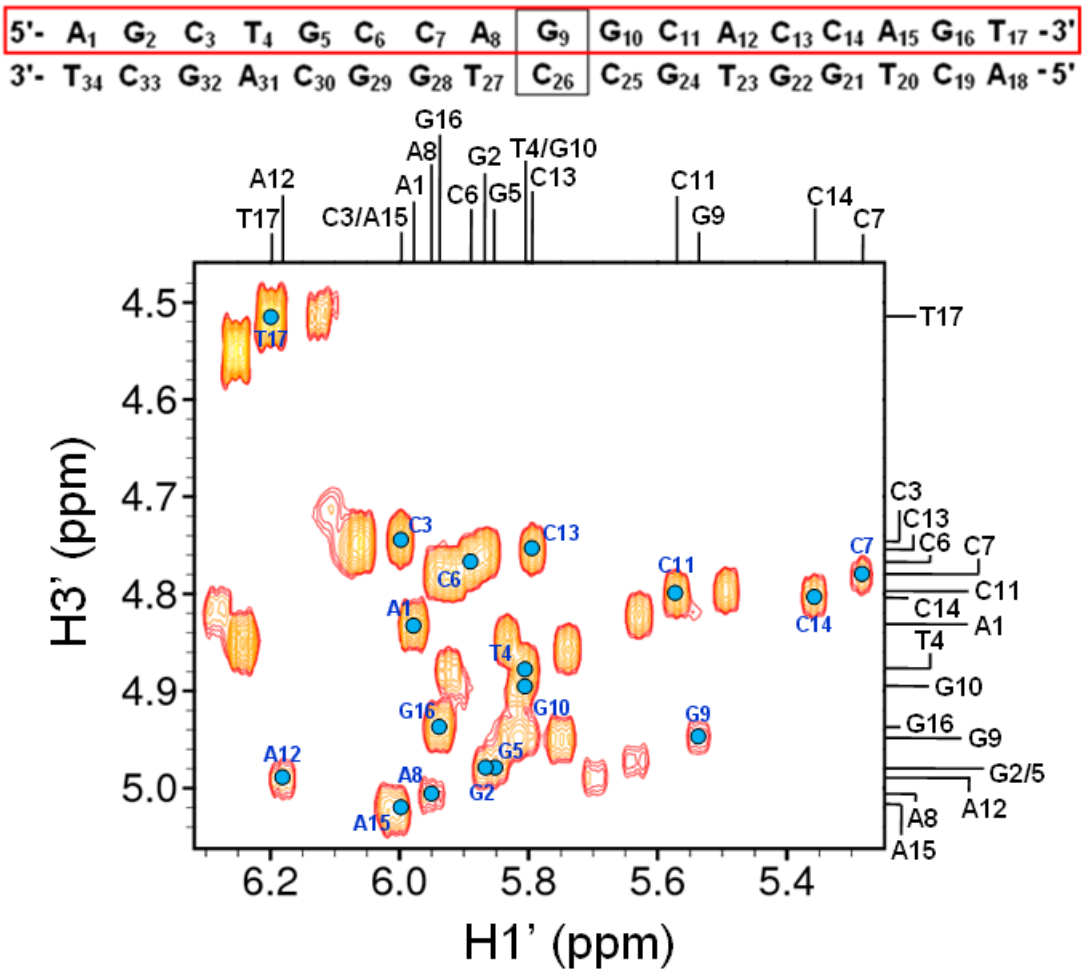
600 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2'' sequential NOE connectivities for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide and internucleotide connectivities are shown by blue and purple circles respectively.



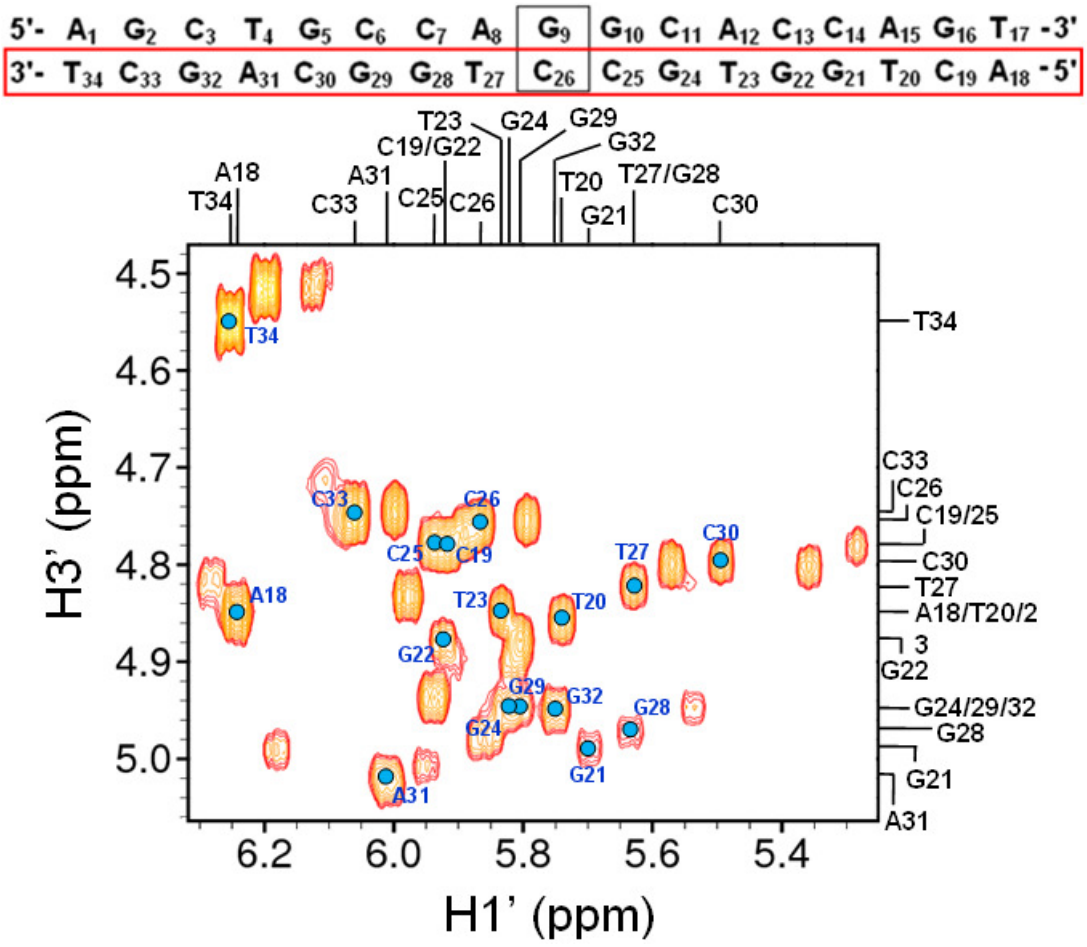
600 MHz ^{1}H-^{13}C HSQC spectrum showing the assignment of C2' resonances for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. H2' and H2'' resonances are shown by blue and red circles respectively; black circles indicate an overlap of resonances.



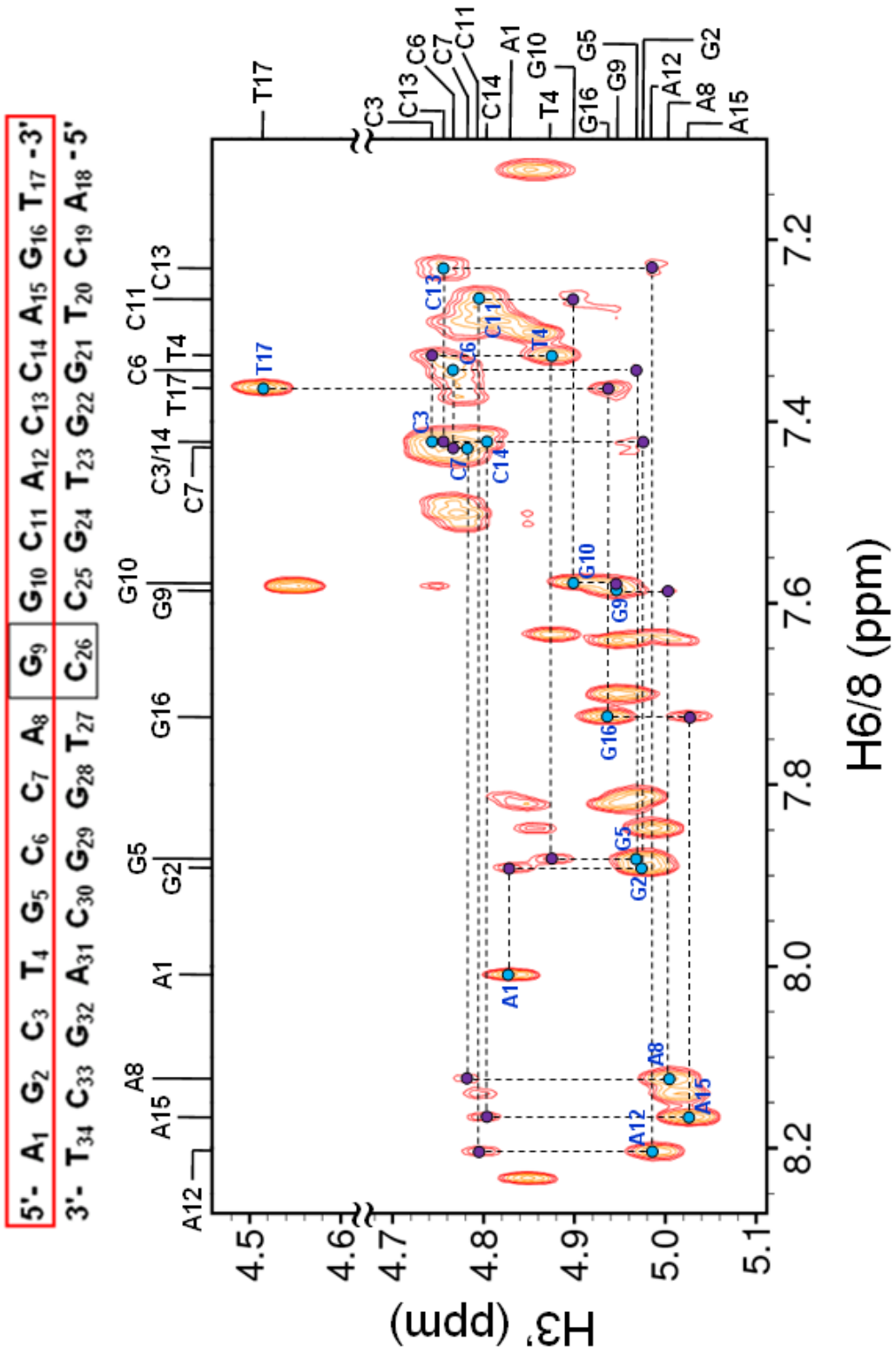
600 MHz ^{1}H-^{13}C HSQC spectrum showing the assignment of C2' resonances for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. H2' and H2'' resonances are shown by blue and red circles respectively; black circles indicate an overlap of resonances.



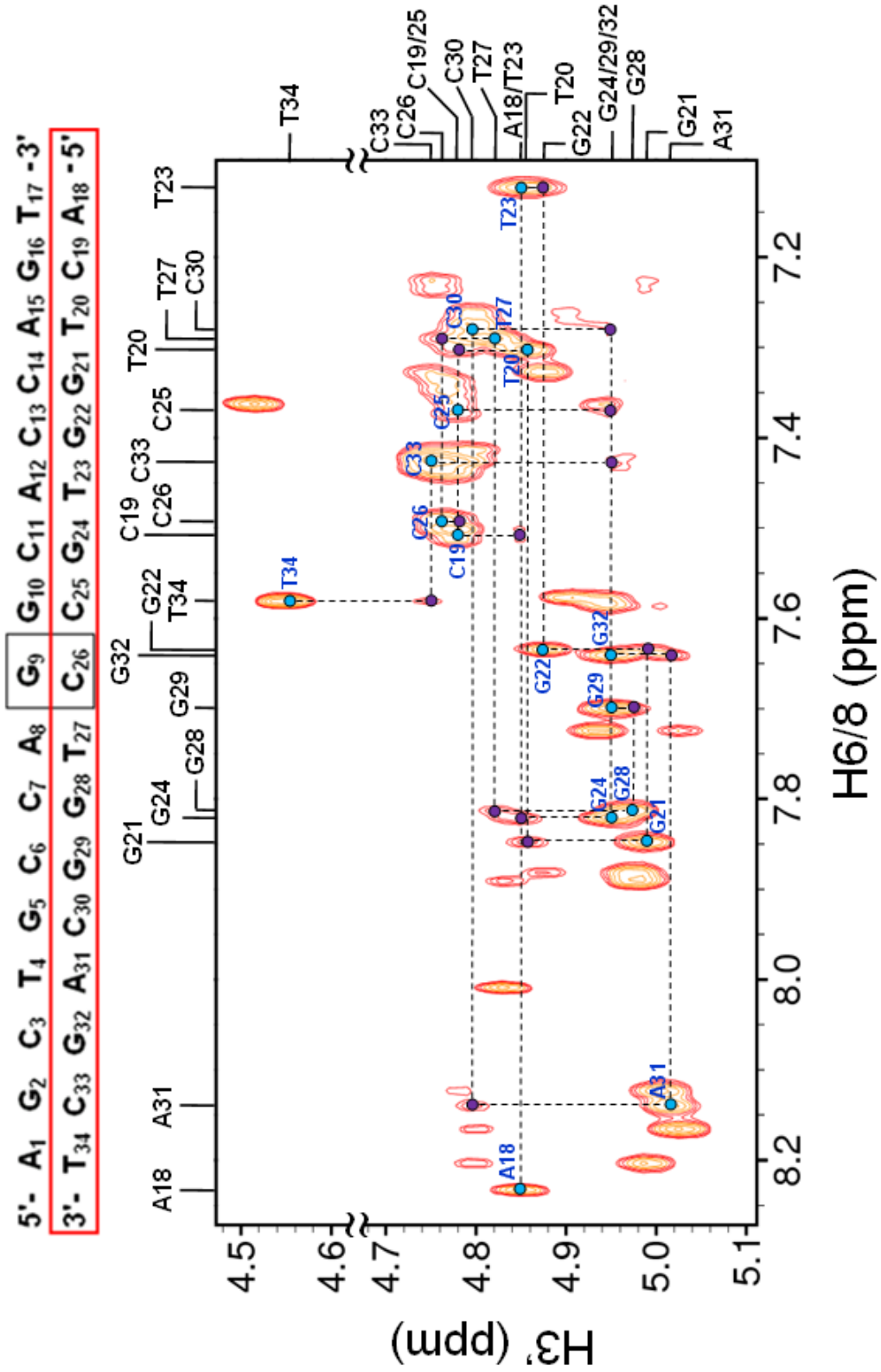
600 MHz TOCSY (SL_{mix} = 75ms) spectrum showing H1'-H3' correlations for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



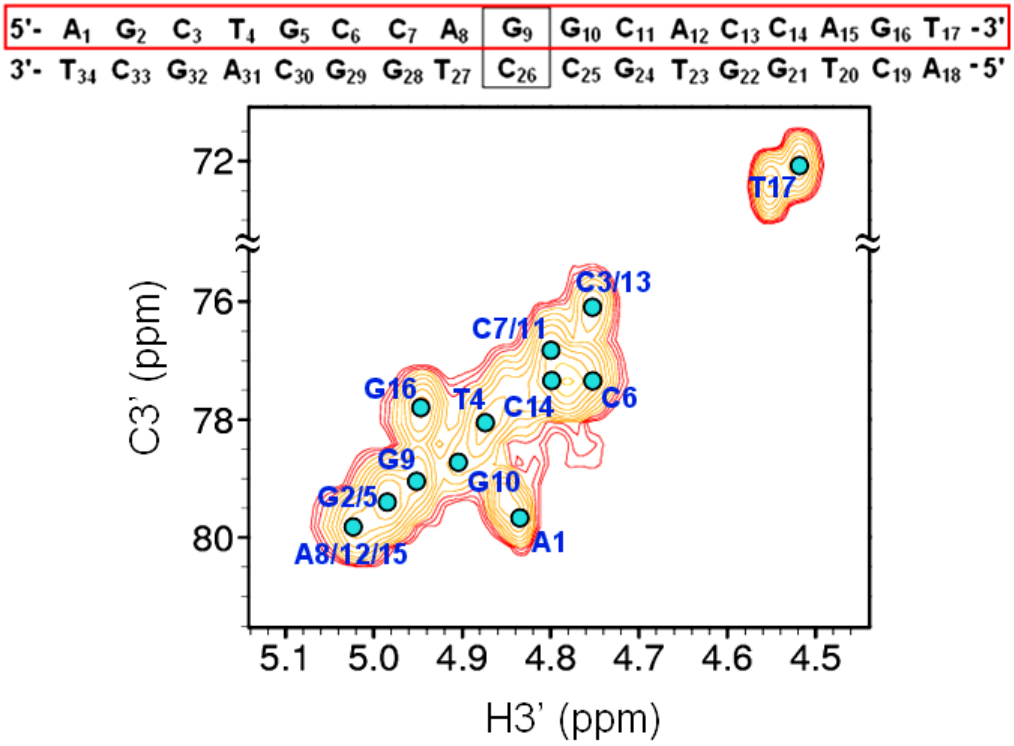
600 MHz TOCSY (SL_{mix} = 75ms) spectrum showing H1'-H3' correlations for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



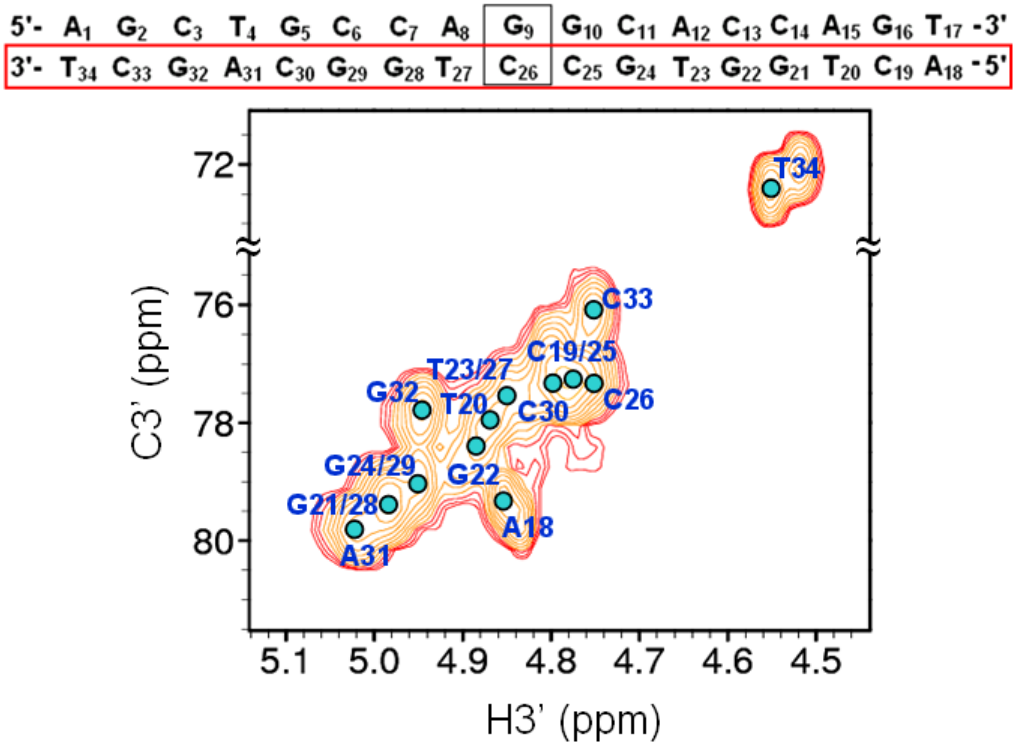
600 MHz NOESY (τ_{mix} = 250ms) spectrum showing H6/8-H3' NOE connectivities for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide and internucleotide connectivities are shown by blue and purple circles respectively



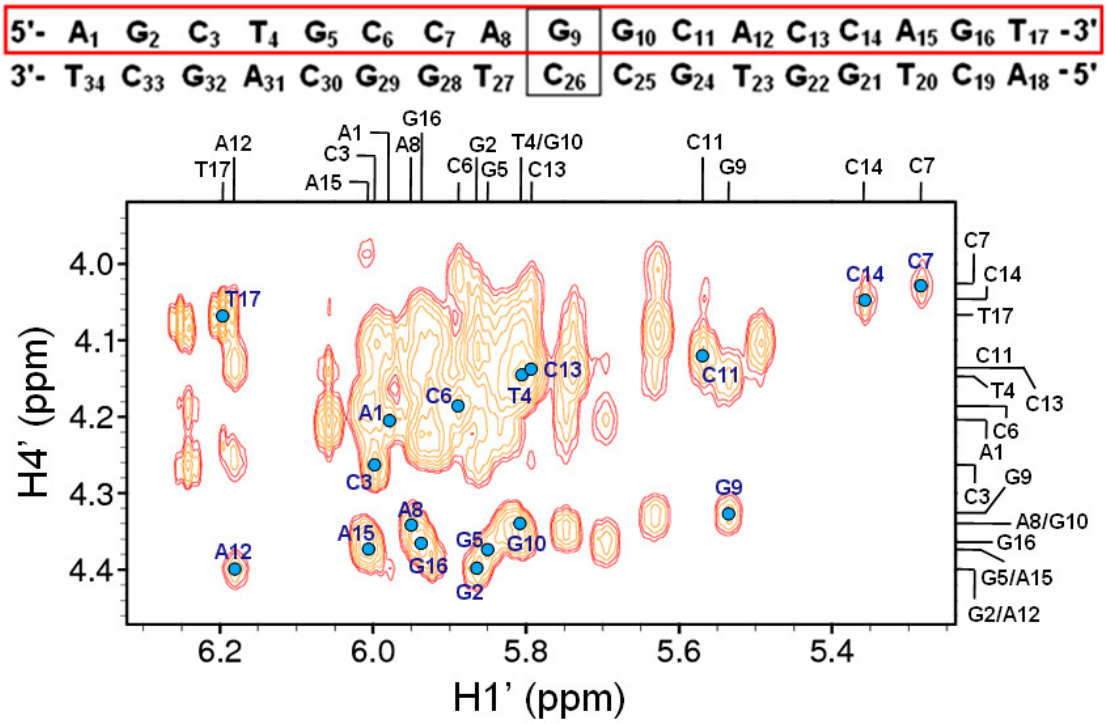
600 MHz NOESY (τ_{mix} = 250ms) spectrum showing H6/8-H3' NOE connectivities for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide and internucleotide connectivities are shown by blue and purple circles respectively



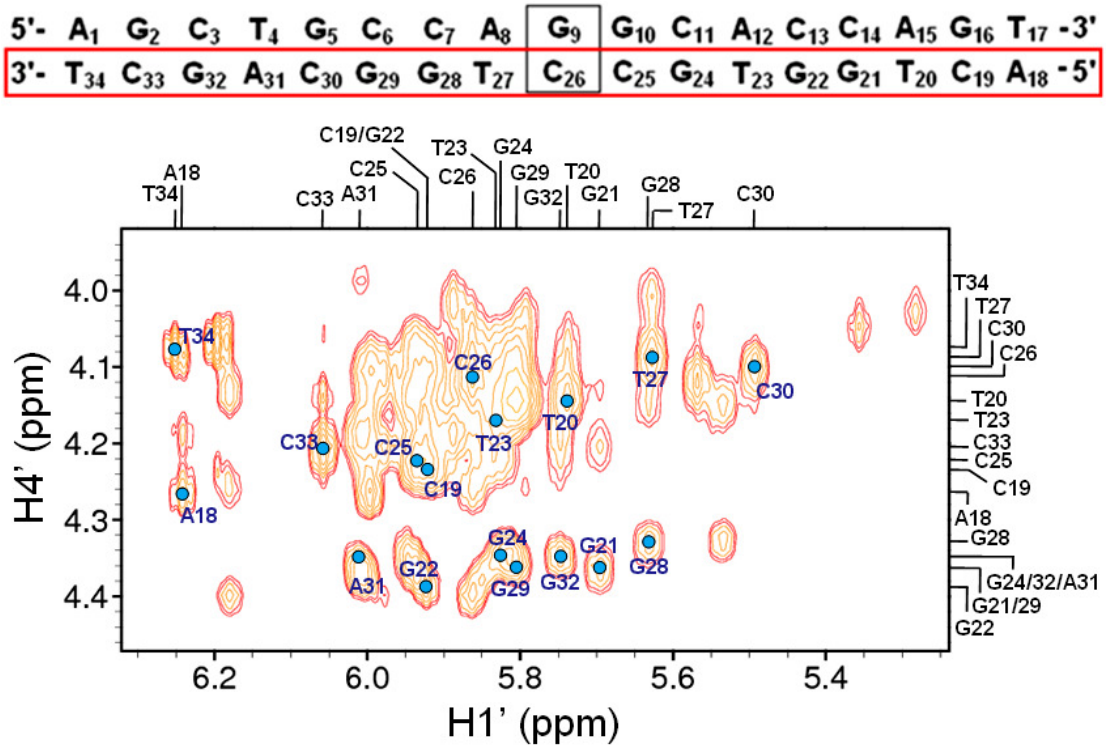
600 MHz ^{1}H-^{13}C HSQC spectrum show the assignment of C3' resonances for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



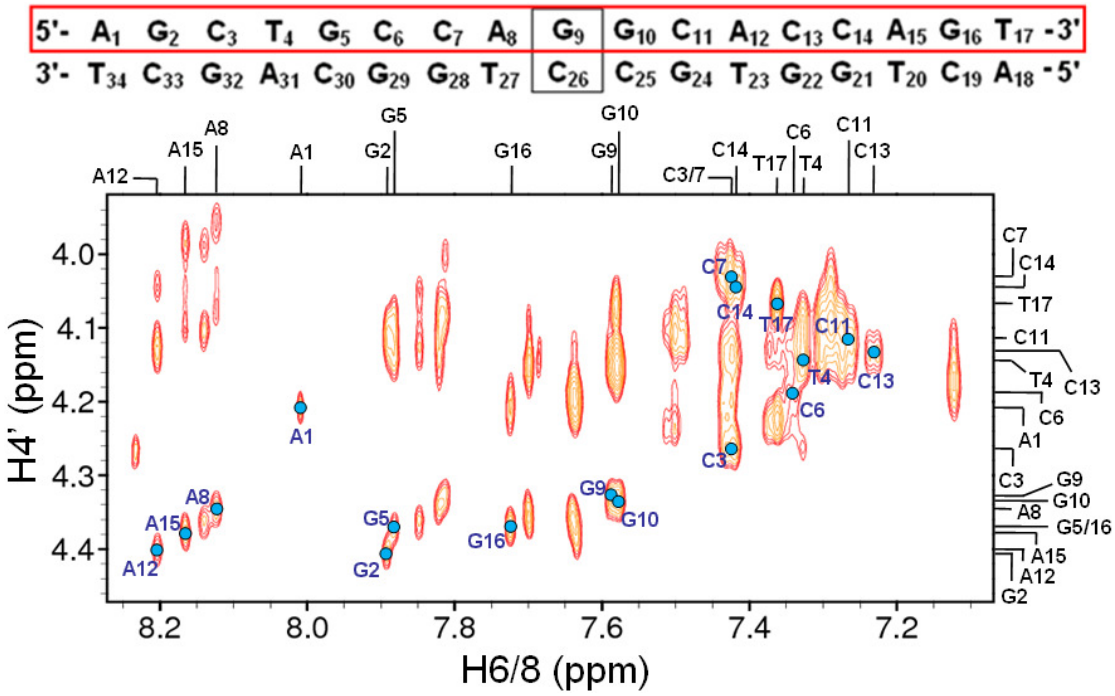
600 MHz ^{1}H-^{13}C HSQC spectrum show the assignment of C3' resonances for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



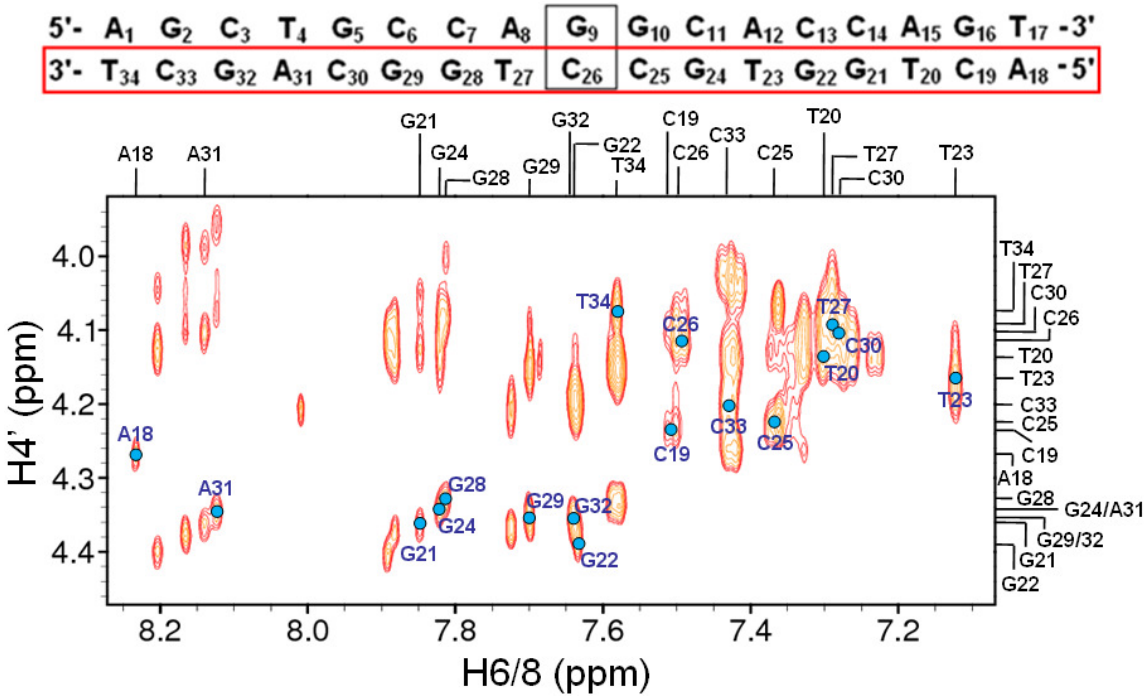
600 MHz NOESY (τ_{mix} = 250ms) spectrum showing H1'-H4' NOE connectivities for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



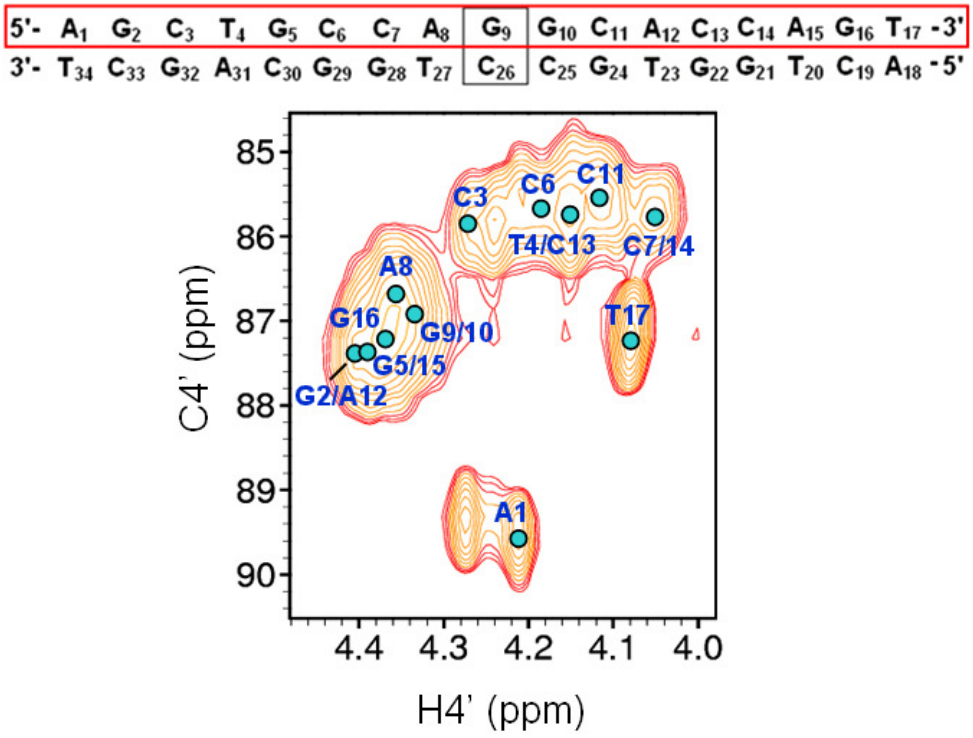
600 MHz NOESY (τ_{mix} = 250ms) spectrum showing H1'-H4' NOE connectivities for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



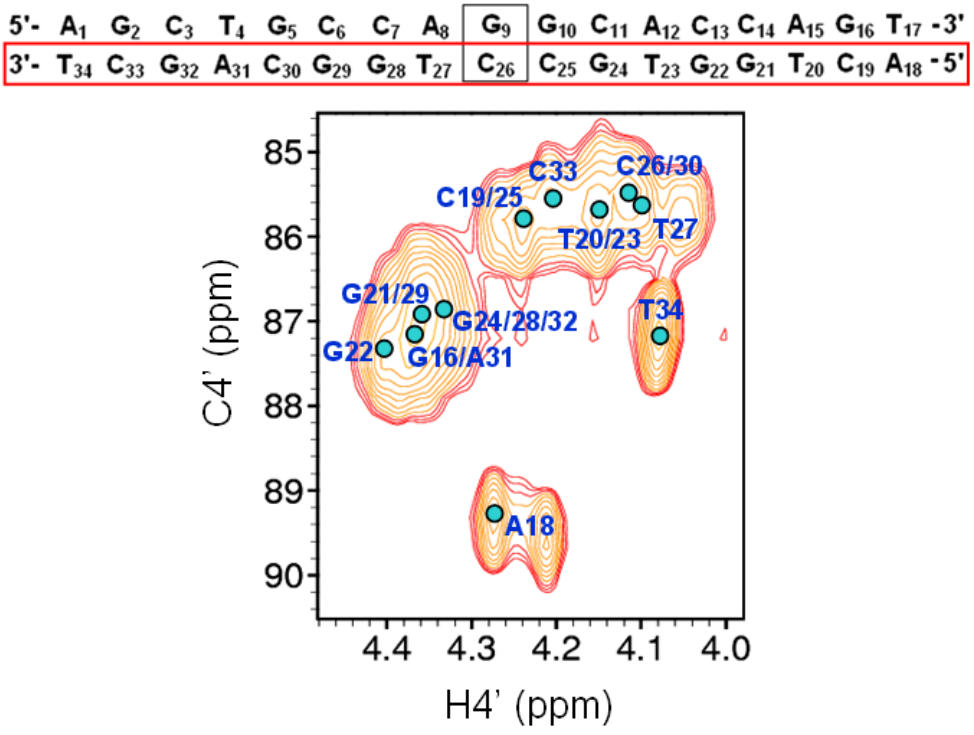
600 MHz NOESY (τ_{mix} = 250ms) spectrum showing H6/8-H4' NOE connectivities for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



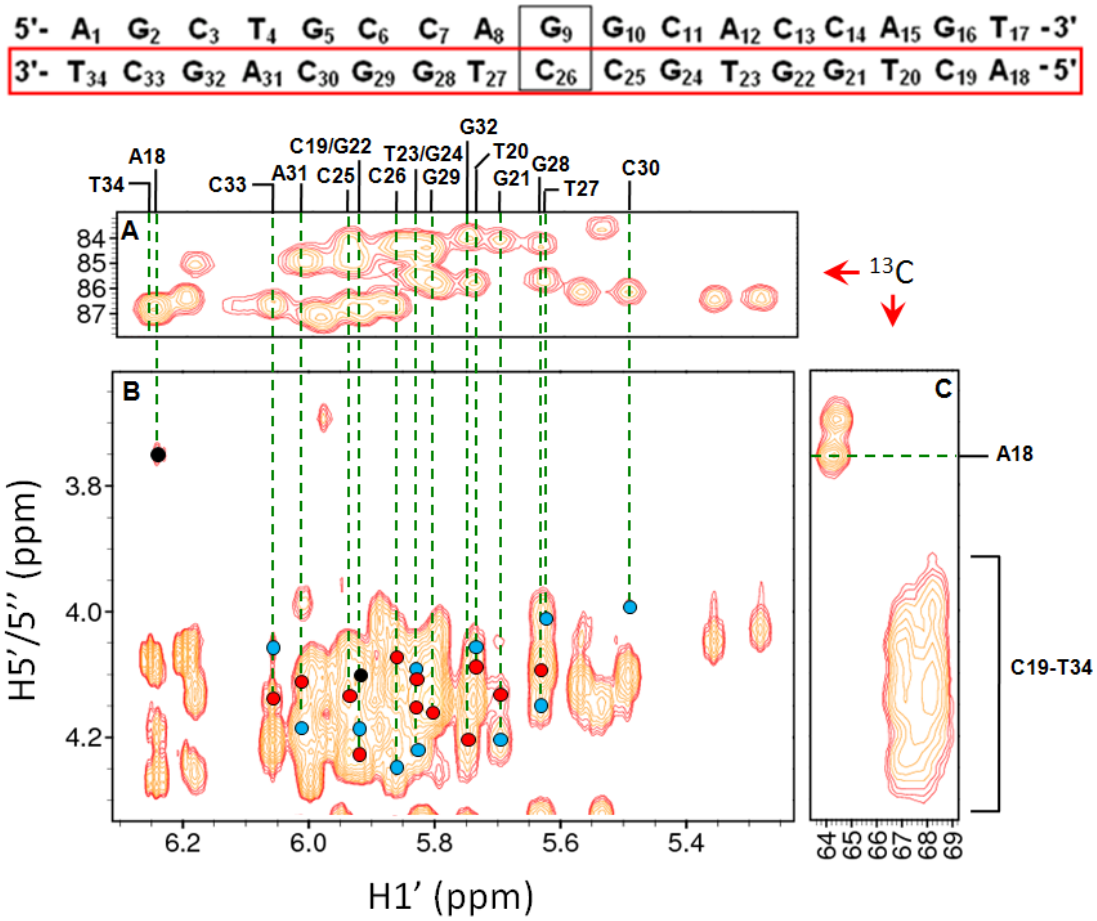
600 MHz NOESY (τ_{mix} = 250ms) spectrum showing H6/8-H4' NOE connectivities for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



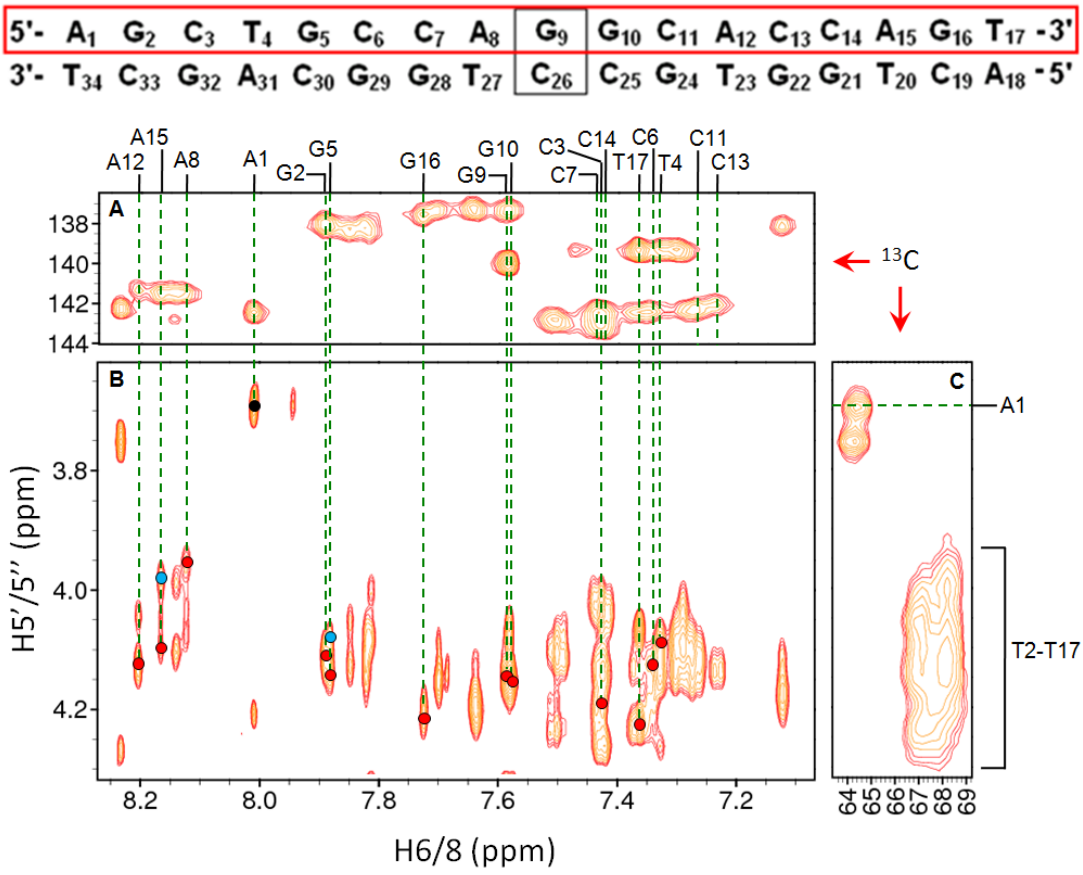
600 MHz ^{1}H-^{13}C HSQC spectrum showing the assignment of C4' resonances for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



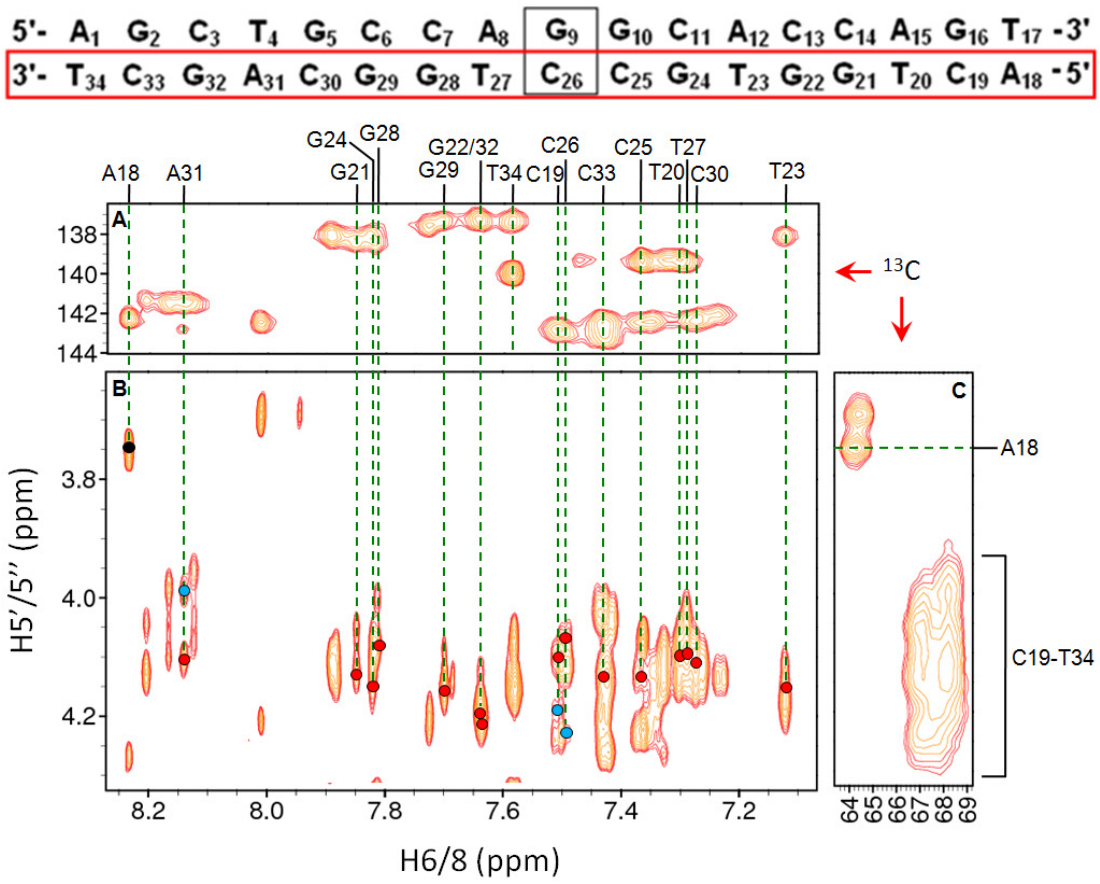
600 MHz ^{1}H-^{13}C HSQC spectrum showing the assignment of C4' resonances for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



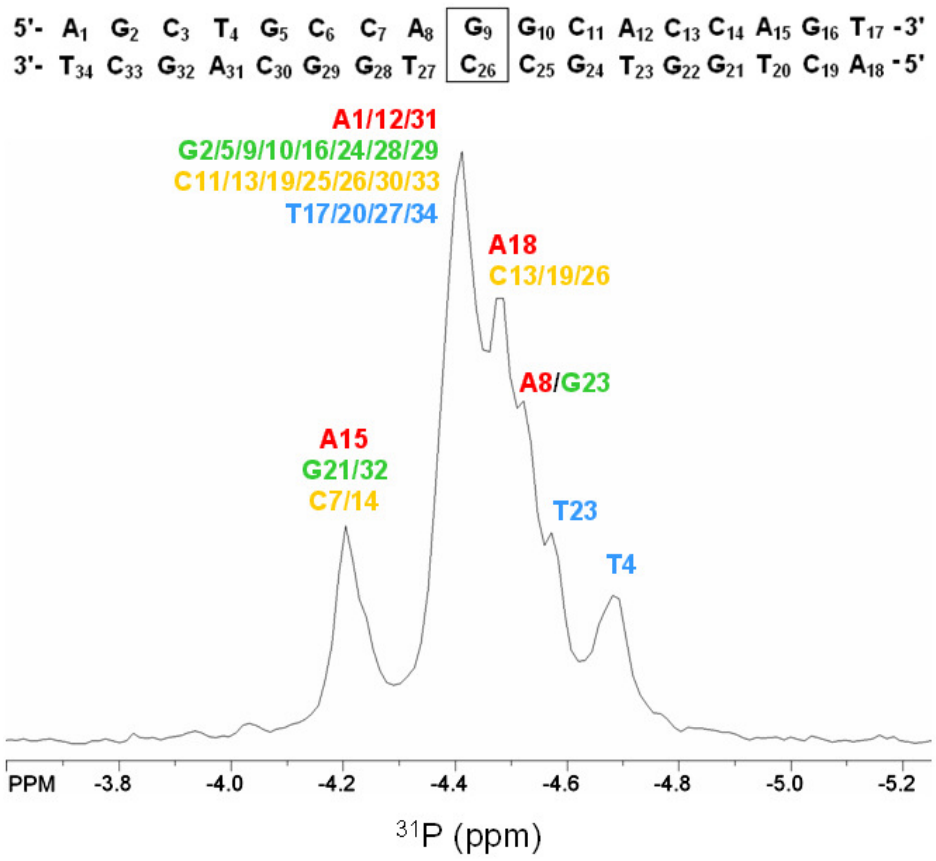
600 MHz NOESY (τ_{mix} = 250ms) spectrum (B) showing the assignment of H1'-H5'/5'' NOE connectivities for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. 600 MHz ^{1}H-^{13}C HSQC panels show the corresponding assignment of C1' (A) and C5' (C) resonances. Blue circles indicate H5' NOEs, red circles indicate H5'' NOEs and black circles indicate an overlap of H5'/5'' NOEs. The C5' resonances of residues C19-T34 are highly degenerate and appear in the heavily overlapped region indicated in the figure.



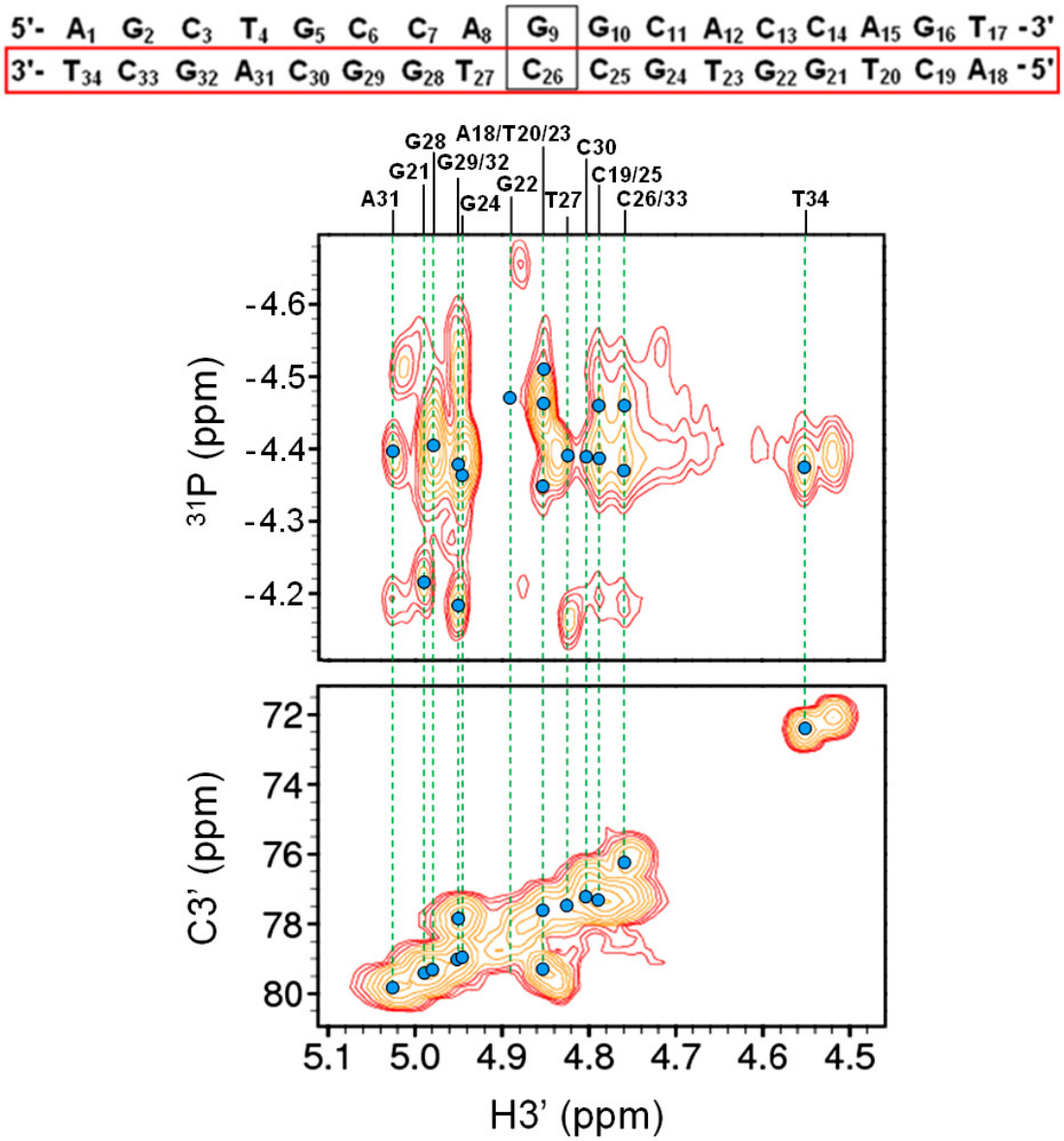
600 MHz NOESY (τ_{mix} = 250ms) spectrum (B) showing H6/8-H5'/5'' NOE connectivities for residues A1-T17 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. 600 MHz ^{1}H-^{13}C HSQC panels show the corresponding assignment of C6/8 (A) and C5' (C) resonances. Blue circles indicate H5' NOEs, red circles indicate H5'' NOEs and black circles indicate an overlap of H5'/5'' NOEs. The C5' resonances of residues C19-T34 are highly degenerate and appear in the heavily overlapped region indicated in the figure.



600 MHz NOESY (τ_{mix} = 250ms) spectrum (B) showing H6/8-H5'/5'' NOE connectivities for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. 600 MHz ^{1}H-^{13}C HSQC panels show the corresponding assignment of C6/8 (A) and C5' (C) resonances. Blue circles indicate H5' NOEs, red circles indicate H5'' NOEs and black circles indicate an overlap of H5'/5'' NOEs. The C5' resonances of residues C19-T34 are highly degenerate and appear in the heavily overlapped region indicated in the figure.

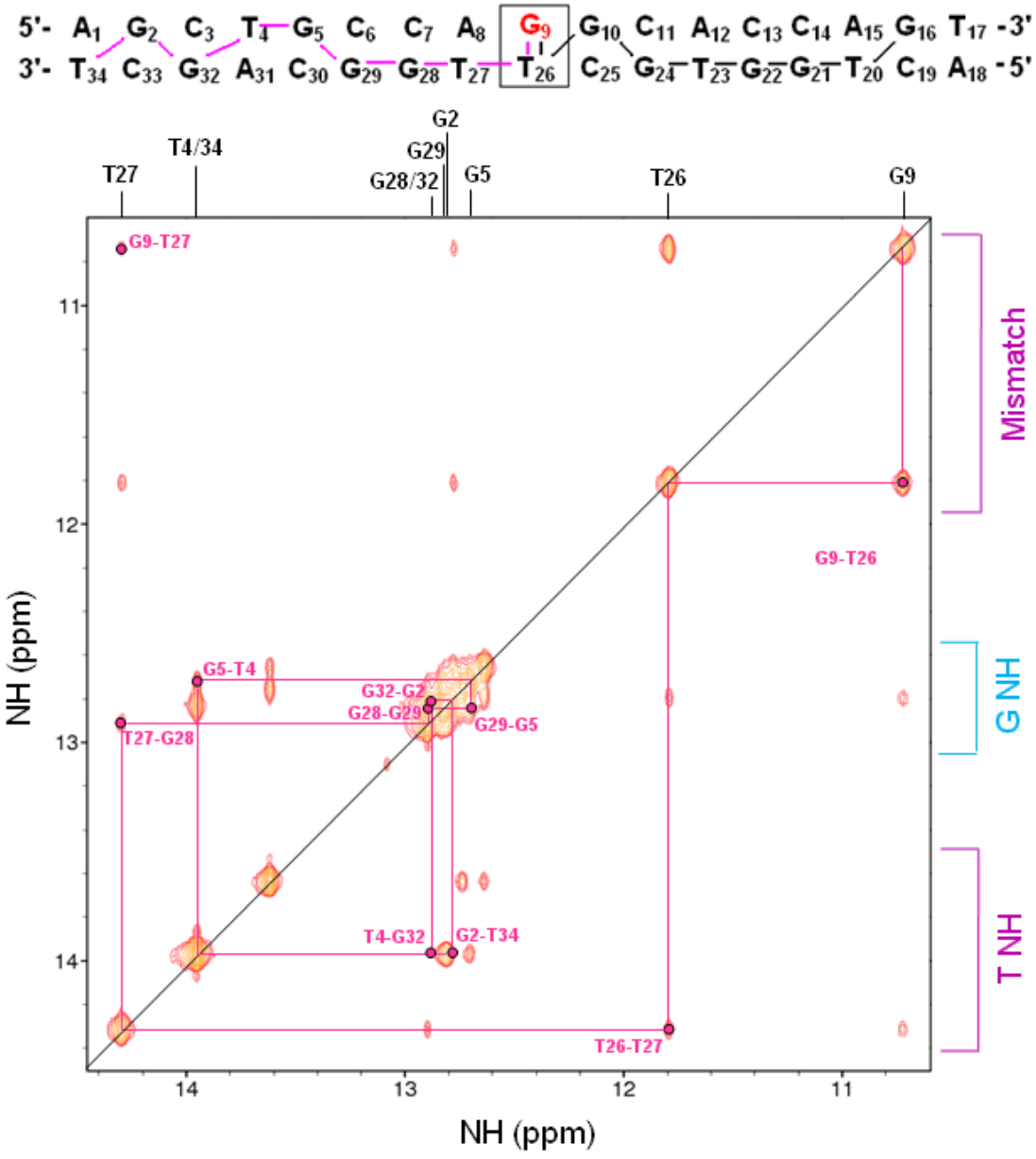


81 MHz ^{1}H decoupled 1D ^{31}P-NMR spectrum showing the assignment of ^{31}P resonances in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.

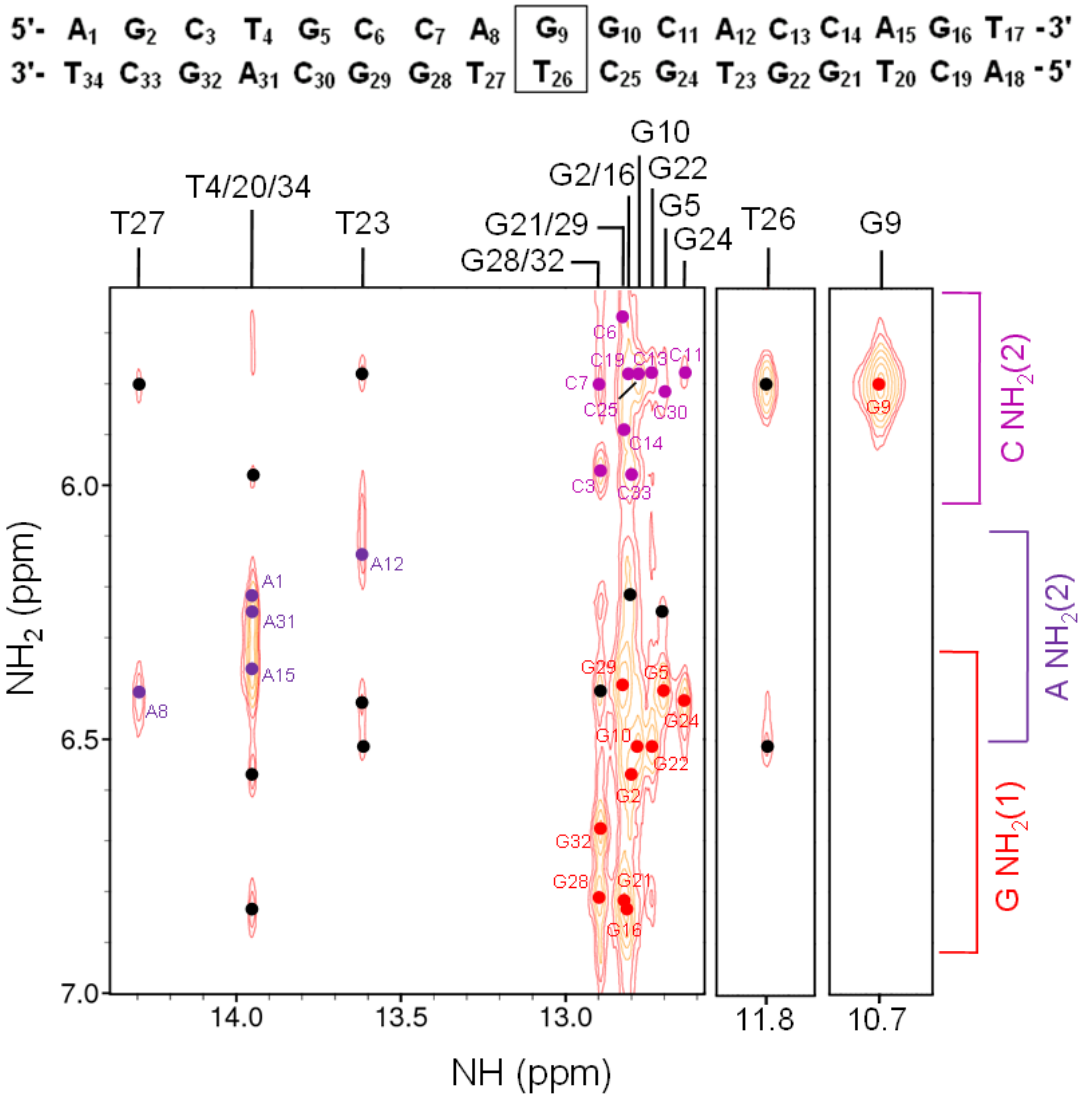


(Top panel) 600 MHz ^{1}H-^{31}P CPMG-HSQC-NOESY (τ_{mix} = 500ms) spectrum and (bottom panel) 600 MHz ^{1}H-^{13}C HSQC spectrum. The assignment of ^{31}P resonances is shown by means of H3'-^{31}P HSQC correlations for residues A18-T34 in 17mer canonical GC DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.

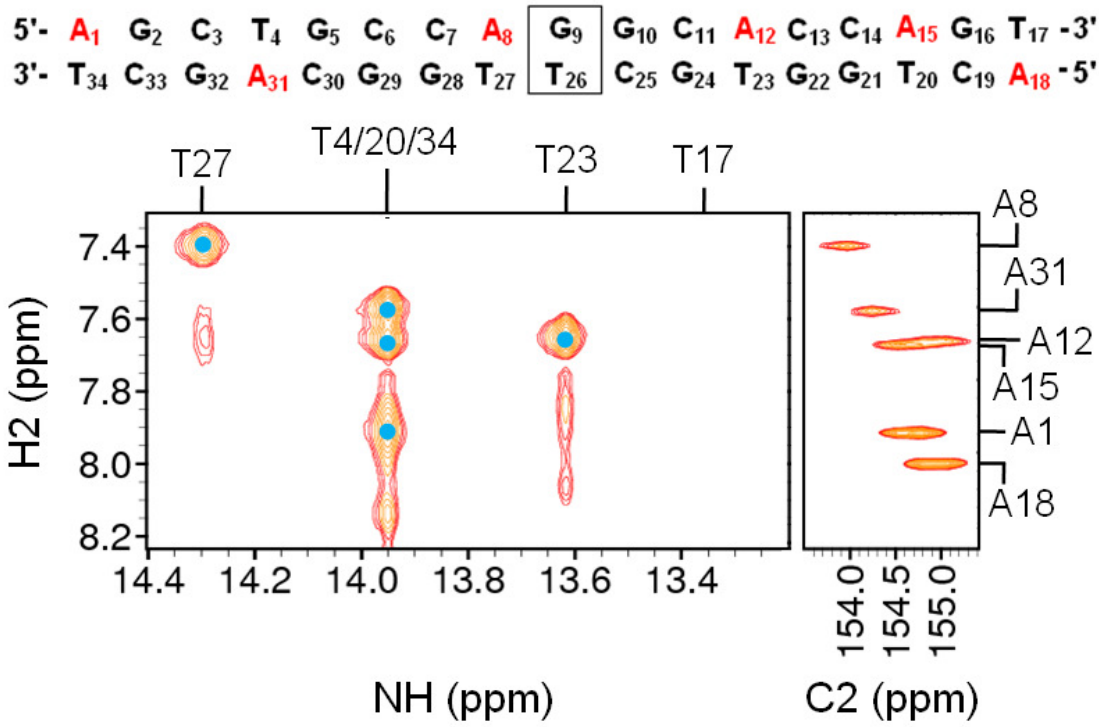
ii. 17mer mismatch GT DNA (800 MHz)



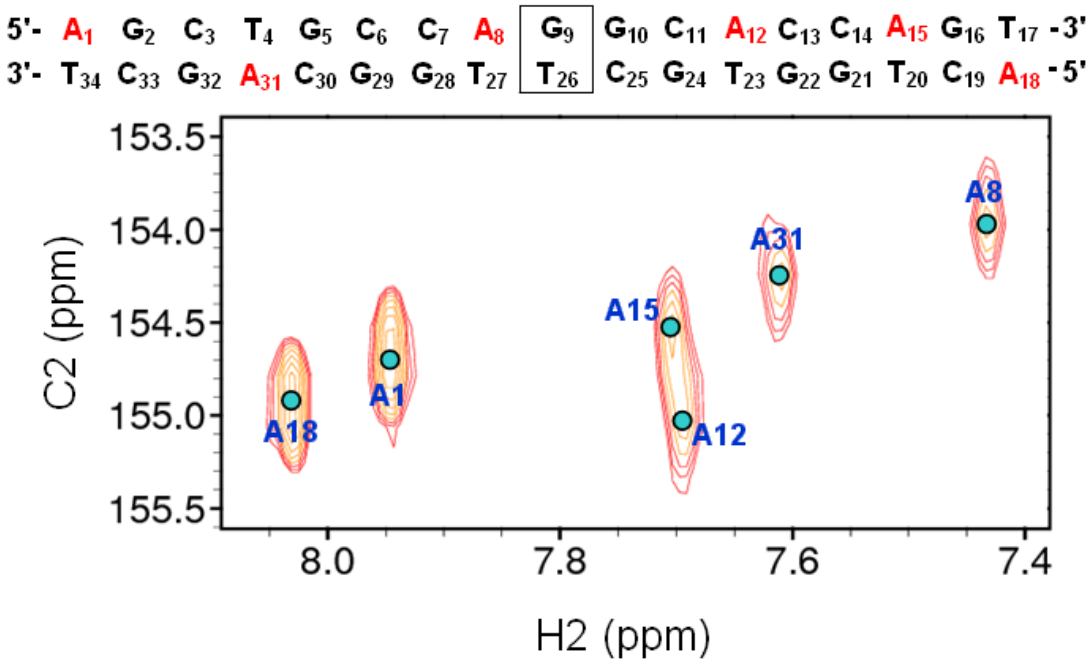
800 MHz NOESY (τ_m = 250ms) spectrum of 17mer mismatch GT DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 1°C. The sequential connectivity stretching from G9 to T34 is indicated by the pink trace.



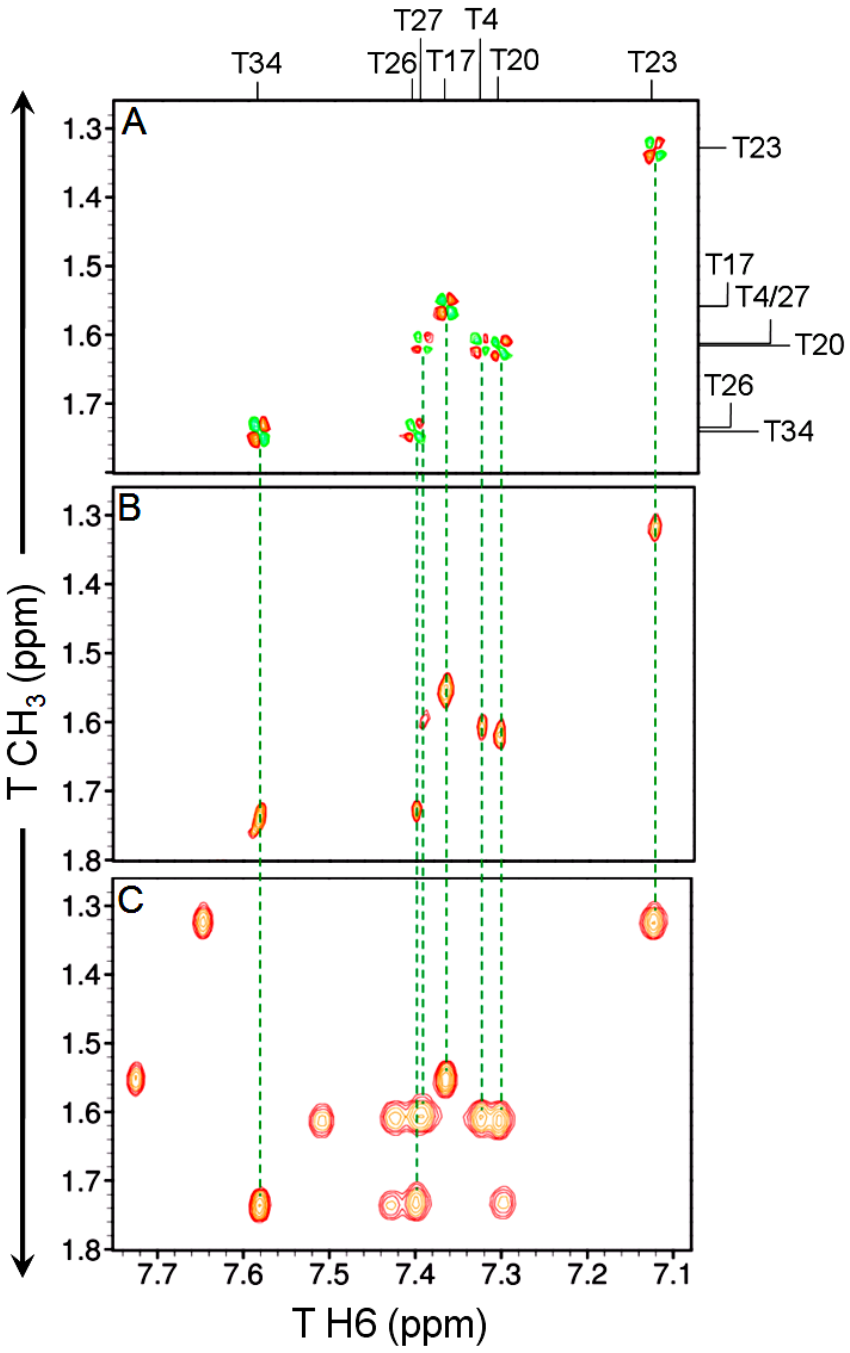
800 MHz NOESY (τ_m = 250ms) spectrum of 17mer mismatch GT DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 1°C. Base pair imino-amino NOE connectivities to hydrogen bonded guanine NH_2 and non-hydrogen bonded cytosine/adenine NH_2 protons are indicated by red, light purple and dark purple circles respectively. Important intra/interstrand NOE connectivities which confirmed assignment are also highlighted by black circles.



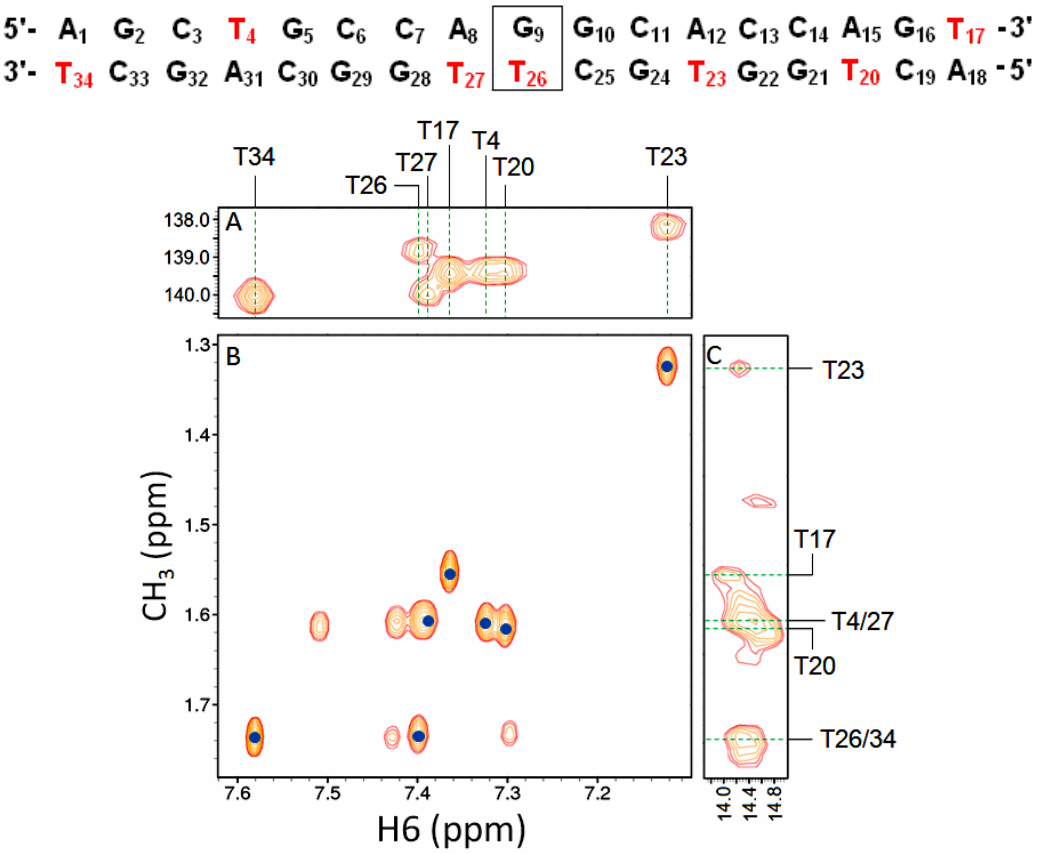
(Left panel) 800 MHz NOESY (τ_m = 250ms) spectrum of 17mer mismatch GT DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 1°C, (Right panel) 800 MHz ^1H-^{13}C HSQC spectrum of 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. The assignment of adenine H2 resonances is based on NOE connectivities to thymine H3 imino resonances. Correlation to the ^1H-^{13}C HSQC spectrum identified the chemical shifts of C2 resonances.



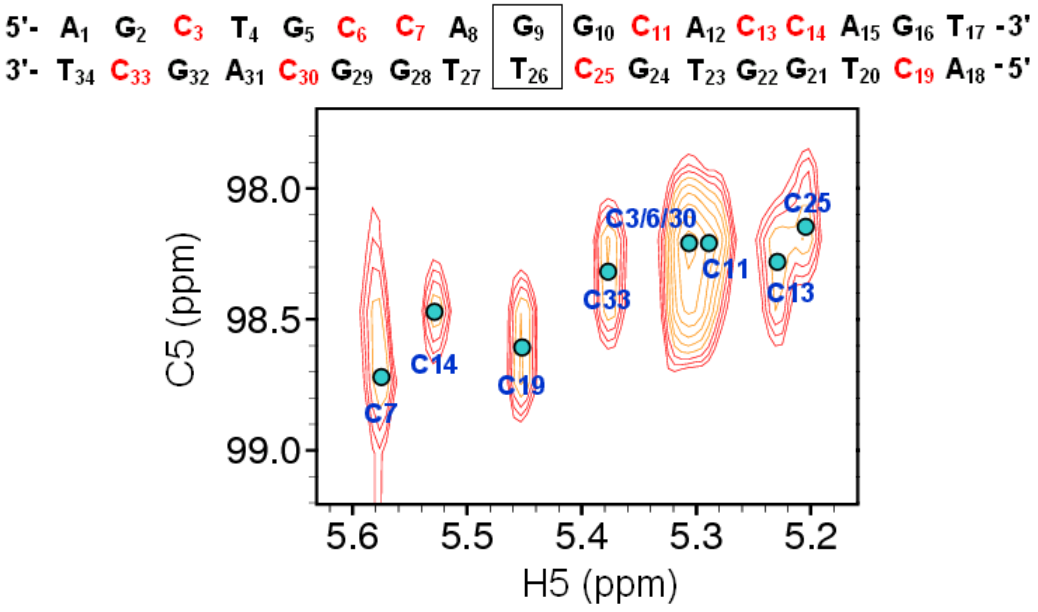
800 MHz ^1H-^{13}C HSQC spectrum of 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C showing assignment of C2 resonances.



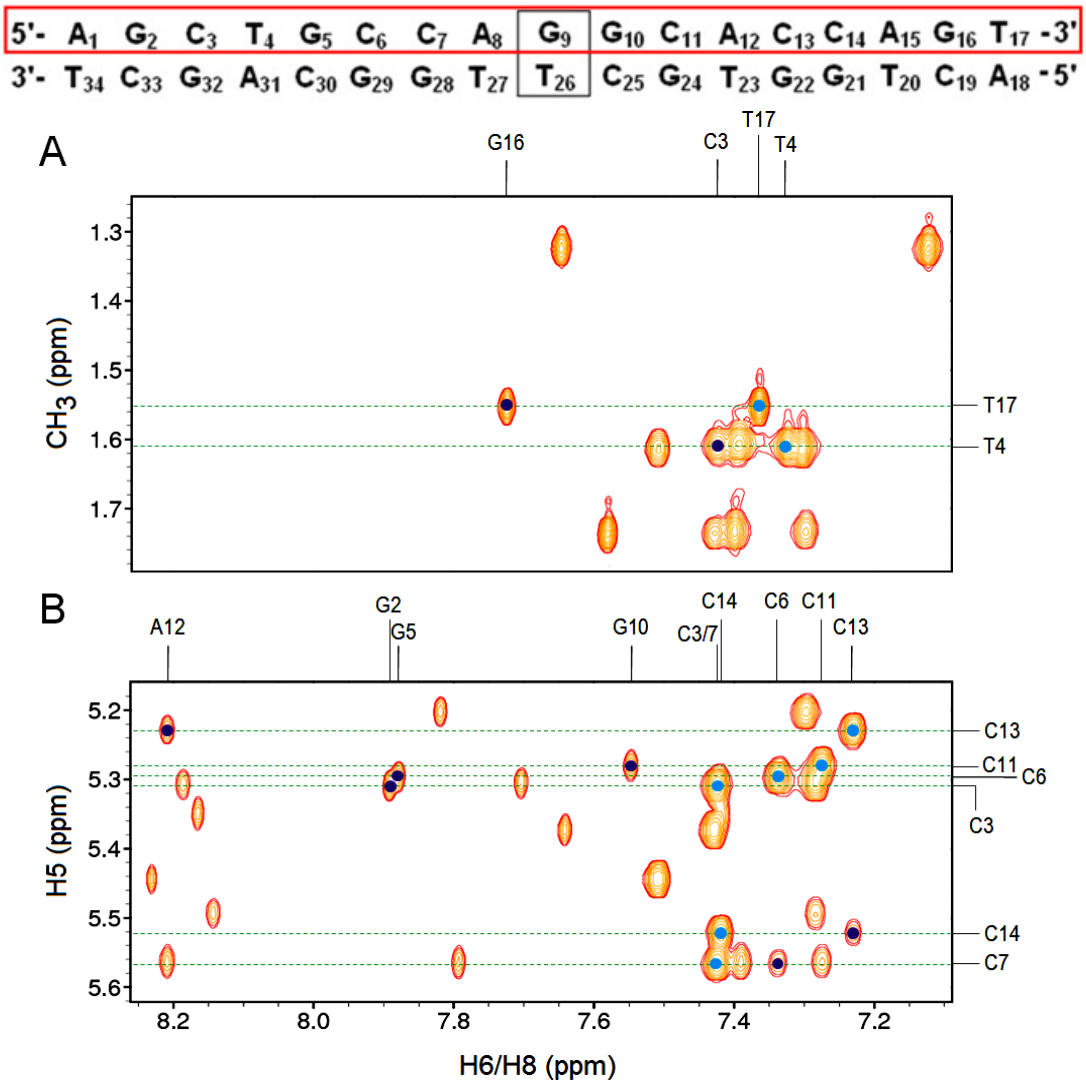
800 MHz spectrum stack showing the assignment of thymine CH_3-H6 correlations in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. DQF-COSY, B. TOCSY (SL_{mix} = 75ms) and C. NOESY (τ_m = 250ms).



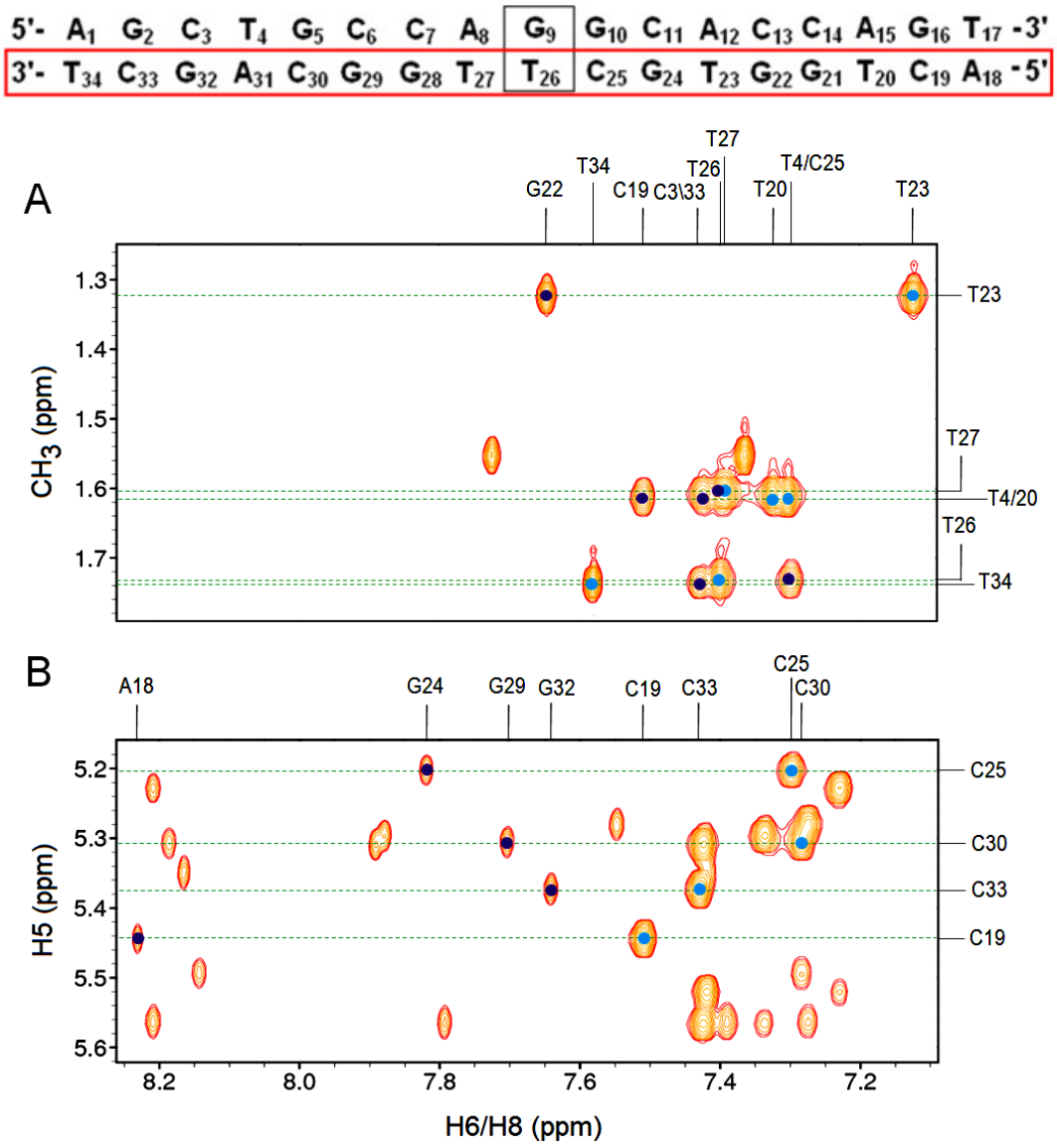
800 MHz NOESY (τ_m = 250ms) spectrum (B) highlighting thymine CH_3-H6 correlations in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. 800 MHz ^1H-^{13}C HSQC panels show the assignment of the cytosine C6 (A) and C5 (C) chemical shifts.



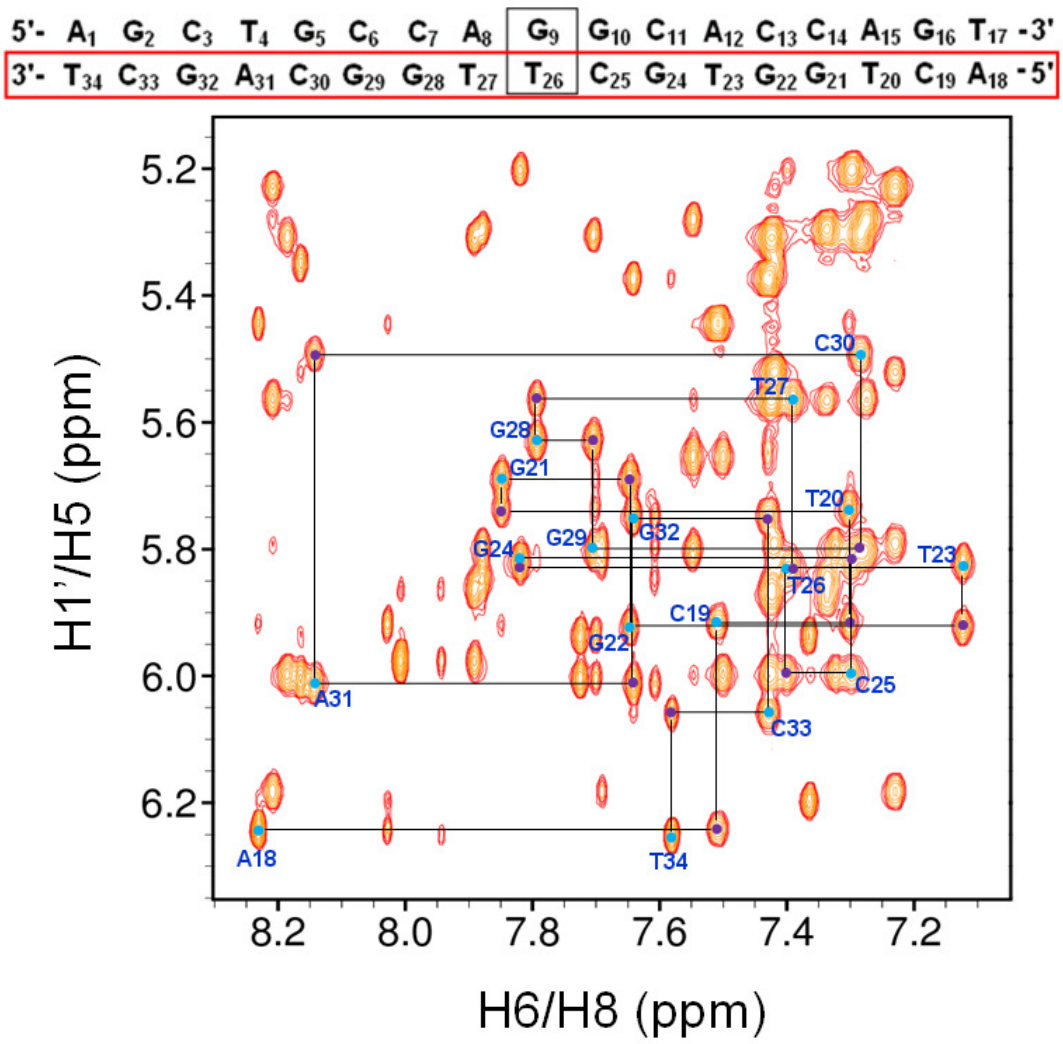
800 MHz ^1H-^{13}C HSQC spectrum of 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C showing assignment of cytosine C5 resonances.



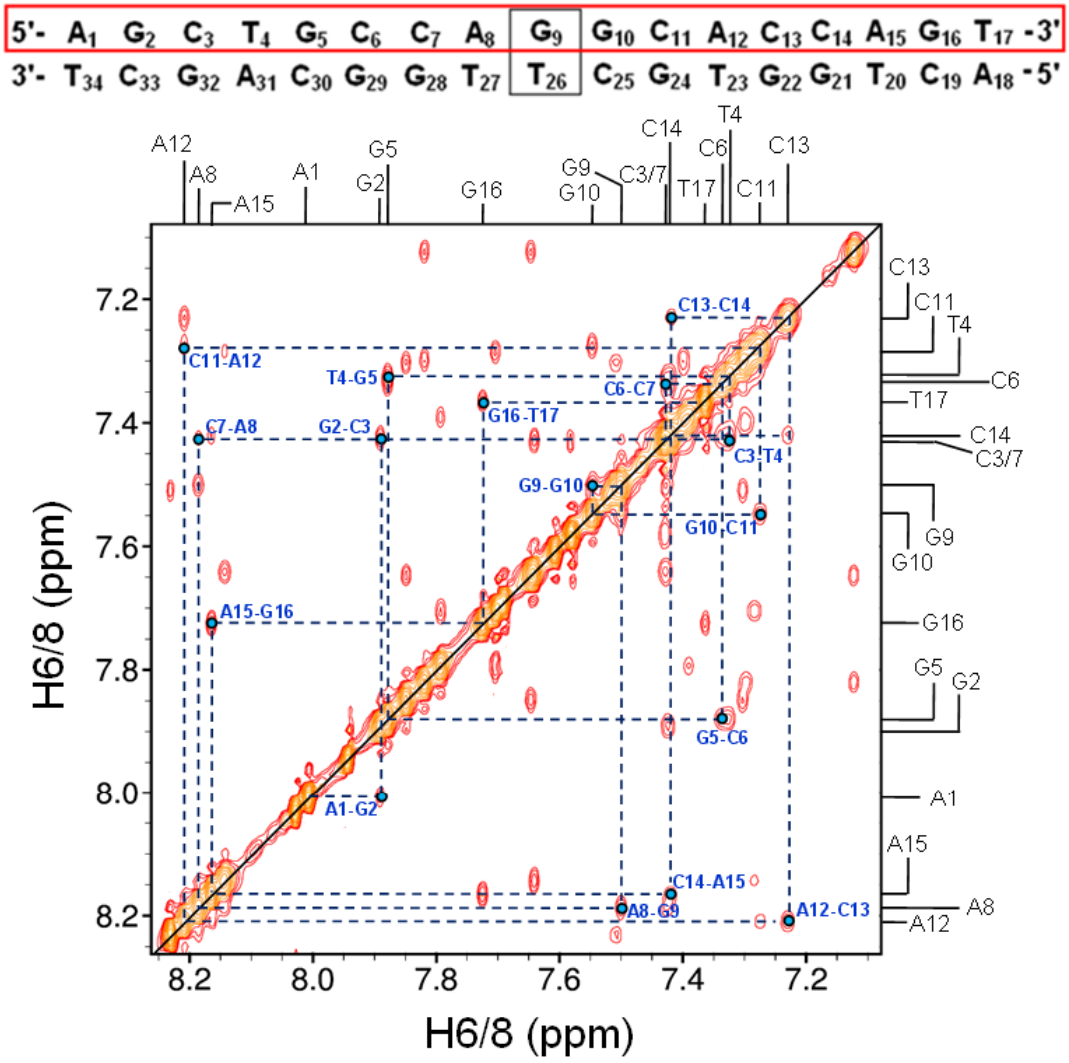
800 MHz NOESY (τ_m = 250ms) spectra showing internucleotide H6/8 (i-1)-H5/CH_3 (i) NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. Thymine and B. Cytosine connectivities. Scalar coupled connectivities are indicated by blue circles and internucleotide connectivities by purple circles.



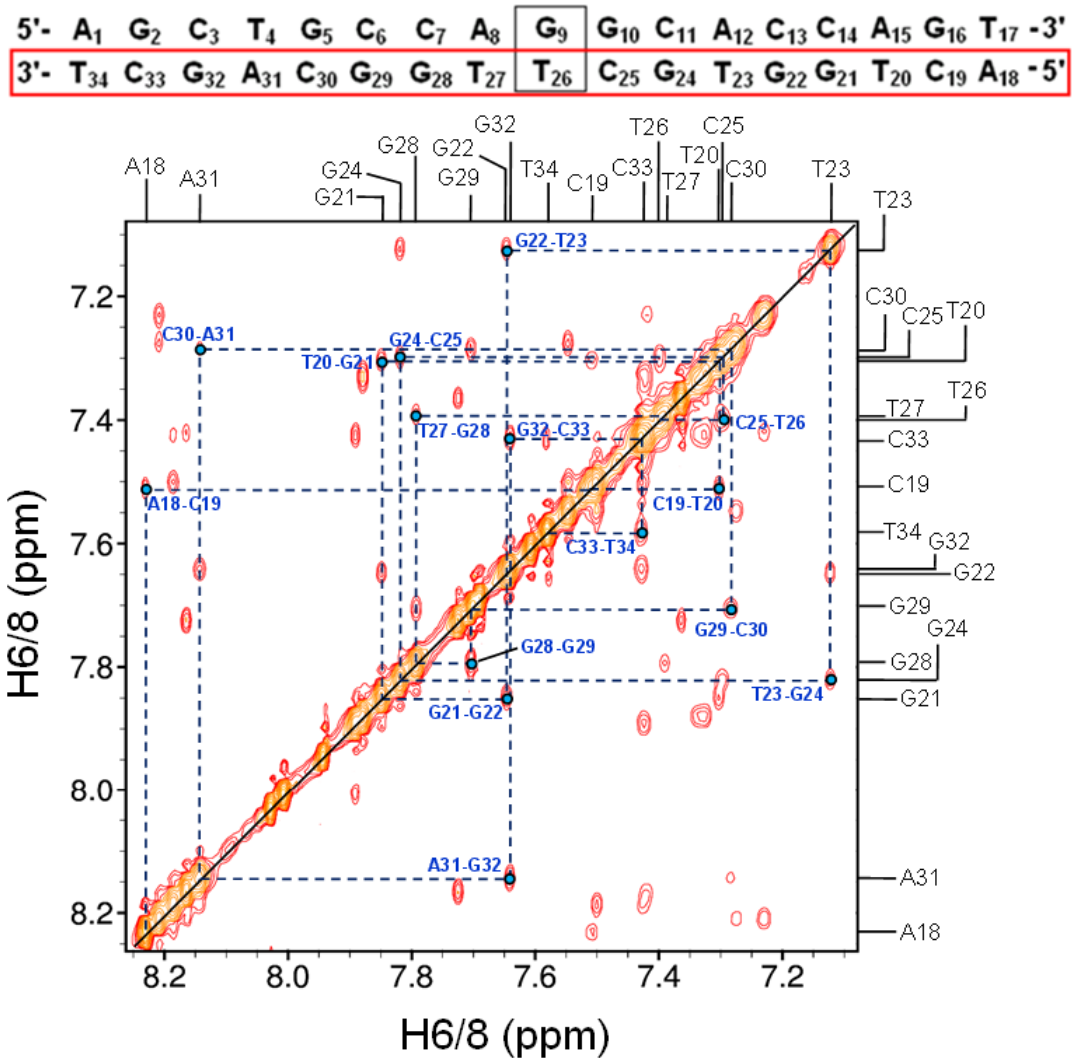
800 MHz NOESY (τ_m = 250ms) spectra showing internucleotide H6/8 (i-1)-H5/CH_3 (i) NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. Thymine and B. Cytosine connectivities. Scalar coupled connectivities are indicated by blue circles and internucleotide connectivities by purple circles.



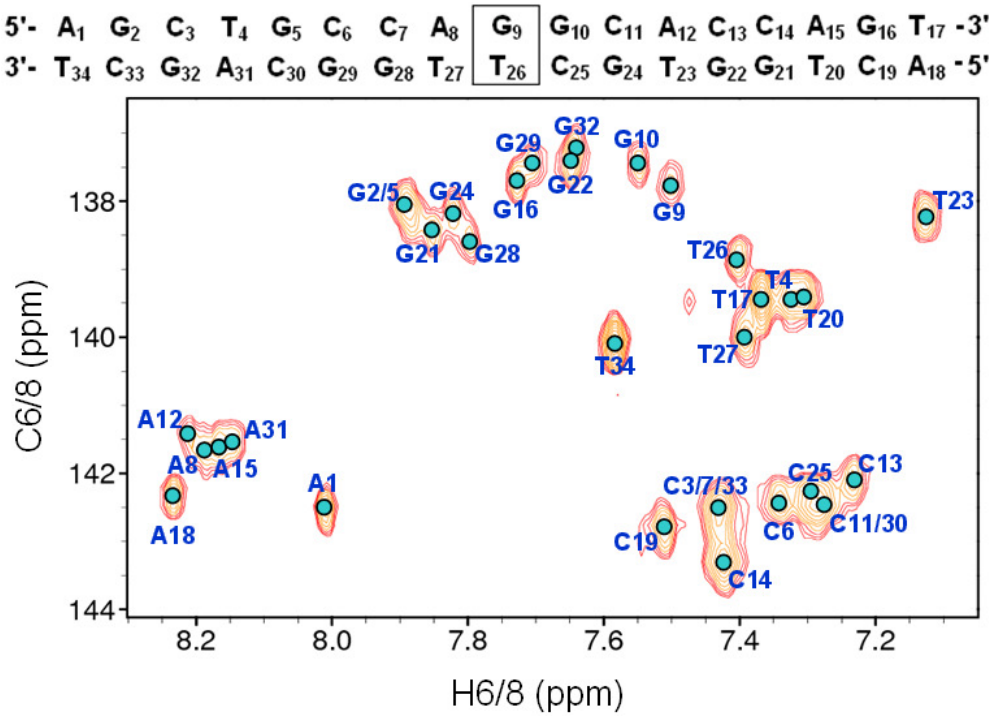
800 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H1' sequential NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide and sequential connectivities are shown by blue and purple circles respectively.



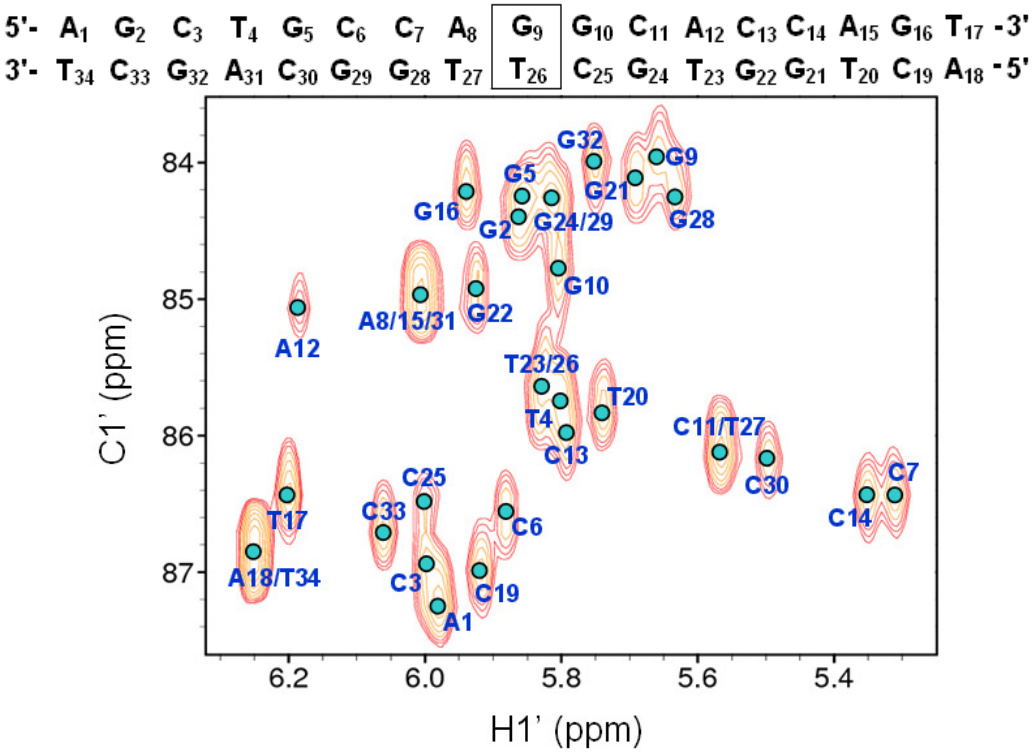
800 MHz NOESY (τ_m = 250ms) spectrum showing internucleotide H6/8-H6/8 NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



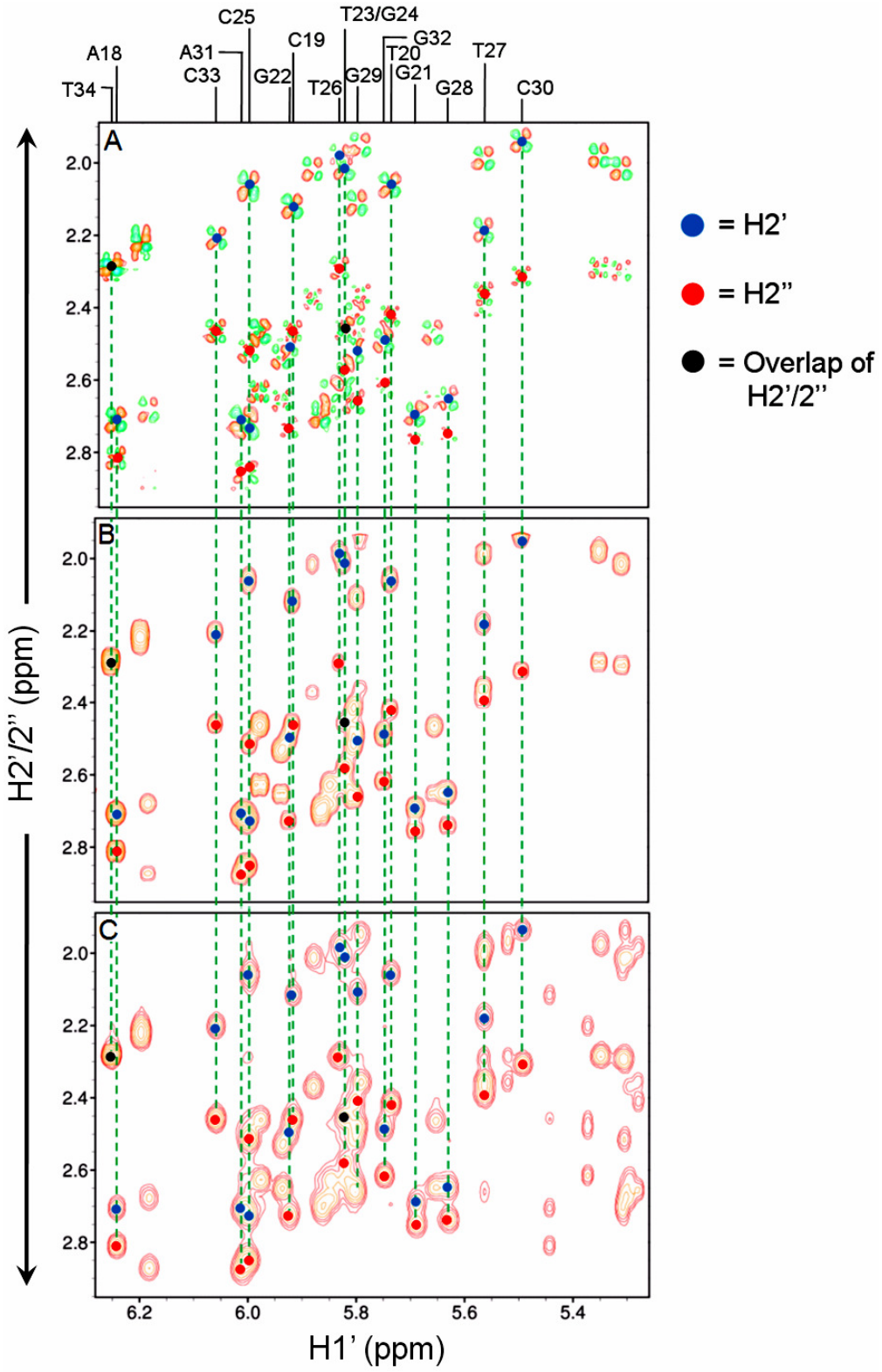
800 MHz NOESY (τ_m = 250ms) spectrum showing internucleotide H6/8-H6/8 NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



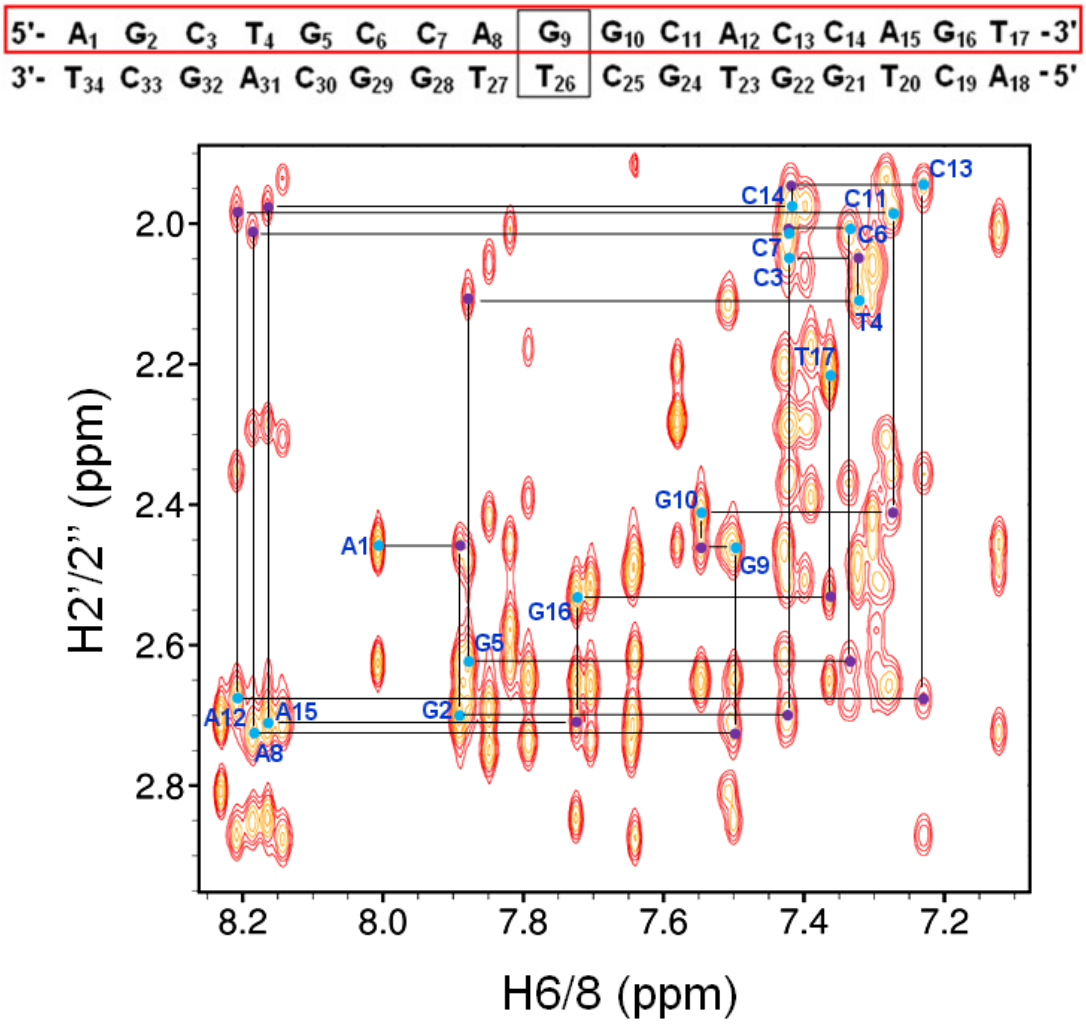
800 MHz ^{1}H-^{13}C HSQC spectrum showing the assignment of C6/C8 resonances in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



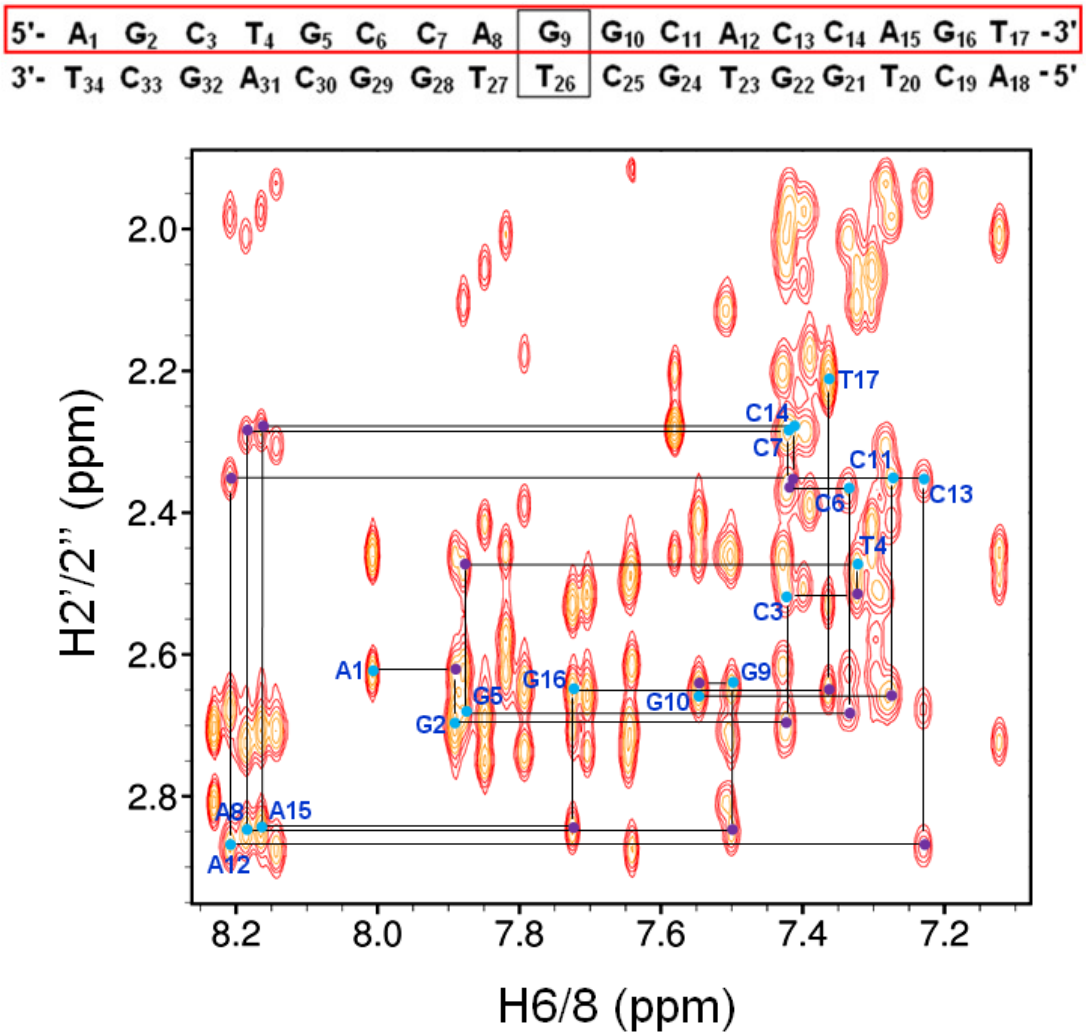
800 MHz ^{1}H-^{13}C HSQC spectrum showing the assignment of C1' resonances in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



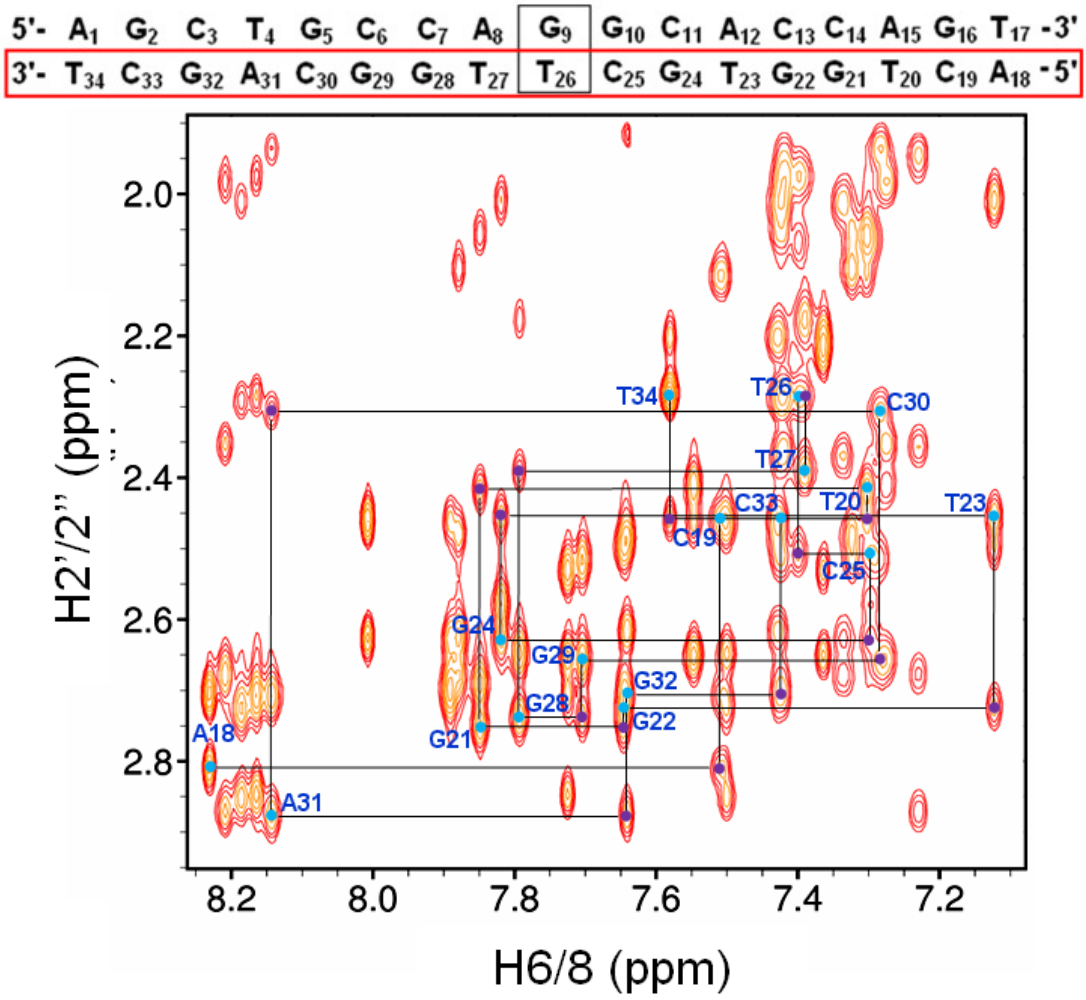
800 MHz spectrum stack showing identification and assignment of H2'/H2'' resonances for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. DQF-COSY, B. TOCSY (SL_{mix} = 75ms) and C. NOESY (τ_m = 250ms).



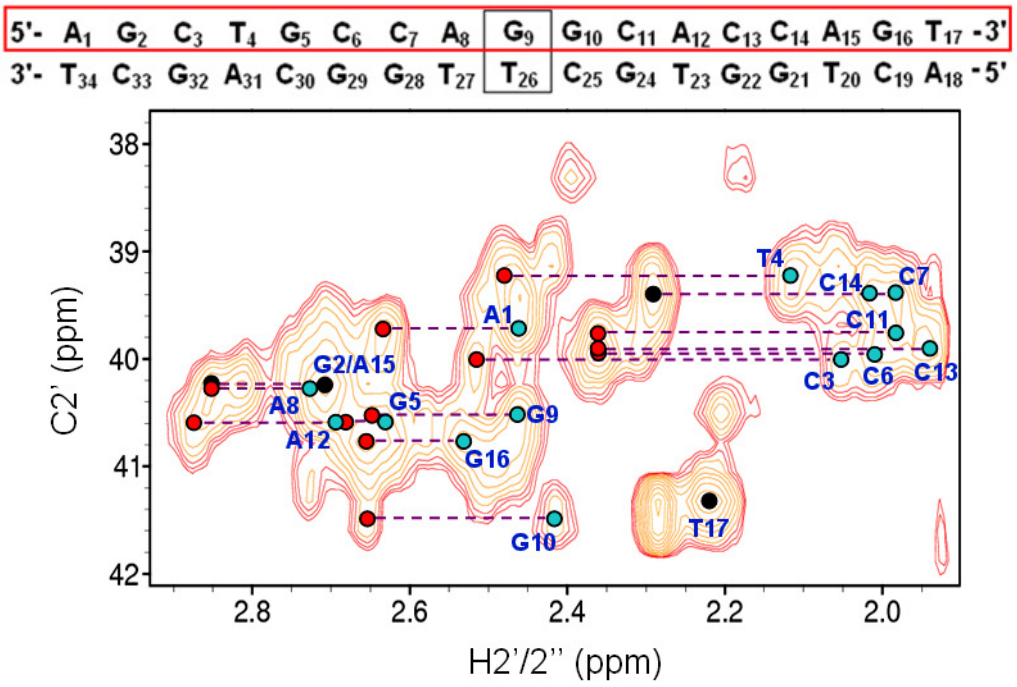
800 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2' sequential NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide and sequential connectivities are shown by blue and purple circles respectively.



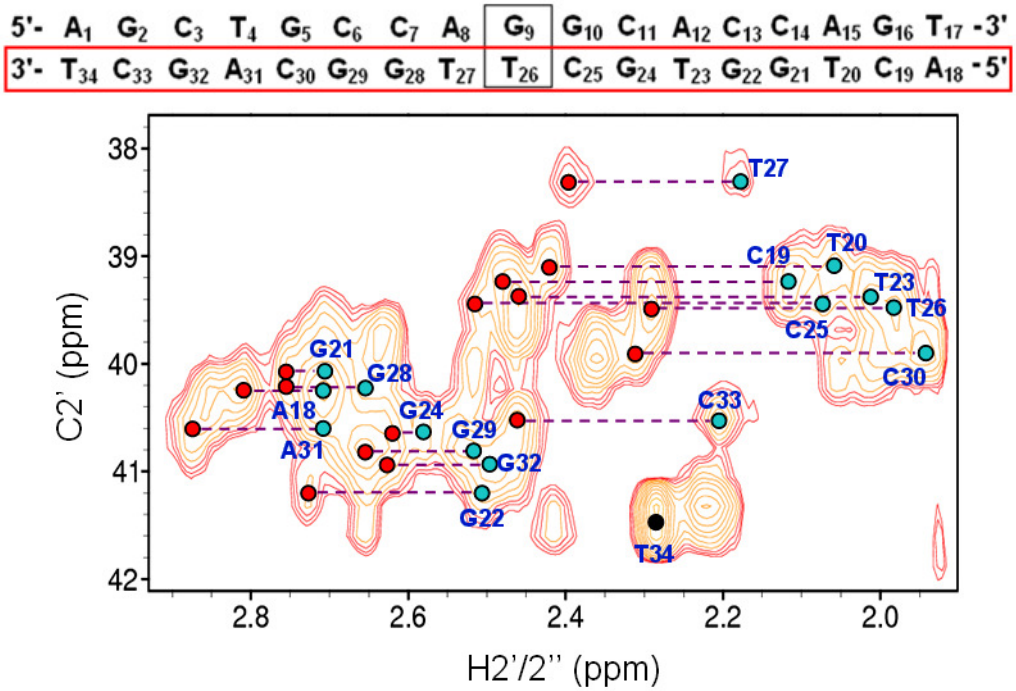
800 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2'' sequential NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide and sequential connectivities are shown by blue and purple circles respectively.



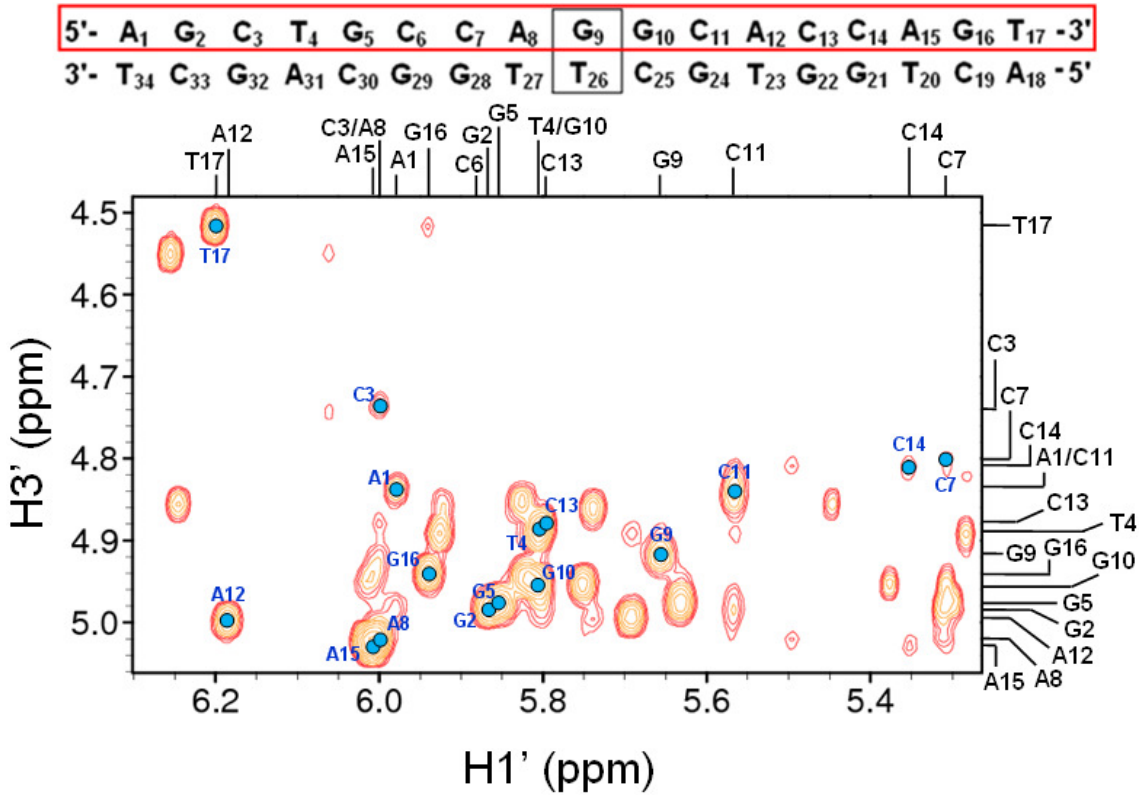
800 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2'' sequential NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide and sequential connectivities are shown by blue and purple circles respectively.



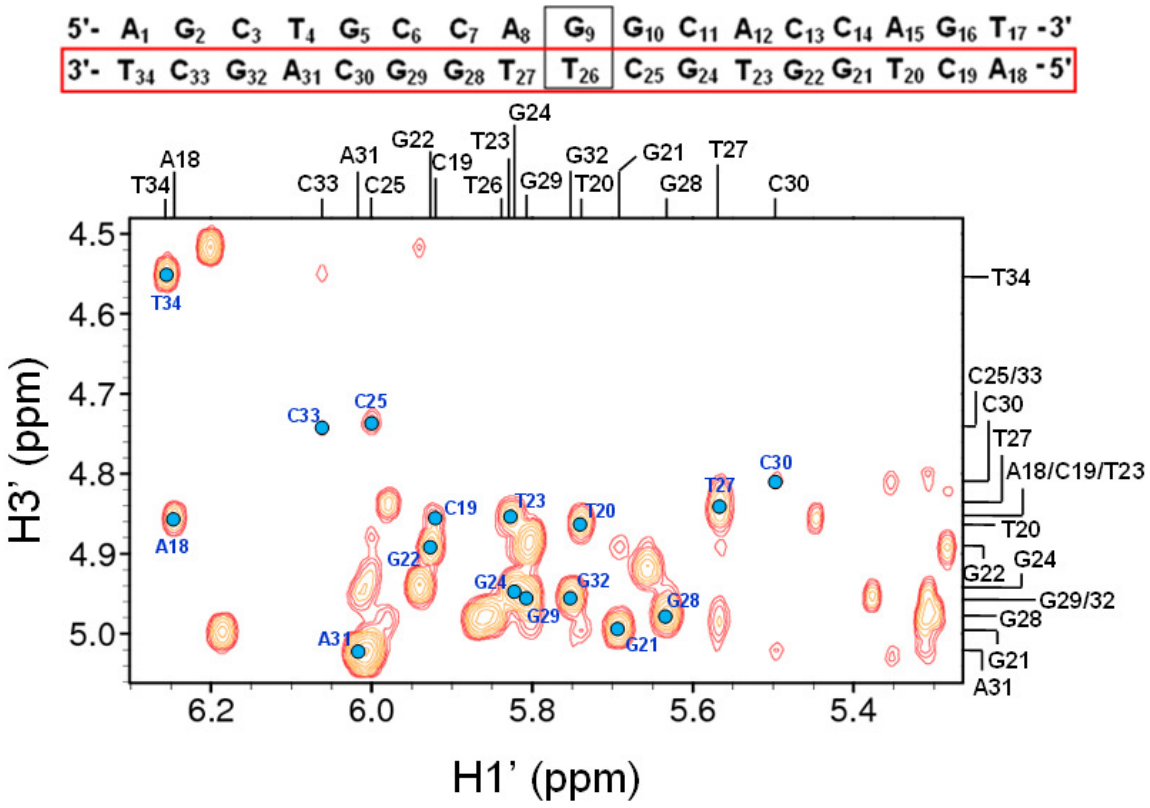
800 MHz ^{1}H-^{13}C HSQC spectrum showing the assignment of C2' resonances for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. H2' and H2'' resonances are shown by blue and red circles respectively; black circles indicate an overlap of resonances.



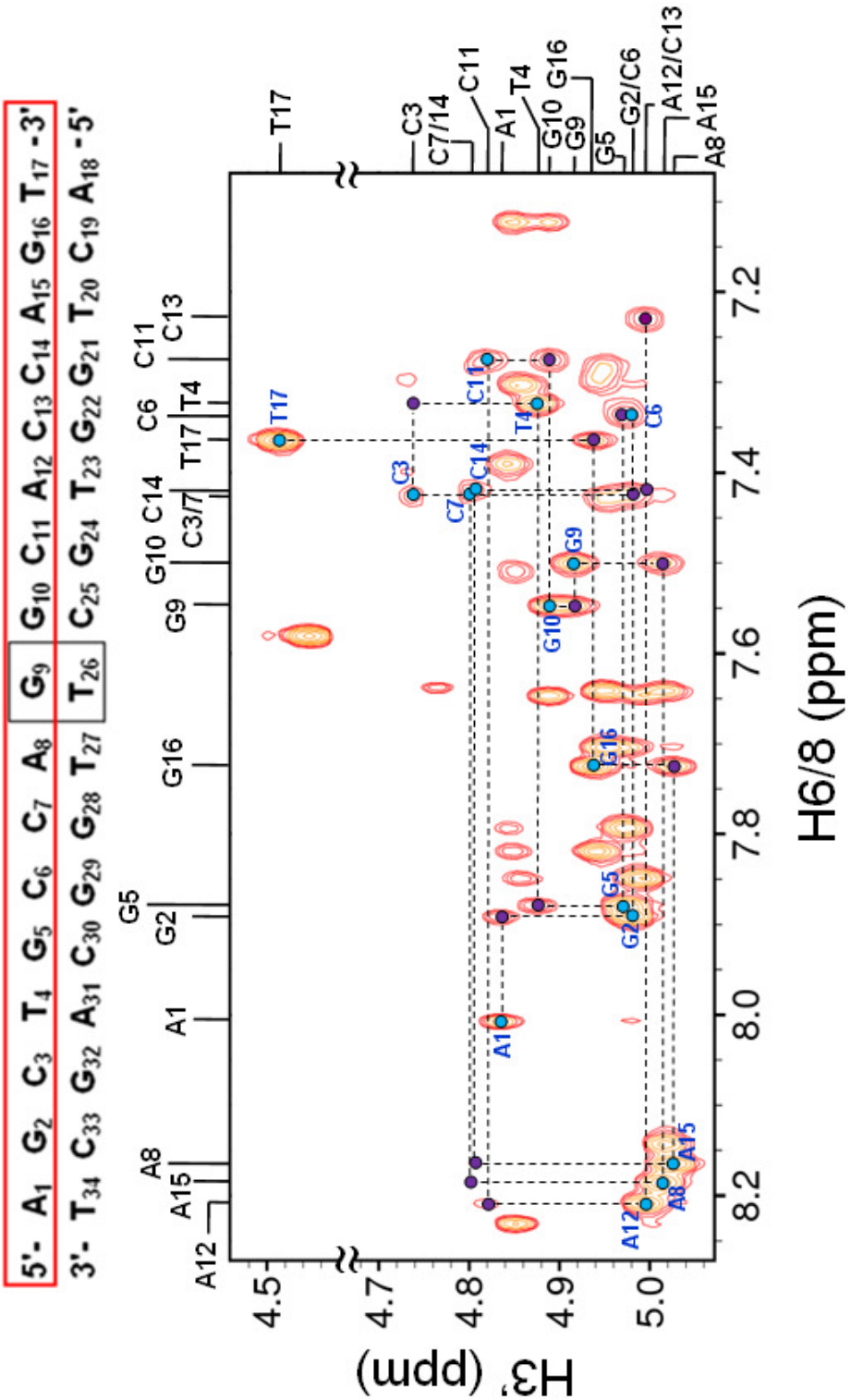
800 MHz ^{1}H-^{13}C HSQC spectrum showing the assignment of C2' resonances for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. H2' and H2'' resonances are shown by blue and red circles respectively; black circles indicate an overlap of resonances.



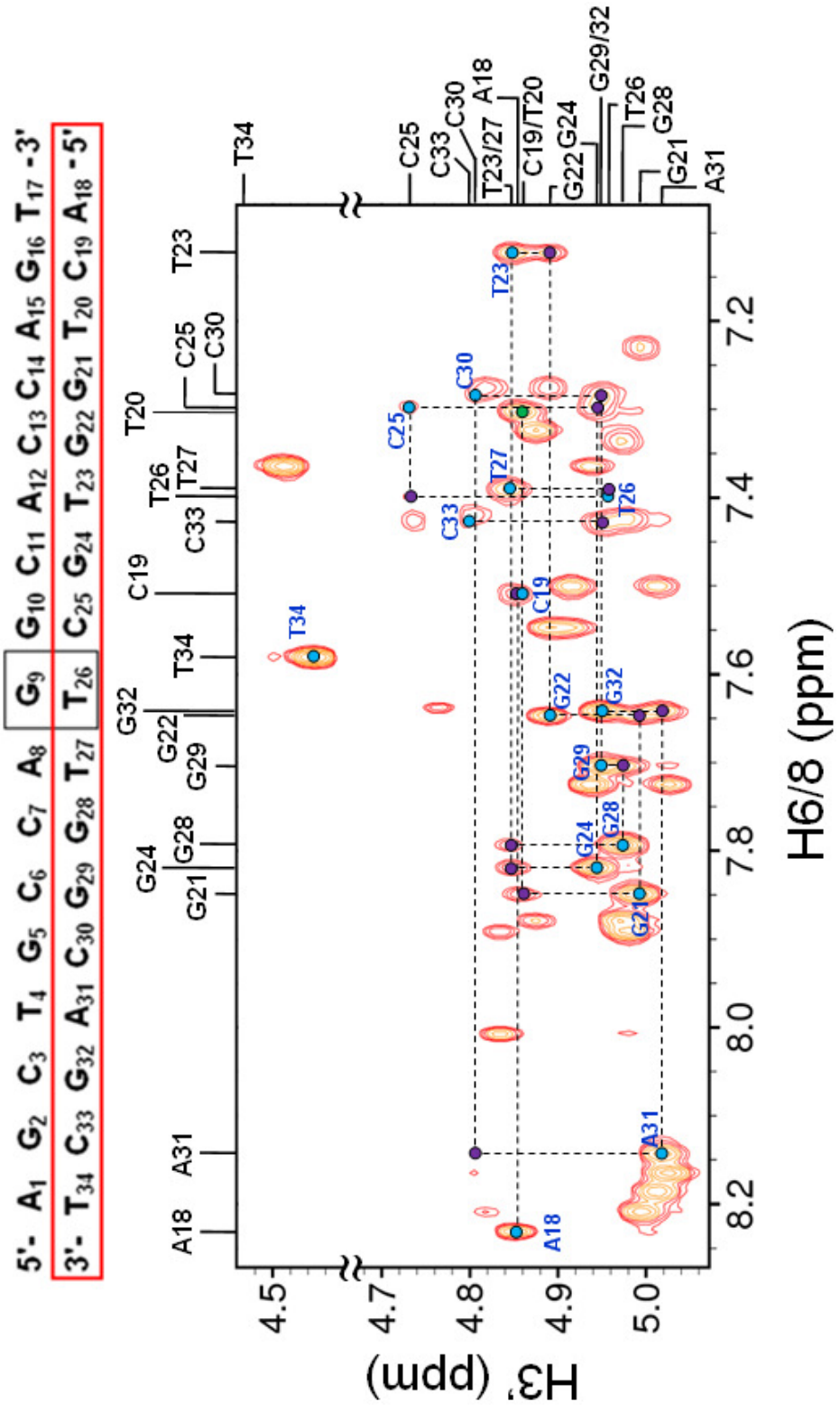
800 MHz NOESY (τ_{mix} = 250ms) spectrum showing H1'-H3' NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



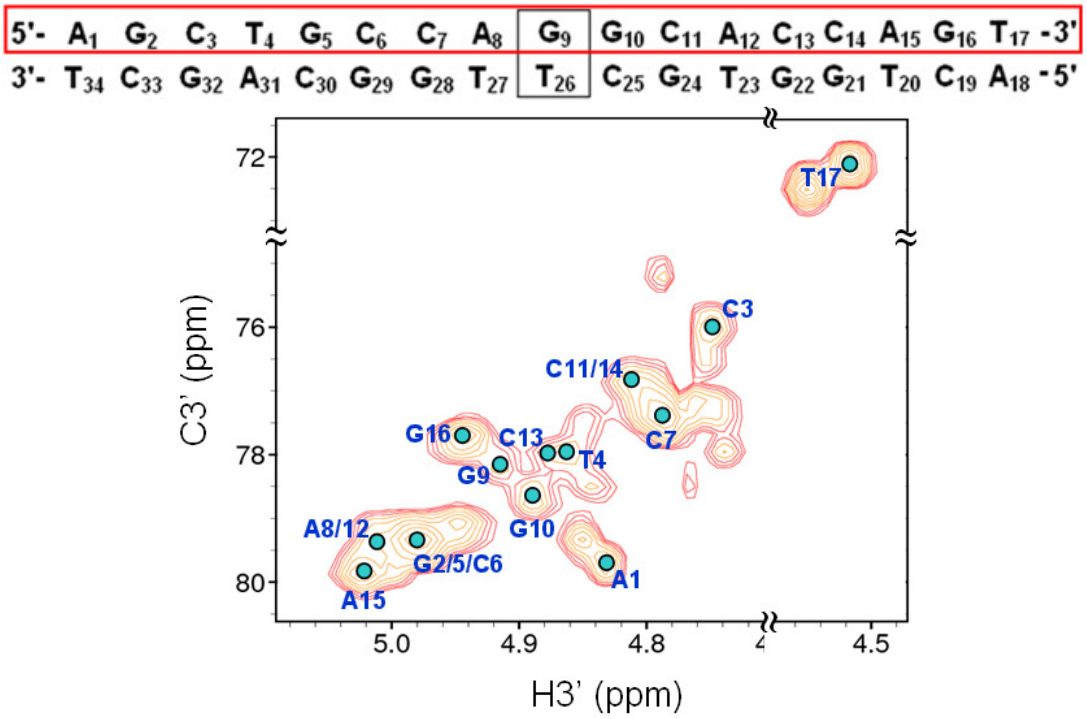
800 MHz NOESY (τ_{mix} = 250ms) spectrum showing H1'-H3' NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



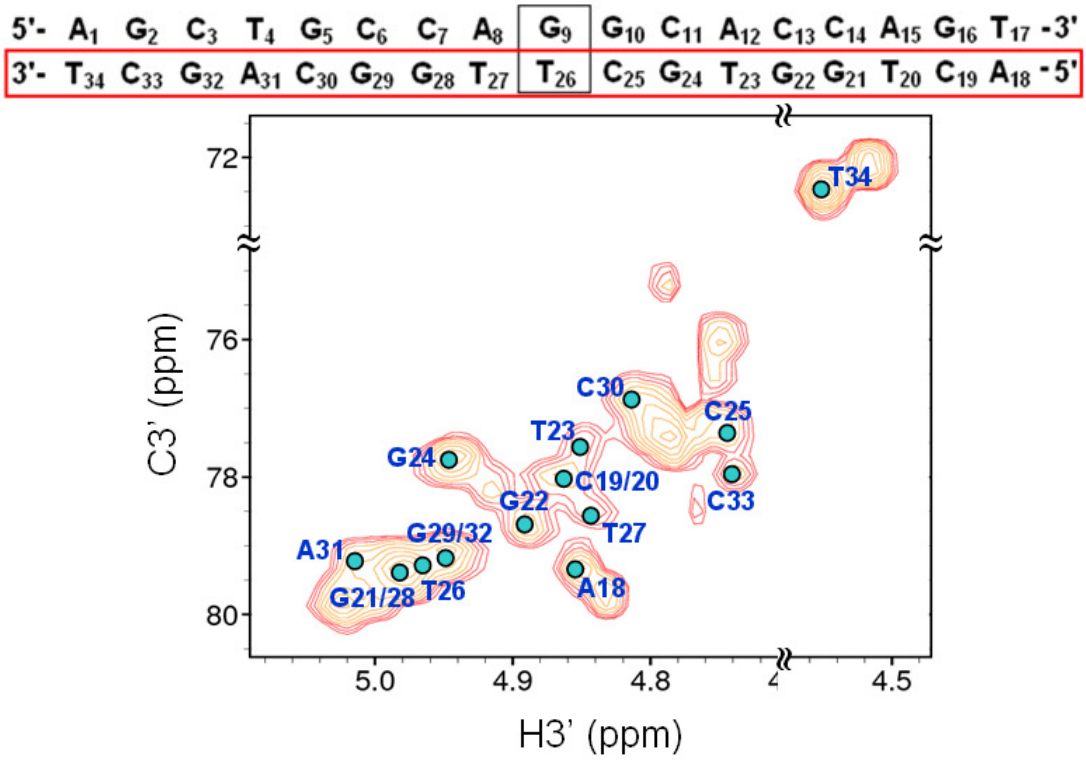
800 MHz NOESY (τ_{mix} = 250ms) spectrum showing H6/8-H3' NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide and sequential NOEs are shown by blue and purple circles respectively.



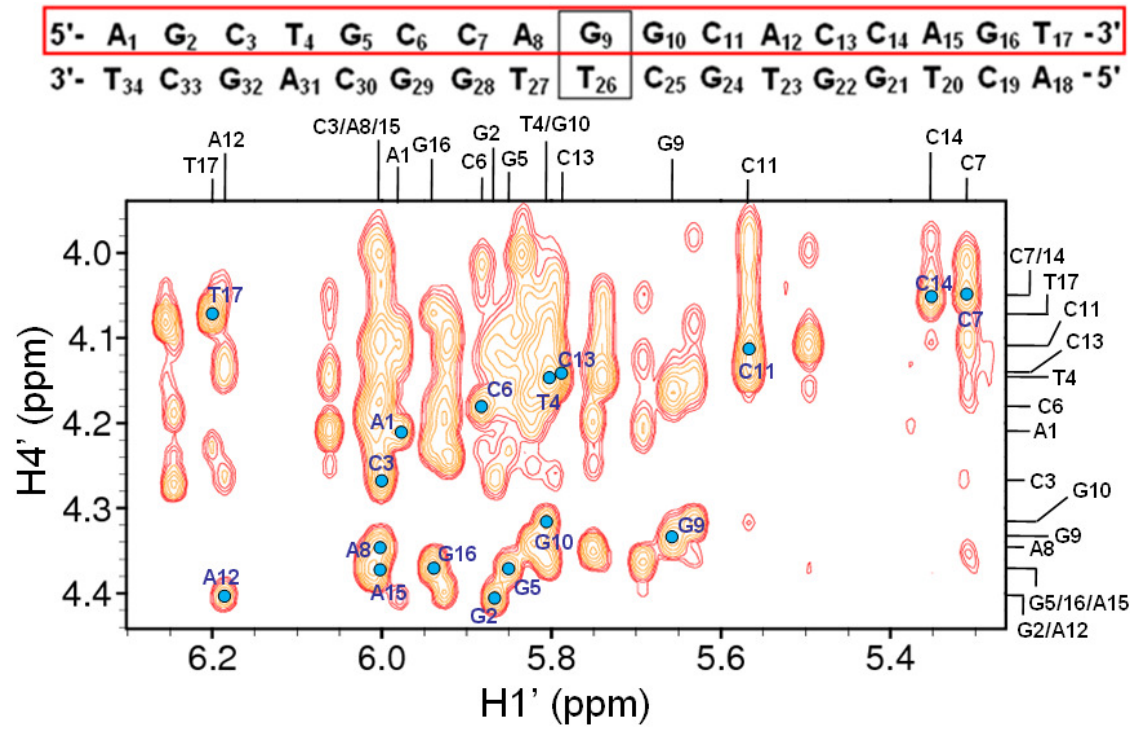
800 MHz NOESY (τ_{mix} = 250ms) spectrum showing H6/8-H3' NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide and sequential NOEs are shown by blue and purple circles respectively.



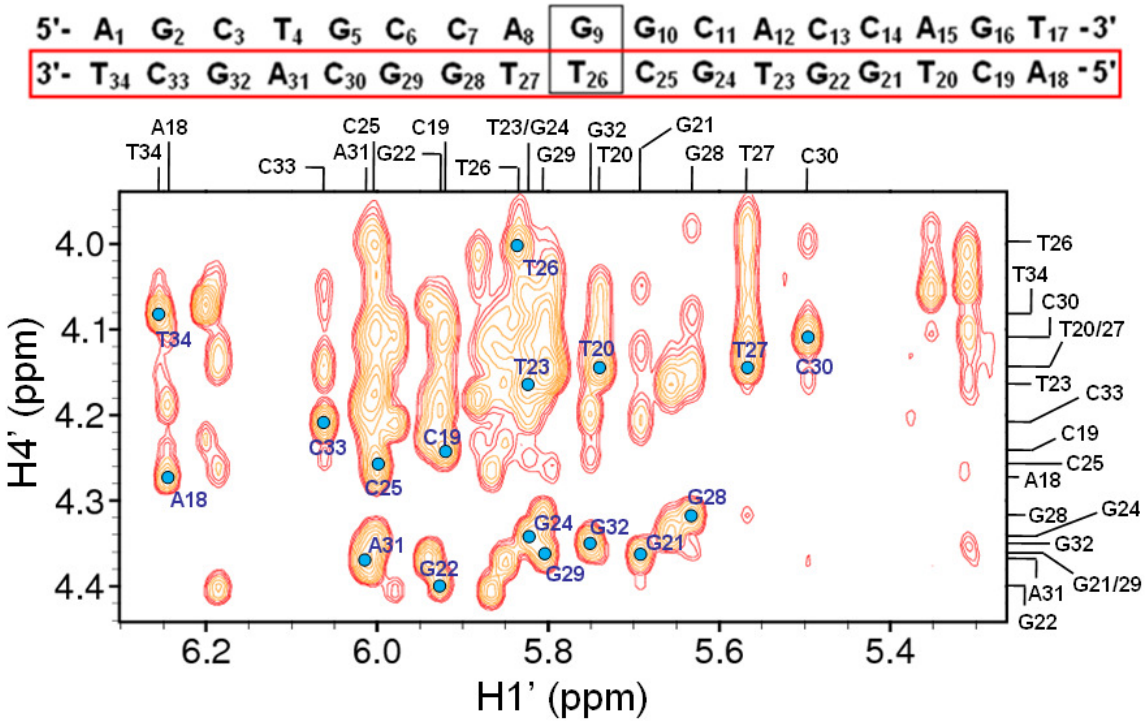
800 MHz ^{1}H-^{13}C HSQC spectrum showing the assignment of C3' resonances for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



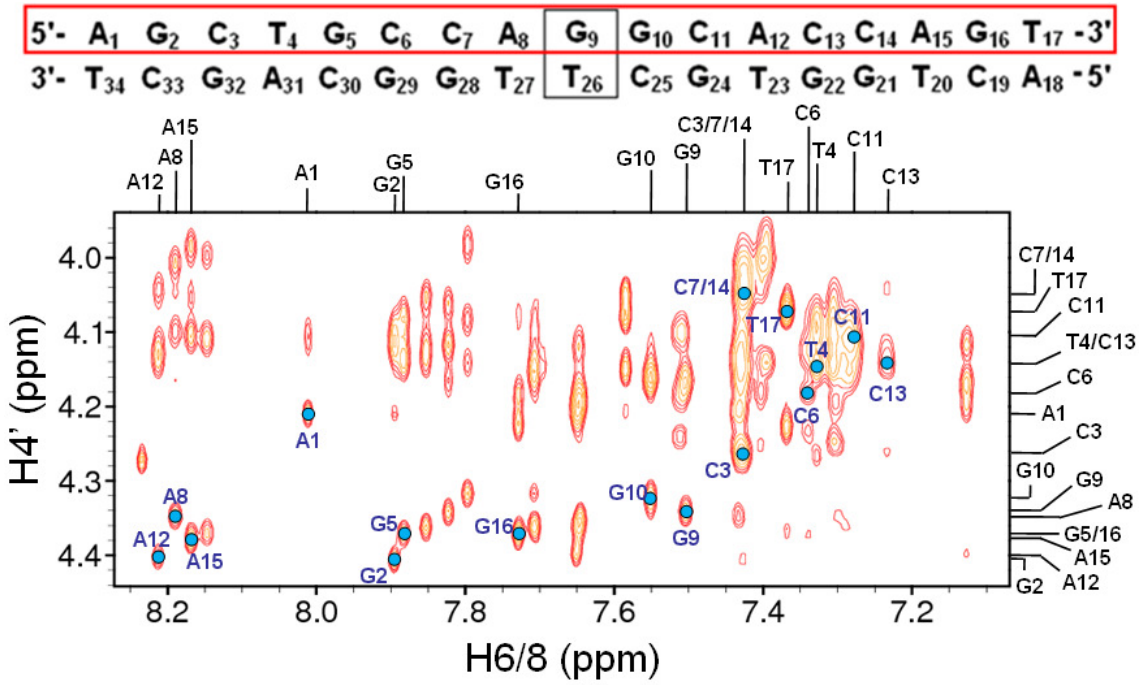
800 MHz ^{1}H-^{13}C HSQC spectrum showing the assignment of C3' resonances for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



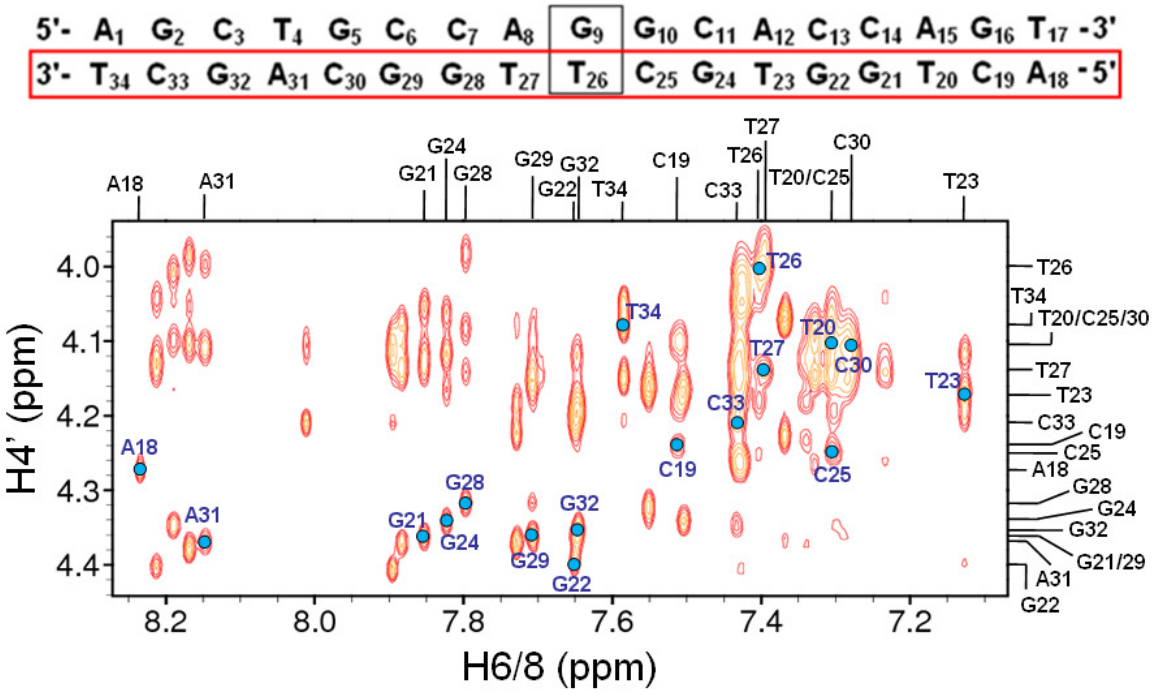
800 MHz NOESY (τ_{mix} = 250ms) spectrum showing H1'-H4' NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



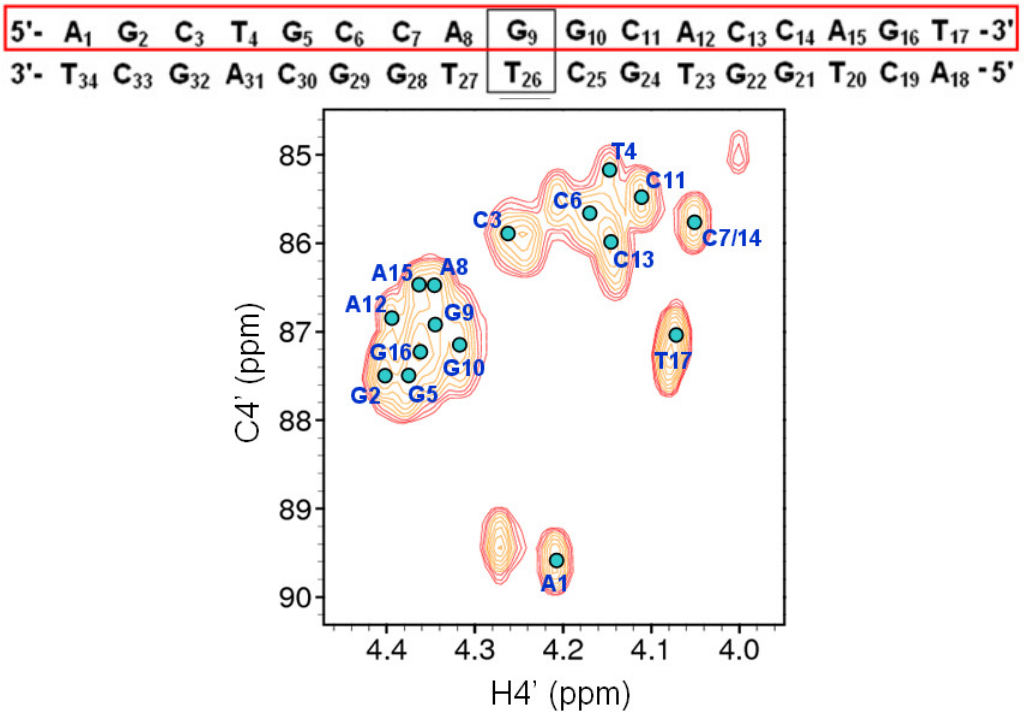
800 MHz NOESY (τ_{mix} = 250ms) spectrum showing H1'-H4' NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



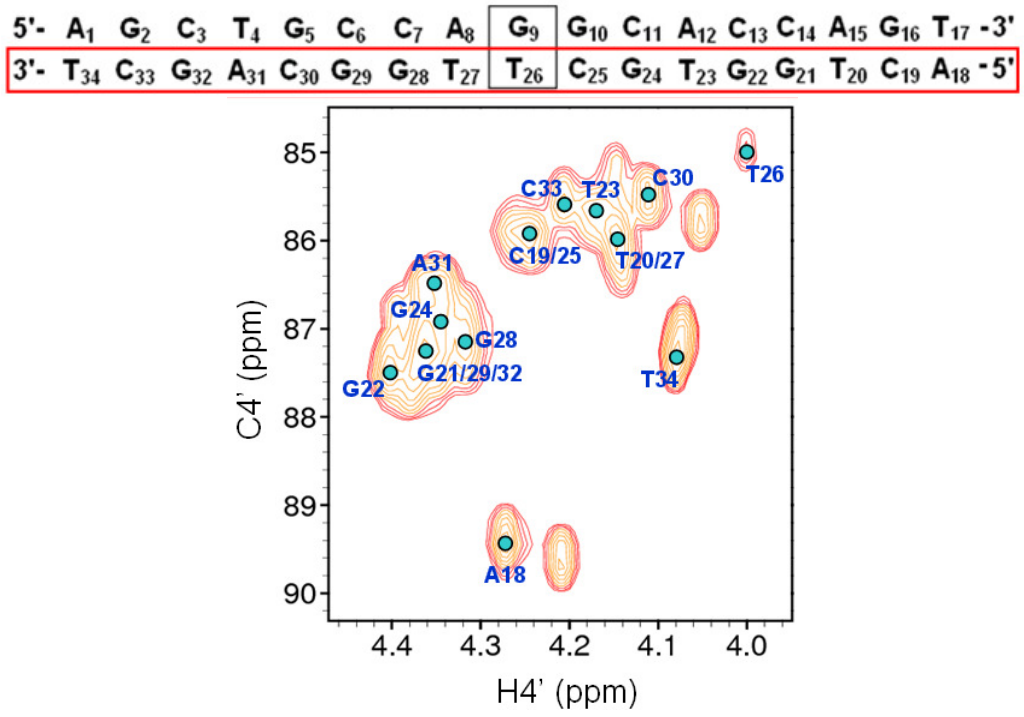
800 MHz NOESY (τ_{mix} = 250ms) spectrum showing H6/8-H4' NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



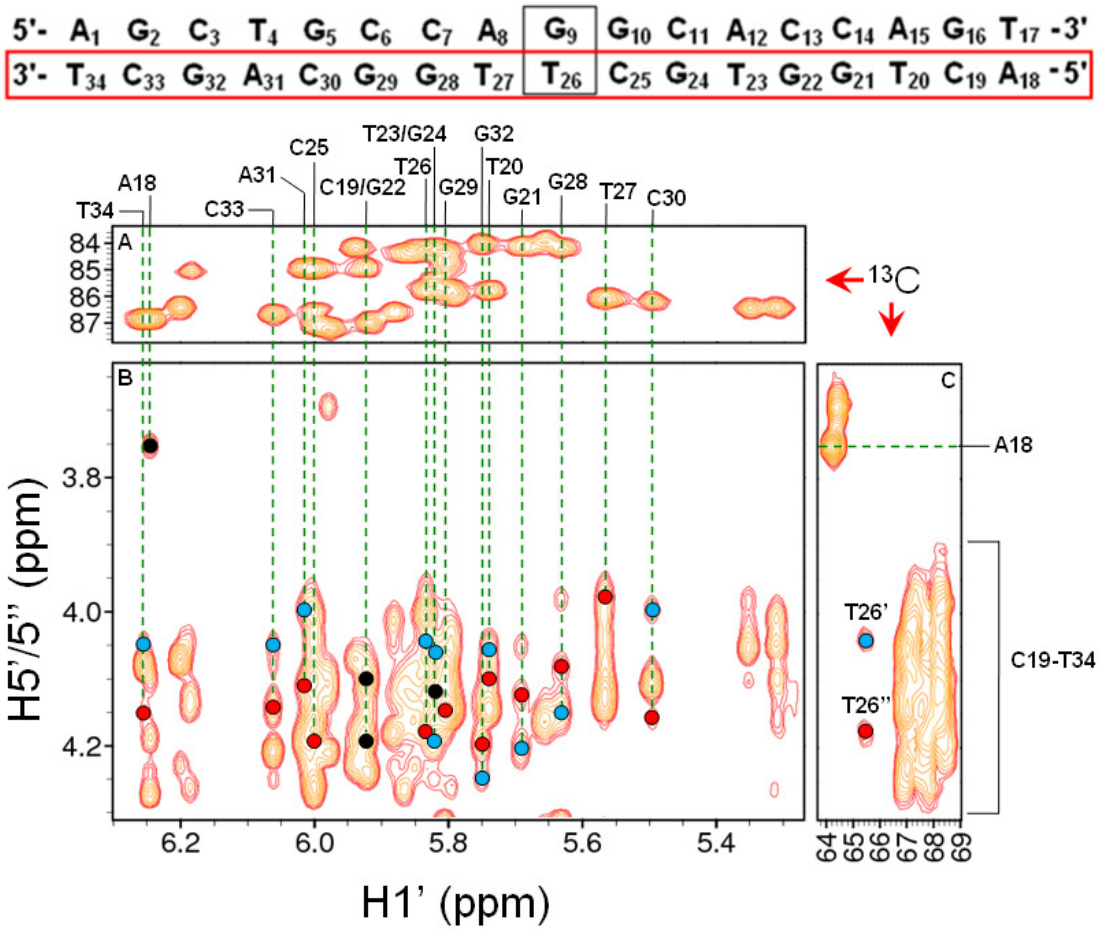
800 MHz NOESY (τ_{mix} = 250ms) spectrum showing H6/8-H4' NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



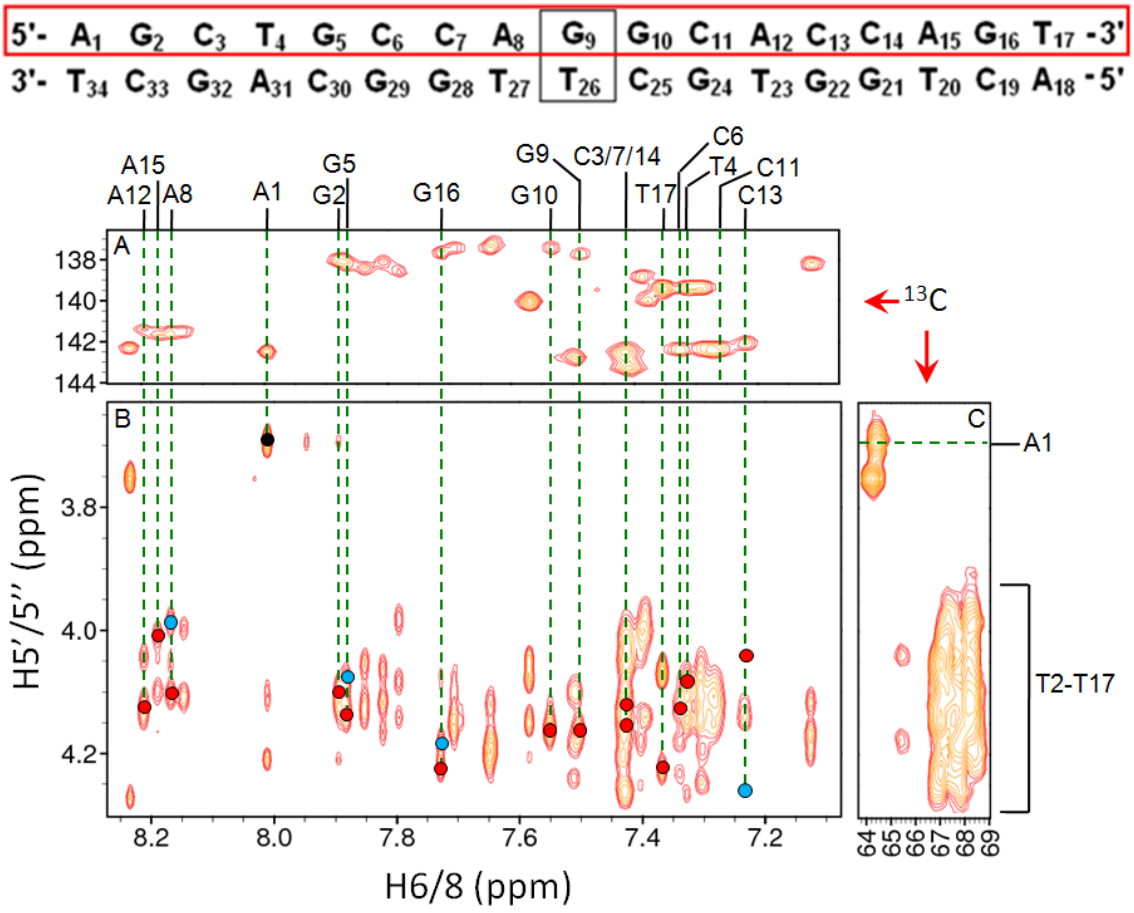
800 MHz ^{1}H-^{13}C HSQC spectrum showing the assignment of C4' resonances for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



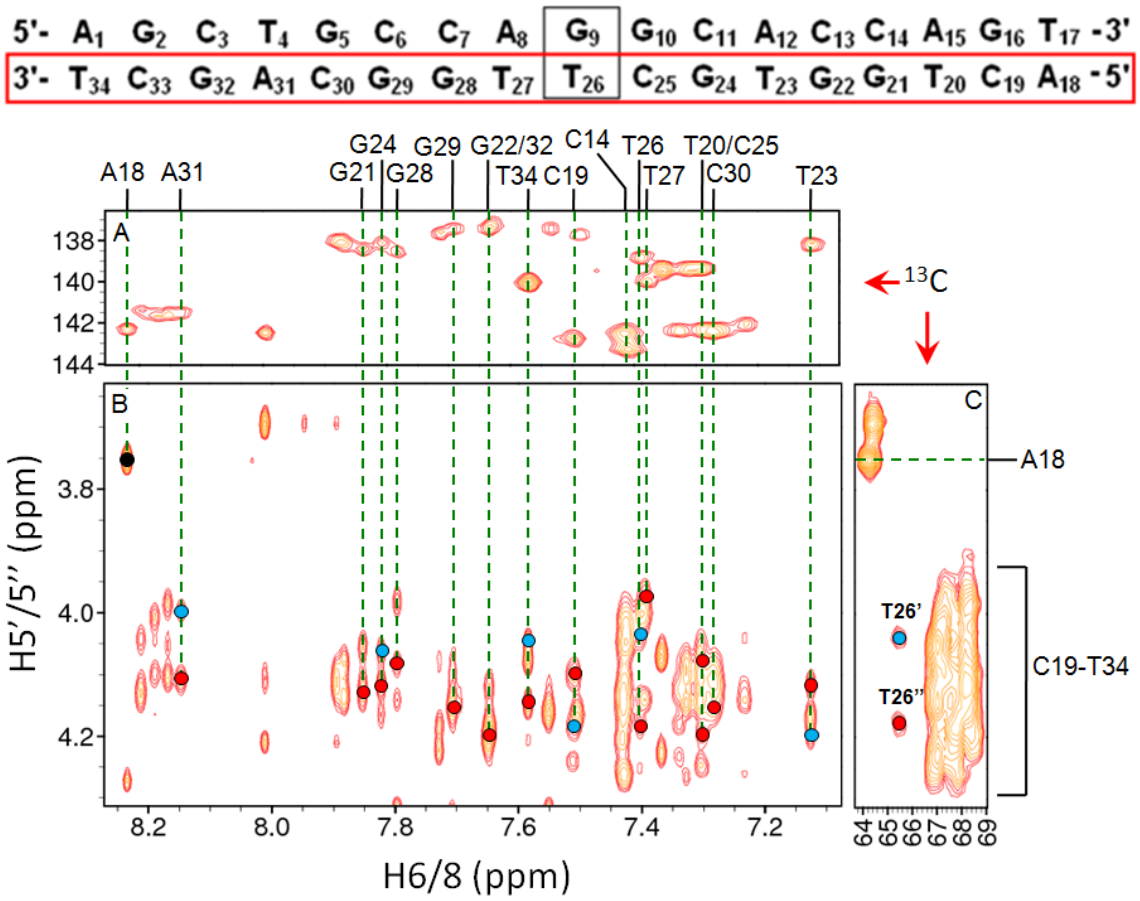
800 MHz ^{1}H-^{13}C HSQC spectrum showing the assignment of C4' resonances for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



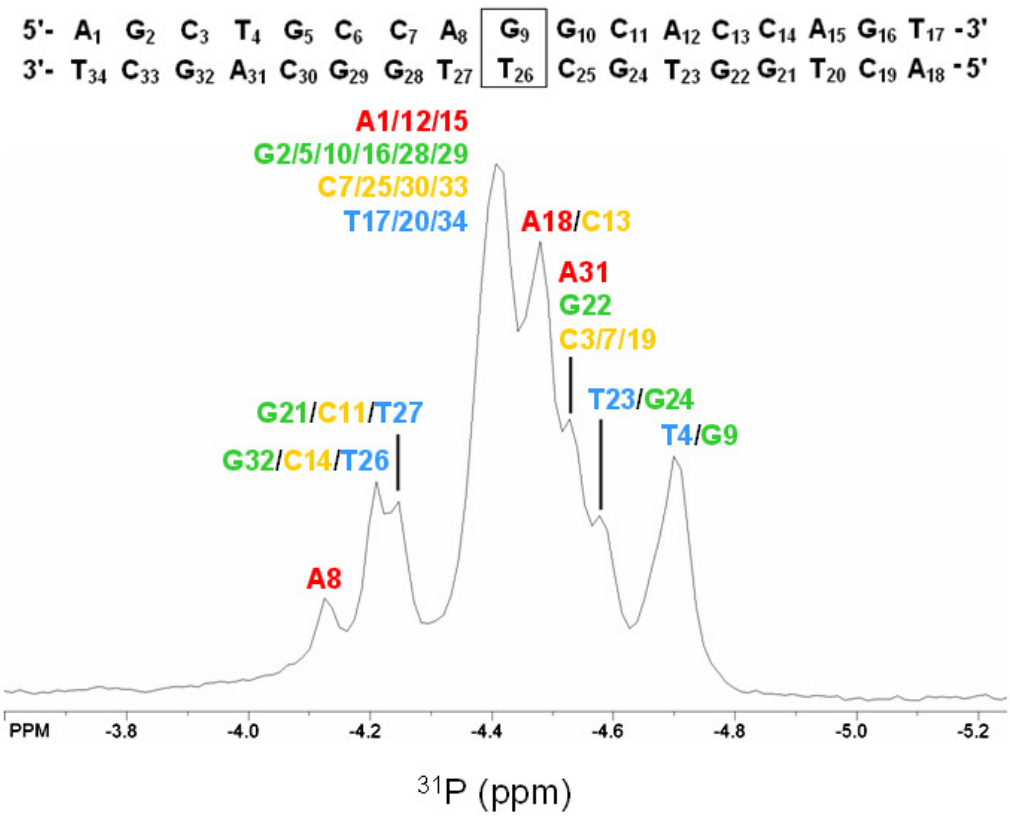
800 MHz NOESY (τ_{mix} = 250ms) spectrum (B) showing H1'-H5'/5'' NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. 800 MHz ^{1}H-^{13}C HSQC panels show the corresponding assignment of C1' (A) and C5' (C) resonances. The C5' resonances of residues C19-T34 are highly degenerate and appear in the heavily overlapped region indicated in the figure, with the exception of T26.



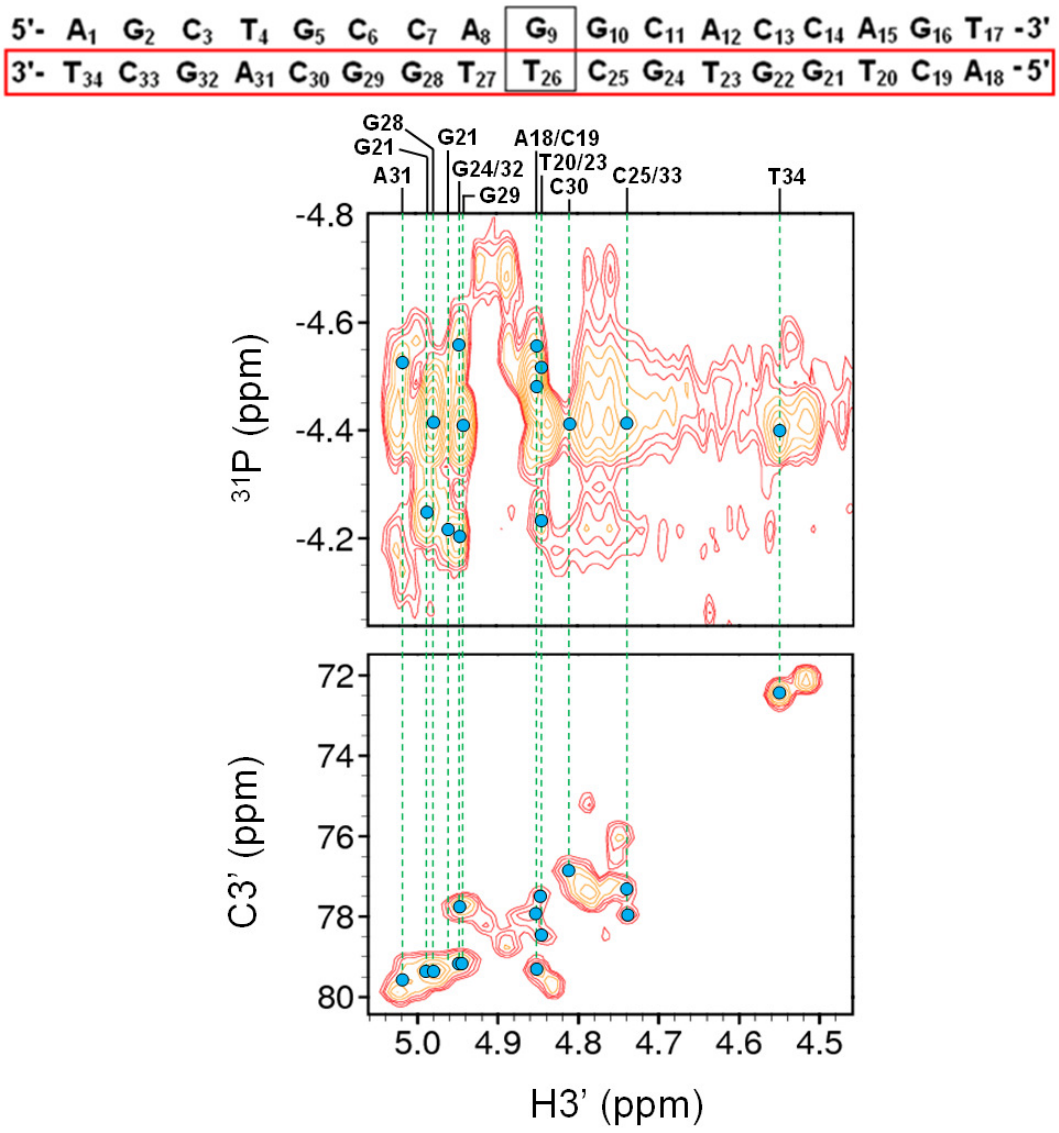
800 MHz NOESY (τ_{mix} = 250ms) spectrum (B) showing H6/8-H5'/5'' NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. 800 MHz ^{1}H-^{13}C-HSQC panels show the corresponding assignment of C1' (A) and C5' (C) resonances. The C5' resonances of residues T2-T17 are highly degenerate and appear in the heavily overlapped region indicated in the figure.



800 MHz NOESY (τ_{mix} = 250ms) spectrum (B) showing the assignment of H6/8-H5'/5'' NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. 800 MHz ^{1}H-^{13}C-HSQC panels show the corresponding assignment of C1' (A) and C5' (C) resonances. The C5' resonances of residues C19-T34 are highly degenerate and appear in the heavily overlapped region indicated in the figure, with the exception of T26.



81 MHz ^{1}H decoupled 1D ^{31}P-NMR spectrum showing the assignment of ^{31}P resonances in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.

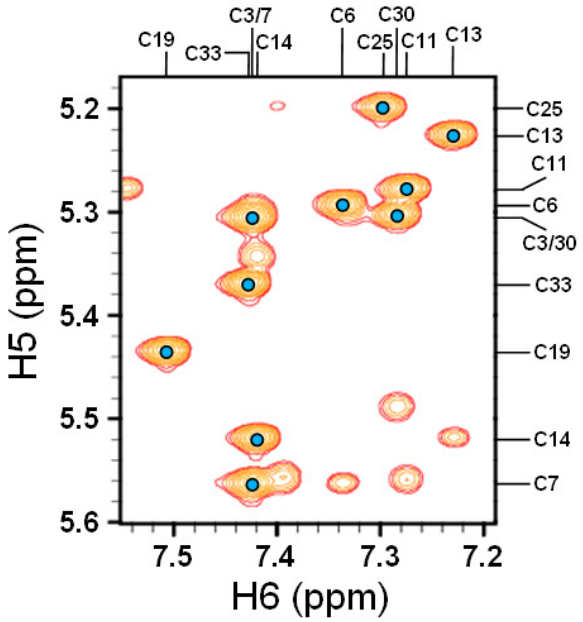


(Top panel) 600 MHz ^{1}H-^{31}P CPMG-HSQC-NOESY (τ_{mix} = 500ms) spectrum and (bottom panel) 800 MHz ^{1}H-^{13}C HSQC spectrum. The assignment of ^{31}P resonances is shown by means of H3'-^{31}P HSQC correlations for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.

iv. 17mer mismatch GT DNA (1000 MHz)

5'- A_1 G_2 C_3 T_4 G_5 C_6 C_7 A_8 G_9 G_{10} C_{11} A_{12} C_{13} C_{14} A_{15} G_{16} T_{17} -3'

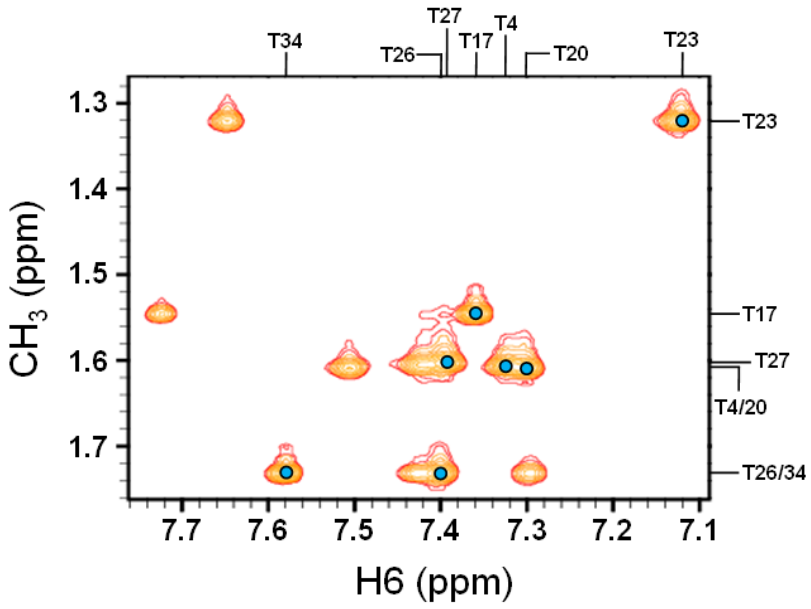
3'- T_{34} C_{33} G_{32} A_{31} C_{30} G_{29} G_{28} T_{27} T_{26} C_{25} G_{24} T_{23} G_{22} G_{21} T_{20} C_{19} A_{18} -5'



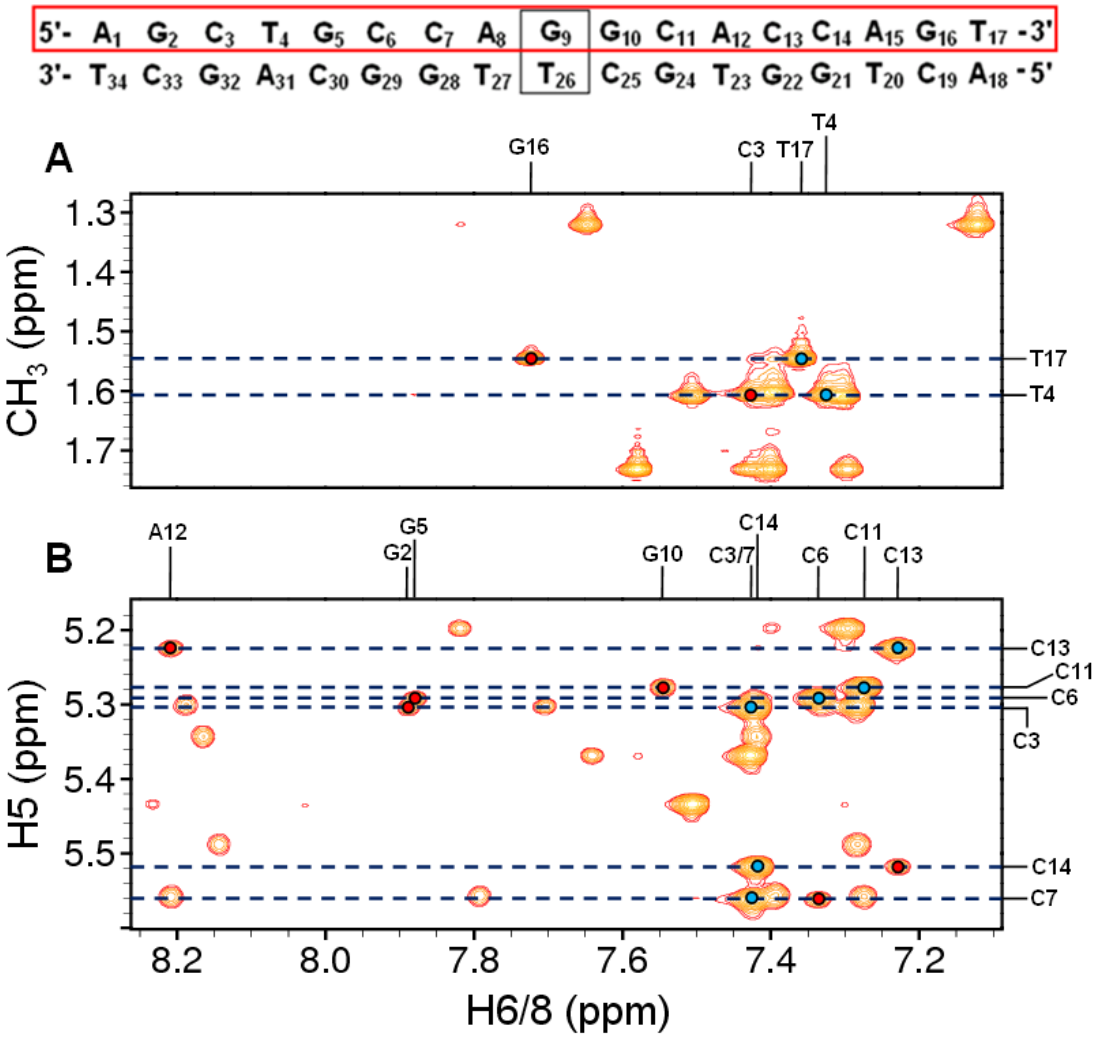
1000 MHz NOESY (τ_m = 250ms) spectrum showing cytosine H5-H6 NOE connectivities in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.

5'- A_1 G_2 C_3 T_4 G_5 C_6 C_7 A_8 G_9 G_{10} C_{11} A_{12} C_{13} C_{14} A_{15} G_{16} T_{17} -3'

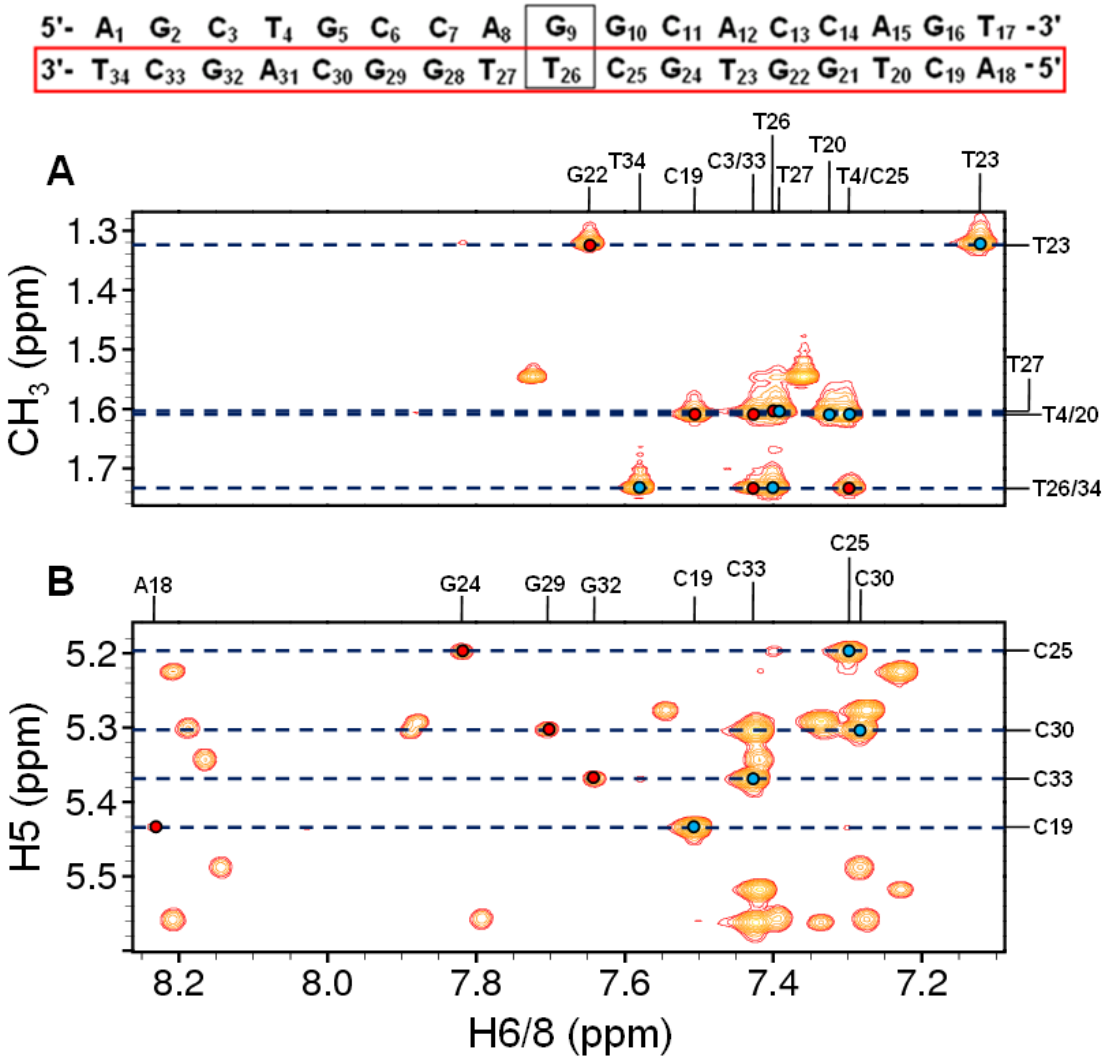
3'- T_{34} C_{33} G_{32} A_{31} C_{30} G_{29} G_{28} T_{27} T_{26} C_{25} G_{24} T_{23} G_{22} G_{21} T_{20} C_{19} A_{18} -5'



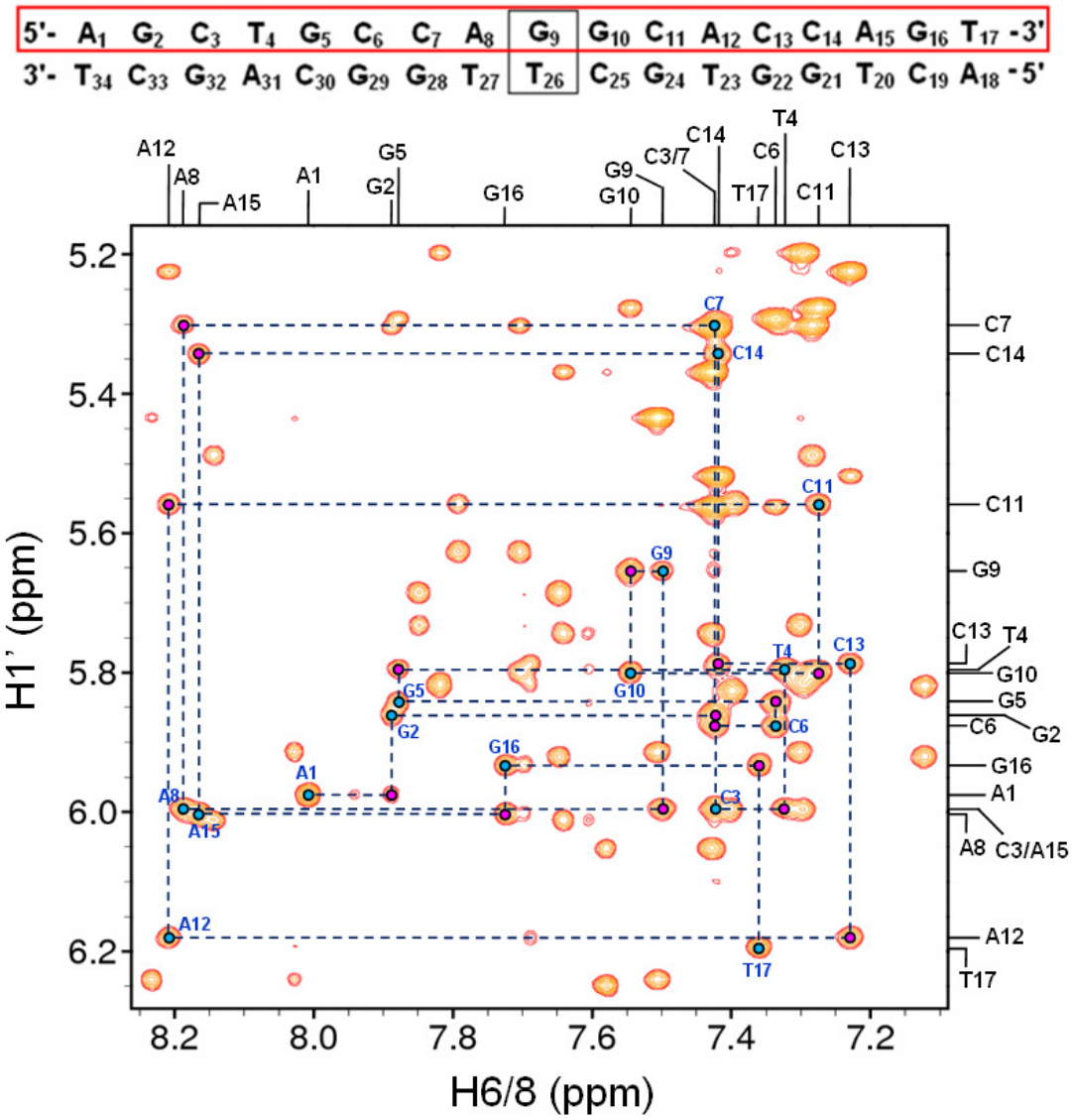
1000 MHz NOESY (τ_m = 250ms) spectrum showing thymine CH_3-H6 NOE connectivities in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



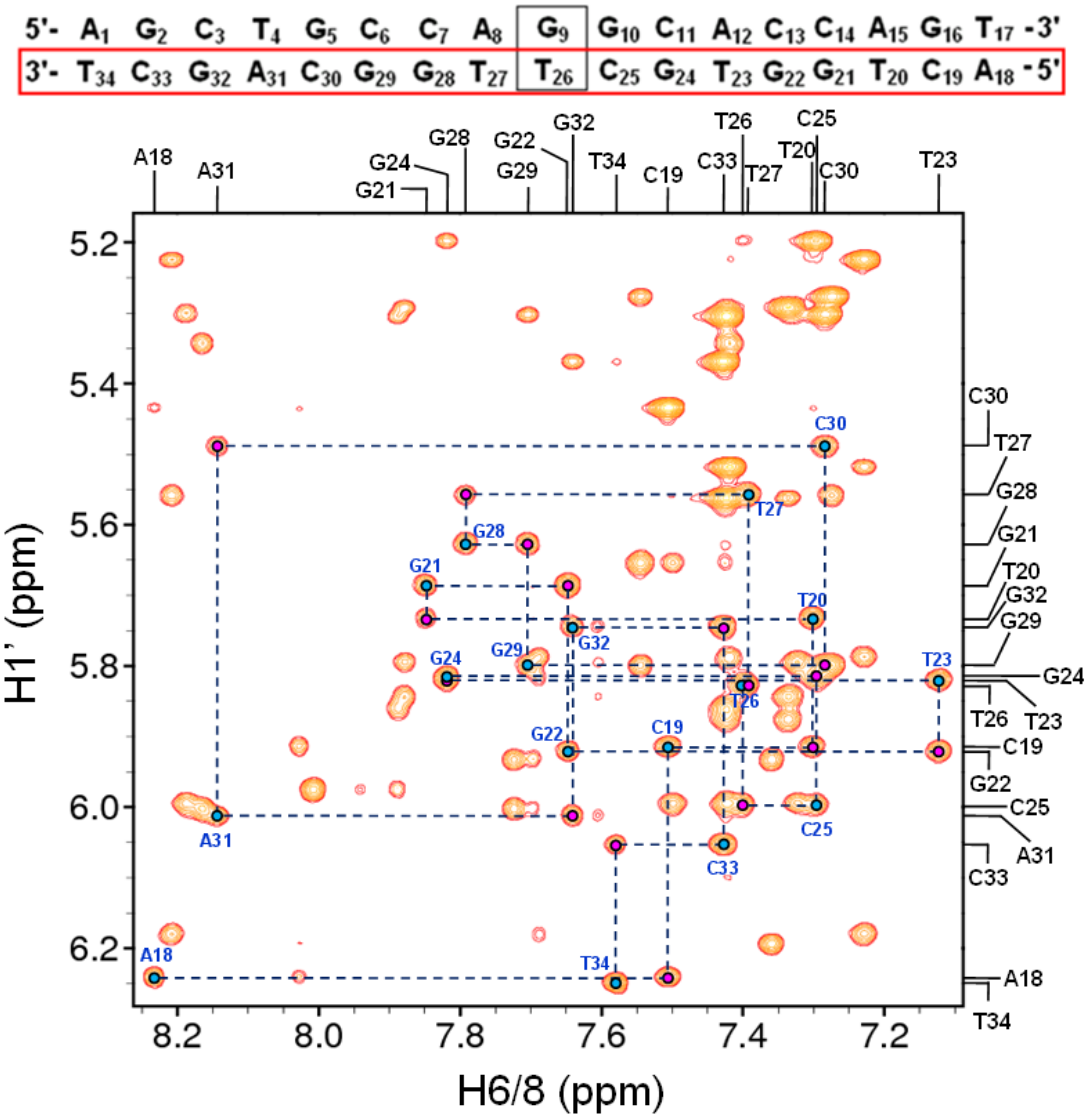
1000 MHz NOESY (τ_m = 250ms) spectra showing internucleotide H6/8 (i-1)-H5/CH_3 (i) NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. Thymine and B. Cytosine connectivities. Scalar coupled connectivities are indicated by blue circles and internucleotide connectivities by red circles.



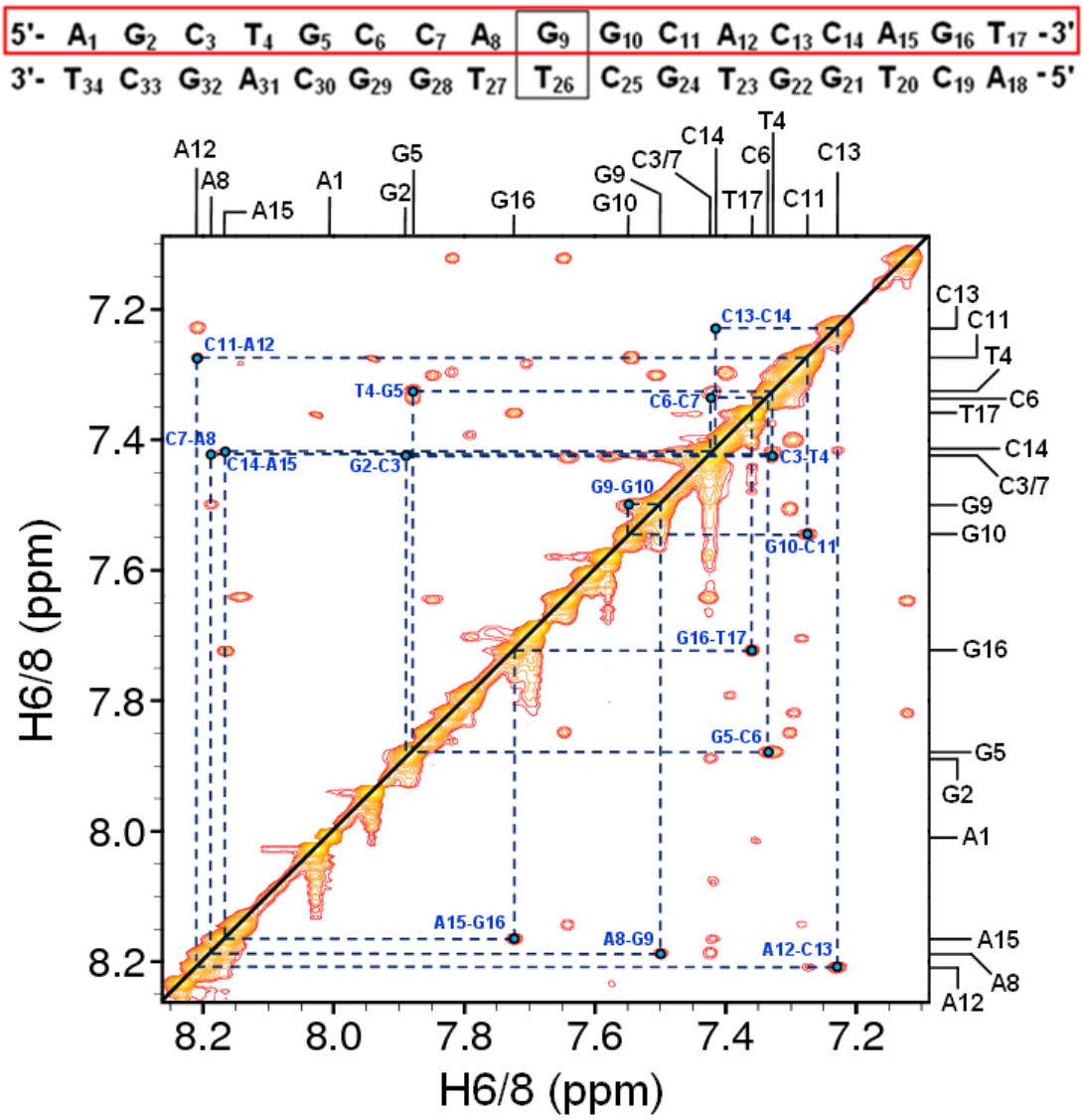
1000 MHz NOESY (τ_m = 250ms) spectra showing internucleotide H6/8 (i-1)-H5/CH_3 (i) NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. Thymine and B. Cytosine connectivities. Scalar coupled connectivities are indicated by blue circles and internucleotide connectivities by red circles.



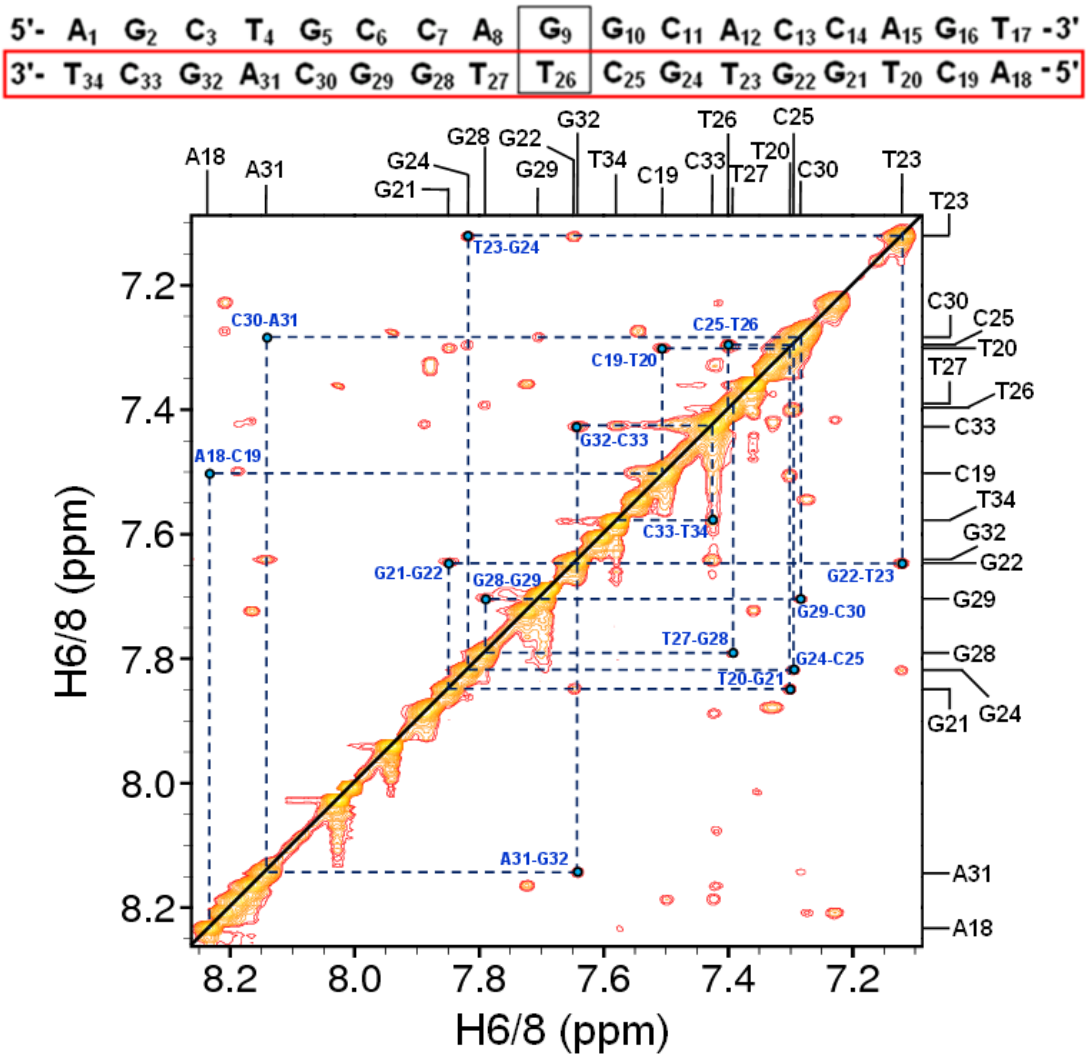
1000 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H1' sequential NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide connectivities are shown by blue circles and internucleotide connectivities by purple circles.



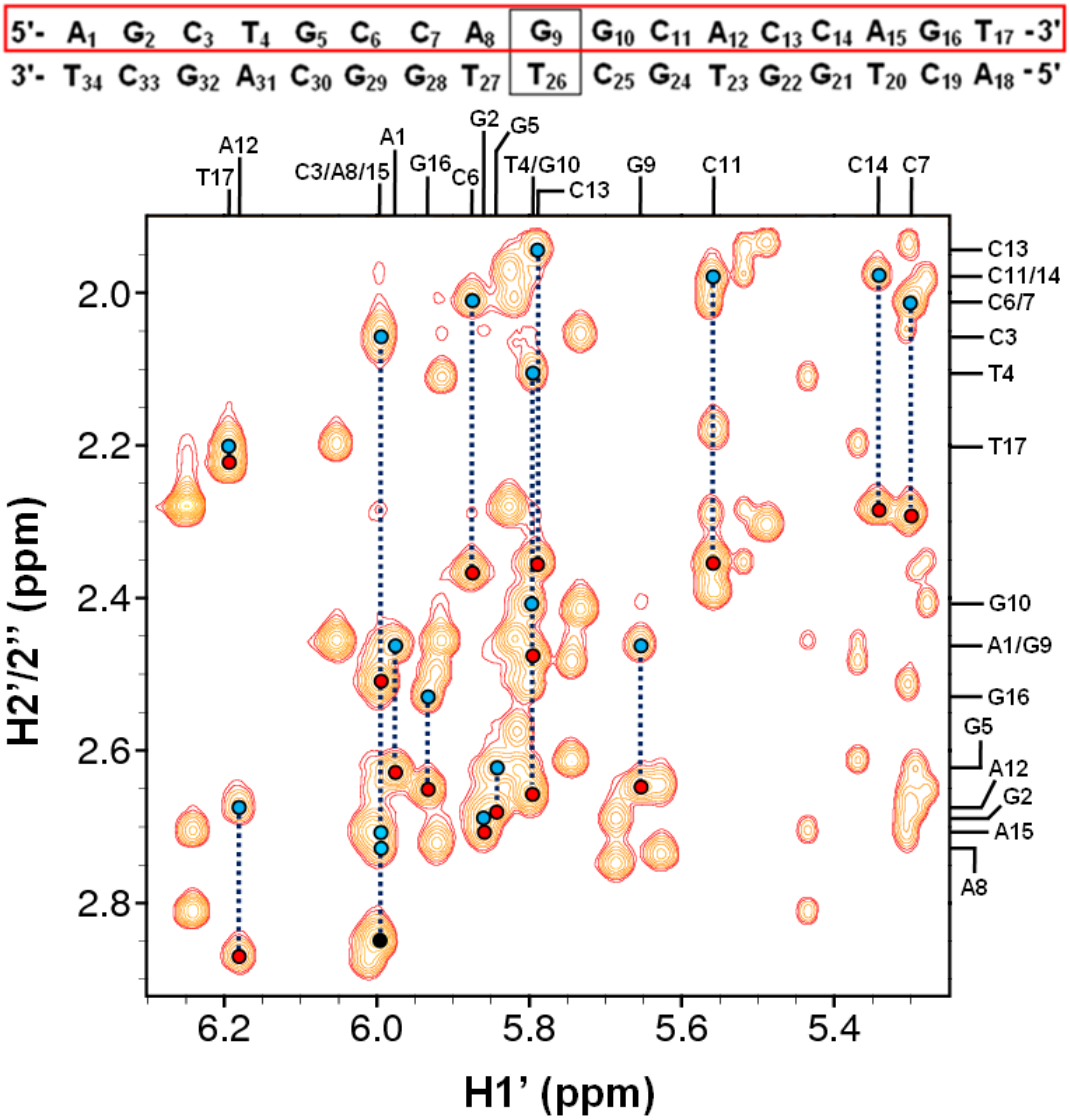
1000 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H1' sequential NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide connectivities are shown by blue circles and internucleotide connectivities by purple circles.



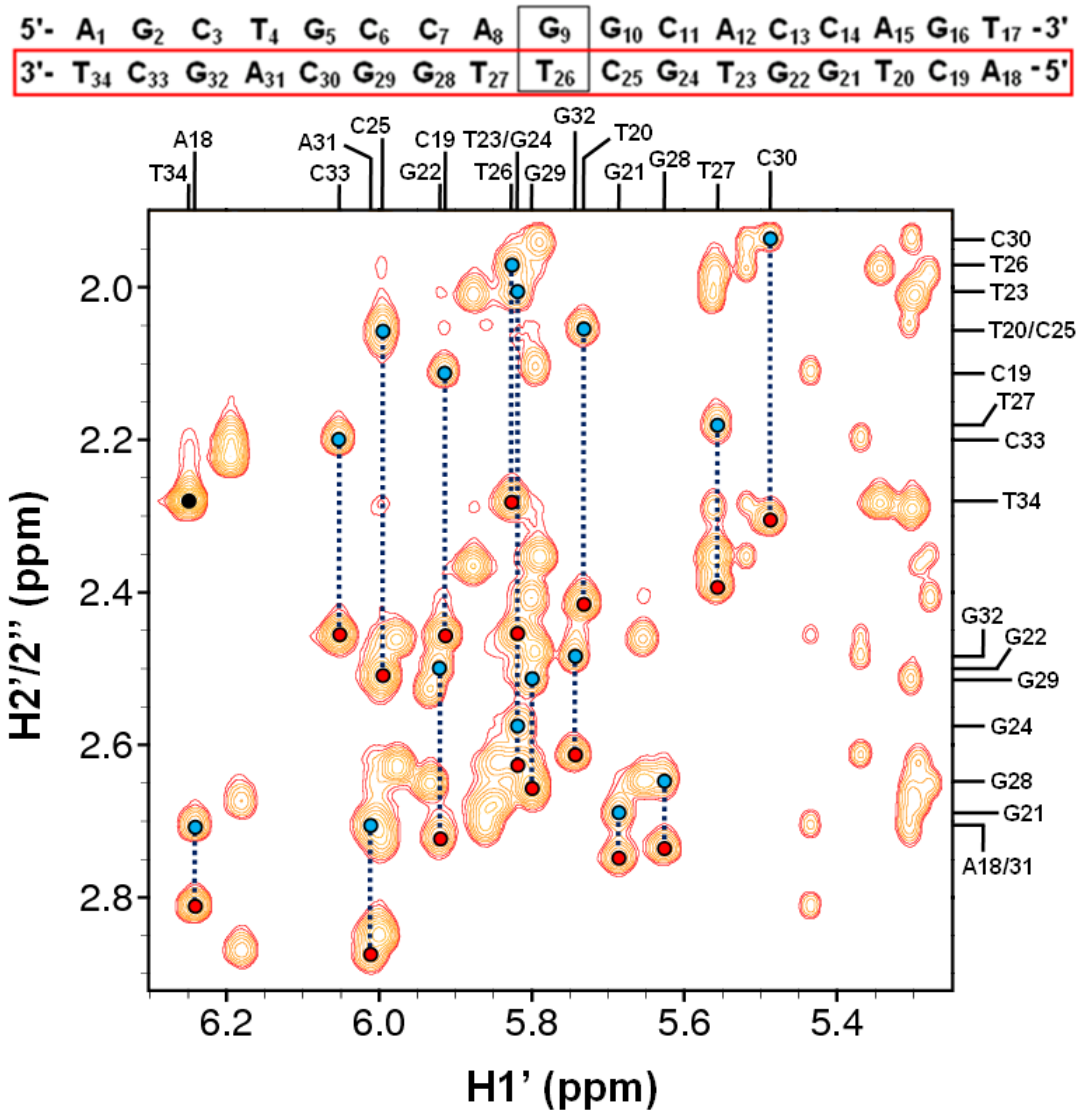
1000 MHz NOESY (τ_m = 250ms) spectrum showing internucleotide H6/8-H6/8 NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



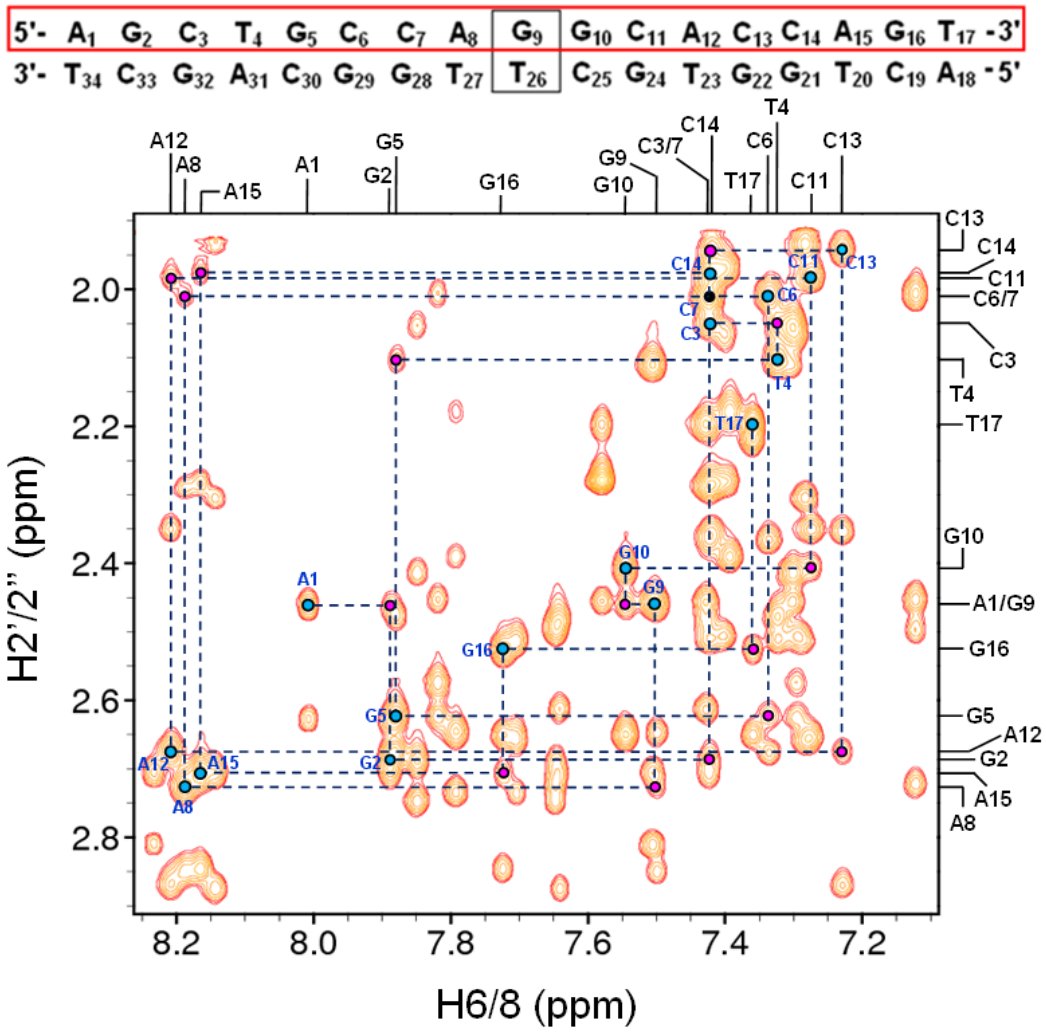
1000 MHz NOESY (τ_m = 250ms) spectrum showing internucleotide H6/8-H6/8 NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



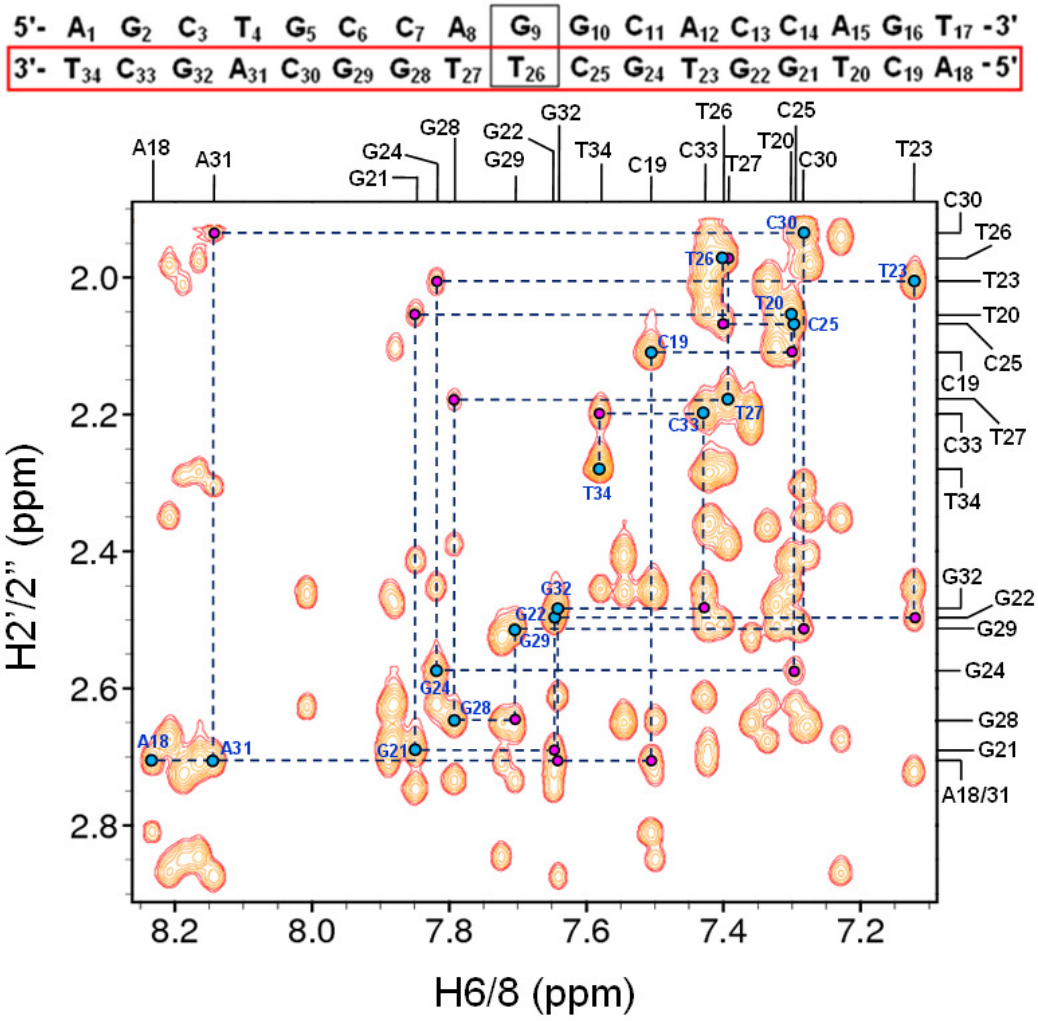
1000 MHz NOESY (τ_m = 250ms) spectrum stack showing H1'-H2'/2'' NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. H1'-H2' and H1'-H2'' NOEs are shown by blue and red circles respectively; those shown by black circles indicate an overlap of the H2 and H2'' resonances.



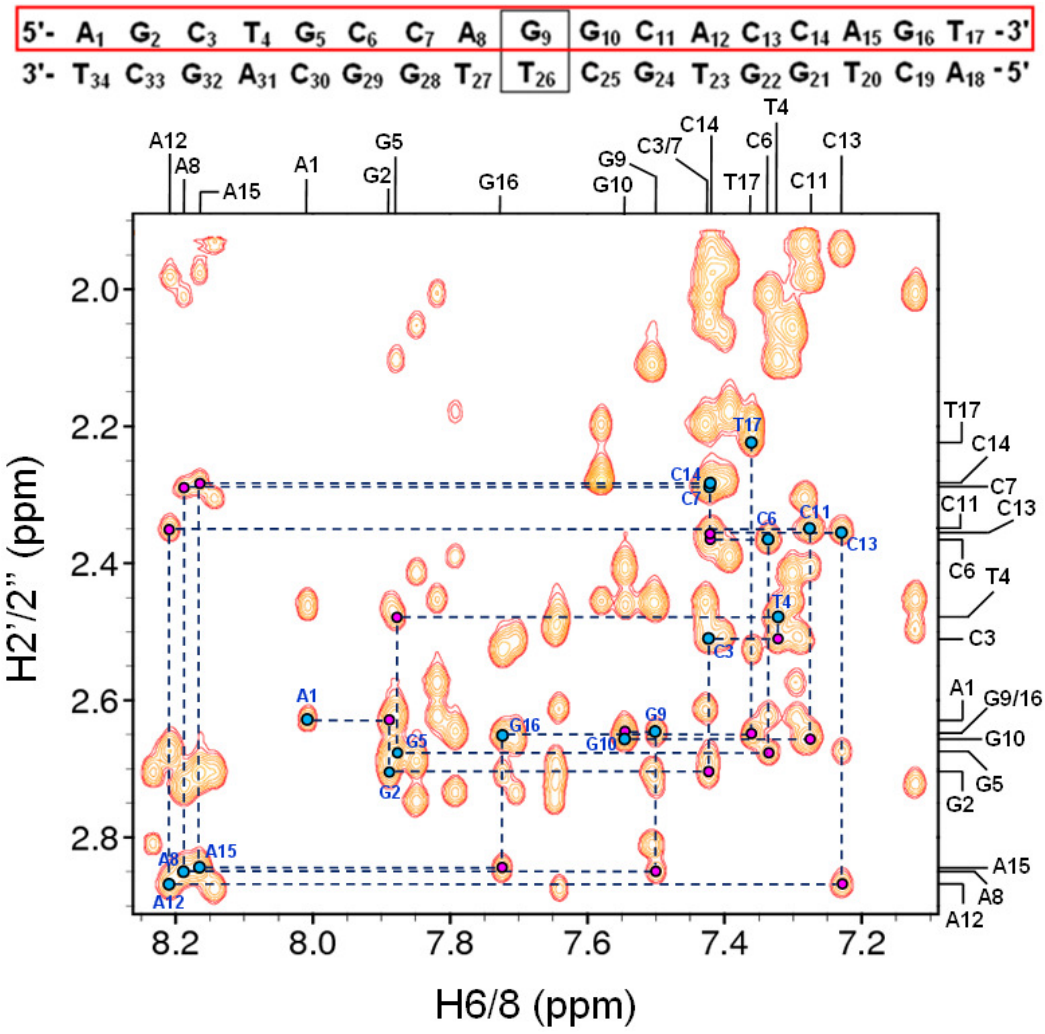
1000 MHz NOESY (τ_m = 250ms) spectrum stack showing H1'-H2'/2'' NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. H1'-H2' and H1'-H2'' NOEs are shown by blue and red circles respectively; those shown by black circles indicate an overlap of the H2 and H2'' resonances.



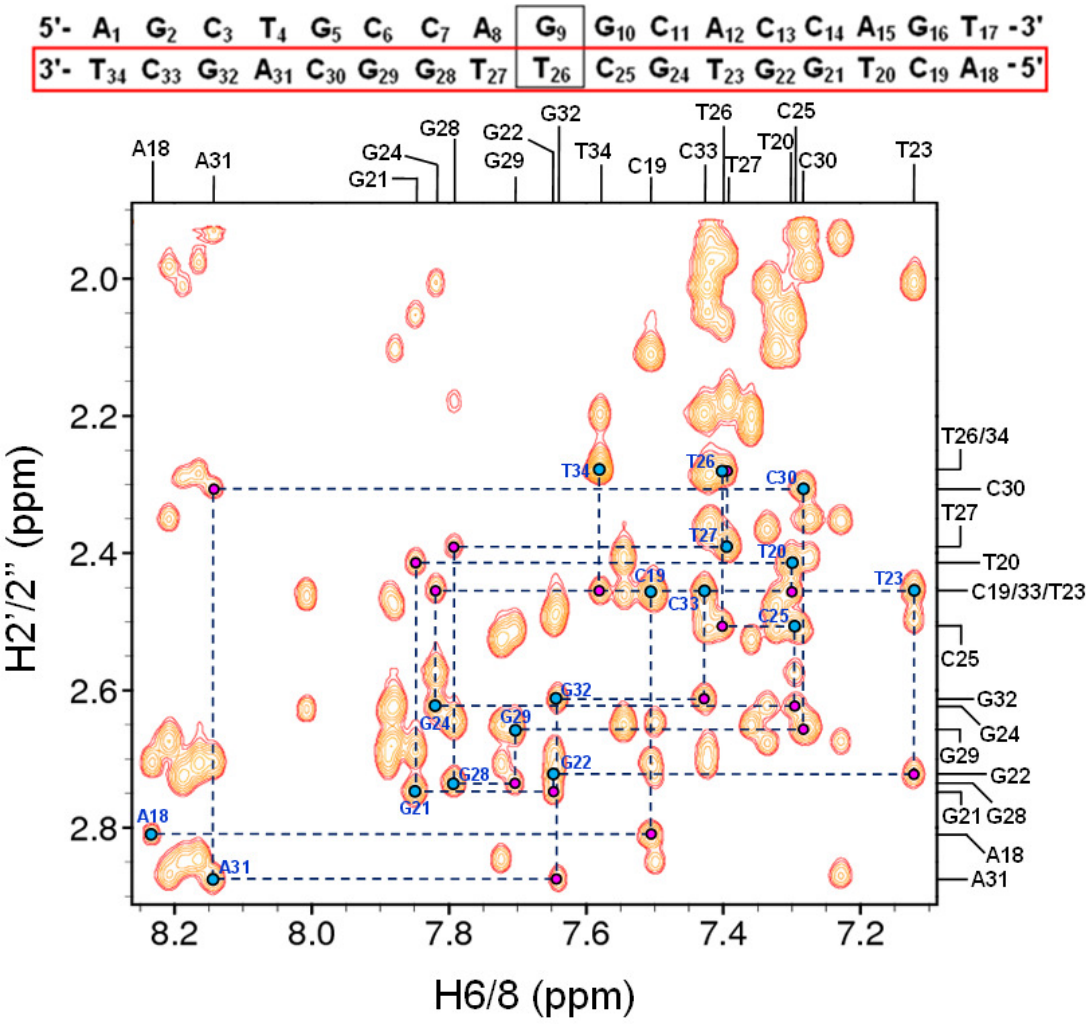
1000 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2' sequential NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide connectivities are shown by blue circles and internucleotide connectivities by pink circles, overlap of connectivities are shown by black circles.



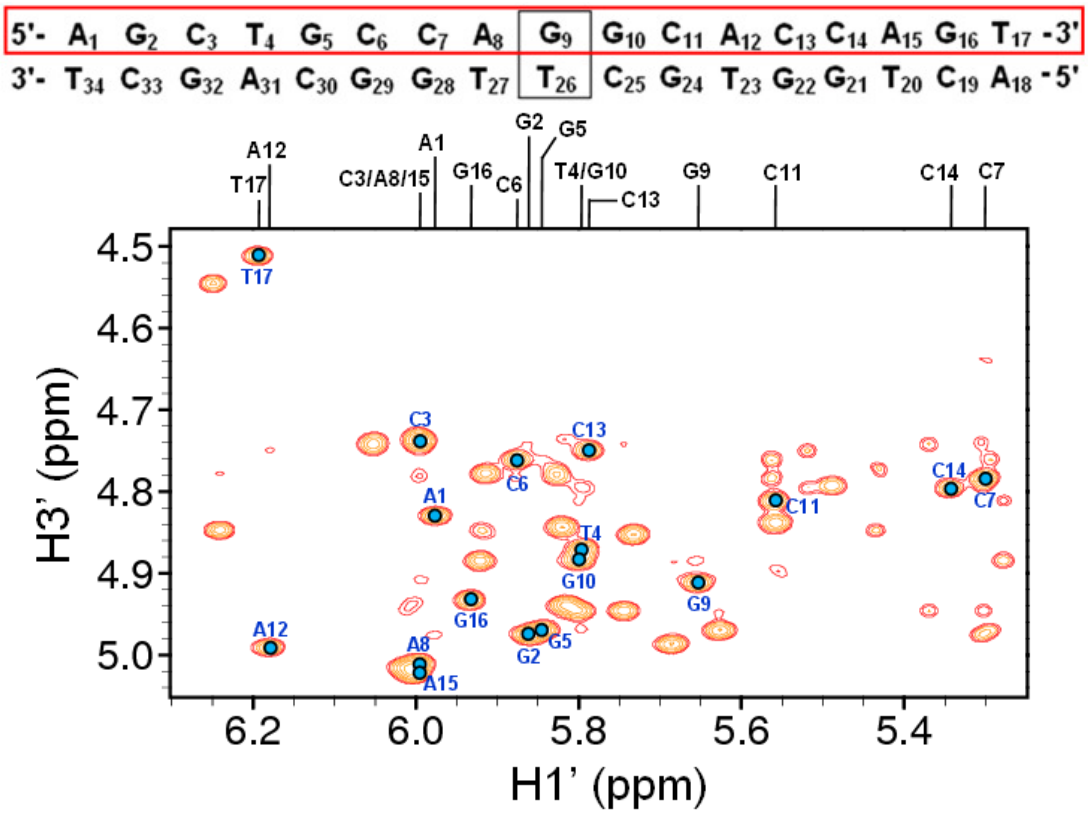
1000 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2' sequential NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide connectivities are shown by blue circles and internucleotide connectivities by pink circles.



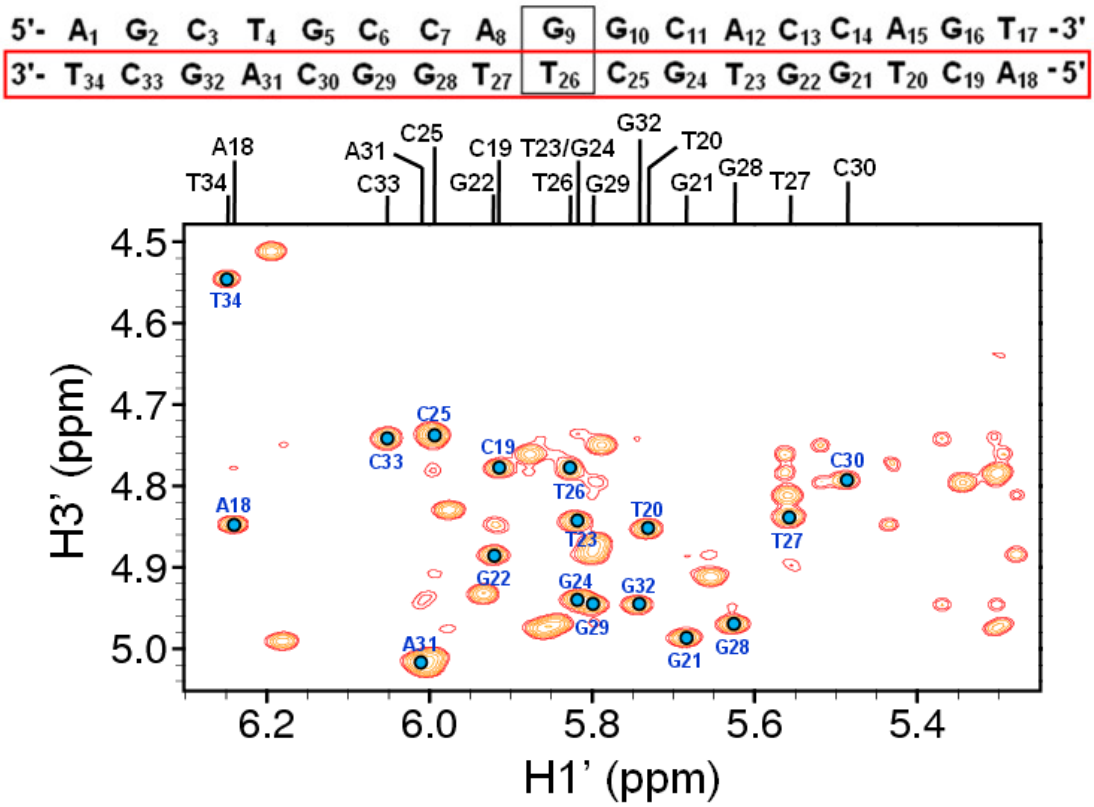
1000 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2'' sequential NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide connectivities are shown by blue circles and internucleotide connectivities by pink circles.



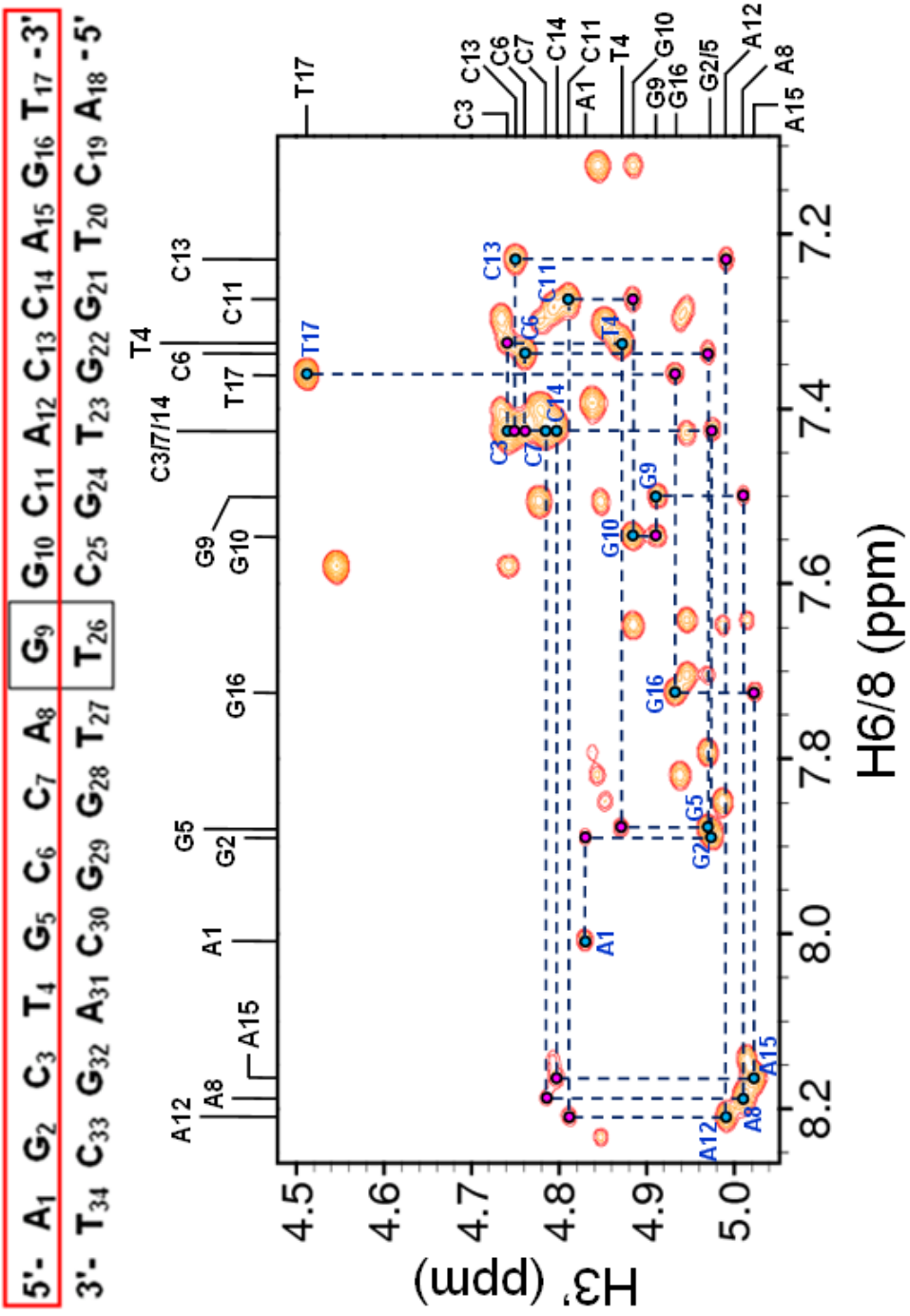
1000 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2'' sequential NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intranucleotide connectivities are shown by blue circles and internucleotide connectivities by pink circles.



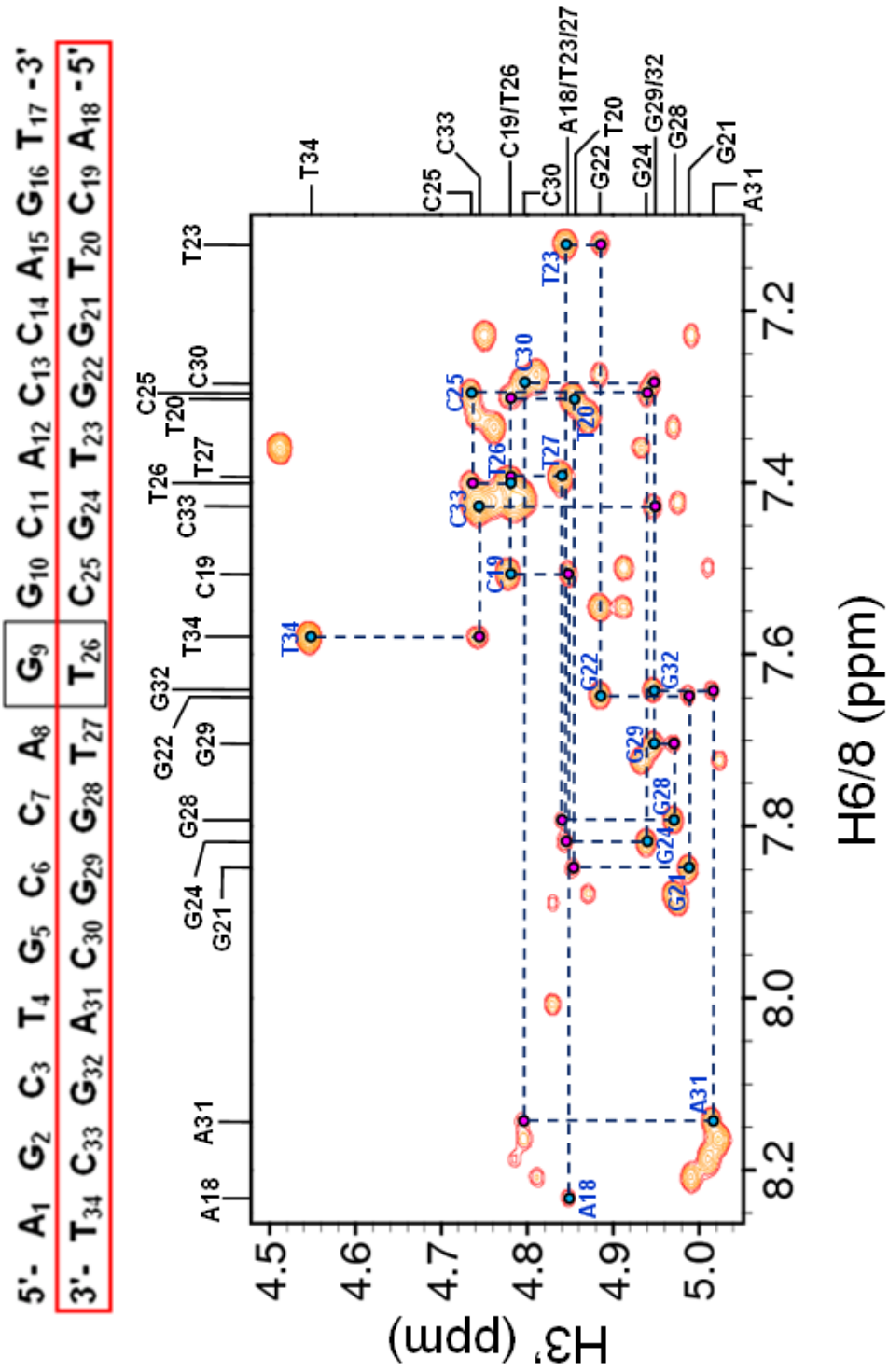
1000 MHz NOESY (τ_m = 250ms) spectrum showing H1'-H3' NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



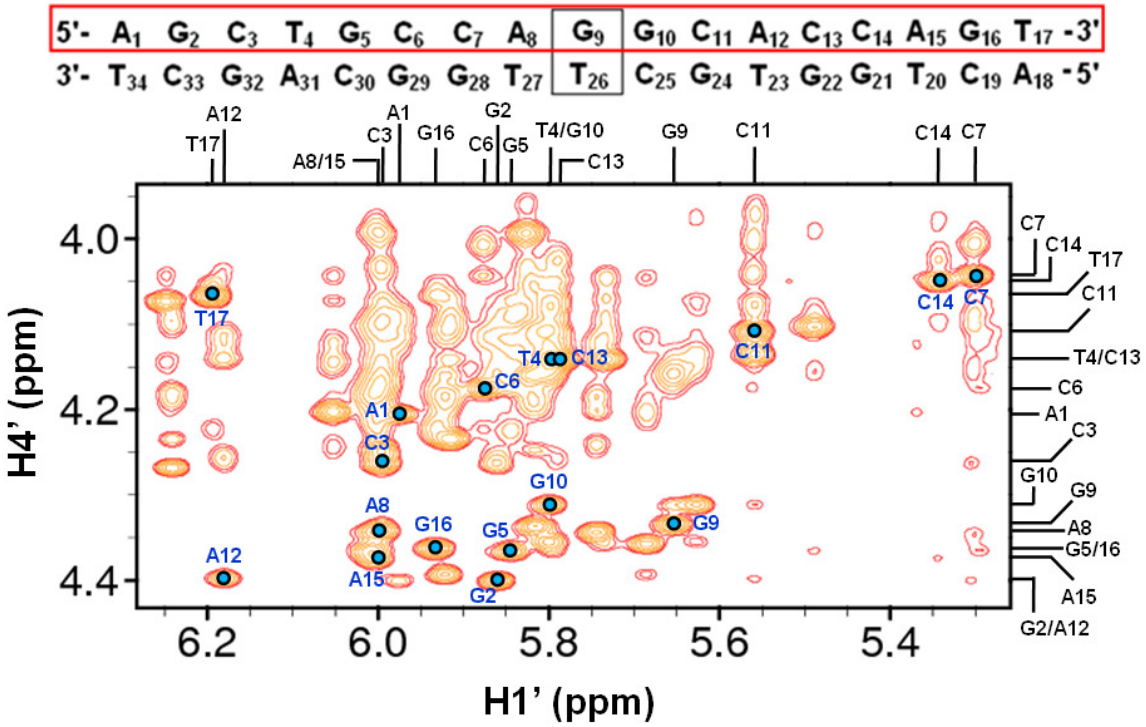
1000 MHz NOESY (τ_m = 250ms) spectrum showing H1'-H3' NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



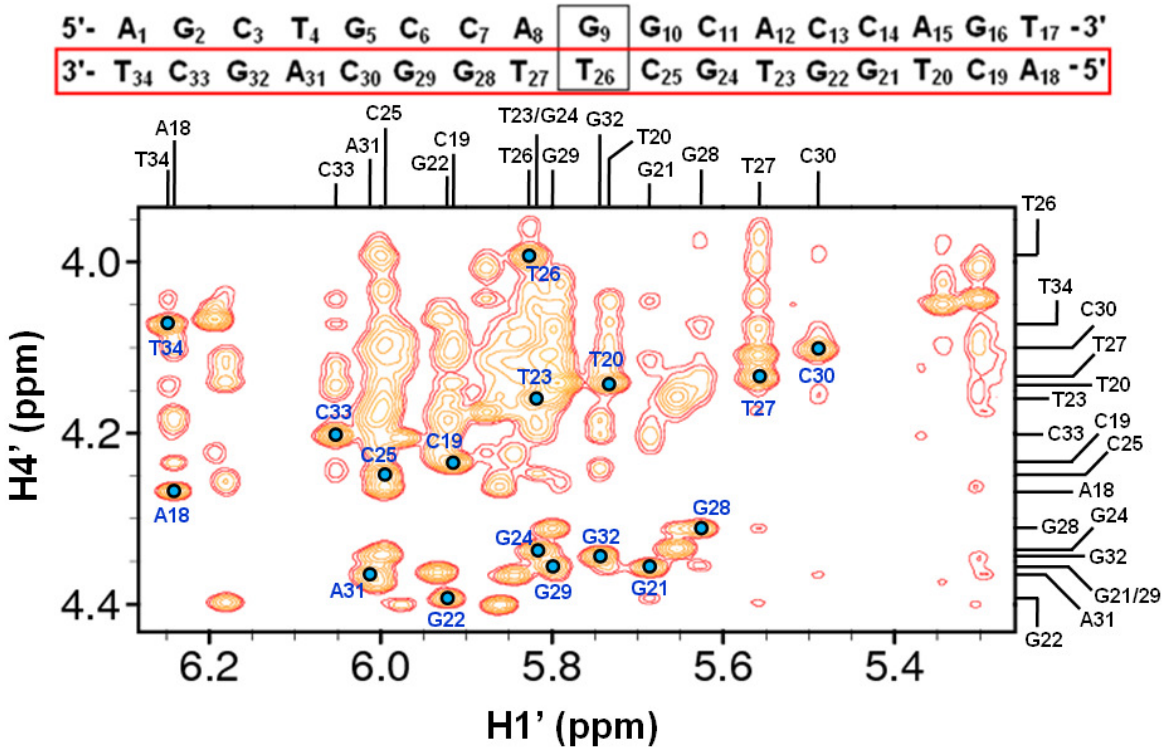
1000 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H3' NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intra- and internucleotide connectivities are shown by blue circles and pink circles respectively.



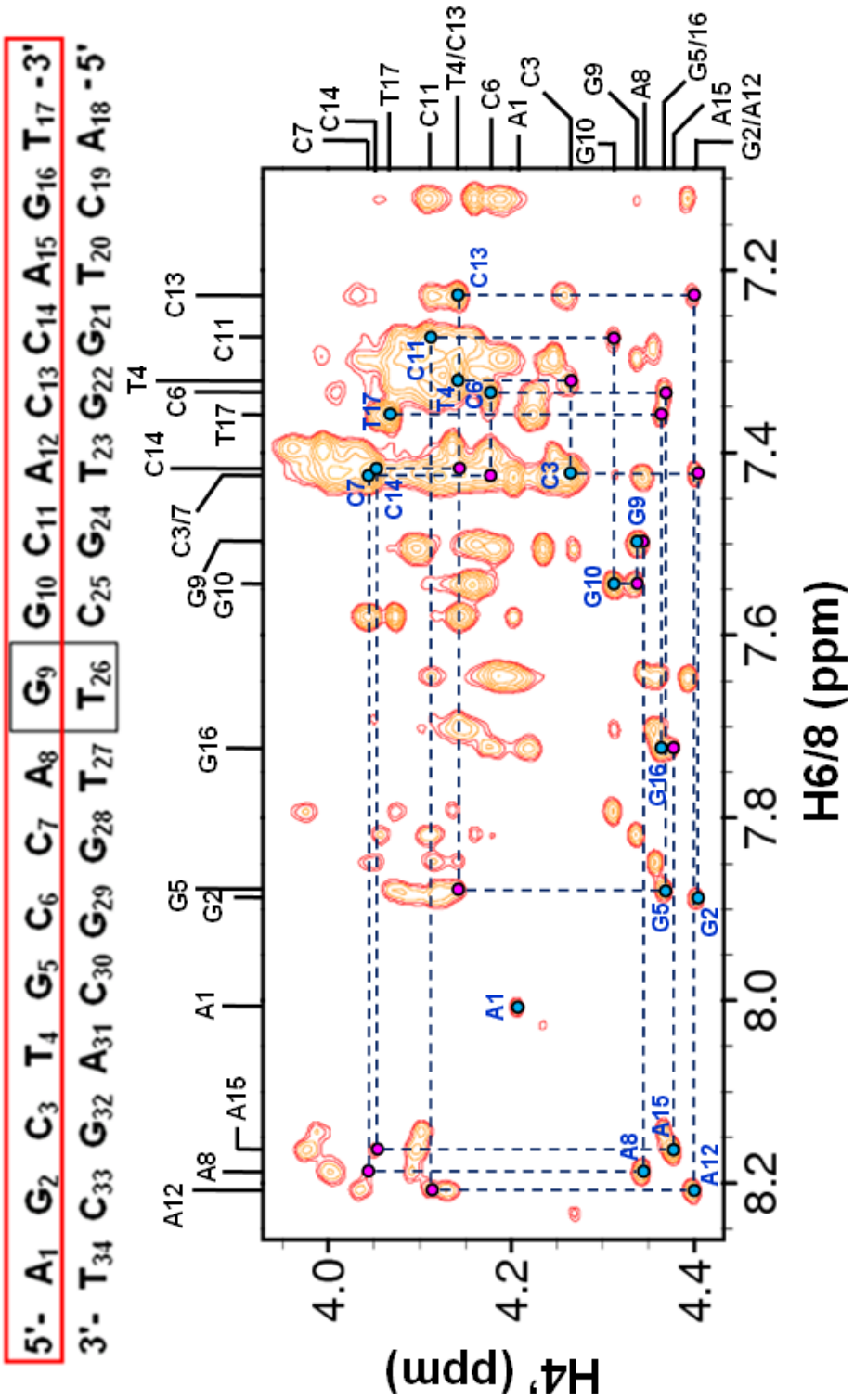
1000 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H3' NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intra- and internucleotide connectivities are shown by blue circles and pink circles respectively.



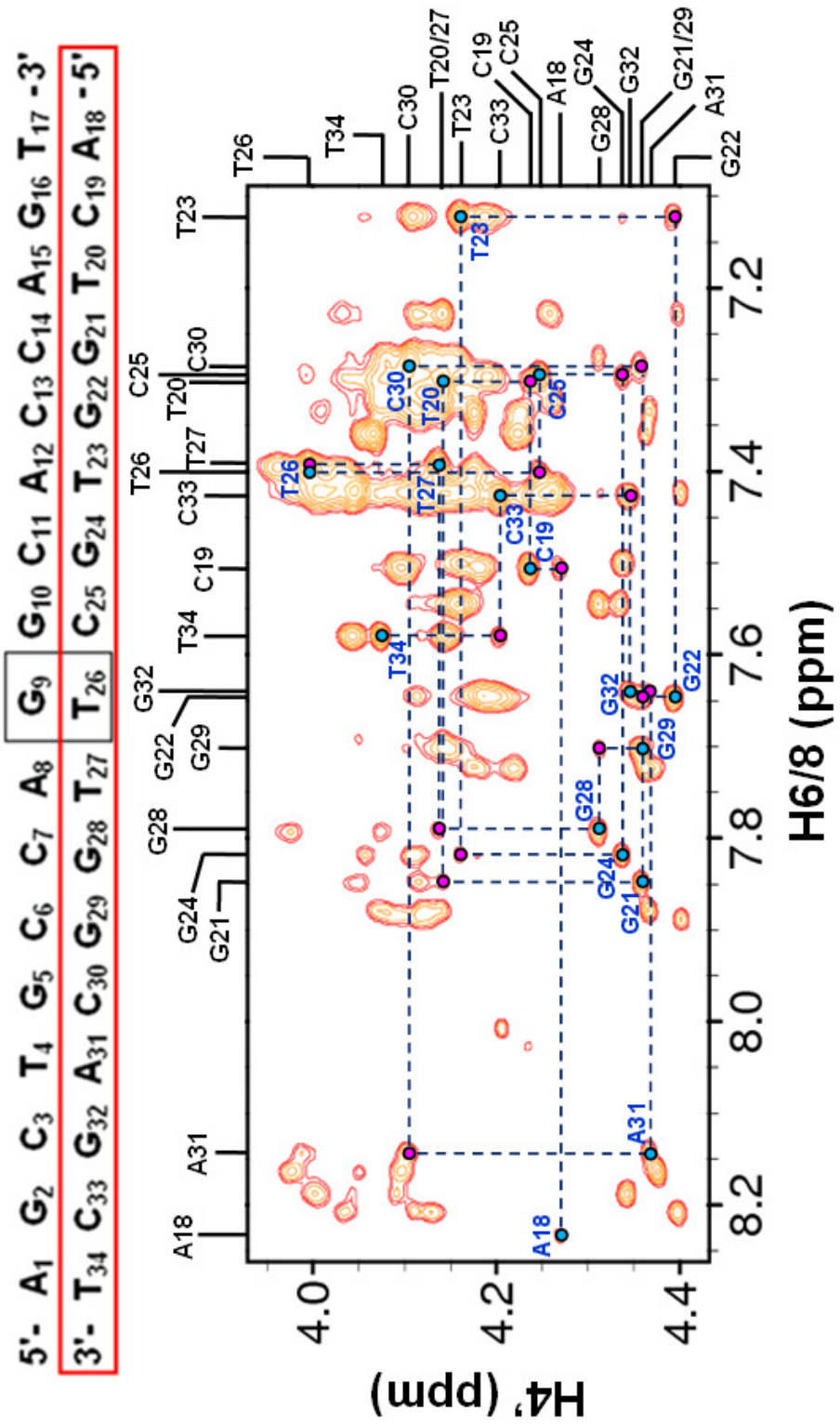
1000 MHz NOESY (τ_m = 250ms) spectrum showing H1'-H4' NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



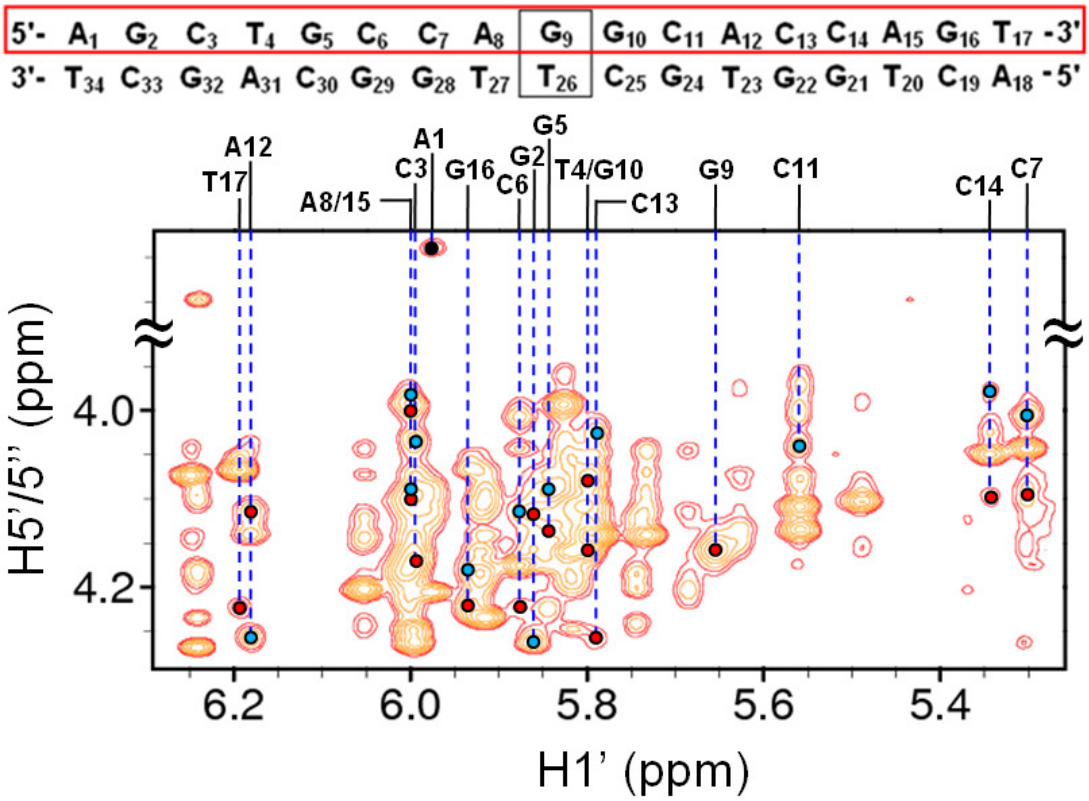
1000 MHz NOESY (τ_m = 250ms) spectrum showing H1'-H4' NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



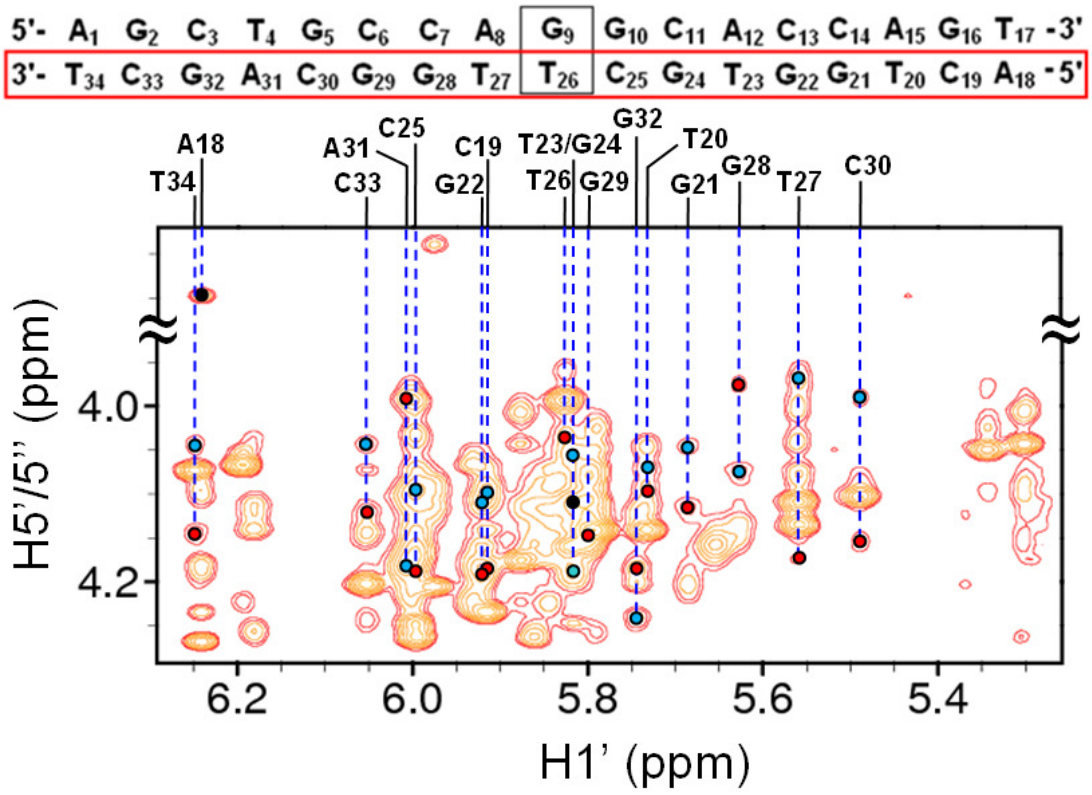
1000 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H4' NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intra- and internucleotide connectivities are shown by blue circles and pink circles respectively



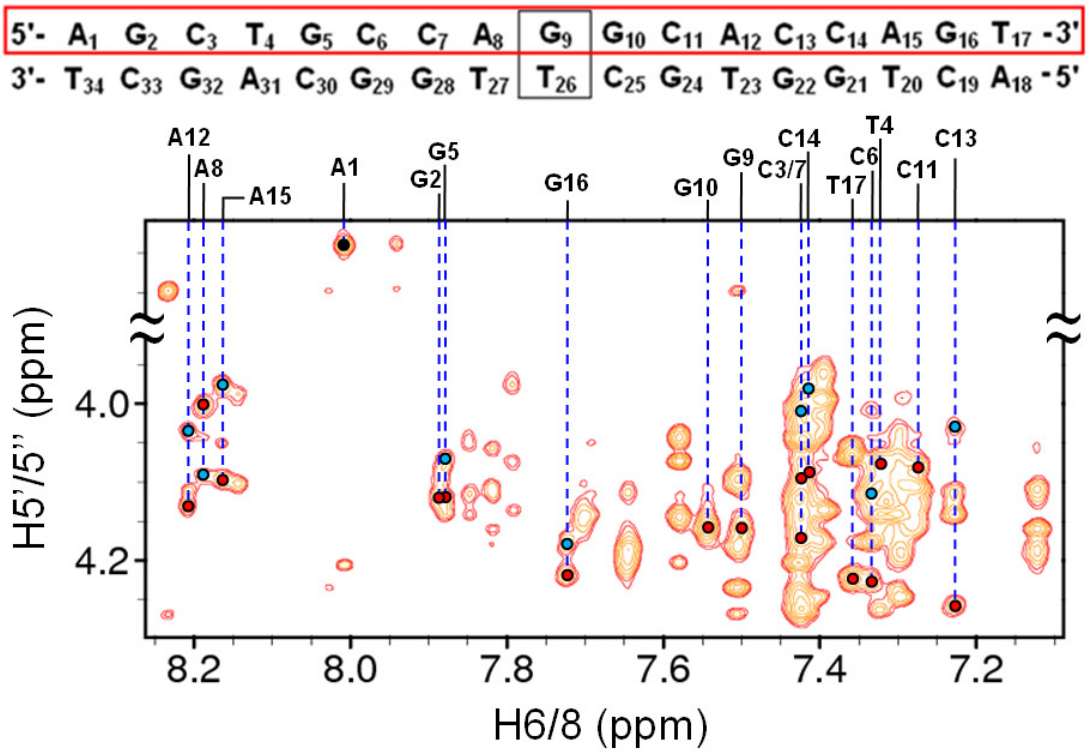
1000 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H4' NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intra- and internucleotide connectivities are shown by blue circles and pink circles respectively



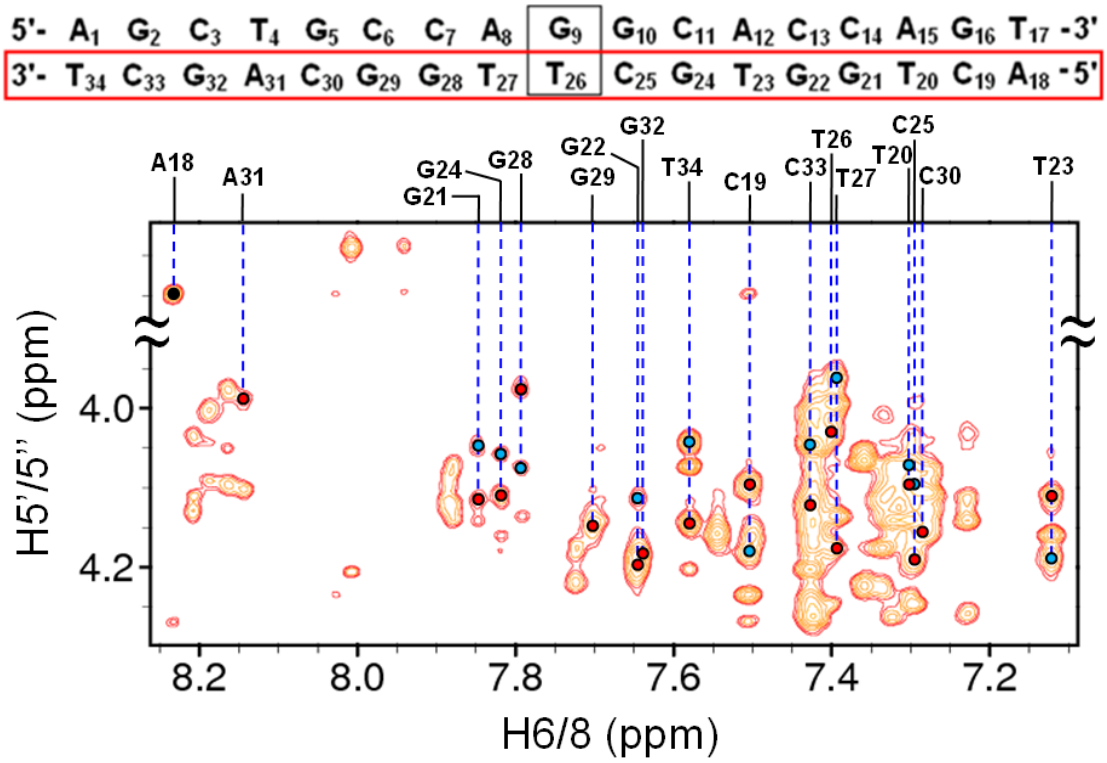
1000 MHz NOESY (τ_m = 250ms) spectrum showing H1'-H5'/5'' NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. H1'-H5' and H1'-H5'' NOEs are shown by blue and red circles respectively and black circles indicate an overlap.



1000 MHz NOESY (τ_m = 250ms) spectrum showing H1'-H5'/5'' NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. H1'-H5' and H1'-H5'' NOEs are shown by blue and red circles respectively and black circles indicate an overlap.

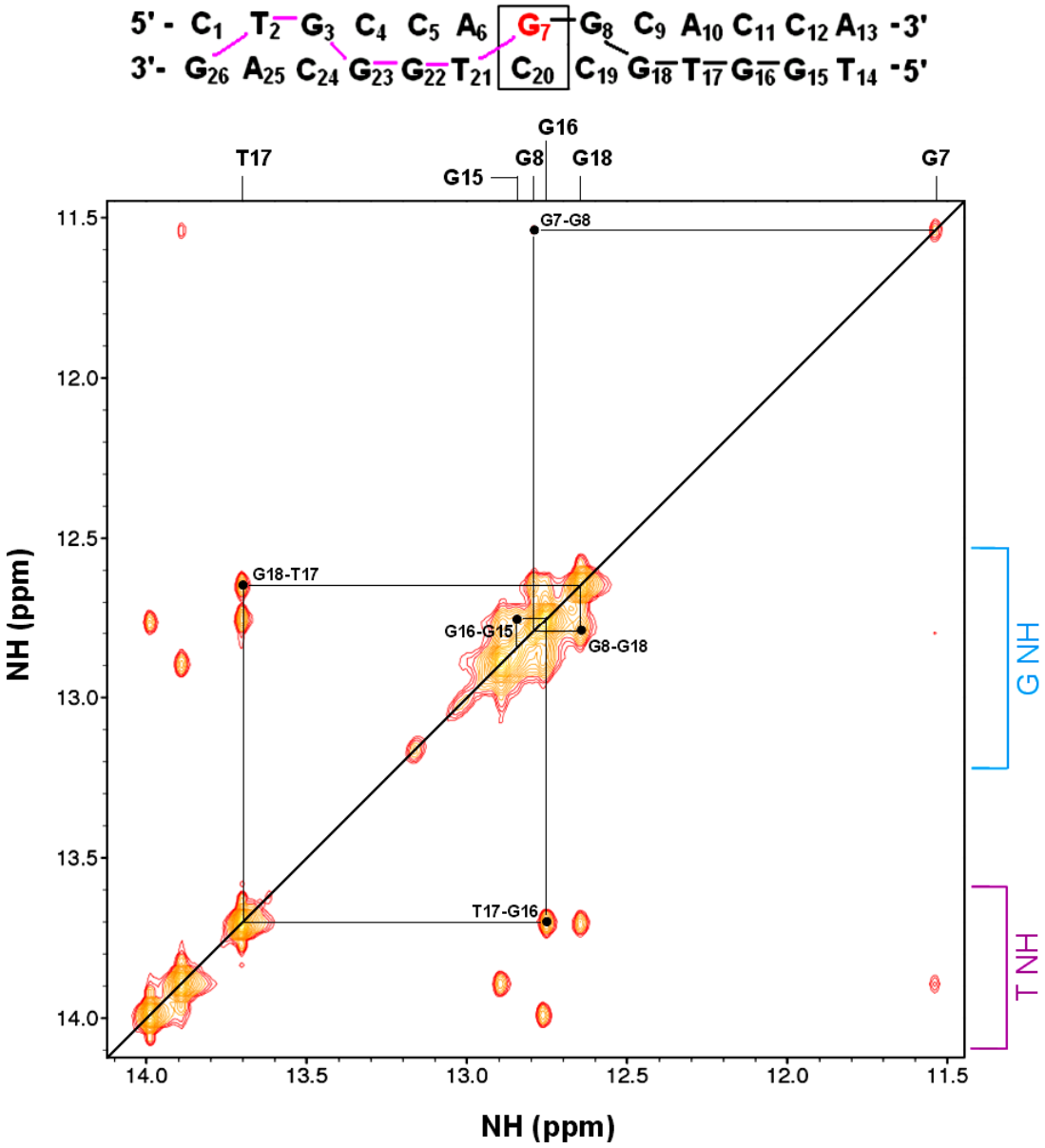


1000 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H5'/5'' NOE connectivities for residues A1-T17 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. H6/8-H5' and H6/8-H5'' NOEs are shown by blue and red circles respectively and black circles indicate an overlap.

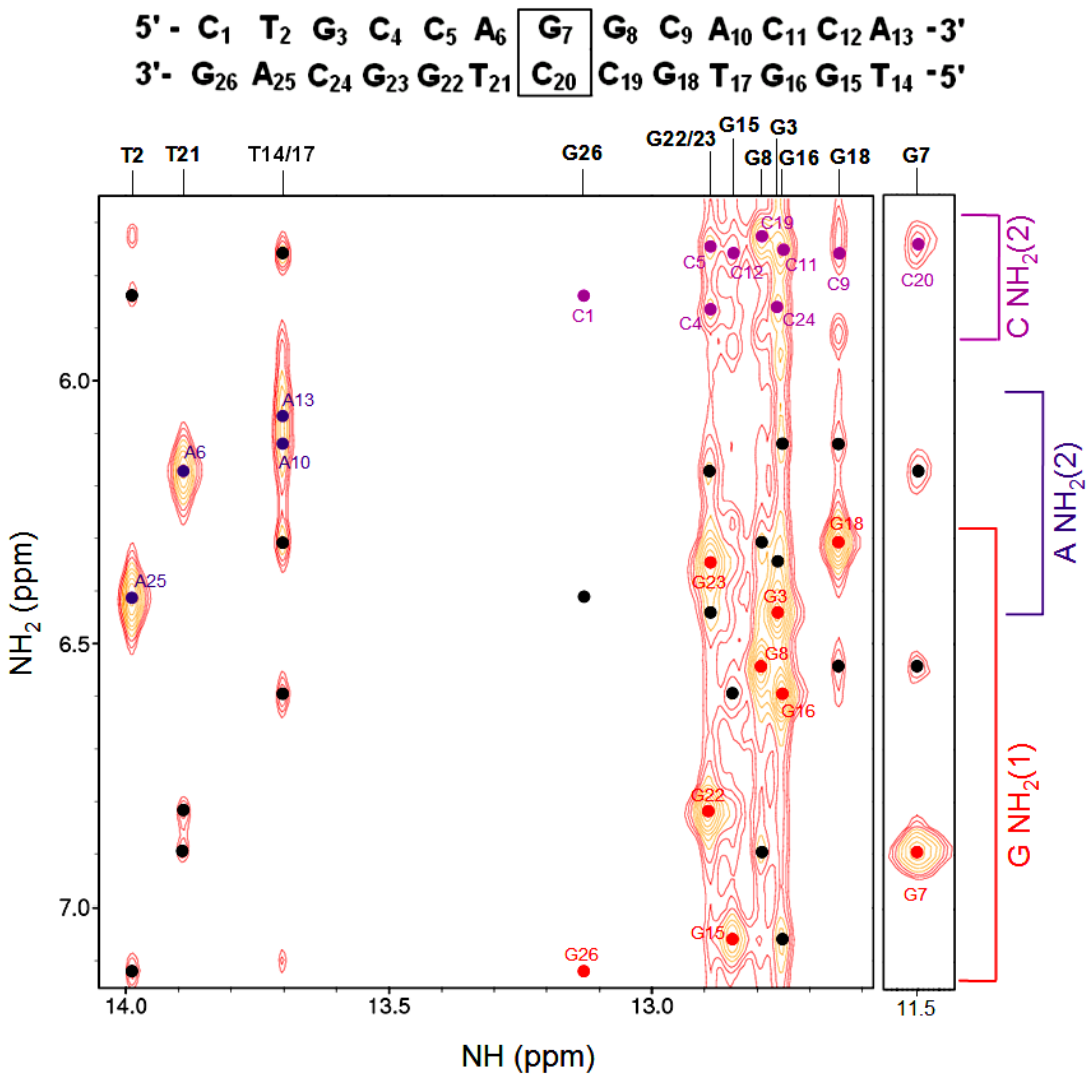


1000 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H5'/5'' NOE connectivities for residues A18-T34 in 17mer mismatch GT DNA (1mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. H6/8-H5' and H6/8-H5'' NOEs are shown by blue and red circles respectively and black circles indicate an overlap.

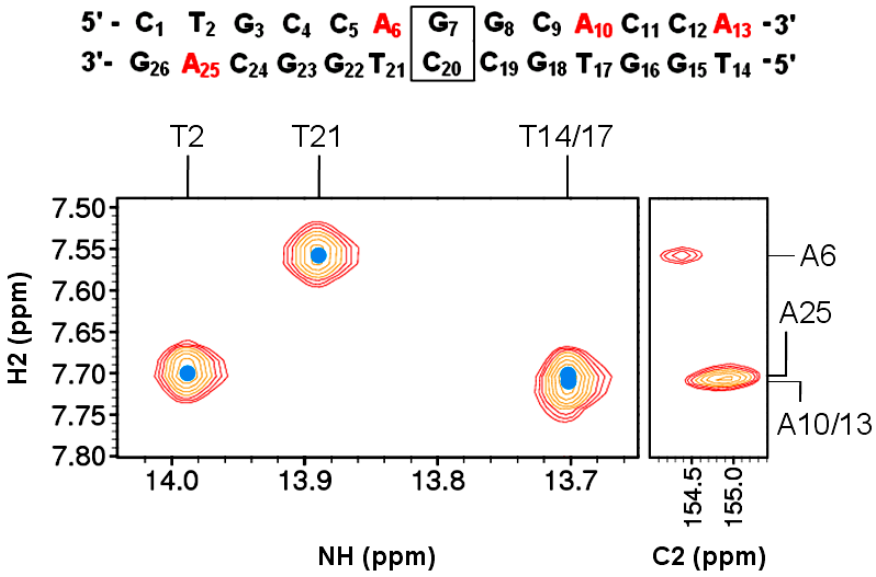
iv. 6-TG modified 13mer GC DNA



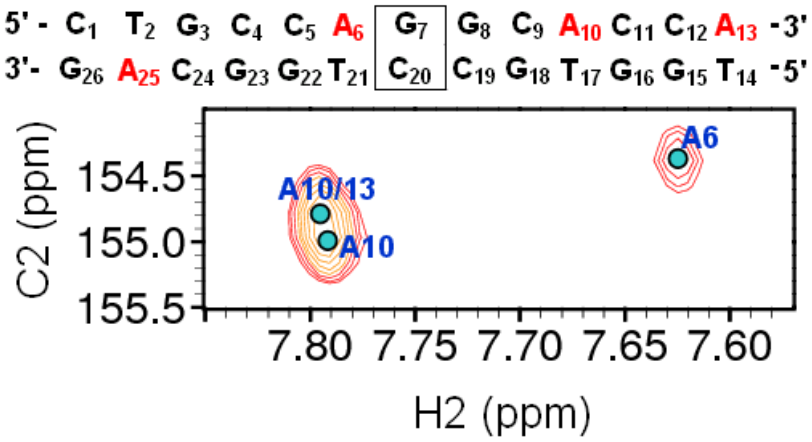
800 MHz NOESY (τ_m = 250ms) spectrum of 6-TG modified 13mer GC DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C. The imino protons were identified and assigned based on the sequential connectivity from G7-G15 as illustrated by the black lines in the sequence.



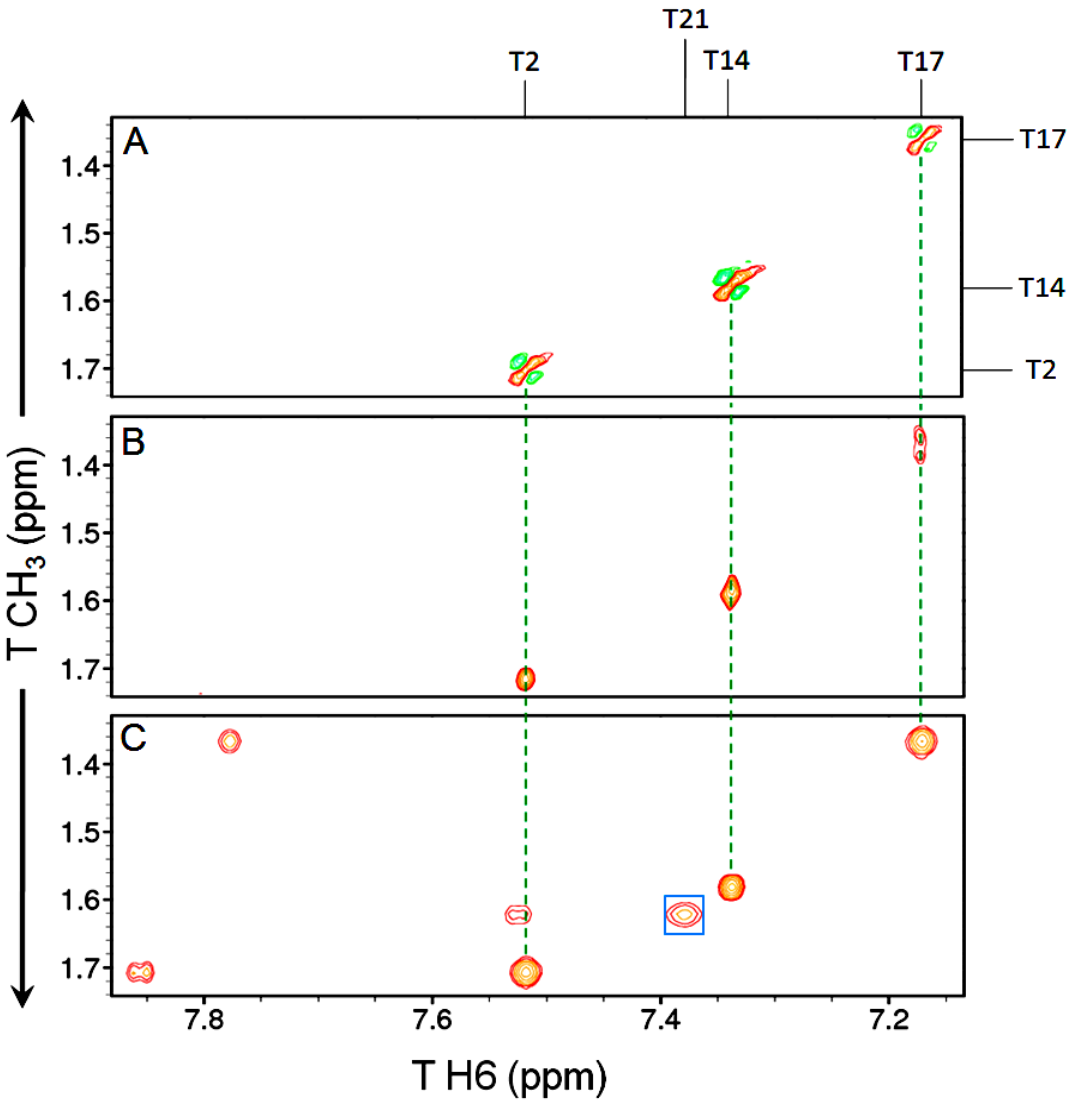
800 MHz NOESY (τ_m = 250ms) spectrum of 6-TG modified 13mer GC DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C. Base pair imino-amino NOE connectivities to hydrogen bonded guanine NH_2 and non-hydrogen bonded cytosine/adenine NH_2 protons are indicated by red, light purple and dark purple circles respectively. Important intrastrand and cross-strand NOE connectivities which confirmed assignment are also highlighted by black circles.



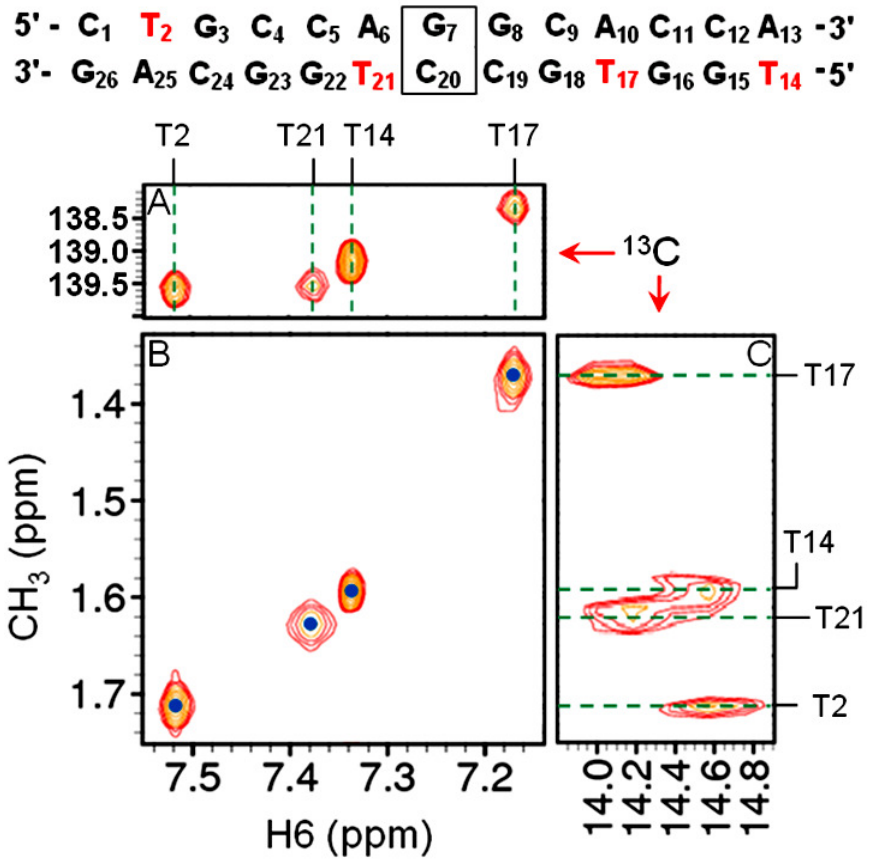
(Left panel) 800 MHz NOESY (τ_m = 250ms) spectrum of 6-TG modified 13mer GC DNA (0.8mM, 90% ^1H_2O + 10% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 2°C, (Right panel) 800 MHz ^1H-^{13}C HSQC spectrum of 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. The assignment of adenine H2 resonances is shown from NOE connectivities to thymine H3 imino resonances. Correlation to the ^1H-^{13}C HSQC spectrum identified the chemical shifts of C2 resonances.



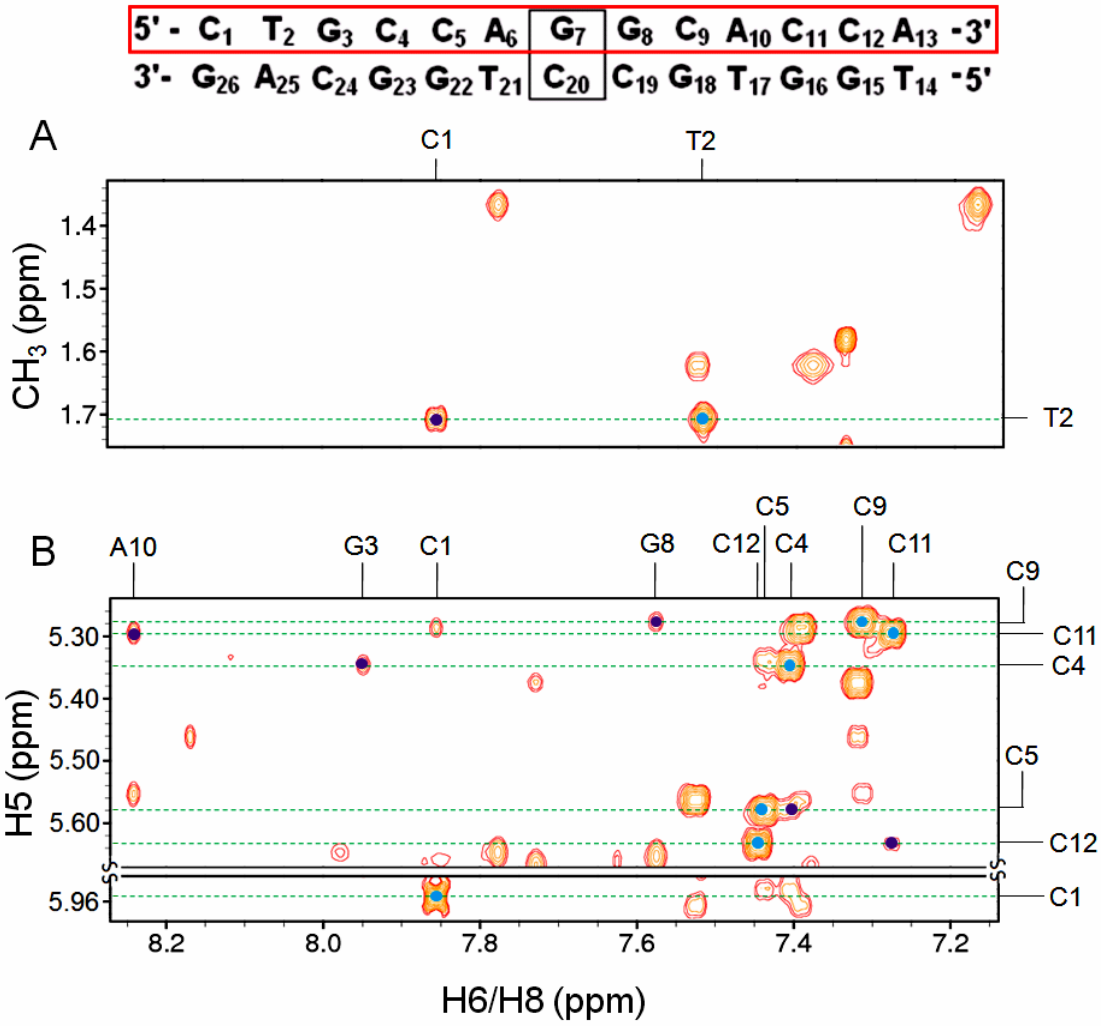
600 MHz ^1H-^{13}C HSQC spectrum of 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C showing the assignment of C2 resonances.



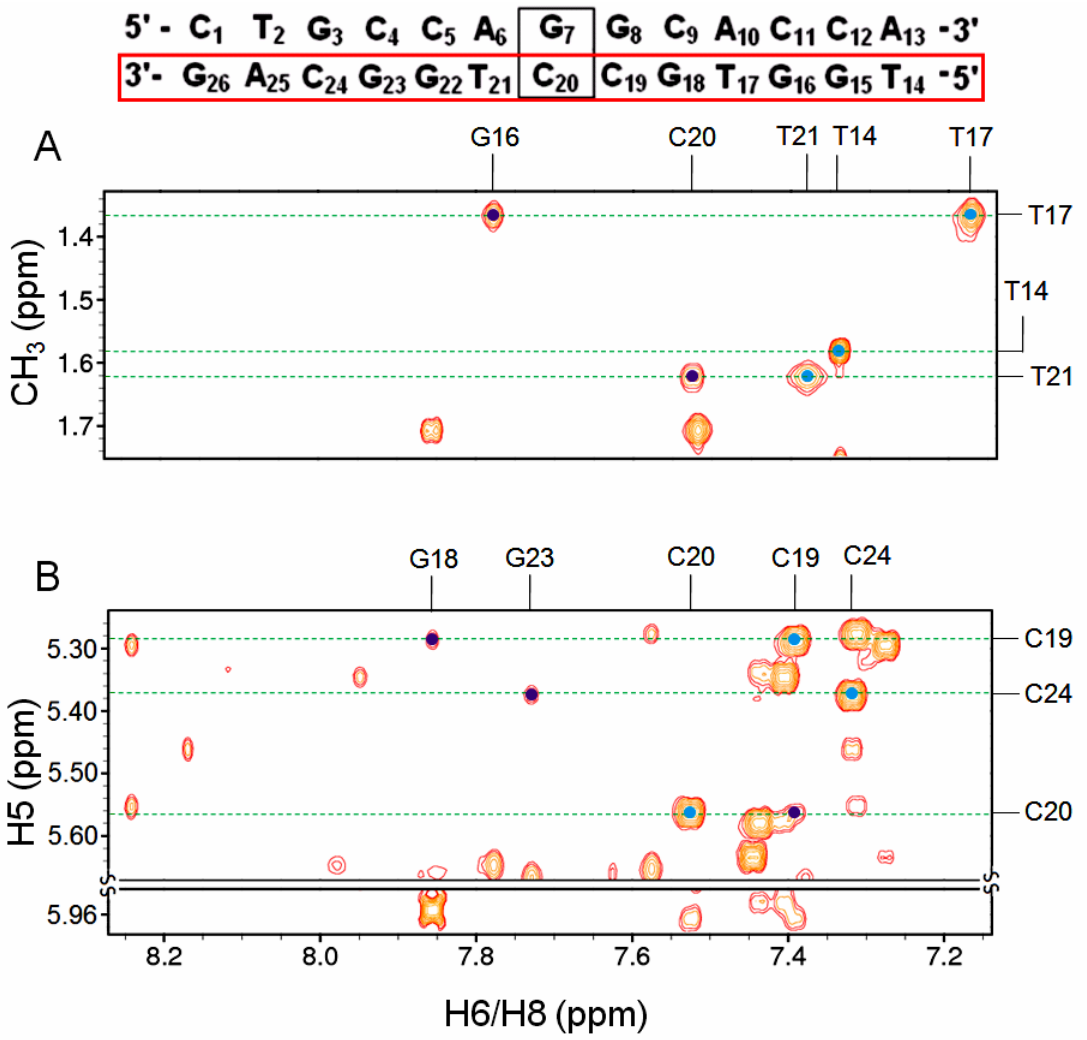
800 MHz spectrum stack showing the assignment of thymine CH_3-H6 correlations in 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. DQF-COSY, B. TOCSY (SL_{mix} = 75ms) and C. NOESY (τ_m = 250ms).



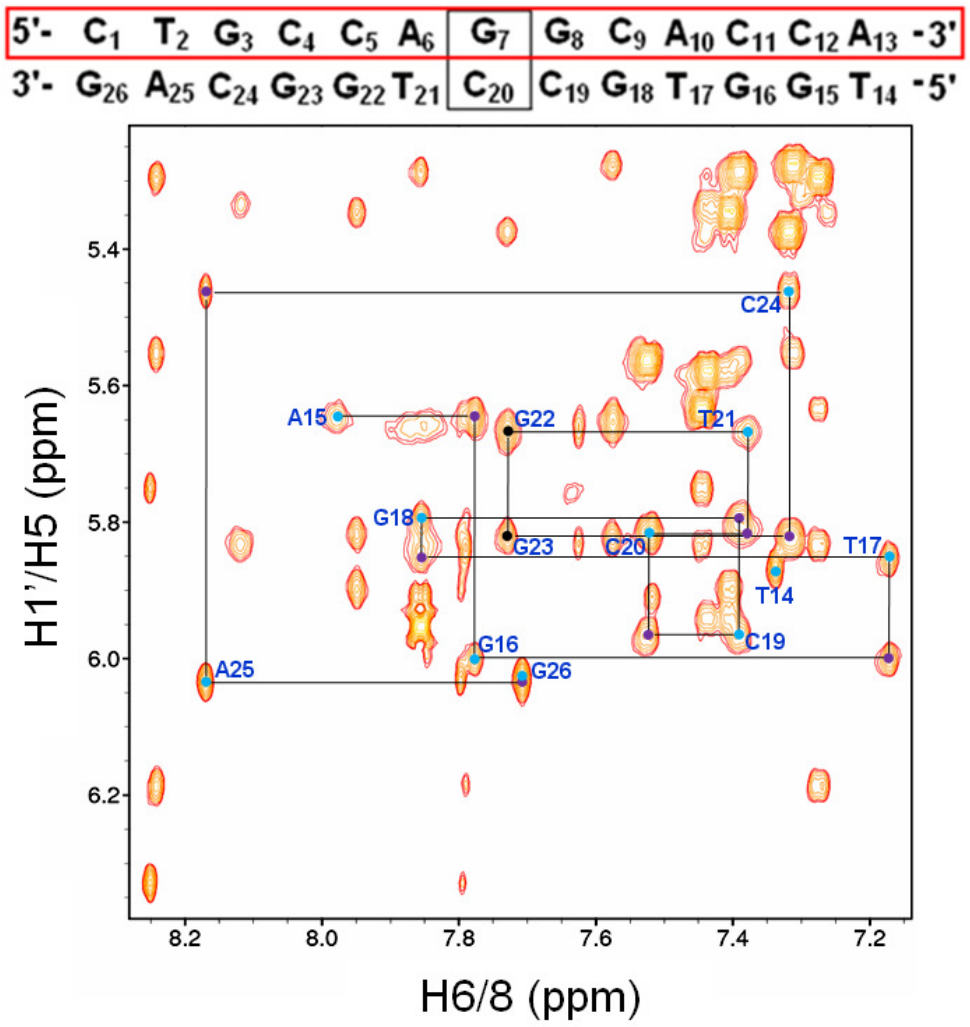
800 MHz NOESY (τ_m = 250ms) spectrum (B) highlighting thymine CH_3-H6 NOE connectivities in 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. 800 MHz ^1H-^{13}C HSQC panels show the assignment of the C6 (A) and CH_3 (C5) (C) chemical shifts.



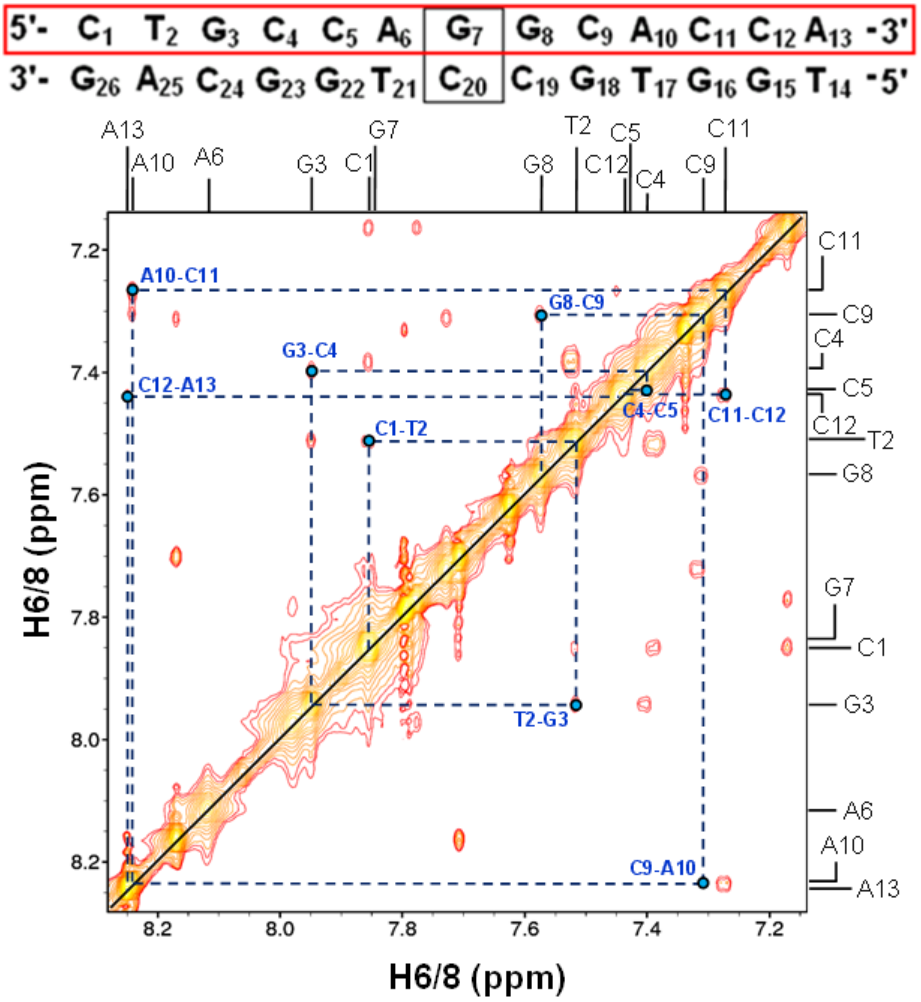
800 MHz NOESY (τ_m = 250ms) spectra showing interresidue H6/8 (i-1)-H5/CH_3 (i) NOE connectivities for residues C1-A13 in 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. Thymine and B. Cytosine connectivities. H5/CH_3(5)-H6 connectivities are indicated by blue circles and interresidue connectivities by purple circles.



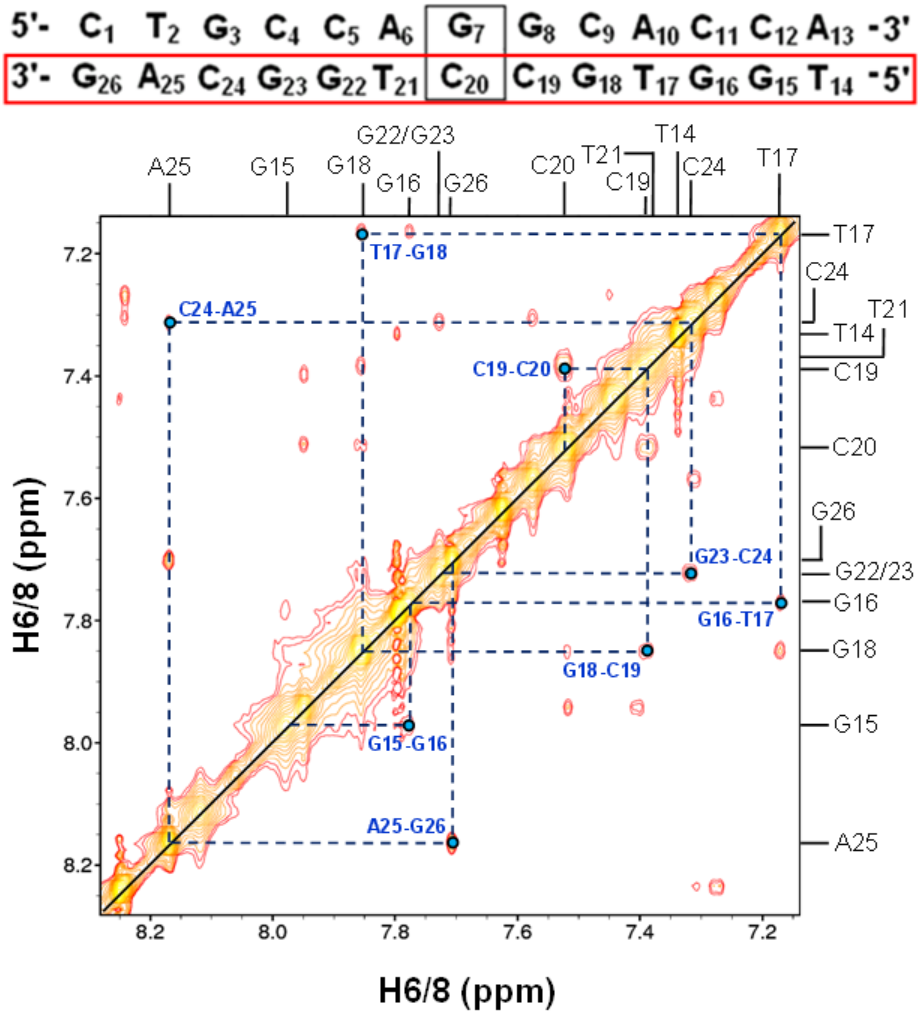
800 MHz NOESY (τ_m = 250ms) spectra showing interresidue H6/8 (i-1)-H5/CH_3 (i) NOE connectivities for residues T14-G26 in 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. Thymine and B. Cytosine connectivities. H5/CH_3(5)-H6 connectivities are indicated by blue circles and interresidue connectivities by purple circles.



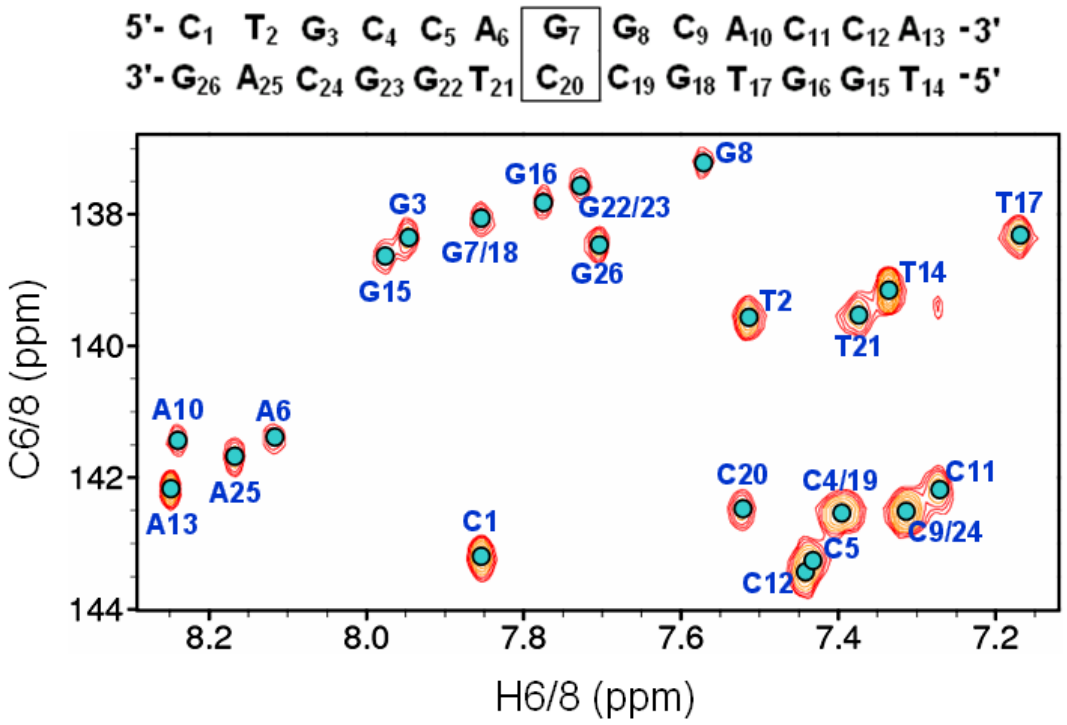
800 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H1' sequential NOE connectivities for residues T14-G26 in 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intraresidue connectivities are shown by blue circles and interresidue connectivities by purple circles; black circles indicate an overlap of connectivities.



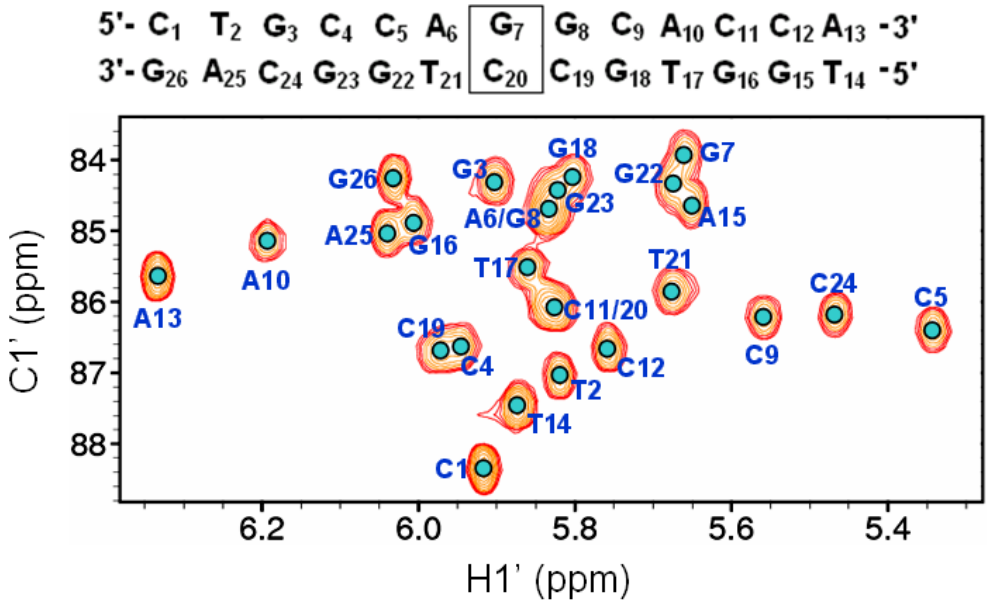
800 MHz NOESY (τ_m = 250ms) spectrum showing interresidue H6/8-H6/8 NOE connectivities for residues C1-A13 in 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



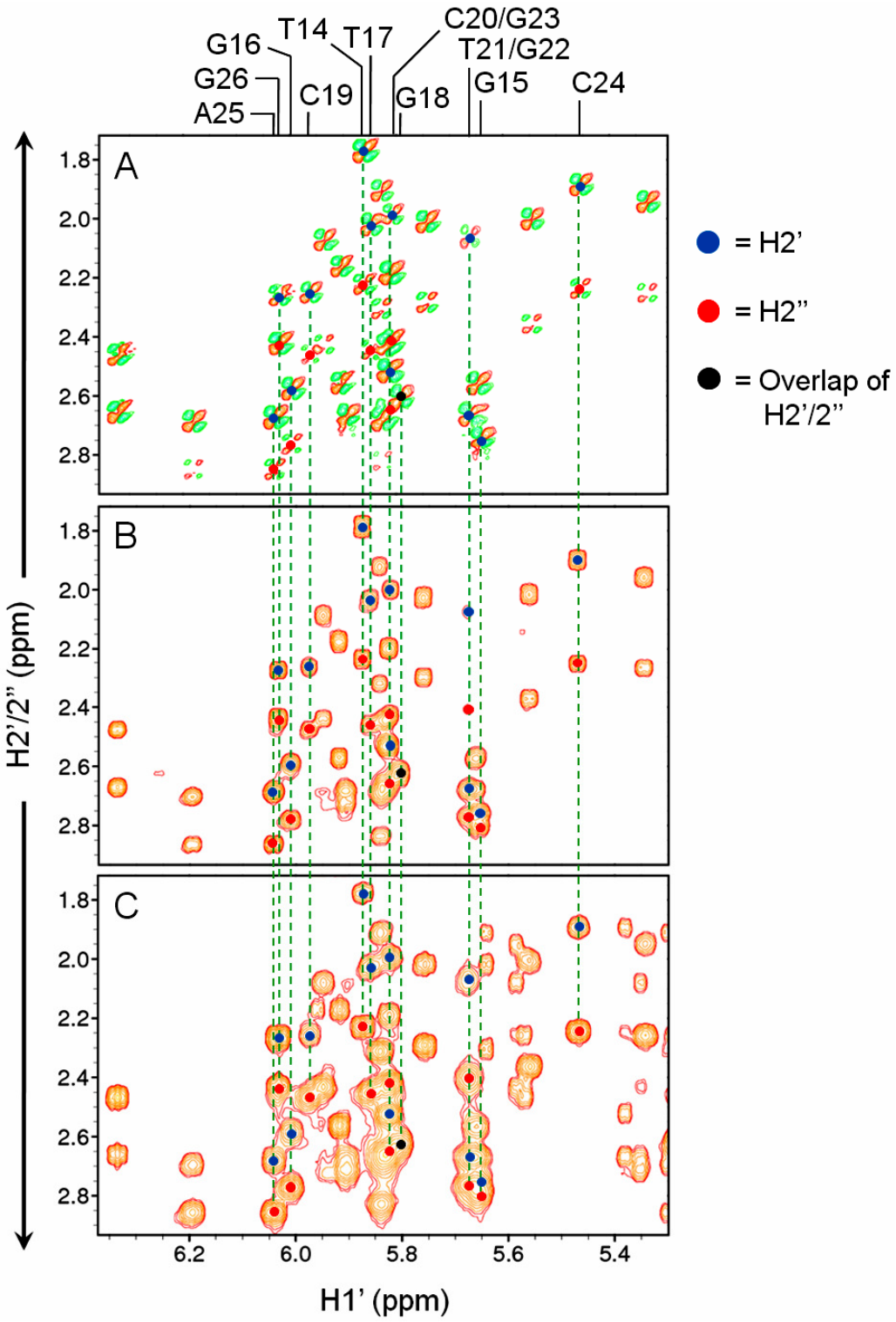
800 MHz NOESY (τ_m = 250ms) spectrum showing interresidue H6/8-H6/8 NOE connectivities for residues T14-G26 in 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



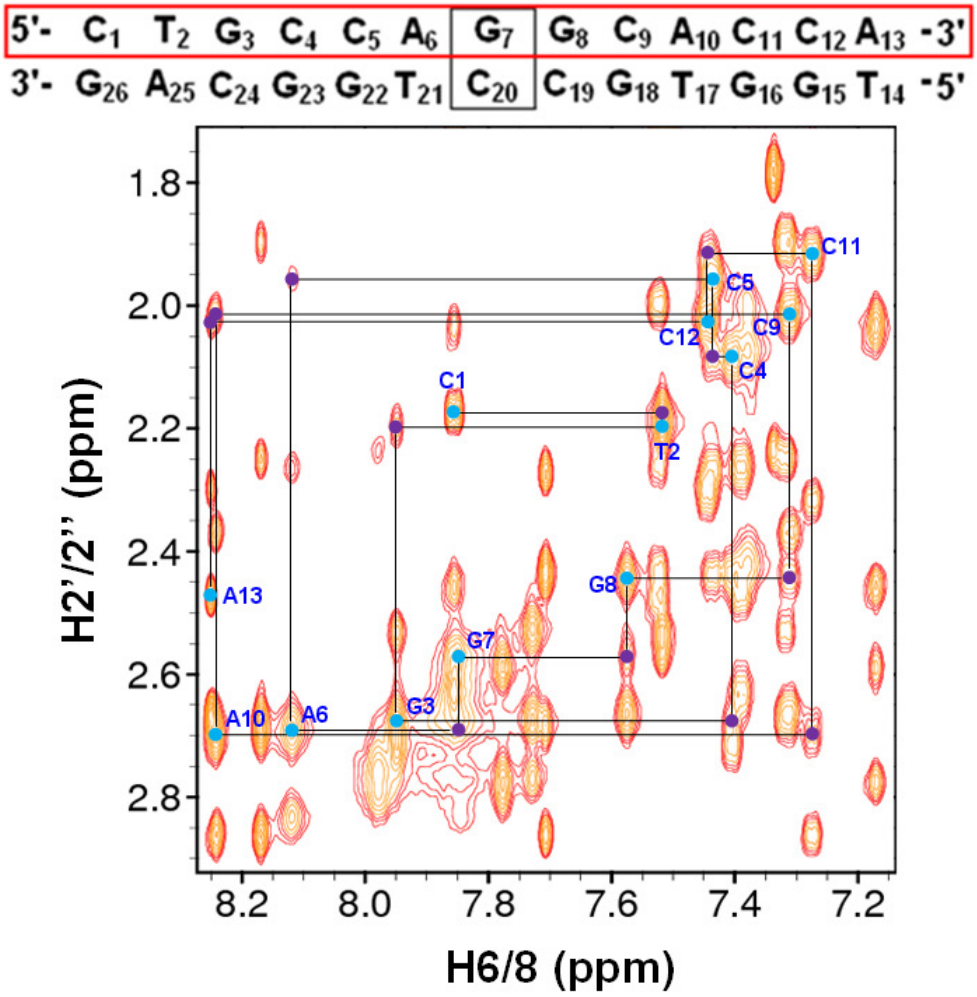
600 MHz ^{1}H-^{13}C HSQC spectrum of 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C showing the assignment of C6/8 resonances.



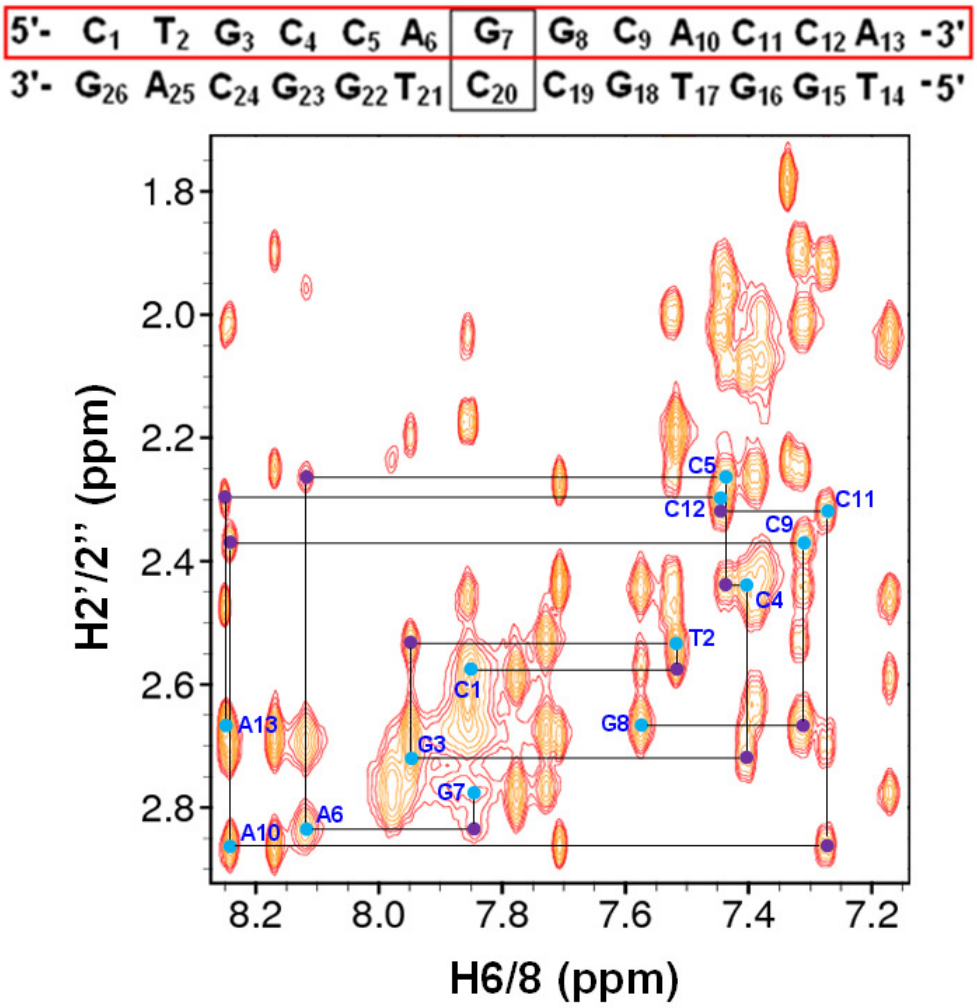
600 MHz ^{1}H-^{13}C HSQC spectrum of 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C showing the assignment of C1' resonances.



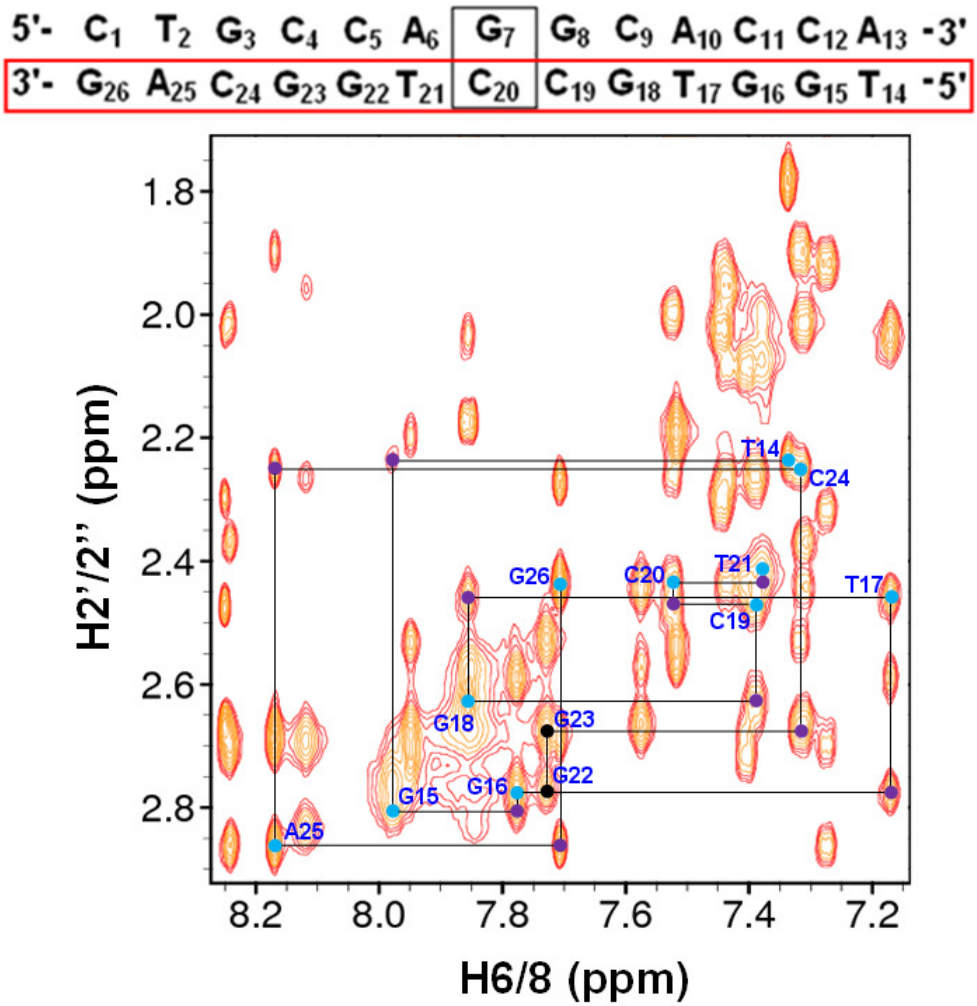
800 MHz spectrum stack showing identification and assignment of H2'/H2'' resonances for residues T14-G26 in 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C where A. DQF-COSY, B. TOCSY (SL_{mix} = 75ms) and C. NOESY (τ_m = 250ms).



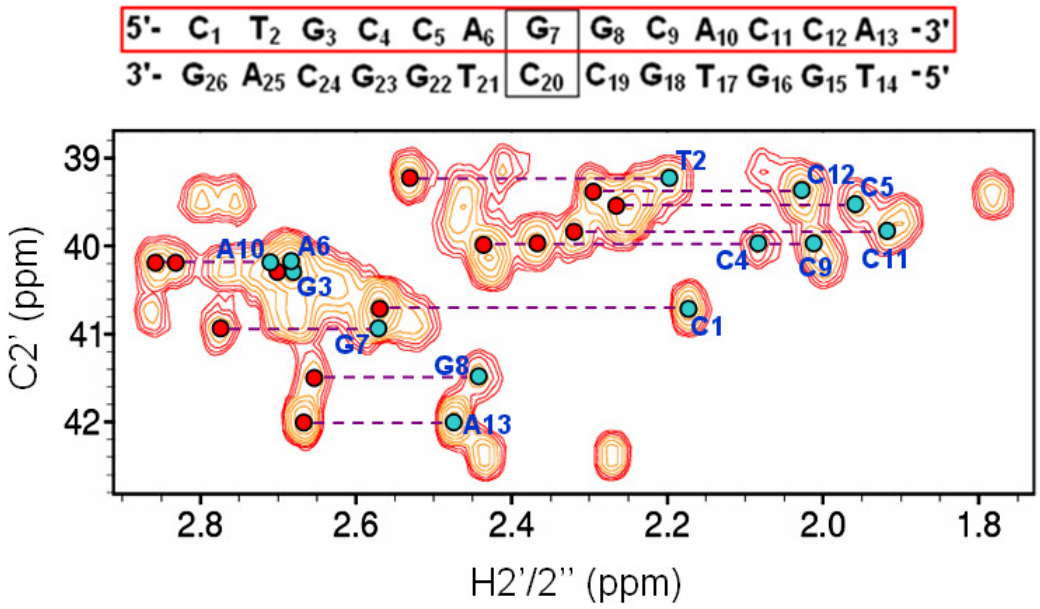
800 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2' sequential NOE connectivities for residues C1-A13 in 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intraresidue connectivities are shown by blue circles and interresidue connectivities by purple circles.



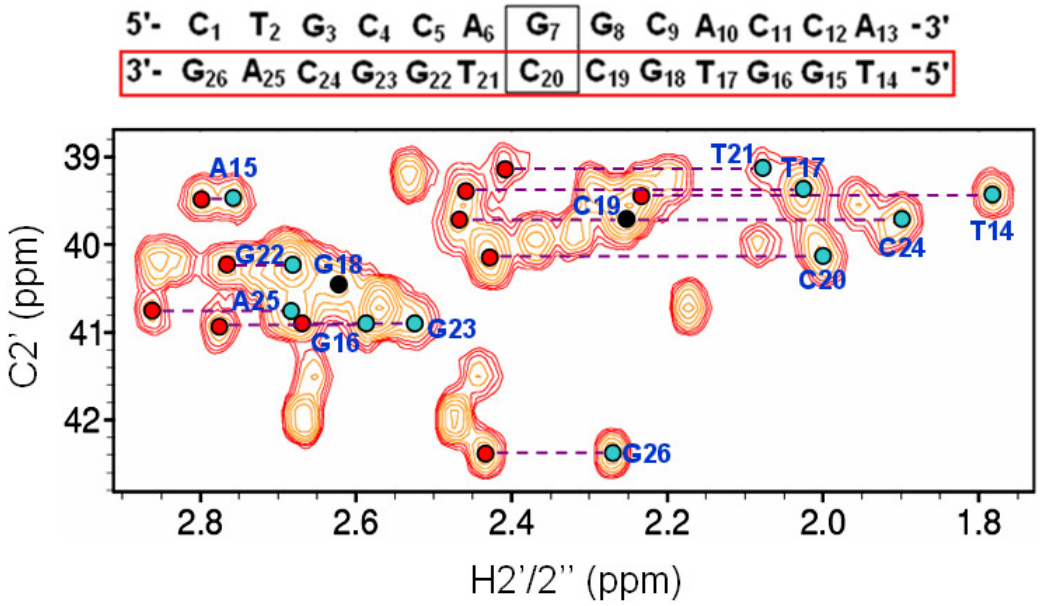
800 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2'' sequential NOE connectivities for residues C1-A13 in 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intraresidue connectivities are shown by blue circles and interresidue connectivities by purple circles.



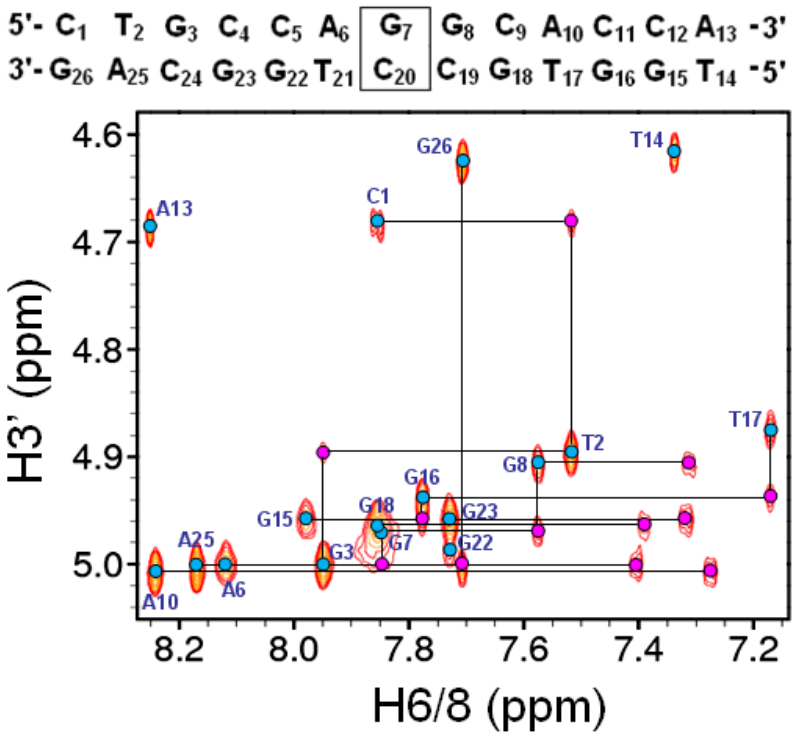
800 MHz NOESY (τ_m = 250ms) spectrum showing H6/8-H2'' sequential NOE connectivities for residues T14-G26 in 6-TG modified 13mer GC DNA (0.8mM, 100% ^2H_2O, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intraresidue connectivities are shown by blue circles and interresidue connectivities by purple circles; black circles indicate an overlap of connectivities.



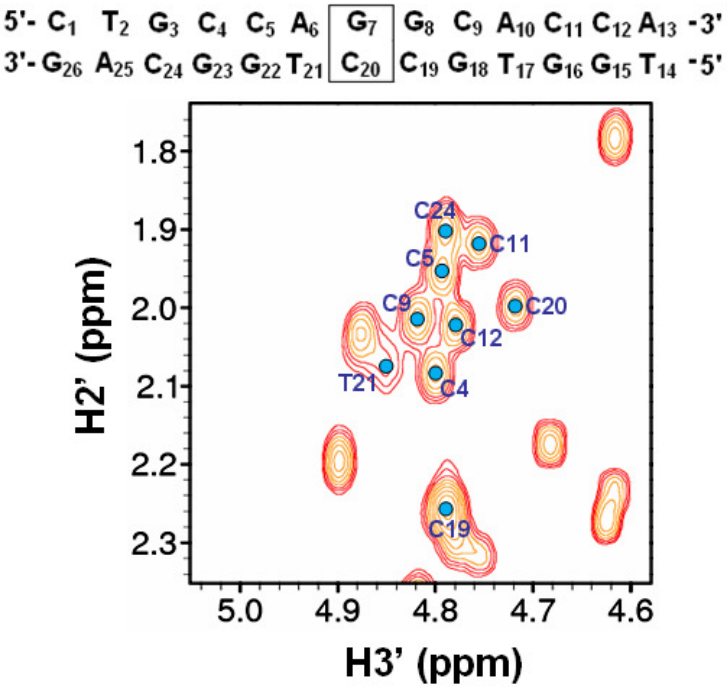
600 MHz ^{1}H-^{13}C HSQC spectrum of 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C showing the assignment of C2' resonances for residues C1-A13. H2' and H2'' are show by the blue and red circles respectively.



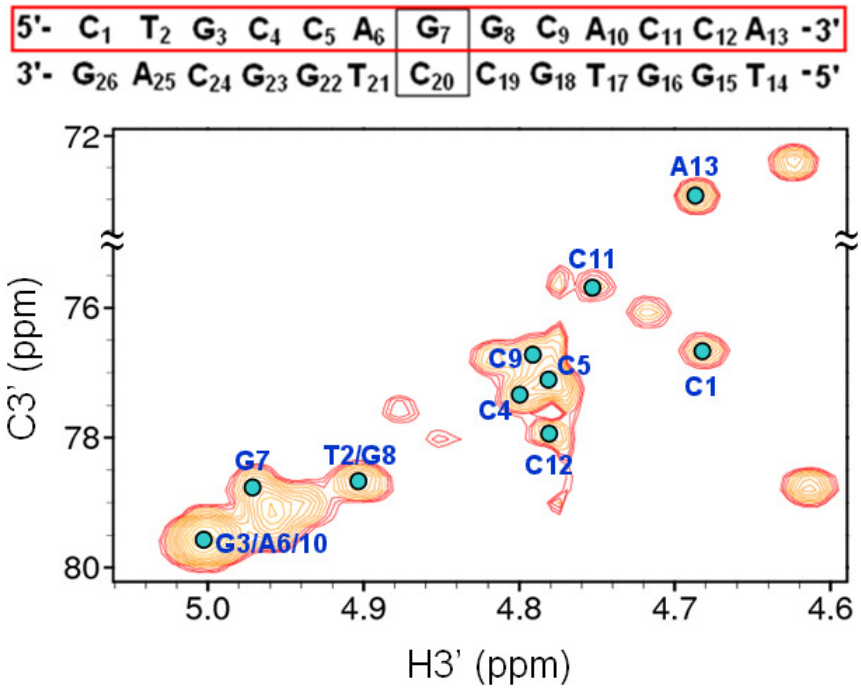
600 MHz ^{1}H-^{13}C HSQC spectrum of 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C showing the assignment of C2' resonances for residues T14-G26. H2' and H2'' are show by the blue and red circles respectively; black circles indicate an overlap of resonances.



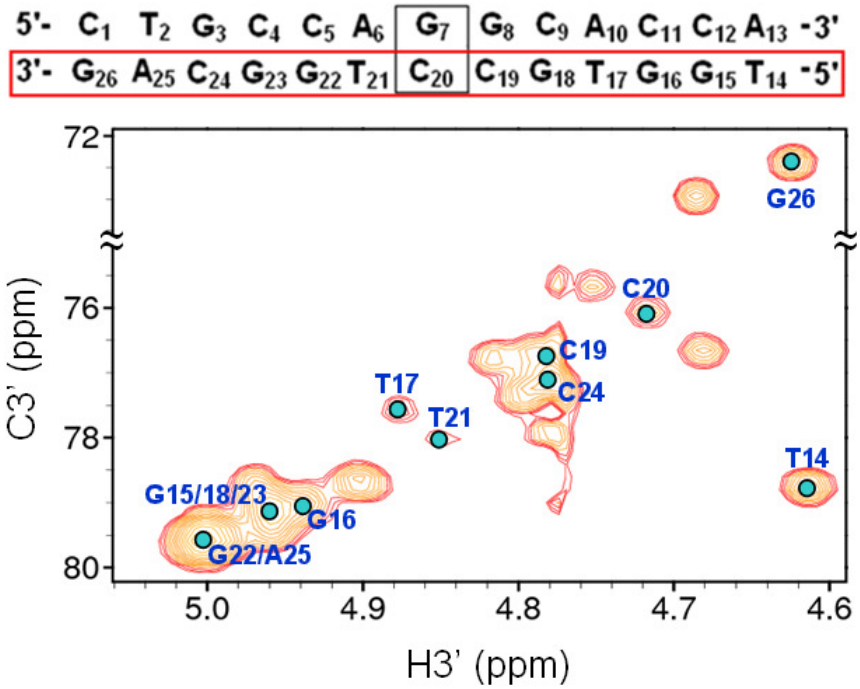
800 MHz NOESY (τ_{mix} = 250ms) spectrum showing H6/8-H3' NOE connectivities in 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C. Intra- and internucleotide NOEs are shown by blue and pink circles respectively.



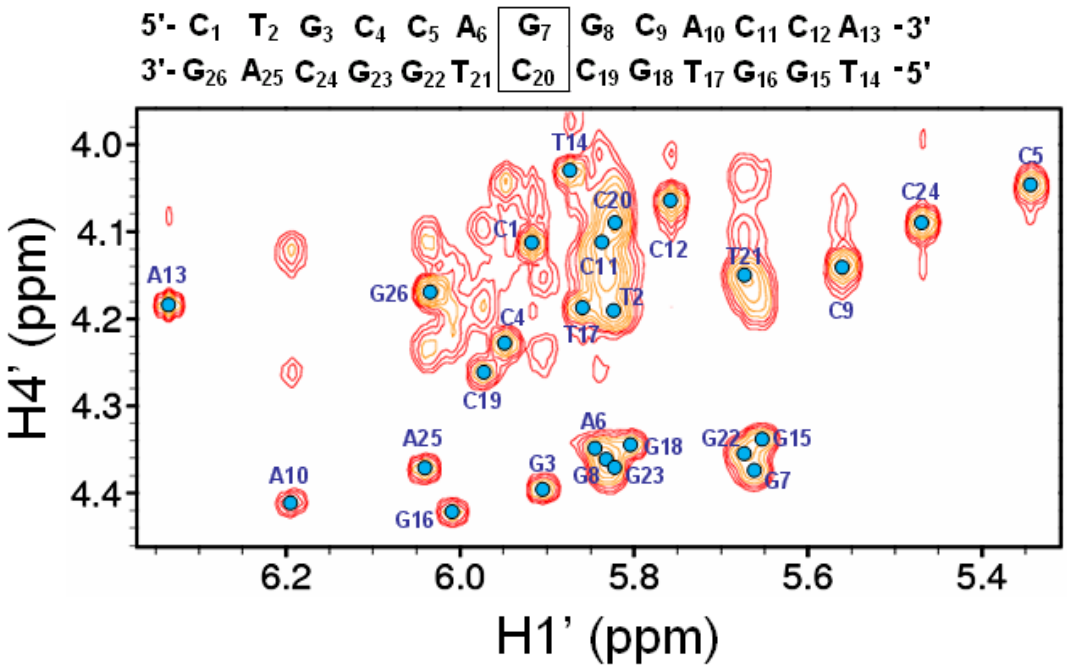
800 MHz NOESY (τ_{mix} = 250ms) spectrum showing H2'-H3' NOE connectivities in 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) for selected residues at 25°C.



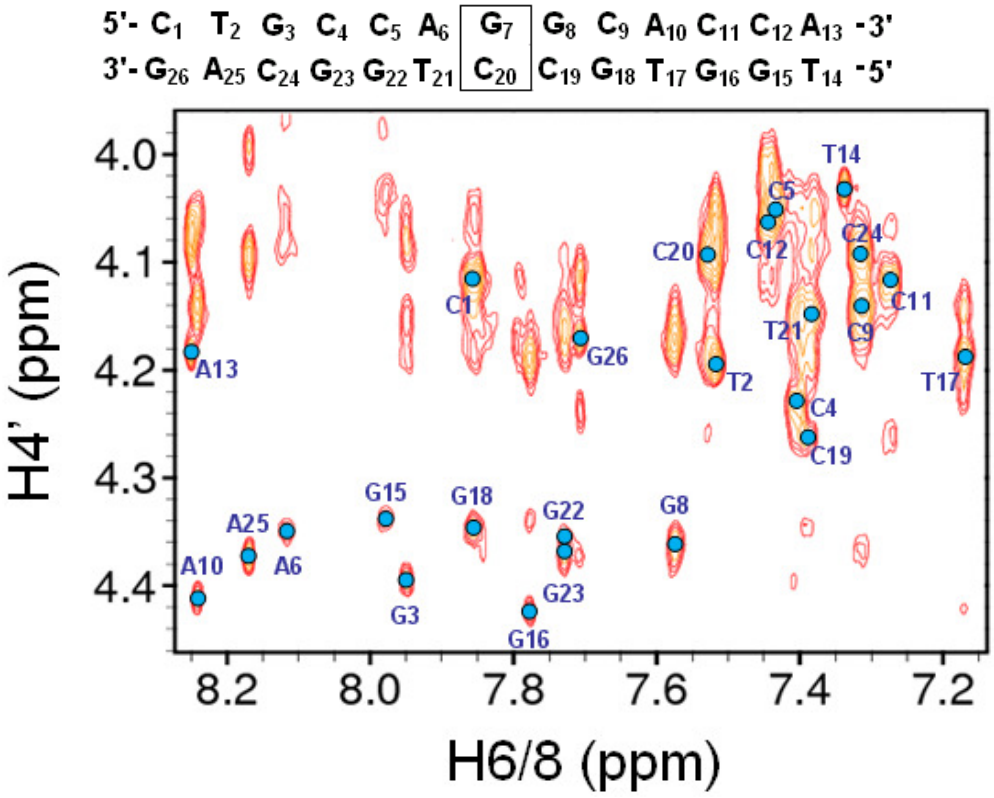
600 MHz ^{1}H-^{13}C HSQC spectrum of 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C showing the assignment of C3' resonances for residues C1-A13.



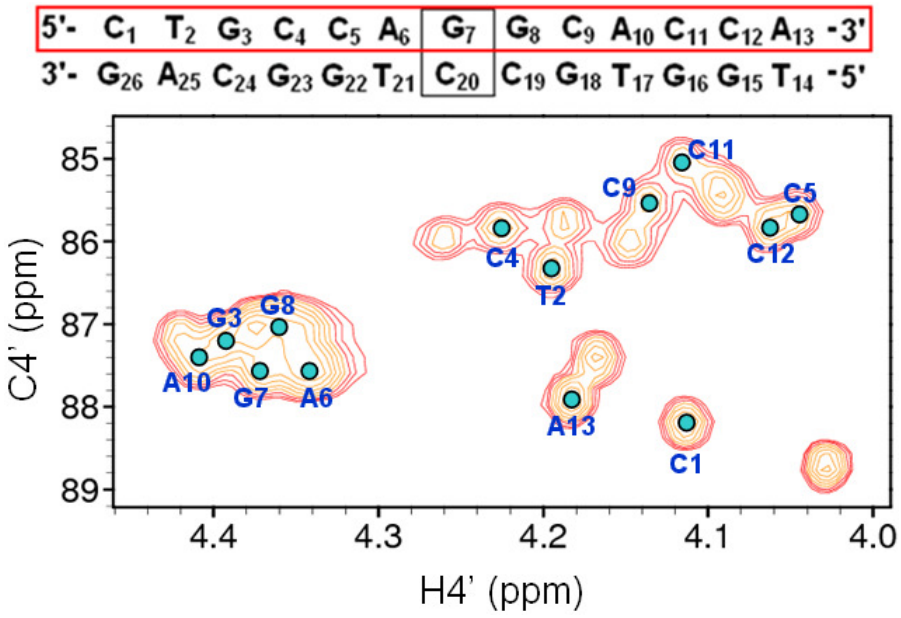
600 MHz ^{1}H-^{13}C HSQC spectrum of 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C showing the assignment of C3' resonances for residues T14-G26.



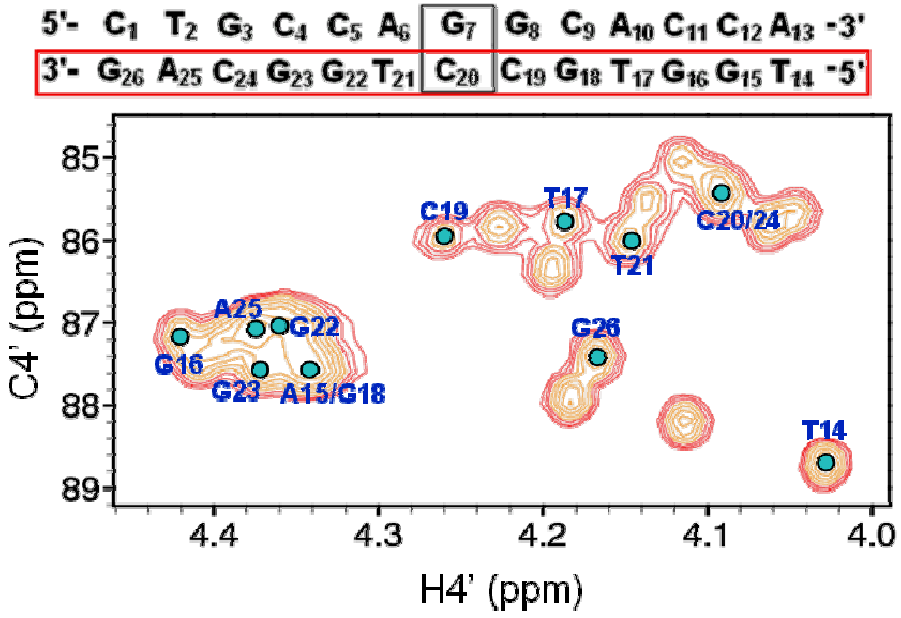
800 MHz NOESY (τ_{mix} = 250ms) spectrum showing H1'-H4' NOE connectivities in 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



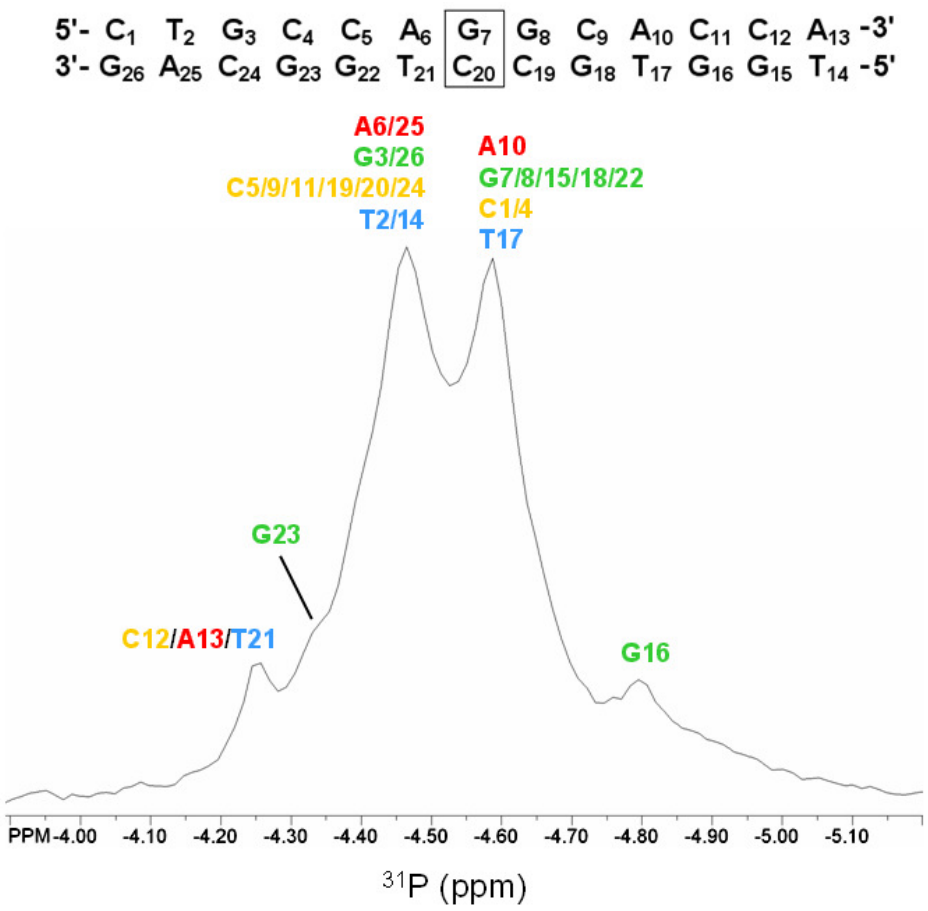
800 MHz NOESY (τ_{mix} = 250ms) spectrum showing H6/8-H4' NOE connectivities in 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C.



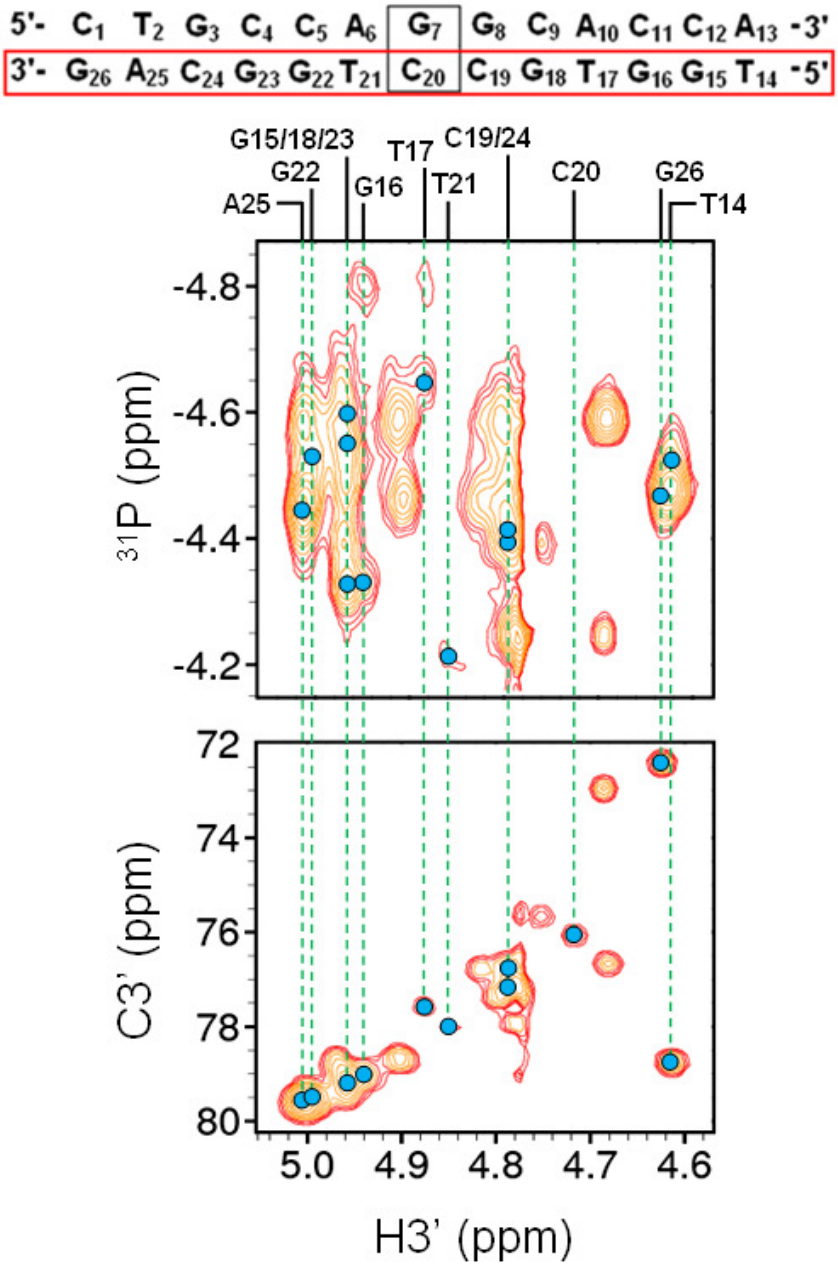
600 MHz ^{1}H-^{13}C HSQC spectrum of 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C showing the assignment of C4' resonances for residues C1-A13.



600 MHz ^{1}H-^{13}C HSQC spectrum of 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) at 25°C showing the assignment of C4' resonances for residues T14-G26.

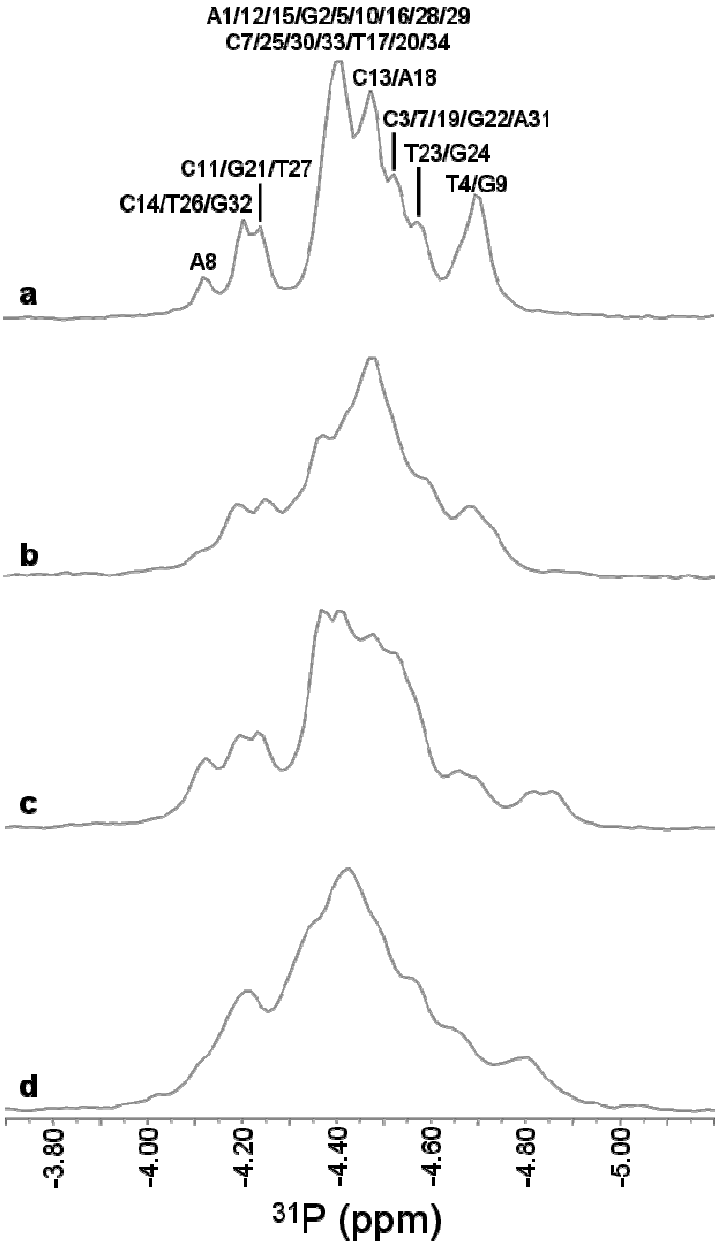


81 MHz ^{1}H decoupled 1D ^{31}P-NMR spectrum showing the assignment of ^{31}P resonances in 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2).

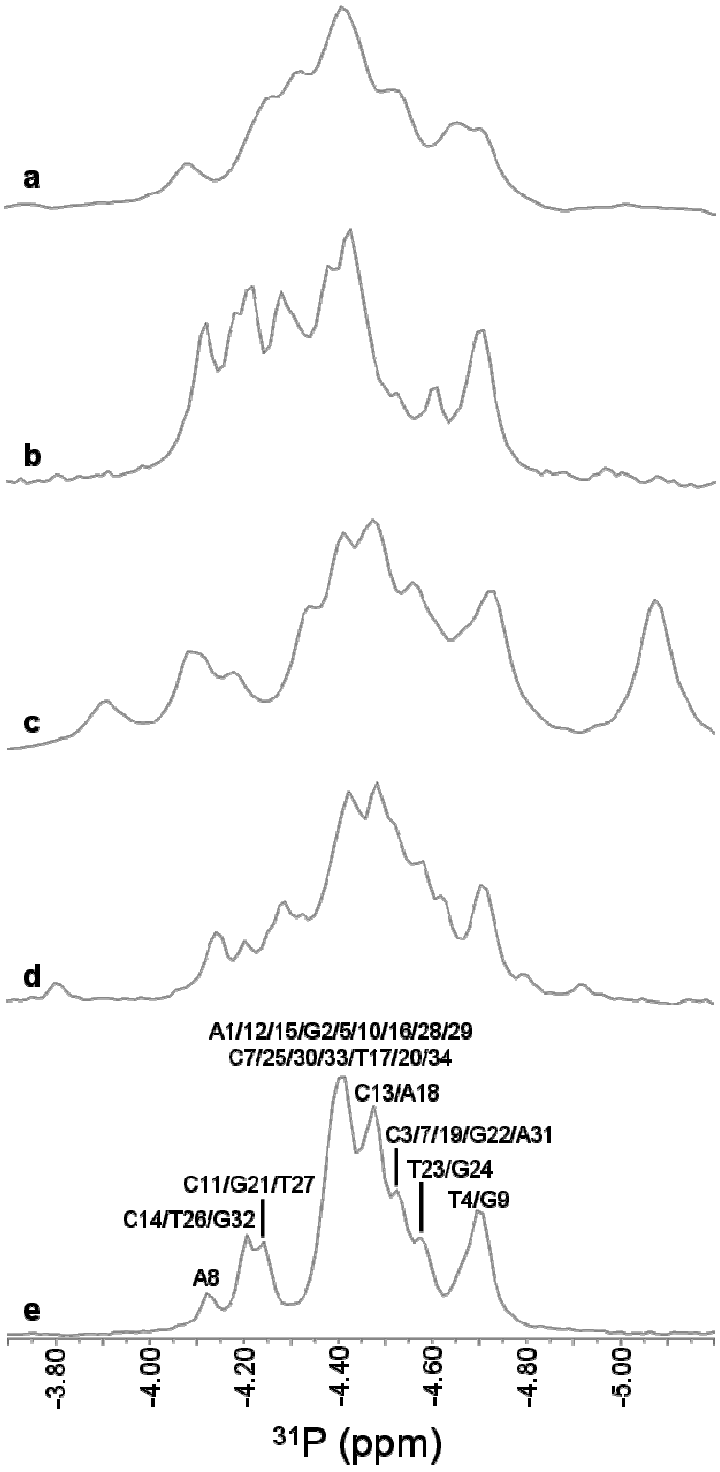


(Top panel) 600 MHz ^{1}H-^{31}P CPMG-HSQC-NOESY (τ_{mix} = 500ms) spectrum and (bottom panel) 600 MHz ^{1}H-^{13}C HSQC spectrum. The assignment of ^{31}P resonances is shown by means of H3'-^{31}P HSQC correlations for residues T14-G26 in 6-TG modified 13mer GC DNA (0.8mM, 100% $^{2}H_2O$, 50mM PO_4^{3-} and 50mM NaCl, pH 6.2) measured at 25°C.

v. Mismatch orientation, sequence context and oligonucleotide analogues



81MHz spectra stack of 1D ^{31}P-NMR 17mer mismatch GT DNA and its mismatch orientation and sequence context analogues where a.17mer GT, b. 17mer GT-FB, c. 17mer TG and d. 17mer TG-FB. The assignment of ^{31}P resonances in 17mer mismatch GT DNA is shown.



81MHz spectra stack of 1D ^{31}P-NMR 17mer mismatch GT DNA and its oligonucleotide length analogues where a.9mer GT, b. 11mer GT, c. 13mer GT, d. 15mer GT and e. 17mer GT DNAs. The assignment of ^{31}P resonances in 17mer mismatch GT DNA is shown.