CRANFIELD UNIVERSITY

William Joshua David Stebbeds

The Effect of Oxidation on the Stability of G-quadruplex DNA: Implications for Oncogene Expression

CRANFIELD HEALTH

MSc by Research
Academic Year: 2010 - 2011

Supervisors: Dr. Lee Larcombe & Prof. Joseph Lunec
October 2011

CRANFIELD UNIVERSITY

CRANFIELD HEALTH

MSc by Research

Academic Year 2010 - 2011

William Joshua David Stebbeds

The Effect of Oxidation on the Stability of G-quadruplex DNA: Implications for Oncogene Expression

Supervisors: Dr. Lee Larcombe & Prof. Joseph Lunec

October 2011

ABSTRACT

G-quadruplexes (G4-DNA) are a class of secondary structures formed from Guanine rich sequences. In recent years these structures have been implicated in both telomere maintenance and oncogene expression, and have been shown to be abundant in upstream promoter regions and at telomeric ends.

The mutagenic properties of oxidative stress on DNA have been widely studied, as has the association with carcinogenesis. The oxidation of deoxyguanosine to 8-oxo-2'deoxyguanosine (8-oxo-dG) is the most common result when DNA is under oxidative stress and as such, the G-rich sequences that form G-quadruplexes can be viewed as potential "hot-spots" for DNA oxidation. We propose that oxidation may destabilise the G-quadruplex structure, leading to its unfolding into the duplex structure, affecting gene expression. This would imply a possible mechanism by which oxidation may impact on oncogene expression.

This project used both *in silico* and *in vitro* methods to observe the effect of oxidation on the G-quadruplex structure and the consequences in oncogene expression, using two biologically relevant G-quadruplex structures, those found in the promoter regions of the proto-oncogenes c-Myc and c-Kit as proof of concept.

Molecular dynamics (MD) simulations were performed (isothermic, isobaric 500ns unrestrained simulation in explicit solvent and counterions) on the c-Kit and c-Myc G-quadruplex structures with and without 8-oxo-dG incorporated into the central tetrad. FRET experiments were performed on these same structures, observing the conformation of sequences known to form G-quadruplexes under near physiological conditions and subjected to oxidative stress, through Fenton chemistry. Gene expression data analyses were also performed to evaluate the prevalence of different G-quadruplex forming motifs (GQMs) in genes affected by oxidation.

Although no relevant information was gained from the FRET experiments, the MD results constitute the longest simulations of this type performed on the c-Myc and c-Kit G-quadruplex structures published to date and predict the high stability of these structures under normal physiological conditions. They also clearly demonstrate a destabilising effect of oxidation on G-quadruplex structures, with the extent of the effect dependent on the structure oxidised.

Furthermore, gene expression data analysis showed that genes whose expression is significantly altered when subjected to oxidative stress are statisticallymore likely to contain a GQM than the remainder of the genome, through the use of significance testing.

These findings demonstrate a differential effect of oxidation on G-quadruplexes, likely dependent on other known characteristics affecting G4 stability such as loop length and sequence. Results also point towards this mechanism affecting gene expression. This is suggestive of a novel route for oxidation mediated carcinogenesis, through upregulation of oncogene expression or possibly downregulation of tumour suppression genes.

Keywords:

G-quadruplexes; Oxidative Stress; oncogene expression; Carcinogenesis; c-Myc; c-Kit.

ACKNOWLEDGEMENTS

Thank you to Dr. Lee Larcombe, the perfect supervisor, who gave me the freedom and responsibility to make this project my own. You know how much you helped me, and it is obvious that I would not have been half as successful nor had half as much fun with any other supervisor.

Thank you to Professor Joseph Lunec, for funding this project and for his advice throughout this year. Your expertise in this field was invaluable.

Thank you to all the guys and gals in the office and lab for all the laughs, help and many hours of procrastination – amongst them, Paul and Adeel stand out as the best of friends and time wasters.

Thanks also to mum, pops and sis for all your love and support.

Last, but certainly not least, thank you Marta. Thank you for letting me vent at you and putting up with my ups and downs and thank you most of all for making it through the last two years with me.

TABLE OF CONTENTS

ABSTRACT iii
ACKNOWLEDGEMENTS v
TABLE OF CONTENTS vi
LIST OF FIGURES viii
LIST OF TABLES xiii
LIST OF ABBREVIATIONS xiv
1 Literature Review 1
1.1 Introduction to G-Quadruplexes 1
1.1.1 G-Quadruplex Structures 3
1.1.2 Biological Relevance of G-Quadruplexes 10
1.2 Modification of G-Quadruplexes 24
1.2.1 Ligand-G-Quadruplex Interactions 24
1.2.2 Modification to G-Quadruplex Nucleotides 27
1.3 Applications and Uses of G-quadruplexes in Supramolecular Chemistry and Nanotechnology 35
1.4 Methods to determine presence of and characterise G-quadruplexes 36
1.4.1 Fluorescence Resonance Energy Transfer 37
1.4.2 Molecular Dynamics Simulations of G-Quadruplex Structures 38
1.4.3 Gene Expression Data Analyses 43
1.5 Aims and Objectives 44
2 Experimental section 46
2.1 Fluorescence Resonance Energy Transfer 46
2.1.1 Determination of Presence of c-Myc G-Quadruplex Structure 48
2.1.2 Initial Oxidation Experiments 50
2.1.3 Determination of Presence of c-Kit G-Quadruplex Structure 52
2.1.4 Optimised Experiments 57
2.2 Molecular Modelling 69
2.2.1 Optimisation of Molecular Modelling Protocol and Parameters 70
2.2.2 Force Field Optimisation 79
2.2.3 Analysis of Results 82
2.2.4 Simulations Performed on c-Kit G-quadruplex Structures 83
2.2.5 Simulations Performed on c-Kit B-DNA Structures 93
2.2.6 Simulations Performed on c-Myc G-Quadruplex Structures 100
2.2.7 Simulations Performed on c-Myc B-DNA Structures 108
2.3 Gene Expression Data Analyses 114
2.3.1 Analysis of the Prevalence of c-Myc and c-Kit Similar Motifs in Genic Regions of the Human Genome 116
2.3.2 Analysis of Expression Profiles of Cells Subjected to Oxidation for the Presence of GQMs 117
3 Discussion 124
3.1 Fluorescence Resonance Energy Transfer Experiments 124
3.1.1 Overview of results 124
3.1.2 Critical Appraisal of Materials and Methods 127
3.2 Molecular Dynamics Simulations 131

3.2.1 Overview of Results 131
3.2.2 Critical Appraisal of Materials and Methods 135
3.3 Gene Expression Data Analyses 142
3.3.1 Overview of Results 142
3.3.2 Critical Appraisal of Materials and Methods 145
3.4 General Overview of Biological Implications 146
4 Conclusions & Further Work 149
5 Bibliography 153
6 APPENDICES 189

LIST OF FIGURES

Figure 1-1: Chemical representation of a G-tetrad (left); and schematic representation of a G-quadruplex consisting of three G-tetrads (right) [5]. ... 1
Figure 1-2: Comparison of Watson-Crick base pairing (G-C) and Hoogsteen base pairing (G-G). ... 3
Figure 1-3: Schematic representations of the different structures tetramolecular G-quadruplexes can adopt, denoting strand polarity and associated glycosidic torsional angles of the guanines. Adapted from [10]. ... 6
Figure 1-4: Schematic representations of the different structures bimolecular G-quadruplexes can adopt, denoting strand polarity and associated glycosidic torsional angles of the Guanines. Adapted from [10]. ... 6
Figure 1-5: G-Quadruplex motif as used by Zhang and co-workers [31] (left) and example of a unimolecular G-Quadruplex (right). ... 7
Figure 1-6: Different topologies adopted by unimolecular G-quadruplexes. Adapted from (10). ... 8
Figure 1-7: Model for the regulation of gene expression by G-quadruplexes in upstream promoter regions. Adapted from (26). ... 17
Figure 1-8: Oncogene associated G-quadruplexes categorised in accordance with the "hallmarks of cancer" [107]. Adapted from (32). ... 18
Figure 1-9: Schematic Representation of the G-quadruplex structure found in the NHE III of c-Myc. Adapted From (67). ... 20
Figure 1-10: Model for the activation and repression of c-Myc expression by stabilisation of a G-quadruplex structure within the NHE III region with the use of ligands. Adapted from (118). ... 21
Figure 1-11: Schematic representation of the c-Kit87up G4 structure. Adapted from (125). ... 22
Figure 1-12: Proposed model for a G-wire. Adapted from (230). ... 35
Figure 1-13: Schematic representation of the smFRET experiment carried out by Shirude and co-workers (2007) [251]. ... 38
Figure 2-1: Sequence of the synthesized c-Myc GQM and complementary sequence used in the FRET Assays. ... 47
Figure 2-2: FRET values observing effect of concentrations of complementary sequence and KCL, plotted against temperature. ... 48
Figure 2-3: Cy5 fluorescence values observing effect of concentrations of complementary sequence and KCL, plotted against temperature. ... 49
Figure 2-4: FRET values observing the effect of H_2O_2 concentration. ... 50
Figure 2-5: Cy5 fluorescence values observing the effect of H_2O_2 concentration. ... 51
Figure 2-6: Sequence of the synthesized c-Kit GQMs and complementary sequences used in the FRET Assays. ... 52
Figure 2-7: FRET values observing the effect of 100 mM KCL. ... 53
Figure 2-8: Cy5 (top) and Cy3 (bottom) fluorescence values observing the effect of 100 mM KCL. ... 54
Figure 2-9: FRET values observing the effect of cDNA concentration. ... 55

Figure 2-10: Cy5 fluorescence values observing the effect of cDNA concentration..... 55
Figure 2-11: Cy3 fluorescence values observing the effect of cDNA concentration..... 56
Figure 2-12: Sequence of the synthesized c-Kit control complementary sequences used in the FRET Assays. 56
Figure 2-13: FRET values observing effect of Hydrogen peroxide and complementary oligonucleotide concentration. 59
Figure 2-14: Cy5 fluorescence values observing effect of Hydrogen peroxide and complementary oligonucleotide concentration raw data (above) and data normalised to control values (sample 1) (below). 61
Figure 2-15: Cy3 fluorescence values observing effect of Hydrogen peroxide and complementary oligonucleotide concentration raw data..... 61
Figure 2-16: FRET values observing effect of Hydrogen peroxide and complementary oligonucleotide concentration, normalised to control and Cy5 fluorescence. 62
Figure 2-17: FRET values comparing the effect of H_2O_2 and Fenton's reaction. 64
Figure 2-18: FRET values observing the effect of ascorbic acid and $FeCl_2$ on fluorescence..... 64
Figure 2-19: FRET values observing effect Fenton's reagent using different concentrations of Hydrogen peroxide..... 66
Figure 2-20: Cy5 fluorescence values for oxidation threshold experiment raw data (above) and data normalised to control values (below). 67
Figure 2-21: Cy3 raw data values from oxidation threshold experiment. 68
Figure 2-22: FRET values from the oxidation threshold experiment, normalised to control and Cy5 fluorescence. 69
Figure 2-23: Molecular Modelling Overview Flow Chart. 70
Figure 2-24: Overview Flow Chart of the Equilibration Protocol 75
Figure 2-25: Comparison of unmodified Deoxyguanosine (left) and 8-oxo-dG (right)..... 81
Figure 2-26: Representative snapshots of the cluster analysis of the non-oxidised c-Kit G-quadruplex. Top left: Cluster of 194 frames representative of the first 100ns; Top centre: Cluster of 99 frames representative of the second 100ns; Top right: Cluster of 256 frames representative of the third 100ns; Bottom left Cluster of 119 frames representative of the fourth 100ns; Bottom right: Cluster of 291 frames representative of the last 100ns. 85
Figure 2-27: Representative snapshots of the cluster analysis of the oxidised c-Kit G-quadruplex. Top left: Cluster of 223 frames representative of the first 100ns; Top centre: Cluster of 108 frames representative of the second 100ns; Top right: Cluster of 448 frames representative of the third 100ns; Bottom left Cluster of 119 frames representative of the fourth 100ns; Bottom right: Cluster of 139 frames representative of the last 100ns. 86
Figure 2-28: Focused snapshot of the Guanines involved in tetrad formation in the final conformation of the oxidised c-Kit G-quadruplex. 86
Figure 2-29: Root mean square deviation (RMSD) comparison of c-Kit G-quadruplex structures with and without 8-oxo-dG incorporated. 87

Figure 2-30: Root mean square deviation (RMSD) comparison of the Guanines involved in tetrad formation in the c-Kit G-quadruplex structures with and without 8-oxo-dG incorporated. 88
Figure 2-31: Root mean square fluctuation (RMSF) comparison of c-Kit G-quadruplex structures with and without 8-oxo-dG incorporated. 89
Figure 2-32: Comparison of the number of Hoogsteen Hydrogen bonds in the c-Kit G-quadruplex structures with and without 8-oxo-dG. 89
Figure 2-33: Solvent accessible surface area of each residue in the non-oxidised and oxidised c-Kit G-quadruplex Structures. 90
Figure 2-34: Graph plotting percentage of motility explained by eigenvalues against eigenvectors (principal components) of non-oxidised and oxidised c-Kit G-quadruplex structure and cumulative percentage. 91
Figure 2-35: Extreme structures and RMSF values of the first (top) and second (bottom) eigenvectors of the non-oxidised c-Kit G-quadruplex..... 92
Figure 2-36: Extreme structures and RMSF values of the first (top) and second (bottom) eigenvectors of the oxidised c-Kit G-quadruplex. 93
Figure 2-37: Representative snapshots of the cluster analysis of the non-oxidised c-Kit B-DNA molecule. Top left: Cluster of 81 frames representative of the first 100ns; Top right: Cluster of 136 frames representative of the second 100ns; Centre left: Cluster of 92 frames representative of the third 100ns; Centre right: Cluster of 49 frames representative of the fourth 100ns; Bottom: Cluster of 75 frames representative of the last 100ns. 94
Figure 2-38: Representative snapshots of the cluster analysis of the oxidised c-Kit B-DNA molecule. Top left: Cluster of 178 frames representative of the first 100ns; Top right: Cluster of 98 frames representative of the second 100ns; Centre left: Cluster of 85 frames representative of the third 100ns; Centre right: Cluster of 87 frames representative of the fourth 100ns; Bottom: Cluster of 146 frames representative of the last 100ns. 95
Figure 2-39: Focused snapshot of the base pairing/orientation of the oxidised Guanine and its complementary Cytosine at the end of the simulation. 95
Figure 2-40: Root mean square deviation (RMSD) comparison of c-Kit B-DNA structures with and without 8-oxo-dG incorporated. 96
Figure 2-41: Root mean square fluctuation (RMSF) comparison of c-Kit B-DNA structures with and without 8-oxo-dG incorporated. 96
Figure 2-42: Comparison of the number of Hydrogen bonds in the c-Kit B-DNA structures with and without 8-oxo-dG. 97
Figure 2-43: Solvent accessible surface area of the non-oxidised and oxidised c-Kit B-DNA molecules. 97
Figure 2-44: Graph plotting percentage of motility explained by eigenvalues against eigenvectors (principal components) of non-oxidised and oxidised c-Kit B-DNA molecule and cumulative percentage. 98
Figure 2-45: Extreme structures and RMSF values of the first (left) and second (right) eigenvectors of the non-oxidised c-Kit B-DNA molecule. 99
Figure 2-46: Extreme structures and RMSF values of the first (left) and second (right) eigenvectors of the oxidised c-Kit B-DNA Molecule. 99

Figure 2-47: Representative snapshots of the cluster analysis of the non-oxidised c-Myc G-quadruplex. Top left: Cluster of 166 frames representative of the first 100ns; Top centre: Cluster of 312 frames representative of the second 100ns; Top right: Cluster of 243 frames representative of the third 100ns; Bottom left Cluster of 249 frames representative of the fourth 100ns; Bottom right: Cluster of 158 frames representative of the last 100ns. 101
Figure 2-48: Representative snapshots of the cluster analysis of the oxidised c-Myc G-quadruplex. Top left: Cluster of 133 frames representative of the first 100ns; Top centre: Cluster of 216 frames representative of the second 100ns; Top right: Cluster of 168 frames representative of the third 100ns; Bottom left Cluster of 158 frames representative of the fourth 100ns; Bottom right: Cluster of 134 frames representative of the last 100ns. 102
Figure 2-49: Root mean square deviation (RMSD) comparison of c-Myc G-quadruplex structures with and without 8-oxo-dG incorporated. 103
Figure 2-50: Root mean square deviation (RMSD) comparison of the Guanines involved in tetrad formation in the c-Myc G-quadruplex structures with and without 8-oxo-dG incorporated. 103
Figure 2-51: Root mean square fluctuation (RMSF) comparison of the c-Myc G-quadruplex structures with and without 8-oxo-dG incorporated. 104
Figure 2-52: Comparison of the number of Hydrogen bonds in the c-Myc G-quadruplex structures with and without 8-oxo-dG. 104
Figure 2-53: Solvent accessible surface area of the non-oxidised and oxidised c-Myc G-quadruplex structures. 105
Figure 2-54: Graph plotting percentage of motility explained by eigenvalues against eigenvectors (principal components) of the non-oxidised and oxidised c-Myc G-quadruplex structure and cumulative percentage. 106
Figure 2-55: Extreme structures and RMSF values of the first (top) and second (bottom) eigenvectors of the non-oxidised c-Myc G-quadruplex. 107
Figure 2-56: Extreme structures and RMSF values of the first (top) and second (bottom) eigenvectors of the oxidised c-Myc G-quadruplex. 108
Figure 2-57: Representative snapshots of the cluster analysis of the non-oxidised c-Myc B-DNA molecule. Top left: Cluster of 59 frames representative of the first 100ns; Top right: Cluster of 114 frames representative of the second 100ns; Centre left: Cluster of 77 frames representative of the third 100ns; Centre right: Cluster of 130 frames representative of the fourth 100ns; Bottom: Cluster of 95 frames representative of the last 100ns. 109
Figure 2-58: Representative snapshots of the cluster analysis of the non-oxidised c-Myc B-DNA molecule. Top left: Cluster of 164 frames representative of the first 100ns; Top right: Cluster of 97 frames representative of the second 100ns; Centre left: Cluster of 73 frames representative of the third 100ns; Centre right: Cluster of 96 frames representative of the fourth 100ns; Bottom: Cluster of 133 frames representative of the last 100ns. 109
Figure 2-59: Root mean square deviation (RMSD) comparison of c-Myc B-DNA structures with and without 8-oxo-dG incorporated. 110

Figure 2-60: Root mean square fluctuation (RMSF) comparison of c-Myc B-DNA structures with and without 8-oxo-dG incorporated. 111
Figure 2-61: Comparison of the number of Hydrogen bonds in the c-Myc B-DNA structures with and without 8-oxo-dG. 111
Figure 2-62: Solvent accessible surface area of the non-oxidised and oxidised c-Myc B-DNA molecules. 112
Figure 2-63: Graph plotting percentage of motility explained by eigenvalues against eigenvectors (principal components) of non-oxidised and oxidised c-Myc B-DNA molecule and cumulative percentage. 112
Figure 2-64: Extreme structures and RMSF values of the first (left) and second (right) eigenvectors of the non-oxidised c-Myc B-DNA molecule. 113
Figure 2-65: Extreme structures and RMSF values of the first (left) and second (right) eigenvectors of the oxidised c-Myc B-DNA Molecule..... 114

LIST OF TABLES

Table 1-1: Newtonian laws of motion 40
Table 2-1: Fluorescence measurement channels used in all experiments on the Rotor Gene Q PCR machine. 47
Table 2-2: Samples used in optimised protocol. 57
Table 2-3: Samples used in experiment evaluating effect of Fenton's reaction components 63
Table 2-4: Samples used in oxidation threshold experiment 65
Table 2-5: Sequences for the c-Myc and c-Kit B-DNA molecules. 71
Table 2-6: Particle Interaction Parameters Used. 76
Table 2-7: Summary of the parameters used in the production simulation runs. Descriptions taken from GROMACS online manual (http://manual.GROMACS.org/) 77
Table 2-8: GROMACS tools used to generate data on the MD simulations. ... 82
Table 2-9: G-quadruplex motifs used in bioinformatics studies. 116
Table 2-10: Number and percentage of the different GQMs found in the different locations in the Gene 117
Table 2-11: Gene expression data analysis of genes significantly affected by both H_2O_2 and Menadione treatment. 120
Table 2-12: Gene expression data analysis of genes significantly affected by H_2O_2 treatment. 120
Table 2-13: Gene expression data analysis of genes significantly affected by Menadione treatment. 121
Table 2-14: Gene expression data analysis of gene clusters not involved in DNA damage repair significantly affected by H2O2 treatment. 123
Table 2-15: Gene expression data analysis of gene clusters not involved in DNA damage repair significantly affected by Menadione treatment. 123

LIST OF ABBREVIATIONS

µs - microsecond

8-oxo-dG - 8-hydroxy-2’-deoxyguanosine/8-oxo-7,8-dihydro-2’deoxyguanosine

A – Adenine

Å – Angstrom

AFM – Atomic Force Microscopy

AMBER – Assisted Model Building with Energy Refinement

BER - Base Excision Repair

BLM - Bloom Protein/Gene

BS - Bloom Syndrome

C - Cytosine/Carbon

CD - Circular Dichroism

Chi^2 – Chi-squared

CNBP - Cellular Nucleic Acid Binding Protein

DNA - Deoxyribonucleic Acid

FapydG - 2,6-diamino-4-hydroxy-5-formamidopyrimidine

FRET - Forster Resonance Energy Transfer/Fluorescence Resonance Energy Transfer

G – Guanosine/Guanine

G4 – G-quadruplex

G_4-DNA - G-Quadruplex DNA

GMP - Guanosine Monophosphate

GQM - G-Quadruplex Forming Motif

GROMACS – Groningen Machine for Chemical Simulations

GSH - Glutathione

H - Hydrogen

H_2O_2 – Hydrogen Peroxide

hnRNPK - Heterogeneous Nuclear Ribonucleoprotein K

hTERT - Human Telomerase Reverse Transcriptase

K^+ - Potassium Ion

LINCS – Linear Constraint Solver

MD - Molecular Dynamics

Mg^{2+} - Magnesium Ion

mRNA - Messenger Ribonucleic Acid

N - Nitrogen

Na^+ - Sodium Ion

NHE - Nuclease Hypersensitivity Element

nm – nanometre

NMM - N-methyl mesoporphyrin IX

NMR - Nuclear Magnetic Resonance

ns – Nanosecond

O – Oxygen

ORF – Open Reading Frame

PAGE – Poly-Acrylamide Gel Electrophoresis

PCA – Principle Component Analysis

PERL – Practical Extraction and Report Language

PME – Particle Mesh Ewald

PNA - Peptide Nucleic Acid

rDNA - Ribosomal DNA

RFU – Relative Fluorescence Unit

RMSD – Root Mean Square Deviation

RMSF – Root Mean Square Fluctuation

RNA - Ribonucleic Acid

ROS - Reactive Oxygen Specie

SASA – Solvent Accessible Surface Area

T - Thymine

TBA - Thrombin Binding Aptamer

TMPyP4 - meso-5,10,15,20-Tetrakis-(N-methyl-4-pyridyl)porphine

TSS - Transcription Start Site

U - Uridine

UTR - Untranslated Region

UV - Ultra-Violet

WRN - Werner Protein/Gene

WS – Werner's Syndrome

γ-GCS - γ-Glutamyl-Cysteine Synthetase

1 Literature Review

1.1 Introduction to G-Quadruplexes

The recognition that DNA sequences rich in Guanine form stable structures has its origins 50 years prior to the Watson-Crick double helical structure [1]. In 1910, Bang reported that highly concentrated solutions of Guanosine monophosphate (GMP) formed a gel (cited in (2) (3) & (4)). Gellert and co-workers [3] expanded on this by inferring that the aforementioned phenomenon was due to helix formation by the Guanosine nucleotides. The authors used X-ray crystallographic data to demonstrate that GMP self-associated into a tetrameric arrangement described as a G-quartet.

This paper demonstrated what has become known as G-quadruplexes (G_4-DNA; G4). These refer to four-stranded nucleic acids formed by the association of four guanine bases to form a G-tetrad; these tetrads then associate vertically to form G-quadruplexes and are linked together by several loops, dependent on the number and orientation of the strands [5] (Fig. 1-1).

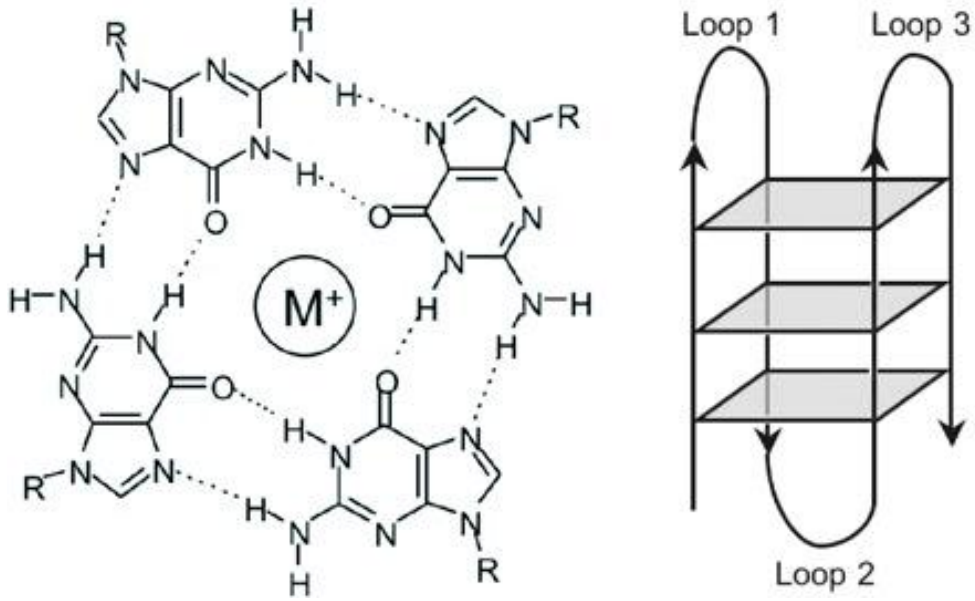

Figure 1-1: Chemical representation of a G-tetrad (left); and schematic representation of a G-quadruplex consisting of three G-tetrads (right) [5].

This phenomenon was thought of for many years as an exasperating artefact of GMP and similar solutions or as a futile intellectual challenge [6] in fact, the intractability of Guanosine derivatives in the laboratory is notorious, and is supported by a list of at least 30 Guanosine derivatives that form a gel in aqueous solution (reviewed in (4)). However, since the 1960s, the tendency for guanine-rich oligonucleotides to form polymers based on the association of these nucleotides into quartets and of these quartets into quadruplexes has been widely studied [2]. The emerging evidence of G-quadruplex structures in vivo and the association with several key biological processes in humans, demonstrate the importance of these structures.

Studies have also shown that the self-association of G_4-DNA is an important phenomenon in other organisms, being present in the form of crystalline plates in specific surface cells, known as guanocytes, of certain spiders which can be retracted to instantly change the spider's colour [7]. Guanine crystals are also found in the eyes of certain deep-sea fish and are used to focus light to the photoreceptors [8].

These structures have also emerged as an important object of study in nanotechnology and supramolecular chemistry [9] (discussed in section 1.3).

This chapter will present the background information required to understand and interpret subsequent chapters. A summary of the unique properties of G-quadruplex structures, their biological relevance, applications and uses will be discussed. Chemical and ligand-based modification of G_4-DNA will also be discussed, highlighting the work carried out in the field of oxidation of G-quadruplexes. This chapter will also present the different methods used to investigate G-quadruplexes and will conclude with the aim and objectives of this project.

1.1.1 G-Quadruplex Structures

The basic unit of G_4-DNA is the G-tetrad, consisting of four guanines in a co-planar arrangement held together by eight Hoogsteen Hydrogen bonds [10] [11] [12] (Fig. 1-1).

Hoogsteen base pairing differs from the traditional Watson-Crick base pairing in the atoms involved in Hydrogen bonding and in the number of Hydrogen bonds between the nucleotides [13](Fig. 1-2).

These Hoogsteen Hydrogen bonds were first observed in crystal structures of monomeric Adenine and Thymine base derivatives [14] and, in the case of G-G base pairing, consist of Hydrogen bonds between the N7 and O6 of one face and the N2 and N1 of another face [10] [12] (Fig. 1-2). As such, these Guanine bases are simultaneously Hydrogen bond donors and acceptors [10] [11].

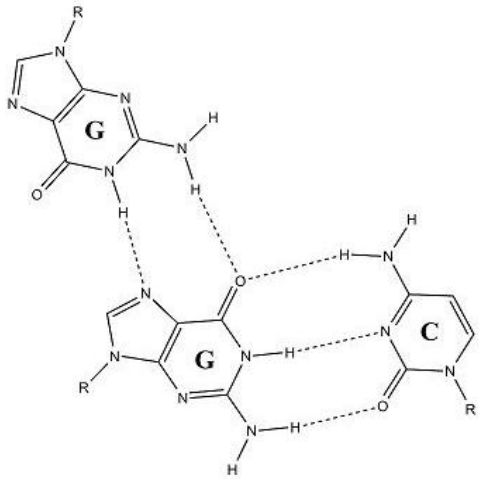

Figure 1-2: Comparison of Watson-Crick base pairing (G-C) and Hoogsteen base pairing (G-G).

The lack of a third Hydrogen bond between two guanines involved in Hoogsteen base pairing could lead to a conclusion that these are less stable than bases involved in Watson-Crick base pairing. However the fact that, in G-tetrads, Hoogsteen base pairing occurs between four nucleotides, with a total of eight Hydrogen bonds and that the energy of these bonds is increased from 0.22 eV in B-DNA to 0.42 eV in G-quadruplexes, explains why these Guanine-rich sequences are capable of forming stable secondary structures [4].

G-tetrads associate vertically to form secondary structures - G-Quadruplexes [15]; these consist of two to five G-tetrads formed by four strands (tetramolecular), two strands (bimolecular), or one strand (unimolecular or intramolecular) [16].

Guanines (as all other nucleotides), reduce the area of the hydrophobic nucleobase exposed to the natural polar environment of the cell, by associating and stacking. In G-quadruplexes this stacking contributes substantially to stability [17]. However, due to the displacement of electrons caused by the attraction of the tetrads, G-quadruplexes are neither perfectly symmetric nor planar [10].

The association of tetrads in the G-quadruplex structure is an additional source of stability as the increase in the number of tetrads is associated with an increase of the association rate constant by a factor of 10 when comparing a two tetrad complex to a three tetrad complex [17].

The G-quadruplex structure is further stabilised by a network of water molecules present in various grooves, ordered around Hydrogen bond donors and acceptors (N7, O6, N2 and N1) and interacting with these [18].

The ionic environment of G_4-DNA is also important to stability. Due to the arrangement of the tetrads, the O6 atoms are orientated towards the centre, forming a negatively charged central channel. Furthermore, the space between two adjacent tetrads is lined by four O6 atoms from one tetrad and four O6 atoms from the tetrad below or above it [19]. This cavity located between tetrads must be neutralised with a cation. The nature of the cation used has a dramatic effect on conformation and stability of the quadruplex [20]. It has been shown that monovalent cations, such as sodium, potassium and lithium, amongst others, have the most stabilising effect on intramolecular G-quadruplexes. Within these cations, the order, in terms of stabilising effect, in decreasing value is $Na^+ > K^+ > Li^+$ [21] [11] [22]. This is potentially due to potassium having the ideal characteristics to effectively coordinate the carboxyl Oxygen atoms, positioned

dead centre of the cavity [10] [19]. Due the enhanced stability offered by K^+ ions, both *in vitro* experiments and *in silico* simulations of G-quadruplex structures are generally performed using this cation.

Bioinformatics analysis of the Human genome (NCBI build 36), have identified a total of 379,000 sequences capable of forming G-quadruplexes, i.e. G-Quadruplex forming Motifs (GQMs) [23] [24]. This, in itself, is not significant as it is a third less than would be expected by chance and may even suggest that there may be evolutionary pressures against G-quadruplex formation [25]. However what is significant is that the vast majority of GQMs are located at telomere ends and in upstream promoter regions of genes [24] [26].

Tetramolecular quadruplexes (those that form from the association of four strands) can adopt five different structures with differences in strand orientation and associated loops and glycosidic torsional angles of the guanines involved in tetrad formation (Fig. 1-3). As, independently of strand orientation, the Guanosines involved in tetrad formation must maintain the same orientation to form Hoogsteen hydrogen bonds with the adjacent Guanosines, they either form *syn* or *anti* glycosidic angles with the backbone [27]. The following figure denotes the simplest topologies that a G-quadruplex can form, by the self-association of four strands each containing a run of three Guanosines [10].

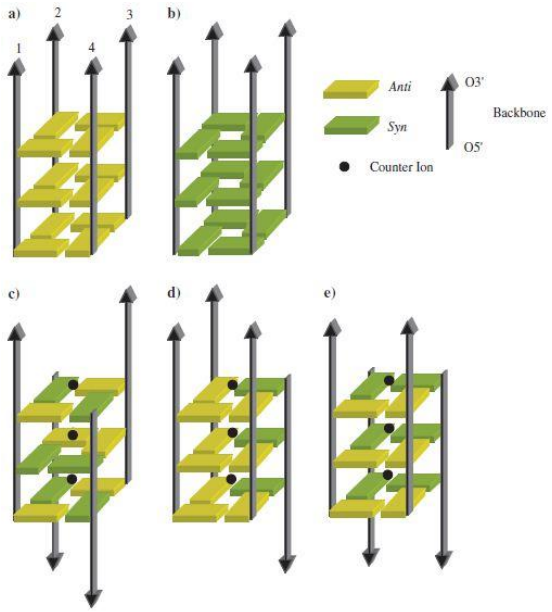

Figure 1-3: Schematic representations of the different structures tetramolecular G-quadruplexes can adopt, denoting strand polarity and associated glycosidic torsional angles of the guanines. Adapted from [10].

Bimolecular G_4-DNA can also adopt five different conformations and are formed by the association of two strands that each contain two runs of four consecutive guanines. Figure 1-4 denotes the conformations a bimolecular complex can adopt when two strands, each with four consecutive guanines, self-associate.

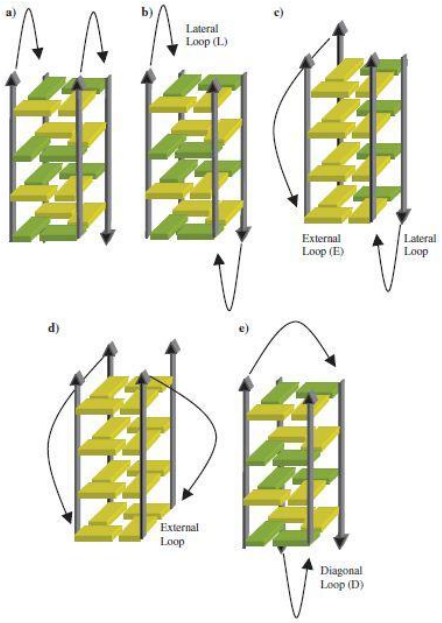

Figure 1-4: Schematic representations of the different structures bimolecular G-quadruplexes can adopt, denoting strand polarity and associated glycosidic torsional angles of the Guanines. Adapted from [10].

Of particular interest for this project are the unimolecular/intramolecular G-quadruplexes, i.e. those formed by the self-association of one strand. These are ubiquitously found in upstream promoter regions that are associated with gene regulation [26] [29]. Sequences capable of forming intramolecular G-quadruplex structures contain, necessarily, runs of three consecutive guanines interspersed with a variable number of nucleotides (up to seven), corresponding to the loops [23] [24] [26]. This property was exploited by Huppert & Balasubramanian (2005) [24] to formulate a motif that represents sequences capable of forming intramolecular G-quadruplexes (Fig. 1-5). It should be noted that both the template and complementary strands containing GQMs can form G-quadruplexes, with equivalent consequences [30].

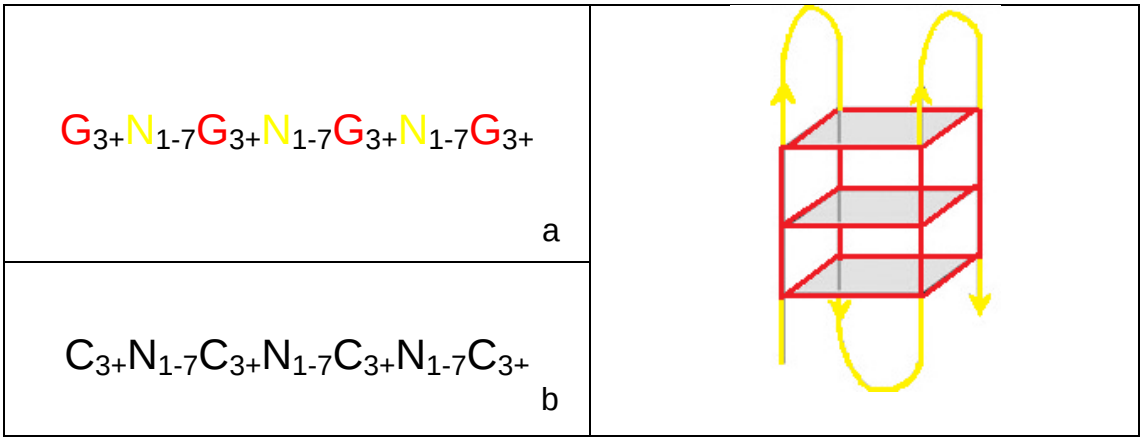

Figure 1-5: G-Quadruplex motif as used by Zhang and co-workers [31] (left) and example of a unimolecular G-Quadruplex (right).

Note: N = any nucleotide. Colours of the nucleotides in the motif correspond to the colours of the strands in the structure on the right.

Using this motif, which identifies G-quadruplex forming motifs on both the template (Fig. 1-5a), and complementary strand (Fig 1-5b), an online database has been constructed, listing potentially G-quadruplex regulated genes, i.e. genes which contain a G-quadruplex motif upstream of one or more promoters [31]. This database, named Greglist (database of G-Quadruplex Regulated Genes (http://tubic.tju.edu.cn/greglist/)), lists 10,277 Human genes with promoter GQMs within 1000 base pairs of the transcription start site (TSS) corresponding to 32.6% of all genes. This database also calculated an average

of almost 2 (1.93) GQMs per gene with an average length of the GQMs of approximately 30 (29.19±13.57).

However, it should be noted that more defined, i.e. more specific algorithms may yield more useful information on the relative distribution of different classes of G-quadruplex structures in upstream promoter regions [32]. To this end, a recent study by Todd and Neidle (2011) [33] has used bioinformatics analysis of the Human genome to categorise GQMs and to map their distribution. Furthermore, the research group of Jean-Louis Mergny are currently working on a new algorithm for detecting GQMs which reduces both the false positives and false negatives observed using the standard algorithm (results not published).

The association between G-Quadruplexes and gene regulation will be discussed in section 1.1.2.

Unimolecular G_4-DNA can adopt a variety of different structures, dependant on strand orientation and associated loops and glycosidic torsional angles of the Guanosines involved in tetrad formation [10].

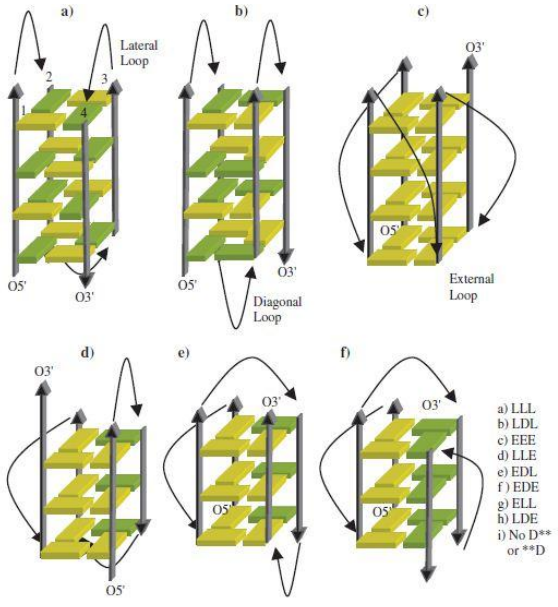

Figure 1-6: Different topologies adopted by unimolecular G-quadruplexes. Adapted from (10).

Note: L=lateral loop; E=External (double chain reversal) loop; D=diagonal loop; *=any form of loop.

As can be observed in Figure 1-6, the topologies adopted by unimolecular complexes are more varied and complex than those adopted by bi- and tetramolecular quadruplexes. This is due to the additional linking nucleotides in the loops [34].

As with bimolecular complexes, these structures have acquired a vulgar terminology based on the loops: chair denotes a structure with only lateral loops (Fig. 1-6a), basket denotes a structure with lateral and diagonal loops (Fig. 1b) and dog-eared denotes a chair conformation with one double chain reversal loop (Fig. 1-6d).

The conformations are highly dependent on the length of the loop sequences; short loop sequences will prevent diagonal loops but can accommodate both lateral and external loops. As such, shorter loop lengths predispose G_4-DNA to form parallel stranded structures [35] [36].

Comparing the relative stability of tetra-, bi- and intramolecular quadruplexes is not a simple task, principally due to the different molecularities associated with the different structures [10]. However several studies [22] [37] [38] [39] have contributed to the conclusion that tetramolecular structures are the most stable form of G-Quadruplex and intramolecular structures are the least stable, with bimolecular complexes having intermediate stability. It has also been shown that parallel conformations are more stable than antiparallel conformations [40].

Thus, simplistically, an all parallel tetramolecular structure is the most stable of form G_4-DNA and intramolecular structures tend to be the least stable. However it should be noted that the conformations presented in this section are simplistic models, with in vivo G-quadruplexes presenting a far more varied range of conformations. Furthermore, specific phenomena that affect the stability of promoter G-quadruplexes will be discussed in section 1.1.3.

1.1.2 Biological Relevance of G-Quadruplexes

When considering the function of G-quadruplexes in biology, it is important to consider the stability of these structures in a cellular environment. In contrast to many other DNA secondary structures observed *in vitro* [41], G_4-DNA is generally stable, with melting temperatures frequently observed in intramolecular G-quadruplexes of over 37°C under near physiological conditions [36] [42] [43].

As stated previously, bioinformatics analysis has identified a large amount GQMs at telomeric ends and in upstream promoter regions. A lot of evidence has accumulated, through mainly *in vitro* methods, on the structure and function of G_4-DNA in the human genome [44] [45]. In recent years, *in vivo* evidence has also been accumulating, pertaining to the stability, structure and function of these structures at telomeric ends [46] and in upstream promoter regions [32].

1.1.2a Telomeric G-Quadruplexes

Human telomeric DNA, located at the end of chromosomes, consists of several thousand bases with a high proportion of Guanosines, with tandem repeats of a 5'-TTAGGG-3' motif, which is a GQM. It terminates in a single-stranded DNA overhang of approximately 200 nucleotides, which, containing GQMs can fold into a variety of different G-quadruplex structures [44] [47]. These ends also tend to attach specific DNA-binding proteins, such as hPOT1, forming a complex which holds the telomeric ends stable and protects them from chromosomal fusions and DNA damage responses. If this complex is disassembled however, these telomeric ends are recognised as DNA damage and the cell undergoes activation of DNA repair pathway and apoptosis [48].

Another important phenomenon is that of base adding by the hTERT component of telomerase, a ribonucleoprotein complex responsible for maintaining telomere length by adding more bases to the receding telomeric tail. This phenomenon evades cellular senescence and death and telomerase activity is widely associated with cancer and its function is now considered as a good anticancer strategy as cancer cells tend to have shorter telomeric lengths

and will therefore require more active telomerase function to survive [15] [49]. Indeed, 85% of the most common cancers present increased telomerase activity [50].

Telomeric G_4-DNA, when in a folded conformation, prevents telomerase from adding bases to the tail. It has been found that many known telomerase inhibitors are, in fact, G_4-DNA ligands which stabilise the folded structure which when applied cause cellular apoptosis, suggesting a possible role as therapeutic agents [51]. Unresolved G-quadruplex structures will also inhibit the transcription of these sequences, whose product is involved in chromatin organisation at the telomere, furthering the importance of G-quadruplexes and ligands that bind to them [15].

In addition, G-quadruplex structures at the end of telomeres have also been shown to increase genomic instability by intervening normal recognition of telomerase-associated proteins with their targets [52] [53] [54].

In vivo evidence for telomeric G-quadruplexes is largely based on investigations into deficiencies in the RecQ family of DNA helicases, specifically WRN and BLM, whose deficiencies cause Werner's and Bloom's syndromes, respectively [55].

The RecQ family of helicases are essential for maintaining genomic integrity and mutations in these enzymes are associated with elevated predisposition to cancer and premature ageing [56] [57] [58].

WRN helicase plays a critical role in repairing damaged DNA, acting as both helicase and exonuclease. It is thought that WRN unwinds DNA and removes abnormal DNA structures that have been generated, maintaining genome structure and stability. This protein has also been shown to play a role in transcription and replication [57] [58]. WRN helicase has also been associated with telomere maintenance and stability, including in the transcription of telomeric-repeat-containing RNA [46]. Werner syndrome (WS) is characterised by premature aging, with a variety of associated maladies, including an

increased incidence of various cancers [59]. This has been demonstrated as due to the loss of WRN helicases, which causes global alterations in expression analogous to the ageing process [60].

BLM helicase also plays a role in replication, transcription and in maintaining genome stability and structure. However, the role of BLM in maintaining genome stability differs from that of WRN; BLM also plays a role in suppressing chromatid exchanges, such as translocations [58] [61], by resolving double strand breaks. Bloom syndrome (BS) is characterised by immunological and developmental abnormalities as well as an increased incidence in the majority of forms of cancer [62].

WRN and BLM have also both been demonstrated to preferentially unwind G-quadruplexes and it has been suggested that this is due to the RQC domain that these helicases contain [55].

A study by Johnson and co-workers (2010) [55], demonstrated that there was a significant association between genes upregulated in WS or BS cells and genes with GQMs. The authors came to the conclusion that the RecQ helicases WRN and BLM regulate transcription by resolving G-quadruplex structures. The authors further hypothesised that upregulation might be due to the quadruplex structures being located in repressor binding sites as such, nucleosomes, for example, which bind to duplex DNA would be excluded in the presence of an unresolved G_4-DNA structure, increasing transcription. However it should be noted that despite the global trend, several proto-oncogenes that had previously been found to be downregulated in the presence of G-quadruplexes were also found to be downregulated in WS and BS cells, such as c-Myc and c-Kit.

Although this project concentrates on G-quadruplexes in upstream promoter regions, it is important to recognise the importance of these structures in telomere maintenance and function. The findings in this report can be generalised, to some extent, to G-quadruplexes in telomeric regions.

1.1.2b G-Quadruplexes in Upstream Promoter Regions

The phenotype, that is to say, the biochemical characteristics of a cell is determined by the genotype, i.e. the genetic information stored in the form of DNA. According to the "central dogma of molecular biology" [63], this transformation of information occurs by transcription of DNA to RNA and then translation of RNA to proteins.

Transcription involves the synthesis of an RNA strand, complementary to the DNA strand, from which it originates, by the DNA-dependent RNA polymerase II. As stated previously, according to Greglist [31], 32.6% of genes in the Human genome contain GQMs in their upstream promoter regions. Furthermore, the GQMs found within these regions demonstrate a strong positional bias towards the transcriptional start site (TSS) [26] [29]. These characteristics are analogous to those observed in traditional transcription factor binding sites [64]. Indeed, bioinformatics analyses have demonstrated that GQMs in upstream promoter regions can, in some cases, be close to [65] or even overlap with [66] traditional transcription factor binding sites. Suggesting a possible mechanism by which GQMs may play a role in transcription factor binding. Studies have investigated several promoter regions of genes that contain GQMs, demonstrating that, for instance, the promoter regions of c-Myc [67], BCL2 [68] and c-Kit [69] form stable G-quadruplex structures.

An important aspect to consider when investigating G-quadruplexes present within the genome is their stability, as these structures will be subjected to a variety of different factors which will affect their stability.

In the telomere, the single stranded overhang enables the formation of the G-quadruplex structure. However, G_4-DNA within the genome must always compete with the duplex or B-DNA form. These two forms therefore exist in an equilibrium that is sensitive to a variety of external factors.

When considering the formation of these structures within contiguous B-DNA, one must also take into account the effect of flanking bases, as DNA G-

quadruplexes in promoter regions will be flanked by numerous bases. However the majority of *in vitro* evidence of the formation of G-quadruplexes is in short sequences with few, if any, flanking bases. A study by Arora and co-workers (2009) [70], describes the intuitive effect of the number of flanking bases on duplex/quadruplex competition in the c-Kit promoter G-quadruplex, with the number of flanking bases proportionate to a decrease in the stability of G-quadruplexes favouring, therefore, the duplex form.

For quadruplexes to form in this unfavourable environment, a variety of phenomena are thought to play a role in promoting the formation of these structures. One such factor thought to play a role in the formation of G-quadruplexes from B-DNA is the phenomenon of DNA Breathing. This process is a transient conformational fluctuation that involves the local denaturation of double stranded DNA to form a bubble of a few tens of base pairs within the chromosome [71] [72], possibly supported by single strand binding proteins [72]. DNA breathing is thought to play a role in several physiological processes such as the initiation of transcription and replication [73]. It is therefore a possibility that this phenomenon plays a role in the formation of G-quadruplexes in promoter regions.

It is also important to note that the C rich, complementary strand to GQMs are often prone to the formation of a quadruplex structure themselves, the i-motif. It is thought that during G-quadruplex folding, the complementary strand also folds into the i-motif structure, in some cases, possibly increasing the stability of the G-quadruplex structure. Although some evidence exists that these structures are biologically active, the evidence is scarcer than with the G-quadruplex, possibly due to the i-motif's reduced stability a physiological pH compared with G_4-DNA [74] [75].

In vitro investigations into quadruplex structures are mainly performed in dilute solutions with low concentrations of DNA. This precludes the effects of molecular crowding that occur within a cell, as biomolecules must function in the presence of high concentrations of other biomolecules. Investigations into the

effects of molecular crowding [76] [77] [78] [79] [80] have shown that this phenomenon has substantial effects on the conformation and stability of G-quadruplexes, with the molecular crowding agents increasing the stability of the G-quadruplex structure and/or decreasing the stability of the duplex structure.

Another important determinant of the stability of G-quadruplexes in promoter regions and, therefore, an important factor in the quadruplex/duplex equilibrium is loop length; with longer loop lengths [81] associated with decreased stability of the quadruplex structure and increased stability of the duplex form. Loop sequence is also suspected to be a determinant of stability, however there is no clear rule as to which nucleotide is preferred [82] [83] [84].

Perhaps the most important and most studied factor that affects the stability of the G-quadruplex structure is the effect of stabilising and destabilising agents, such as proteins and small molecules. Investigations into the use of synthetic agents will be discussed more fully in section 1.2.1; however there are a number of naturally occurring molecules that interact with G4-DNA, with direct consequences on their stability.

In vivo evidence for the formation of G_4-DNA in upstream promoter regions originates from investigations into loss-of-function mutations in the Fanconi anaemia complementation group J (FANCJ) family of DNA helicases which provided a direct link between GQMs and genomic instability [85] [86]. These distinct helicases were found to be of high importance for the stability of G-tracts in the genome. This was first detected in *C. elegans*; as worms defective in DOG-1, the FANCJ orthologue, presented an increased incidence of deletions in regions of the genome with runs of consecutive Guanosines [87] [88], and when sequences that contained GQMs were inserted into the genome, these were removed in worms defective in DOG-1 [89].

FANCJ helicases have been associated with Fanconi anaemia (FA), which is a heritable cancer-susceptibility disorder that is characterised by sensitivity to DNA-crosslinking agents [90]. A study by London and co-workers (2008) [91],

demonstrated the importance of FANCJ helicases in the unwinding of G-quadruplexes through *in cellulo* investigations on cell lines lacking FANCJ helicases, which suffer a large number of deletions at GQM sites. Additionally, these authors used *in vitro* studies to demonstrate that FANCJ preferentially unwinds G-quadruplexes over B-form DNA, suggesting that FANCJ helicase has an important role in resolving folded DNA G-quadruplexes that, if not resolved, will result in deletions of these sequences in the genome.

Furthermore, a study by Cogoi and co-workers (2008) [92], demonstrated that three nuclear proteins, hnRNP-A1, Ku70 and PARP-1, bind to the G-quadruplex structure of k-RAS, suggesting a possible role of these nucleoproteins in G-quadruplex associated transcriptional control. A further study by this research group [93] demonstrated that hnRNP-A1 also binds to the G-quadruplex structures in the c-kit promoter. The authors also showed that hnRNP-A1 unfolds the G-quadruplex in k-RAS, increasing transcription by converting the structure to the duplex form.

More recently, studies on Pif1 helicase in yeast and human cells have shown that this helicase preferentially unwinds G-quadruplexes and cells deficient in this helicase show a slowed replication and stimulated DNA strand breaks [94] [95]. Pif1's role in unwinding has also been shown to play an important role in in preventing G-quadruplex associated genomic instability [96]

Furthermore, it has also been shown that nuclease hypersensitive sites, which are indicative of non-duplex DNA structure, in the human genome, that is to say sequences highly susceptible to deletions from nucleases, have a high prevalence of G-quadruplexes (approximately 230 times higher) than the rest of the genome [26]. The nuclease hypersensitive element III (NHE III) found upstream of the proto-oncogene c-Myc, is an example of a sequence containing GQMs and susceptible to nuclease deletions [67].

There are several models that describe how G-quadruplexes could control gene activity; one is that expounded by Johnson and co-workers (2010) [55]

(described in the previous section), in which unresolved G-quadruplex structures upregulate transcription by excluding repressors. Another model has been associated with a large number of proto-oncogenes, including c-Myc; in this model, the presence of an unresolved G-quadruplex structure silences the gene by impeding transcription (Fig. 1-7).

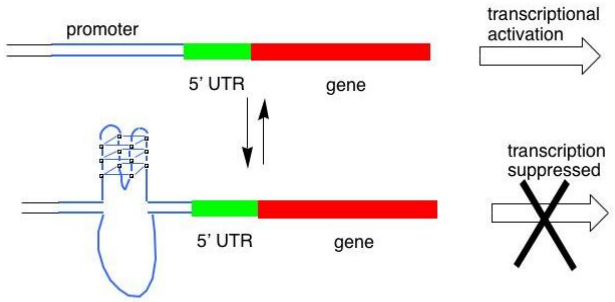

Figure 1-7: Model for the regulation of gene expression by G-quadruplexes in upstream promoter regions. Adapted from (26).

Another model for G-quadruplex–associated regulation of transcription was demonstrated in vivo by Hershman and co-workers (2008) [97] in *S. cerevisiae*. These authors demonstrated a model that can be considered a hybrid of the previous two models described; using a highly selective G-quadruplex ligand, N-methyl mesoporphyrin IX (NMM), which stabilises the structure and compared the expression of genes with GQMs in the open reading frame (ORF) or upstream promoter region with and without treatment with NMM. They also proceeded to compare GQM location and histone occupancy.

These authors hypothesised that when G-quadruplexes occur in the promoter region upstream of the TSS, they exclude the action of histones, which bind to B-DNA and thus upregulating transcription. However, if unresolved quadruplexes occur after the TSS, i.e. within the open reading frame (ORF), the structures impede the transcription complex from progressing, thus downregulating transcription.

As stated previously, G-quadruplex forming sequences are present in a large number promoters of genes such as retinoblastoma susceptibility [98], diabetes

susceptibility [99], vascular endothelial growth factor (VEGF) [100] and fragile X mental retardation genes [101] [102] [103] and this includes those of many proto-oncogenes [104], with many of these forming stable G-quadruplexes *in vitro*, such as c-Myc [67], BCL-2 [68], c-Kit2 [69], RET [105] and k-Ras [106], amongst many others. It appears likely that these structures play a role in the activation of proto-oncogenes to oncogenes by increasing their expression, when unfolded.

This suggests a possible role of G-quadruplexes in oncogene regulation. This topic has been extensively studied. A review by Brooks and co-workers (2010) [32] categorised several well-known oncogenes with G-quadruplexes in the promoter region in accordance with the “hallmarks of cancer” [107] (Fig. 1-8).

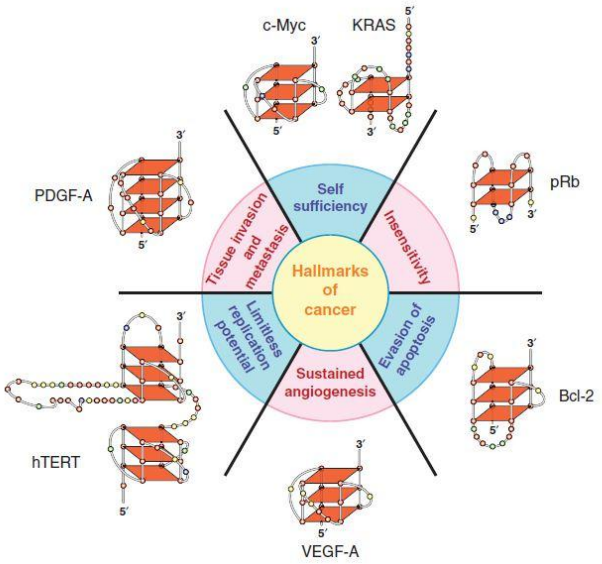

Figure 1-8: Oncogene associated G-quadruplexes categorised in accordance with the "hallmarks of cancer" [107]. Adapted from (32).

Note: although not represented, the authors also identify c-Kit as being associated with self-sufficiency [32].

As can be observed from the above figure, these quadruplexes present very distinct conformations, making them potential drug targets for anticancer therapy [108] [109].

Of special interest to this project is the association of G-quadruplex structures with c-Kit and c-Myc expression; these are discussed in the following sections.

G-Quadruplex Mediated Regulation of c-Myc Transcription

The result of c-Myc expression is a protein which functions as a critical transcription factor that regulates the expression of several genes associated with cellular proliferation, differentiation and apoptosis [110] [111]. Its expression is tightly controlled and the association between alterations in the regulation and a significant number of human cancers is well documented [110] [112]. This, in conjunction with the relatively short half-life of the protein product [113], make c-Myc transcriptional control an attractive target for anticancer chemotherapeutics. The activation of this proto-oncogene is a complex mechanism involving multiple promoters and transcription start sites and can arise from a variety of different mechanisms, including by simple upregulation of transcription [113].

An important element in c-Myc activation is the previously mentioned NHE III which plays a role in allowing access to the promoter by chromatin structure alteration [114], being responsible for between 75 to 85% of c-Myc transcription [115] [116].

NHE III is a C rich, 27 base pair long sequence and is located approximately 100 bps upstream of the P1 promoter [113]. It has been demonstrated that the complementary strand, which is G rich, is capable of forming a variety of different G-quadruplex structures at near physiological conditions [28] [117] [118] [119]. This variety is due to the NHE III sequence containing five runs of three or four contiguous guanines separated by a single A or T. However the most convincing evidence has demonstrated two parallel-stranded G-quadruplex structures within the c-Myc NHE III region [67] [120]; one formed by four 3'-end G-tracts is the kinetically favoured structure, while the thermodynamically favoured structure is that formed by two 3' – and 5'-end G-tracts, with a 6 nucleotide central loop. The former structure has been determined by NMR spectroscopy [67] (Fig. 1-9).

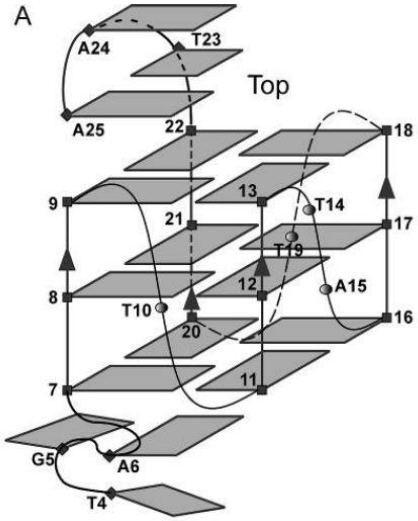

Figure 1-9: Schematic Representation of the G-quadruplex structure found in the NHE III of c-Myc. Adapted From (67).

Note: to accurately determine the structure, two G-T mutations at positions 14 and 23 were performed to avoid a mixture of loop isomers.

The c-Myc G-quadruplex structure represented in Figure 1-9 is characterised by three G-tetrads (top tetrad: G9, G13, G18, G22; central tetrad: G8, G12, G17, G21; bottom tetrad: G7, G11, G16, G20). These tetrads are formed by four parallel DNA strands, linked by three double-chain reversal side loops, two of which are single nucleotide loops (T10 and T19) and one which is a double nucleotide loop (T14, T15) [(67)].

The NHE III region has been shown to repress transcription of c-Myc when in a non-canonical state [(121)]. This was then shown to be due to G-quadruplex formation by the use of a G_4-DNA binding ligand, TMPyP4, which stabilised the G-quadruplex in the NHE III region and decreased c-Myc expression [(122)]. On the other hand, NM32-H2, a metastases suppressor protein has been shown to resolve the G-quadruplex structure, leading to c-Myc activation [(123)].

The wide research into the mechanism for NHE III silencing of c-Myc transcription led to the formulation of a model that describes the way in which the NHE III G-quadruplex can control transcription [(118)].

The chair G-quadruplex structure, represented in Figure 1-10, is known as the paranemic form, which is not transcriptionally active. This can be converted into the transcriptionally active form by the action of NM23-H2. The cationic porphyrin derivative TMPyP4 however, stabilises the G-quadruplex, silencing expression of c-Myc [118] [124]. This occurs due to the inability of transcription factors such as cellular nucleic acid binding protein (CNBP) and heterogeneous ribonucleoprotein K (hnRNPK) to bind to the NHE region when it is in the G-quadruplex form, thus preventing transcription [113].

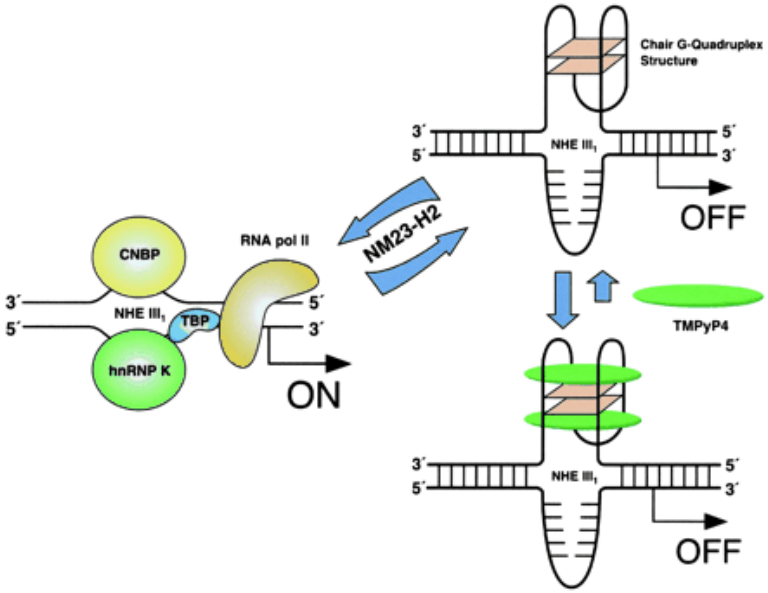

Figure 1-10: Model for the activation and repression of c-Myc expression by stabilisation of a G-quadruplex structure within the NHE III region with the use of ligands. Adapted from (118).

This general model, although not representative of the global trend [55], is ubiquitous within G-quadruplex regulation of oncogene expression, such as c-Kit [69], c-Myc [67] and k-Ras [106].

The role of G_4-DNA in both telomere maintenance and gene regulation is still relatively novel, with *in vivo* and *in cellulo* evidence in human cells only being demonstrated in the last few years. These structures are also an attractive target for anticancer chemotherapeutics as many ligands specifically target G-quadruplexes over canonical DNA. This substantial interest has led to a very good characterisation of the biological and biophysical aspects of this structure. And it was this and the fact that the c-Myc oncogene is associated with

oxidative stress that led to this structure being chosen to be studied in this project.

G-Quadruplex Mediated Regulation of c-Kit Transcription

The c-Kit proto-oncogene encodes a tyrosine-protein kinase, also known as CD117, which regulates key signal transduction cascades associated with cell survival, proliferation and differentiation. However, the overexpression of this gene is also believed to be associated with the self sufficiency of cancer cells, increased in a variety of cancers [32].

The promoter region of this proto-oncogene is a G-rich sequence capable of forming at least two G-quadruplex structures [125] [69]. However within these, the most interesting structure is an unprecedented scaffold in which a Guanine thought to be part of the loop sequence is involved in tetrad formation [125], named c-Kit87up, formed from the sequence GGGAGGGCGCTGGGAGGAGGG.

Figure 1-11 represents the NMR elucidation of this structure. This structure was selected to be studied in this project due to its structural uniqueness, the fact that it has been well studied and that it presents longer loop sequences than the c-Myc G-quadruplex.

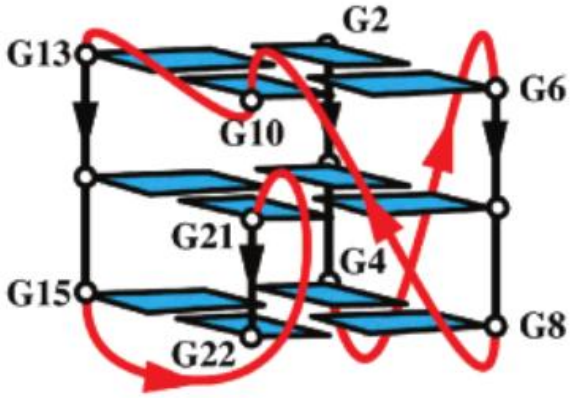

Figure 1-11: Schematic representation of the c-Kit87up G4 structure. Adapted from (125).

The c-Kit G-quadruplex structure represented in this figure consists of three tetrads (from top to bottom: G2, G6, G10, G13; G3, G7, G21, G14; G4, G8, G22, G15), with all guanines in the *anti* conformation. This structure contains

four loops; two of which are single nucleotide double-chain reversal loops (A5 and C9); one double nucleotide loop (C11, T12) which links two adjacent corners (G10 and G13); and one 5 nucleotide long loop (A16, G17, G18, A19, G20) which extends to the G15 on the bottom tetrad and twists to enable the positioning of G21 and G22 into the central and bottom tetrads, respectively [125].

The model of regulation of this quadruplex has been determined to be identical to that of the c-Myc G-quadruplex, discussed in the previous section, as the use of G-quadruplex stabilising ligands downregulate c-Kit expression [126].

Bioinformatics analysis and molecular dynamics simulations have been used to suggest that this structure is unique amongst promoter G-quadruplexes (within 100 bp of a TSS) [127]. This, in conjunction with the association with a wide variety of tumours, makes the c-Kit87up promoter G-quadruplex an attractive target for anticancer chemotherapeutics. Currently, a large effort has been placed on the development of small molecules that target the c-Kit G-quadruplexes. An example of this is the selective down-regulation of c-Kit expression in human gastric carcinoma cells by a G-quadruplex binding ligand [128].

1.1.2c RNA Quadruplexes

G-quadruplex structures have also been demonstrated to form from RNA sequences. Indeed, the vast majority of RNA G-quadruplexes have been shown to have greater stability in the presence of potassium than their DNA counterparts [129]. Quadruplex forming mRNAs have been demonstrated as intervening in gene expression in both prokaryotes and eukaryotes [130] [131] [132] [133] [134] and in telomere maintenance [135], with greater incidence in non-translating regions [136].

The principal process by which RNA quadruplexes intervene in expression has been demonstrated by several research groups; by Joachimi and co-workers (2009), in a variety of mRNAs [129] by Kumari and co-workers (2010) [131], in k-

Ras mRNA and by Shahid and co-workers (2010) [137], in BCL-2 mRNA. These research groups described a model in which unresolved G-quadruplex structures contained within mRNAs impede ribosomal interaction thus impeding translation. This hypothesis was strengthened by in vivo investigations into translation of fragile X mRNA which was increased by the use of G-quadruplex destabilising proteins [103].

However it has also been shown that, in some cases, G-quadruplexes play a role in alternative splicing events during transcription. Two examples, both relevant in in carcinogenesis, are the GQM in intron 3 of TP53, which modulates the splicing of intron 2, inducing the expression of distinct p53 isoforms [138]; and the GQM in the hTERT gene which, when folded, induces the expression of an inactive form of the transcript, downregulating telomerase [139].

As with DNA G-quadruplexes, RNA G-quadruplexes have been associated with oncogene expression, examples being G-quadruplex formed in the 5' UTR of the BCL-2 mRNA [140] and n-Ras mRNA [131]. Indicating that the RNA quadruplex structure could also be an attractive target for anticancer chemotherapeutics [27].

1.2 Modification of G-Quadruplexes

1.2.1 Ligand-G-Quadruplex Interactions

The association of G-quadruplexes in both upstream promoter regions and at telomeric ends with a wide variety of different biological processes, including cancinogenesis make these structures attractive therapeutic targets [108]. In addition, the ubiquity of different ligands that preferentially bind to G_4-DNA structures over canonical DNA has led to a plethora of research into different G-quadruplex ligands and their effects on these structures [141].

The stabilisation of G_4-DNA occurs principally through π-π stacking and electrostatic interaction of, in most cases, a flat aromatic molecule with the folded G-quadruplex structure. This generally occurs through external stacking on one of the terminal G-tetrads. This structural feature represents a key

difference between the G-quadruplex structure and B-DNA and enables the differential interaction with these structures [142].

Although the interactions between G-quadruplexes and ligands are far less understood than those between B-DNA and ligands, it is known that "successful" ligands contain a large aromatic surface and are positively charged [141]. However, other types of G-quadruplex ligands are also under investigation based on modified biomolecules such as short peptide nucleic acids (PNA) sequences that have been used to target the c-Kit87up G-quadruplex structure [143].

The majority of current anticancer chemotherapeutic agents target DNA and although some positive clinical outcomes have been observed in a select few, the benefits are counterbalanced by high toxicity and rapidly acquired resistance mechanisms [108]. G-quadruplex targeting agents provide an alternative route for anticancer therapy due to the high potential for differential molecular recognition by G-quadruplex specific ligands [54] [141].

The use of quadruplex ligands in telomerase inhibition of cancer cells, with the desired consequence of senescence and apoptosis has shown promising results [108] [144]. BRACO-19, RHPS4, telomestatin and quindoline derivatives, for example, have all been demonstrated to stabilise G-quadruplexes in telomeres and induce cellular senescence and apoptosis. Several mechanisms are thought to be involved in this phenomenon, including the displacement of hPOT1 and subsequent activation of DNA repair pathways including the p53, ATM and $p16^{INK4a}$ pathways [145] [146] [147] [148].

Although the majority of the extensive research in recent years has been into G-quadruplex ligands at telomeres, the study of promoter G-quadruplexes and their potential role in gene regulation has led to investigations into the use of G_4-DNA binding agents to differentially regulate transcriptional activity of disease-related genes, particularly of proto-oncogenes [108].

Studies of ligand interactions with the G-quadruplex structure in the NHE III upstream of c-Myc are the most prevalent. The cationic porphyrin derivative TMPyP4 [118] and berbine derivatives [149] have been demonstrated, for instance, to bind to and stabilise this G-quadruplex structure, reducing transcriptional activity. Studies on the two G-quadruplex structures that form in the core promoter of c-kit have demonstrated that an iso-alloxazine molecule reduces transcriptional activity of this oncogene [150].

One G-quadruplex-targeting small-molecule drug (Quarfloxin) binds to G-quadruplex structures in ribosomal DNA (rDNA), preventing the formation of the Nucleolin/rDNA G-quadruplex complex, preventing, therefore, transcription to ribosomal RNA (rRNA), which is critical in many cancer cells, inducing apoptosis [151] [152]. Quarfloxin has recently completed phase II clinical trials (results not published) to determine the safety and efficacy of this compound in the treatment of carcinoid/neuroendocrine tumours.

The formation of G-quadruplexes in mRNA and their association with RNA regulation is also an attractive therapeutic target [153]. However, thus far there has been little interest, when compared to DNA G-quadruplex targets.

These findings all support the notion that the differential binding of small molecules to G_4-DNA alters stability and can even induce the formation of these structures, with consequences in telomere maintenance and gene expression. This field has opened up a potential new avenue for anticancer chemotherapeutics through both the induction of senescence and apoptosis in cancer cells and the regulation of oncogene expression through the use of G-quadruplex ligands at telomeric ends and in upstream promoter regions.

Although the use of G-quadruplex ligands is outside of the scope of this project, it nonetheless highlights the importance of these structures and their potential in as chemotherapeutic targets.

1.2.2 Modification to G-Quadruplex Nucleotides

Investigations into the modification of nucleic acids are well established and in vivo evidence for these alterations is ubiquitous. These have been demonstrated as crucial in many biological processes, and have been strongly associated with a variety of detrimental physiological processes, including carcinogenesis.

The involvement of G-quadruplex structures in these processes and diseases has led to interest in the effects of modifications that occur naturally in DNA on the structure and stability of G-quadruplexes. Furthermore, from a chemical perspective, the modification of G_4-DNA is a potential avenue to discover and create complexes with potential for use in supramolecular chemistry and nanotechnology [9].

1.2.2a Base Substitutions

Changes to the genomic sequence by one or more nucleotides are known as mutations and these are associated with a plethora of different diseases through many different processes.

Investigations into the effect of base substitutions on the structure and stability of G_4-DNA have revealed that quartets can be formed not only from the association of Guanines (G*G*G*G), but also from the association of Adenines (A*A*A*A), Guanines and Cytosines (G*C*G*C) and Adenines and Thymidines (A*T*A*T). Furthermore, pentads, hexads, heptads and octads have also been detected in DNA quadruplexes [154] [155] [156] [157] [158] [159] [160] . Adenine [161], Uracil [162] and mixed Guanosine and Cytosine [163] quartets have also been identified in RNA quadruplexes.

These findings point to high versatility of quadruplexes to accommodate a wide variety of different quartets. However, it should be noted that quadruplexes with non-G-quartets present lower stability. This is thought to be due to the fact that

only one Hoogsteen Hydrogen bond is observed between most of these quartets [164].

A study by Lee and Kim (2009) [165] on base substitution in human telomeric quadruplexes revealed that the degree of destabilisation depends on the position of the mutated nucleotide within the G-quadruplex. Demonstrating that a G-T mutation in the central quartet induces a different conformation, with decreased stability at near physiological conditions. Whereas G-T mutations in a terminal quartet induces a variety of different conformations with a less pronounced effect on stability.

Mutations in the G_4-DNA region involved in insulin transcription destabilise the structure, leading to diminished transcriptional activity [166]. Mutations in the c-Myc promoter have also been shown to destabilise the G-quadruplex structure in this region, leading to over-expression of this oncogene [167].

Furthermore, a study by Halder and co-workers (2004) [168] determined that a C to T mutation in the C-rich (sense) strand of the NHE III region of the c-Myc oncogene promoter increases the stability of the duplex form, unfolding the quadruplex. This is thought to be due to the destabilising the i-motif which, as stated before, is a contributing factor to G-quadruplex stability.

Synthetically modified nucleotides such as locked nucleic acids (LNA) and peptide nucleic acids (PNA) have also been demonstrated to form G-quadruplexes *in vitro* [164], with thermal stabilities comparable to the DNA analogues, although conformational distinctions have been observed [169] [170] [171]. These novel structures may have potential applications as design elements in nanoscale assemblies [169].

1.2.2b Chemical Modification of G-Quadruplexes

A study by Mekmaysy and co-workers (2008) [172] demonstrated that the methylation of Guanosines involved in tetrad formation at the O^6 atom not only decreases stability but can induce alterations to the conformation of the folded

structure. These authors presented results analogous to those of Lee and co-workers [165], as methylation of Guanines involved in the central tetrad induces a more pronounced decrease in stability than methylation of Guanines involved in one of the terminal tetrads. This is possibly due to the central guanines affecting both adjacent tetrads through base stacking, as well as possibly affecting the central cations [173]. Modifications to a Thymine contained in the loop sequence also destabilise the quadruplex structure [174].

This has important biological consequences as methylation is one of the most common form of epigenetic alteration and abnormally low or high levels of methylation (hypomethylation and hypermethylation, respectively) are associated with changes in gene expression and genomic instability, observed in the development of numerous cancers (reviewed in (175) (176)).

The association of G-quadruplexes and CpG methylation has been well studied, which, when occurring in the G-quadruplex appears to destabilise the structure [26] [177]. Furthermore, a bioinformatics study performed by De and Michor (2011) [178] demonstrated the enrichment of GQMs at DNA breakpoints associated with somatic copy-number alterations, which are themselves sites of abnormal hypomethylation. They proposed that abnormal hypomethylation in regions enriched for G-quadruplexes acts as a mutation inducer in carcinogenesis.

The effects of a large number of other chemical modifications have been studied, including of methylation, oxidation and addition of bromide at the 8' Carbon, amongst many other modifications [173] [179] [180].

These studies, performed by Jean-Louis Mergny's research group, demonstrated again an influence of the position of the modification on the effect on stability.

The authors concluded that the vast majority of substitutions are extremely detrimental to quadruplex stability, with unmodified Guanines forming the most stable tetrads, and therefore the most stable quadruplexes. They also confirmed

that alterations to guanines in the central tetrad had a more deleterious impact on stability than in any other position.

However, these authors also demonstrated that the 8-oxo-guanine and 8-methyl-guanine accelerate the formation of the quadruplex when located on the 5' side and that alterations to Guanines that do not directly influence the Hoogsteen Hydrogen bonding (such as modifications to the 8' Carbon) were well tolerated and could effectively replace Guanines.

It should however be noted that these studies were carried out *in vitro* on a synthetic all parallel tetramolecular G-quadruplex, not known to form *in vivo* and that has been demonstrated to be more robust and stable than intramolecular quadruplexes [164].

Systematic analyses of the effect of chemical modifications give insights in to the energetic, kinetics and dynamics of quadruplex folding and are an indispensable avenue for shedding light on some of the subtle features of the formation and stabilisation of these structures.

1.2.2c Oxidative Stress

Approximately 1-4% of oxygen used in aerobic respiration is released into the cell in the form of incompletely reduced reactive oxygen species (ROS) [181]. These naturally occurring oxidants, which are also generated from exogenous environmental factors [182], lead to endogenous oxidative DNA damage, which is implicated in a wide variety of detrimental physiological effects including cancer and aging [183] [184] [185] [186], as well as changes in gene expression, aimed at preventing cellular demise [187].

The biomarker used to detect DNA oxidation is 8-hydroxy-2'-deoxyguanosine or 8-oxo-7,8-dihydro-2'deoxyguanosine (8-oxo-dG), as this is the most commonly observed oxidised nucleotide and has been observed in aged organisms and cells, as well as in a number of diseases including neurodegeneration, cardiac disease and cancer [183] [185] [188] [186] [189].

The presence of oxidised guanine in the genome is a mutagenic phenomenon as 8-oxo-dG preferentially pairs with Adenine through Hoogsteen Hydrogen bonding over Cytosine with conventional Watson-Crick Hydrogen Bonding [190] and is considered to be a genuine threat to the integrity of the genome [183].

The cellular response to oxidative stress is a complex and multifaceted process involving a large number of gene products, and is as yet, poorly understood. At the biophysical level, the incorporation of 8-oxo-dG appears to destabilise the duplex to modest extent and even forming a characteristic bend, facilitating the binding of glycosylases [191], such as oxo-guanine DNA glycosylase (OGG1). This glycosylase initiates the base excision repair (BER) pathway, the major repair mechanism to combat the threat from 8-oxo-dG and repress the mutagenic process. This pathway excises one of the nucleotides involved in the mispair [185] [192] [193]. However, incorporation of oxidised guanine has been shown to, in some cases, cause DNA strand breaks, and even apoptosis [194].

However a host of other gene products have also been found to be affected by high levels of ROS, as part of a multitude of different pathways, aimed at adapting to or resisting the stress and repairing or removing damaged molecules and even apoptosis (reviewed in (195)).

In terms of the association of DNA oxidation and carcinogenesis, the mutagenic properties of this process are associated with various stages of tumour progression; the mutations may be within the coding region of the DNA, leading to a dysfunctional or even absent protein product, which, in the case of tumour suppression proteins, may increase the chances of developing cancer [196] [197]; the mutations may also occur within the promoter region of genes associated with cancer, affecting transcription by either upregulation or downregulation [198] [199].

Furthermore, studies have demonstrated that oxidative stress alters gene expression at a global level in both Prokaryotes and Eukaryotes [200] [201] [202] [203]. This has been hypothesised to occur indirectly by interfering with physiological

signals (hormones, cytokines, etc.) and/or by interfering with the environmental stimuli pathway (physical parameters, xenobiotics, etc.) [196] [204].

The p53 repair pathway has been demonstrated to be associated with many of the alterations in the transcriptome [187], with eleven p53-regulated genes induced by oxidative stress [205]. Interleukins 2 and 6 show converse responses to oxidative stress, with IL-2 being repressed [196] and IL-6 being induced [205].

The overexpression of c-Myc, a crucial gene involved in transcription regulation, has been shown to induce genomic instability [206] [207] and has been demonstrated to over-express in the presence of hydrogen peroxide in mouse, rabbit and human cell lines [208] [209] [210] [211]. The pathway by which this gene over-expresses has been shown to occur through Src-linked signal transduction, which is induced in the presence of hydrogen peroxide [212].

Several studies have also pointed towards Myc over-expression inducing oxidative stress [207] [213] [214] [215]. Indeed, c-Myc over-expression has been shown to generate sufficient ROS to induce DNA strand breaks, activate the p53 pathway and reduce clonogenicity, however apoptosis is not induced [207]. The exact mechanism by which Myc expression induces the generation of ROS is unknown, however it is clear that c-Myc induces the expression of a number of genes involved in intermediary metabolism, and ROS may result from biochemical imbalances generated from the irregular expression of the many c-Myc target genes' products [216].

c-Myc has also been demonstrated as determining redox balance, through the regulation of γ-glutamyl-cystein synthetase (γ-GCS) which, in turn, catalyses glutathione (GSH), a leading molecule mediating redox balance. Exposure to Hydrogen peroxide enhances c-Myc recruitment to γ-GCS, as part of the cell's response to oxidative stress [217].

The over-expression of several Myc proteins, including c-Myc has also been demonstrated to decrease the ability of Human cells to reduce peroxide and hydroperoxides [218].

G-Quadruplexes and Oxidative Stress

As Guanine is the most vulnerable nucleobase to oxidation, the G-rich sequences that form quadruplexes are likely to be “hot-spots” for oxidative damage. This could have implications for the involvement of G-quadruplexes with gene regulation as well as their viability as anti-cancer targets.

However, current information on the effect of oxidative stress on the G-quadruplex structure is largely based on chemical studies on parallel tetramolecular G-quadruplexes *in vitro* [179] [180]. A study by Huang and co-workers (2009) [219] observed that the G_4-DNA structure appeared to offer protection to the Guanines from oxidation when compared to duplex DNA.

Conversely, studies on anti-parallel bimolecular and tetramolecular G-quadruplexes have demonstrated preferential oxidative damage to Guanines in a Quadruplex structure over those in a duplex structure [220] [221] [222] [223].

These discrepancies may indicate, in the opinion of the author, that the more stable the structure, the less susceptible the Guanines are to oxidation. Furthermore, these differences could also be caused by the decreased accessibility of the Guanines in the G-quadruplex structures when in the *anti* conformation as opposed to the *syn* conformation when in a tetramolecular quadruplex.

However these studies have arguably little biological relevance to G-quadruplex-regulated gene expression, as the structures studied are not known to form in the promoter regions of genes and are more stable than the more common unimolecular quadruplexes.

To the knowledge of the authors, the only investigations into the effect of oxidation on G-quadruplexes from a biological perspective have been in telomeres. These studies [222] [224], found that the G-quadruplex structure in telomeres is particularly sensitive to oxidation, when compared to B-DNA and hypothesise that this would lead to telomere shortening. However neither make reference to the effect of oxidation on the stability of the structure, nor to the

possibility that the BER pathway may excise the oxidised nucleotide, which would likely preclude the formation of the quadruplex structure.

Generation and Detection of Reactive Oxygen Species

To artificially induce oxidative stress *in cellulo* and in vivo, a variety of ROS can be generated, including singlet oxygen, hydroxyl radicals, superoxide, hydroperoxides and peroxides. Another method, commonly used to study the effect of oxidative stress on nucleic acids, is the use of Fenton's reagent, a solution of hydrogen peroxide and ferrous Iron(II) which produces hydroxyl radicals and anions by the following reaction:

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH\cdot + OH^-$$

The hydroxyl radical, in particular, is highly reactive and, consequently, has a half-life of approximately 2 ns in aqueous solution and diffusion radius of only about 20 A [(225)].

Fenton's reaction has been used to observe the effect of oxidative stress in DNA, observing that this reaction can induce single and double strand breaks [(226)] and cross-link lesions [(227)].

To detect oxidative stress, one may detect the presence of ROS, by using reagents which will react specifically with one or more ROS, producing a detectable product. One may also detect the presence of ROS by detecting their effects. As mentioned earlier, guanine oxidation is by far the most common oxidation in DNA thus, an appropriate method could be to detect the presence of 8-oxo-dG by the use of specific antibodies.

In summary, oxidative stress is a significant phenomenon, associated with a variety of adverse physiological effects. The effect of oxidation on G-quadruplexes is still not fully understood and our understanding currently relies on the *in vitro* study of sometimes biologically irrelevant structures. The study of the effects of oxidation on intramolecular quadruplexes could increase the

understanding of the mechanism of the global changes to gene expression observed in oxidative stress.

1.3 Applications and Uses of G-quadruplexes in Supramolecular Chemistry and Nanotechnology

Aside from their importance in many biological processes and their potential use as targets for anticancer chemotherapeutics, the unique properties of G-quadruplexes and their constituents, G-quartets, have also been investigated for their exploitation in the fields of supramolecular chemistry and nanotechnology [(9)].

The formation of ordered structures by molecular self-assembly has been pursued as a potential strategy to develop biomaterials for nanotechnology applications [(228)] [(229)]. G-quadruplex DNA, with the ability to form Hydrogen bonds with both its complementary pair and its twin, enhances their potential to form useful nanostructures. G_4-DNA has been used as a system to create large, self-assembling superstructures stabilised by the G-quartets, with self-complementary 5' and 3' ends, known as G-wires [(230)] (Fig. 1-11).

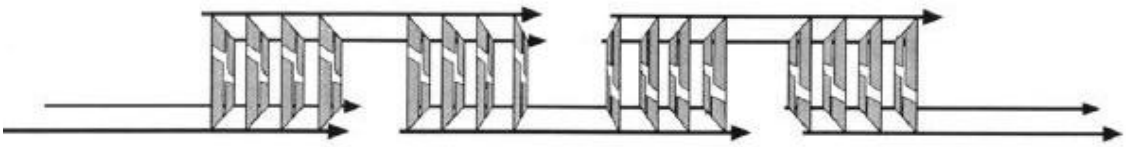

Figure 1-12: Proposed model for a G-wire. Adapted from (230).

These nanostructures can have a length of up to tens of micrometres [(231)] with a diameter of 2.5 nm and is dependent, as G-quadruplex formation, on the presence of Na^+, K^+ and/or Mg^{2+} and are resistant to denaturation [(164)]. G-wires have also been shown to be easily functionalised opening up a vast array of possibilities of these structures [(232)], including the creation of molecular electronic technologies through the incorporation of a switch, sensitive to external stimuli [(233)].

Another important field for the application of the G-quadruplex structure is in the creation of biosensors [9]. A G_4-DNA structure (thrombin binding aptamer, TBA) has been shown to selectively bind to the protease thrombin, inhibiting the formation of fibrin clots [234] [235].

The high affinity and differential stability in the presence of different ions also make G_4-DNA structure an attractive prospect as potential ion channels and as possible detectors and quantifiers of ion concentration. A biomimetic Potassium responsive nano-channel capable of determining the concentration of Potassium has been created using a G-quadruplex, using the conformational changes that occur in the structure in response to ion concentration [236].

The potential of these structures in nanotechnology and supramolecular chemistry is remarkable. Other applications include their use as DNAzymes for colorimetric detection of cocaine [237], their use in the optical detection of proteins [238] amongst many other applications.

1.4 Methods to determine presence of and characterise G-quadruplexes

The presence of G_4-DNA structures in nucleic acid sequences has been inferred by the presence of a GQM, as discussed previously. However a combined biophysical and structural methods approach, using ultra violet (UV) melting analysis and circular dichroism (CD) spectroscopy as well as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy has become indispensable in the investigations into G-quadruplex structure and function [239].

In vitro methods for characterising G-quadruplexes are well established and have been ubiquitously used since the 1960s to investigate the physico-chemical properties of these structures.

These include UV melting experiments, in which the melting of DNA is monitored at increasing temperatures for a hyperchromatic shift at 295 nm, a

maximum for G-quadruplex motifs [240]. This method allows thermodynamic measurements to be obtained from unimolecular and bimolecular quadruplexes but is somewhat ineffective when measuring tetramolecular quadruplexes [10]. CD spectroscopy has also been used, which compares structures to known structural motifs to determine quadruplex topology [241] [242]. Carefully controlled polyacrylamide gel electrophoresis (PAGE) can also be useful in the identification of folded motifs against unfolded single stranded DNA [10].

The most potent methods for elucidating the structure of these complexes have been NMR [243] and X-ray crystallography [244]. These methods can directly and fully determine the structure of G-quadruplexes and have enabled the structural determination of a plethora of these structures, including telomeric [245] [246] and gene regulation-associated [67] [68] [69] G-quadruplexes.

1.4.1 Fluorescence Resonance Energy Transfer

Of particular interest to this project is the use of Fluorescent (or Forster's) Resonance Energy Transfer (FRET) to directly study the conformational of G-quadruplex forming sequences. FRET consists of the measurement of the non-radioactive energy transfer between two dye molecules, termed donor and acceptor, which can be used to report on the intervening distance (within 10-80 Å) and can be estimated by the use of specific formulae [247]. Simply put, when the acceptor dye emits fluorescence, this is detected and indicative of a decrease in inter-dye distance. Thus the distance between two molecules or the dynamic conformation of a molecule can be detected via FRET.

This technique may also use a fluorophore/quencher pair, in which at close distance, the emitting fluorescence from the fluorophore is absorbed by the quencher. In this case, and discounting all other factors, the fluorescence can be seen as proportional to the distance between the fluorophore and quencher.

In terms of the G-quadruplex structure, a variant of FRET, termed single molecule FRET (smFRET) that observes FRET between a single donor and acceptor [248], has been used as both a method to detect the formation of G-

quadruplexes and to gain insight into the folding process [249] [250]. Figure 1-12 is an example of the application of this method to the study of the G-quadruplex structure that occurs in the promoter region of the proto-oncogene c-Kit [251].

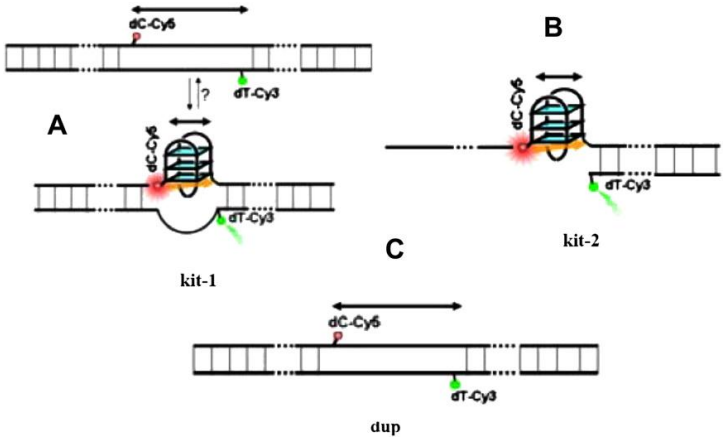

Figure 1-13: Schematic representation of the smFRET experiment carried out by Shirude and co-workers (2007) [251].

The above figure shows a dual-labelled double stranded system containing the GQM sequence and its complementary strand with flanking sequences. Donor (Cy3) and acceptor (Cy5) fluorophores were incorporated on opposite strands, flanking the GQM. The formation of the G-quadruplex would reduce the separation of the fluorophores, leading to an increase in FRET, whereas in the absence of quadruplex formation, FRET would maintain a lower value [251].

Although smFRET is the predominant fluorescent technique used to investigate G-quadruplexes, ensemble FRET, that is to say, the observation of FRET in a solution with numerous donors and acceptors has been used to observe the kinetics of G-quadruplexes [252].

1.4.2 Molecular Dynamics Simulations of G-Quadruplex Structures

Molecular modelling is the science and art of model building and computation to study molecular structure and function. It encompasses a range of different techniques, including *in silico* techniques, ranging from *ab initio* to empirical model building [253].

The field of biomolecular modelling started in the 1960s, gaining momentum in the 1980s and is now advancing rapidly with the advent of more powerful platforms, higher resolution instruments, the creation of genomic and structure databases as well as improvements in the molecular modelling software, i.e., better force field algorithms and computer graphics [254].

Although *in silico* molecular modelling techniques cannot substitute empirical data, they have been used successfully as complementary methods to empirical data; aiding in the interpretation and validation of these results [255].

There are various methods employed for the *in silico* study of molecules; these include both statistical methods (such as the Monte Carlo method [256]) and quantum mechanical methods, such as energy minimisations (EM) and classical molecular dynamics (MD). Recently, hybrid statistical/quantum mechanical simulations have become widely used thanks to increased computational resources; replica exchange molecular dynamics (REMD) is one such method.

The molecular mechanics method, also called force field method or energy minimisation (EM) is a theoretical method based on the use of a force field. This force field is an algorithm that corresponds to parameter sets and the functional form used to describe the potential energy of a system of atoms [253].

These methods calculate the potential energy of a system as a function of the nuclear positions, and attempts to generate the coordinates that correspond to the minimum potential energy. They are iterative descent series methods in which the atomic coordinates are modified in each iteration in order to decrease the potential energy of the system [256]. This method has the advantage of requiring far less computational time than other methods [253]. However its major disadvantage is that it is generally unable to find the global energy minimum of molecules of significant complexity, finding only the local minimum, closest to the starting set of coordinates [256].

There are many force field algorithms that may be used in molecular modelling, which vary mainly in accordance with the molecules and/or systems they were developed to investigate. A subset of algorithms was developed specifically for use on the AMBER molecular dynamics package and is primarily used to study biopolymers, such as nucleic acids [257] [258] [259] [260] [261]. These AMBER force field algorithms were subsequently adapted for use in most other molecular modelling platforms.

Molecular Dynamics is a method of studying the motions and the conformational space of molecular systems by use of classical Newtonian laws of motion (Table 1-2) given a potential energy function and its associated force field [256].

Table 1-1: Newtonian laws of motion

1	A body continues to move in a straight line at constant velocity unless a force acts upon it.
2	Force equals the rate of change of momentum
3	To every action there is an equal and opposite reaction

In essence, this method represents the computer approach to statistical mechanics, used to estimate equilibrium and dynamic information on analytes that are either too complex, or have such a brief lifespan so as to make structure determination by empirical methods inaccurate, or even unviable [254].

Such a method is of high importance as even high resolution X-ray crystallographic data is only a “freeze-view” of a certain molecule, fostering an erroneously static and rigid view of molecules, albeit that this has improved over recent years with refinement methods [262]. However in nature, these molecules are in constant motion, interacting with their environment and amongst themselves [254].

In practical terms this method is far more valuable than molecular mechanics since it allows the system to reach the global minimum, as opposed to the closest local minimum as in MM [256].

In terms of the use of MD for modelling nucleic acids, this method has become ever more important with the rising importance of precise knowledge of not only nucleic acid structure, but also their interactions amongst themselves and with a variety of other molecules [262].

Molecular dynamics simulations of G-quadruplexes have been particularly useful at elucidating some of the subtle features of these structures, including the conformation of the loops [263], stability [264] [265], interactions with ligands [266] [267] [268] [269] and the role of cations [270] [271].

A comprehensive review by Sponer & Spackova (2007) [272], discusses the advantages and disadvantages of the use of molecular dynamics simulations to study quadruplex nucleic acids, concluding that quadruplex simulation studies provide unique insights into the properties of G_4-DNA.

It should be noted that despite this, the current methodology and/or technology do not enable the simulation of the folding pathway of G-quadruplexes, although a recent study claimed to have discovered a new unfolding pathway for a parallel G-quadruplex structure, no empirical data was used to corroborate this statement [273].

There are many software packages designed for molecular modelling, each distinct with different advantages and disadvantages. These include AMBER Molecular Dynamics Software Package, currently in version 11 [274], SYBYL™ Expert Molecular Modelling Environment, currently in version 7.3 [275], MacroModel Molecular Modelling Platform [276] and GROMACS Molecular Dynamics Package, whose version 4.5.3 [277], was used to perform all molecular modelling simulations presented in this report.

GROMACS in an open source engine, designed to perform molecular dynamics simulations and energy minimisation. It's main advantages are summarised in the slogan "Fast, Flexible, Free"; it is highly parallelisable, indeed, it is the base of the Folding@Home project, it is also highly flexible as it is able to perform most mainstream functions in molecular modelling and enables both users and

developers to adapt the algorithms according to convenience or improved knowledge.

The software was primarily designed for the study of biochemical molecules such as proteins, lipids and nucleic acids: large and complex molecules that involve both short range and long range interactions. Its efficiency at calculating non-bonded interactions is particularly appealing, as these tend to be the most computationally expensive part of a system during a molecular dynamics simulation [278].

GROMACS achieves this, in part, through the implementation of particle mesh Ewald (PME) summation of long range forces. Ewald summation is a method for computing the interaction energies of periodic systems, originally developed for use in theoretical physics it is now the standard for calculating long range forces that interact via an inverse square force law, as is the case of electrostatic interactions.

In molecular dynamics, PME calculates the direct-space interactions within a user defined radius using a modification of Coulomb's Law and in reciprocal space using a Fourier transform, building a "mesh" of charges, interpolated onto a grid using Gaussian charge distributions. This charge interpolation is what is used to calculate and incorporate long-range forces into the non-bonded interactions in a simulated system. The estimation of reciprocal forces through this method as opposed to conventional direct measurement methods increases the efficiency of this calculation, requiring far less computational overhead [279] [280].

Another, important aspect to the GROMACS package, is, as mentioned before, its flexibility, achieved through the sheer breadth of programs that comprise the package, allowing, amongst other things, the MD, EM, and mixed QM/classical simulation of a wide variety of systems independent of equilibrium, allows the rapid generation of solvents consisting of various different molecules.

1.4.3 Gene Expression Data Analyses

As described in section 1.1.1, bioinformatics analyses have a played an important role in the understanding of the functions and implications of G-quadruplexes. There is an enormous variety of bioinformatics analyses that can be performed to investigate hypotheses pertaining to the location and effect of motifs on gene expression.

However due to the structural diversity observed in G-quadruplexes, predicting the likelihood of a given sequence to form in vivo is not a simple matter, principally due to chromatin factors, such as proteins and helicases that can stabilise or unwind these structures (reviewed in (281)).

Having said this, current methods can be used to identify motifs with the potential to form G-quadruplexes; for instance, gene expression data of cells subjected to a carcinogen can be analysed for the presence of the general G-quadruplex motif, developed by Huppert and Balasubramanian. If the genes affected by this carcinogen present a statistically significant increase in the presence of potential G-quadruplexes forming motifs, when compared to the prevalence across the entire genome, it is likely that G-quadruplexes play a regulatory role in these genes, and that the chosen carcinogen affects expression levels by affecting the G-quadruplex structure. The result of such analyses can help to support results achieved through other experimental methods.

1.5 Aims and Objectives

G-quadruplexes have been shown to play a role in transcription regulation of certain genes, amongst other biological processes. Several oncogenes, including c-Myc and c-Kit, have been demonstrated as being subject to G-quadruplex regulation of expression. The effect of oxidative stress on G-quadruplexes is not fully understood, however considering the GQM consists of runs of consecutive Guanines, which are particularly susceptible to oxidation, it is reasonable to assume that oxidative stress may have an adverse effect on G-quadruplex structure, with consequences on gene expression.

The aim of this project is to investigate a possible novel route for oxidation mediated oncogenesis through investigating the effect of reactive oxygen species on the stability of G-Quadruplexes and the consequences on gene expression using both *in vitro* and *in silico* methods.

The proposed mechanism by which this would occur would be as follows:

- Oxidation modifies deoxyguanosine to 8-oxo-dG;
- This modification destabilises the G-quadruplex structures found in upstream promoter regions of oncogenes;
- This destabilisation causes the unfolding of the quadruplex to the duplex form;
- In this form, transcription is unimpeded, causing an upregulation of these oncogenes;
- The increased expression of these oncogenes would lead to or predispose a cell to become cancerous.

This aim will be achieved through the following objectives:

Studying whether reactive oxygen species destabilise the G-quadruplex structure *in vitro* using a FRET based assay and Fenton chemistry to observe the effect of oxidative stress on the conformation of an oligonucleotide known to fold into a G-quadruplex at near physiological conditions;

Modelling the effect of 8-hydroxy-2-deoxyguanosine substitutions on the stability of G-Quadruplexes *in silico* through 500 nanosecond classical molecular dynamics simulations comparing the oxidised and non-oxidised G-quadruplex structures of the promoter regions of c-Kit and c-Myc;

Analysis of gene expression data to investigate the effect of oxidation on the expression of genes suspected to be regulated by G-quadruplexes, using gene expression data of cells subjected to oxidation and evaluating the prevalence of G-quadruplex forming motifs in genes whose expression is significantly altered when subjected to oxidation.

2 Experimental section

2.1 Fluorescence Resonance Energy Transfer

As described previously, FRET is a fluorescence intensity-based measurement technique, using a donor and acceptor fluorophore.

The FRET assays in this project were initially attempted on a Thermo Fischer Scientific Varioskan® Flash spectral scanner, however it was established that the instrument was not appropriate for the intended experiment, as it was not possible to determine whether the oligonucleotide under investigation was in the folded G-quadruplex state or in any other conformation; nor was it possible to determine any alteration in the conformation through observation of the FRET values. This was determined to be due to the lack of precision at low concentrations of the sample oligonucleotide.

Due to inability to determine the origin of the fluorescence as being due to the folding/unfolding of the G-quadruplex structure, and that the Varioskan® Flash spectral scanner proved inefficient at detecting fluorescence at the concentrations desired, the decision was made to attempt to perform the experiments in the Quiagen® Rotor-Gene Q qPCR machine.

This instrument uses a rotary format which spins all samples at 3000 rpm and enables detection in specific channels ranging from UV to infrared wavelengths. For the intended experiment this instrument has the advantage of using small volumes (approximately 20ul) has a very accurate thermal cycler and absence of well-to-well light contamination.as well as being very sensitive, in that the limit of detection is very low, as the instrument can reliably detect down to a copy number of ten, when performing PCR experiments.

The experiments were performed using 1 second cycles with fluorescence being measured in all samples using three channels (table 2-1).

Table 2-1: Fluorescence measurement channels used in all experiments on the Rotor Gene Q PCR machine.

Channel	Excitation Wavelength	Emission Wavelength
Cy3	530nm	555nm
Cy5	625nm	660nm
Cy3-Cy5 FRET	530nm	660nm

Although numerous experiments were run in this instrument, only a few examples will be described, in which all the most salient observations can be made. Data, descriptive statistics and graphs of all experiments performed can be found in the Appendix DVD.

The following experiment (figures 2-1 and 2-2) is a representative example of the initial experiments aimed at determining the conformation of the nucleotide and controlling the shift between unfolded and folded G-quadruplex states.

This experiment involved observing FRET, Cy3 and Cy5 fluorescence at 25 C, in samples with and without 100 mM KCl and with and without complementary sequence.

Performing experiments with and without 100 mM KCl were intended to establish a significant increase in fluorescence in the presence of KCl, thus establish K^+ dependent FRET, a characteristic used to determine the presence of a G-quadruplex.

The experiment was also adapted to use the complementary sequence to the GQM (Figure 2-1) as a method of unfolding the quadruplex in a predictable manner, to monitor the decrease in FRET.

GQM	5′ – [Cy3]TATGGTGAGGGTGGGTAGGGTGGGTAAGGTGG[Cy5] – 3′
Comp	3′ – ATACCACTCCCACCCATCCCACCCATTCCACC – 5′

Figure 2-1: Sequence of the synthesized c-Myc GQM and complementary sequence used in the FRET Assays.

2.1.1 Determination of Presence of c-Myc G-Quadruplex Structure

Due to the problems encountered in the previous instrument, the decision was made to include a temperature ramping step. This would enable the identification of the origin of fluorescence as that of a stable secondary structure (be it a G-quadruplex or any other DNA secondary structure, as well as any GQM dimers) within the oligonucleotide. It was expected that by raising the temperature, FRET would be reduced due to destabilisation of the quadruplex structure at high temperatures.

The experiment was performed therefore, with and without 100 mM KCl and with and without excess (5 µM) complementary DNA. The thermocycler was set to run at 25°C for 45 cycles, and the temperature was then raised to 90°C and held there for 45 cycles and cooled to 25°C and held for 45 cycles.

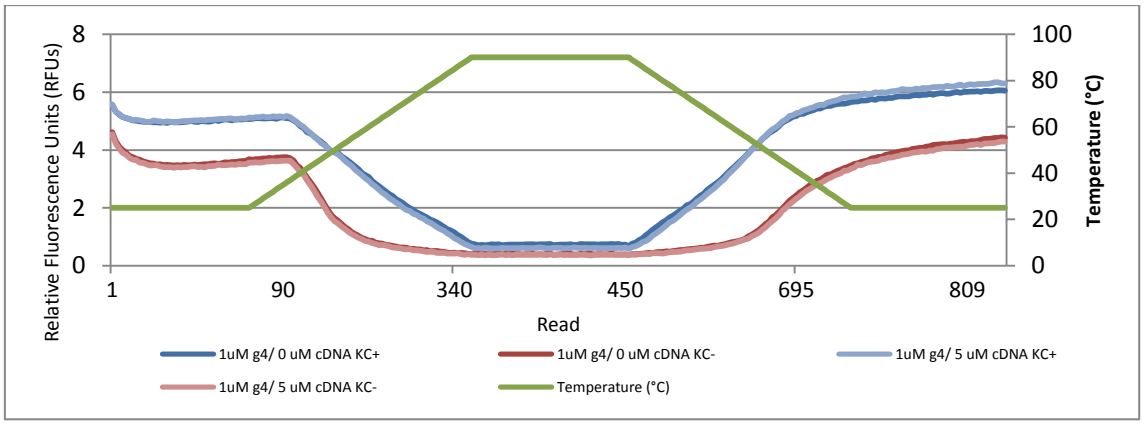

Figure 2-2: FRET values observing effect of concentrations of complementary sequence and KCL, plotted against temperature.

Immediately noticeable from Figure 2-2, when comparing with the results obtained from the Varioskan spectral scanner (see digital Appendix), is the precision of repeated fluorescence readings at constant temperature.

Within the first 45 cycles at 25 C, fluorescence in all four tubes show a logarithmic descente curve, with an increase towards the end. As temperature increases from 25°C to 90 C, fluorescence in all four tubes approaches 0 RFUs and remains there throught the next 45 cycles at 90 C. As temperature

decreases from 90 C back down to 25°C , fluorescence in all four tubes increases and continues to increase throught the 45 cycles at 25°C .

When observing the effect of K^+ concentration on FRET, it can clearly be seen that both samples without potassium have a reduced fluorescence throughout the experiment when compared to the two samples with 100 mM KCl.

When observing the effect of complementary sequence on FRET, no discernable difference can be seen when comparing samples with and without 5 µM of complementary sequence.

As a control measure for this experiment, the Cy5 flurescence values were analysed (figure 2-3).

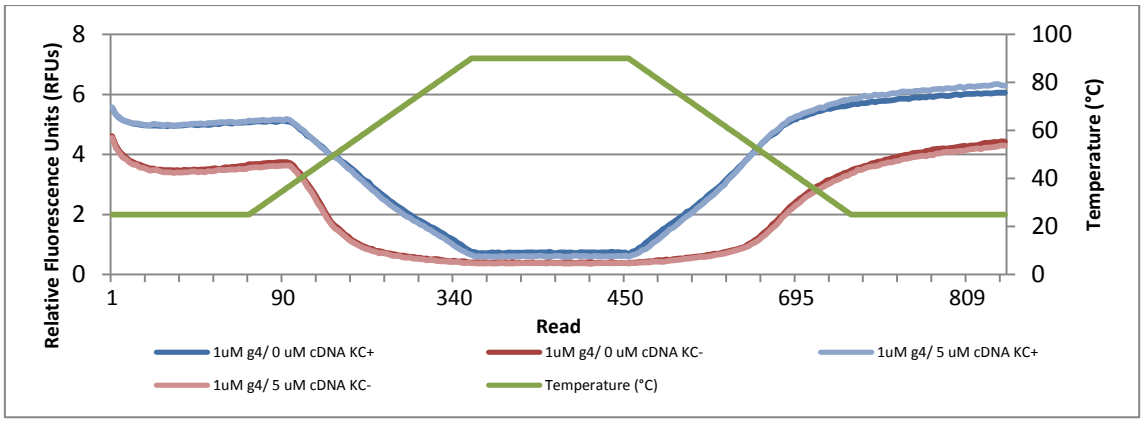

Figure 2-3: Cy5 fluorescence values observing effect of concentrations of complementary sequence and KCL, plotted against temperature.

In general terms, figure 2-3, shows Cy5 fluorescence values with similar behaviour throughout the experiment, to the FRET values observed in figure 2-2, with fluorescence of all four samples staying at 100 RFUs for the first 45 cycles at 25°C , the falling sharply to 20 RFUs when temperature is increased to 90°C and rising again to 100 RFUs when the sample is cooled back to 25°C . However no significant difference can be observed between samples with and without 100 mM KCL or 5 µM of complementary sequence.

This experiment therefore revealed that that FRET reduced with an increase in temperature; however, when observing Cy3 and Cy5 fluorescence, these were

also reduced by the same factor. As such, the temperature denaturing step was considered inappropriate, and removed.

Observing the effect of complementary sequences however, showed that the G-quadruplex formed was remarkably stable, as in the near physiological conditions previously described (100 mM KCl, 10mM Sodium Cacodylate at pH 7.4) fluorescence was not substantially affected by the addition of complementary DNA.

2.1.2 Initial Oxidation Experiments

The next group of experiments, of which figure 2-4 is a representative example, aimed to test the effect of oxidation on the G-quadruplex structure; samples were run in parallel with and without the corresponding complementary sequence in the same conditions described above, although taking into account the observations made in the previous experiments, fluorescence readings were taken only at 25°C , and the cycles at this temperature were increased from 45 to 60. The temperature ramping step was maintained, however no readings were performed during its execution.

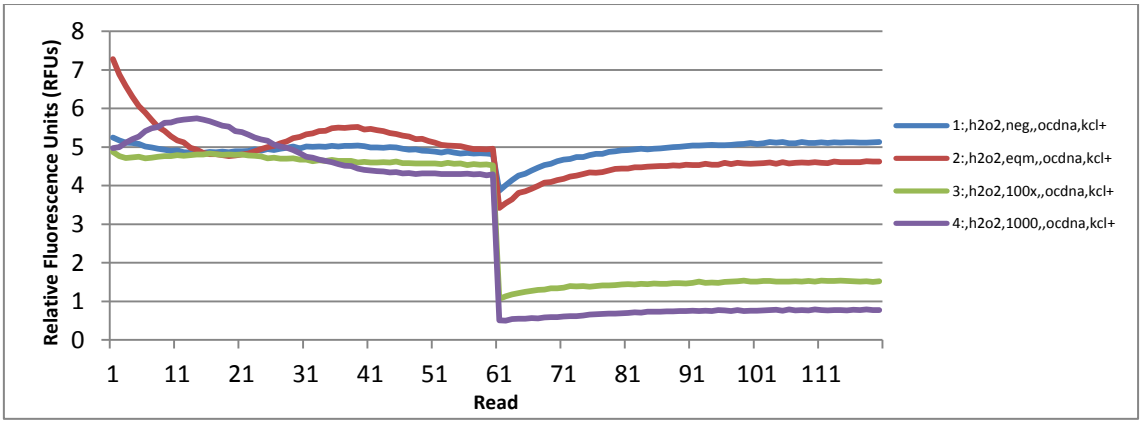

Figure 2-4: FRET values observing the effect of H_2O_2 concentration.

Observing figure 2-4, in the first section, fluorescence values for the samples 1 and 2 ([H_2O_2] = 0 and 1µM, respectively), fluorescence fluctuates and appears to stabilise at values similar to those of samples 3 and 4 ([H_2O_2] = 100 and 1000µM, respectively). After the thermal denaturing step, fluorescence in all

four samples decreases in a proportionate manner to H_2O_2 concentration. As a control measure, Cy3 and Cy5 fluorescence were also measured; figure 2-13 denotes Cy5 fluorescence values for this experiment.

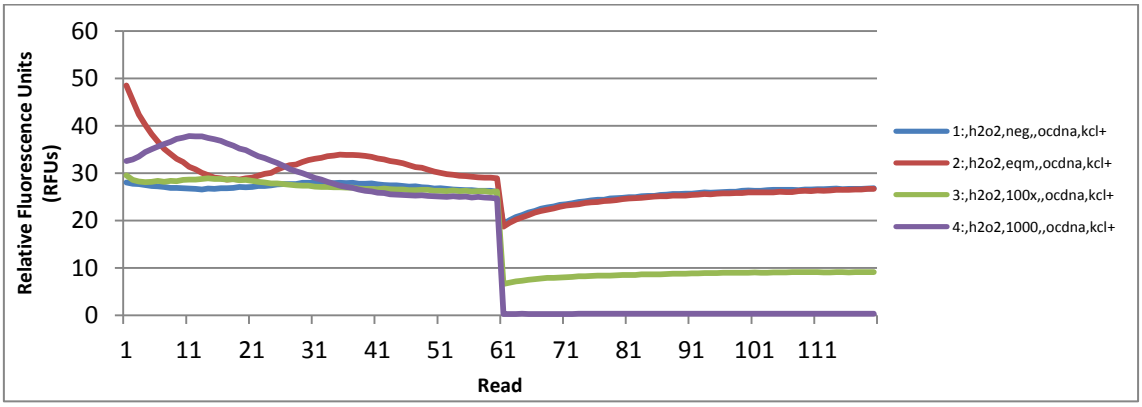

Figure 2-5: Cy5 fluorescence values observing the effect of H_2O_2 concentration.

Results in figure 2-5 are broadly similar to those in figure 2-4, with samples 1 and 2 fluctuating in the first 60 cycles and appearing to stabilise at similar values to samples 3 and 4; and samples decreasing in proportion to H_2O_2 concentration after the thermal denaturing step. The most prominent difference was that after the denaturing step, in samples 1 and 2, Cy5 values coincide whereas FRET values for these samples differed slightly with a mean difference of 0.38 RFUs.

In general terms, the experiment was unable to detect changes in the G-quadruplex structure at low Hydrogen peroxide concentrations and high concentrations lead to a rapid and irreversible decrease in Cy5 fluorescence.

These results, combined with the inability to induce formation of duplex DNA indicated that the c-Myc G-quadruplex structure was extremely stable, this theory was given credence by a study by Kumar and co-workers (2008) [(282)], which demonstrated that quadruplexes with short loop lengths (as is the case of c-Myc) tend to be more stable than those with longer loop lengths.

This also led to a change in the direction of the project. Considering a possible differential stability of different G-quadruplexes, this would likely lead to a

differential effect of oxidation on different G4 structures. This could even include the possibility of the quadruplex form being favoured by oxidation. This hypothesis was added credence by studies by Mergny and co-workers (2007) [180], who observed accelerated formation of an artificial tetramolecular G-quadruplex structure, under certain conditions, when 8-oxo-dG was incorporated.

It was therefore decided that experiments aimed at studying in more detail quadruplex/duplex competition should be performed on a sequence known to favour the duplex form at near physiological conditions.

2.1.3 Determination of Presence of c-Kit G-Quadruplex Structure

For this purpose, a different G-quadruplex structure was selected to study the effect of oxidation. The GQM found in the upstream promoter region of the proto-oncogene c-Kit was used, as this structure has been elucidated through NMR spectroscopy [125] and the thermodynamic competition with duplex DNA has also been studied [70]. The GQM and complementary sequences used are represented in Figure 2-5, these oligonucleotides were synthesised by Thermo Fisher Scientific (UK).

Two oligonucleotides were synthesised, with 8 and 12 flanking bases of the GQM (figure 2-6). The duplex/quadruplex competition of these two sequences has been studied [70], and revealed that the duplex form is favoured when the complementary sequence is present in equimolar concentrations.

ckit8	5′-[Cy3]CGCGCAGAGGGAGGGCGCTGGGAGGAGGGGCTGCTGC[Cy5]-3′
Comp8	3′-GCGCGTCTCCCTCCCGCGACCCTCCTCCCCGACGACG-5′
ckit12	5′-[Cy3]CCGGCGCGCAGAGGGAGGGCGCTGGGAGGAGGGGCTGCTGCTCGC[Cy5]-3′
Comp12	3′-GCGAGCAGCAGCCCCTCCTCCCAGCGCCCTCCCTCTGCGCGCCGG-5′

Figure 2-6: Sequence of the synthesized c-Kit GQMs and complementary sequences used in the FRET Assays.

Using these sequences, experiments were performed using similar parameters to those used previously: tubes were filled with 1 µM c-Kit GQM with 10 mM Sodium Cacodylate at pH 7.4. The baseline temperature was altered to 37°C ,

as this is physiologically relevant, and preliminary results indicated that fluorescence levels were stable at this temperature.

Although c-Kit GQM containing oligonucleotides were synthesised with 8 and 12 flanking bases, only experiments related to the 8-flanked GQM will be described, as these presented more accurate results. Results for the 12 flanked GQM and all other omitted experiments can be found in the Appendix DVD.

Due to the spike in fluorescence observed in the first few readings after the temperature ramping, this step was removed. This appeared to have little effect on the correlation previously described. As such, the temperature ramping step was considered unnecessary and all subsequent experiments were performed at a constant temperature of 37°C .

As with the c-Myc GQM, it was important to establish K^+ dependent FRET, for this reason an experiment was run at 37°C for 100 1 second cycles (figures 2-7 and 2-8), with two samples: one with 100 mM KCL (green) and one without KCl (pink). Note that data for the first 10 cycles is not show as mixing occurred during this time and the fluorescence values are not representative.

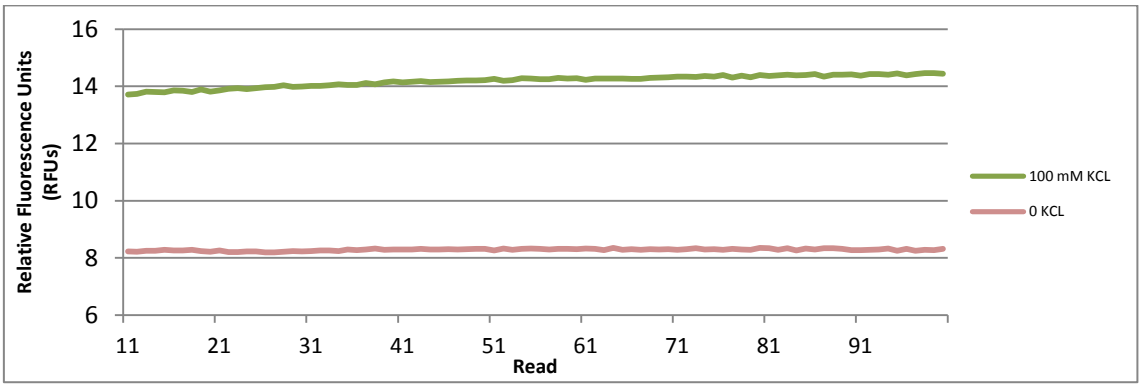

Figure 2-7: FRET values observing the effect of 100 mM KCL.

As can be observed from figure 2-7, fluorescence is stable in both samples over the 90 cycles and there is a clear increase to fluorescence (difference in means of 5.9 RFUs) when the sample is in the presence of 100 mM KCL. As a control measure, Cy3 and Cy5 fluorescence were also measured and the data is plotted in figure 2-8.

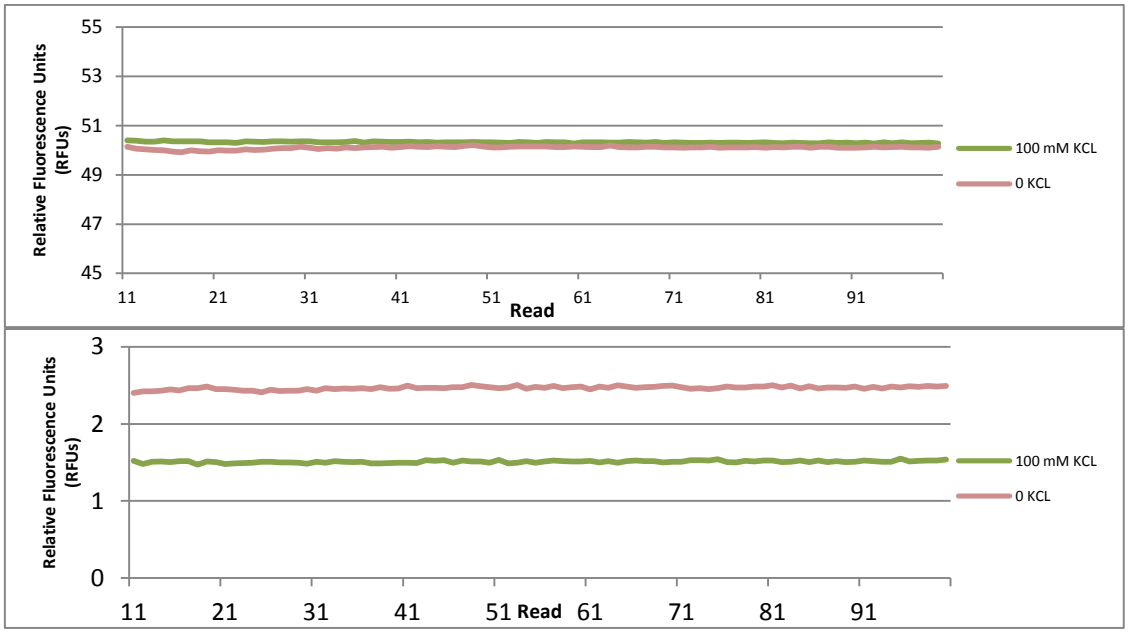

Figure 2-8: Cy5 (top) and Cy3 (bottom) fluorescence values observing the effect of 100 mM KCL.

Observing Cy5 fluorescence in figure 2-8, the fluorescence values are very similar, to the point that they almost coincide throughout the last 60 cycles; whereas Cy3 fluorescence values are markedly different between samples, with the sample without KCl having a higher fluorescence than the sample with 100 mM KCl.

The following experiment was devised to establish annealing curves for the GQM with its complementary sequence. The instrument was set to run at 37°C for 30 cycles to establish a baseline fluorescence level for all three channels (FRET, Cy5 and Cy3) and then the complementary oligonucleotide was added to samples 2, 3 and 4 in varying concentrations (0.5, 1 and 10 µM) and reading for all three channels were taken for 200 cycles (Figures 2-9, 2-10 and 2-11).

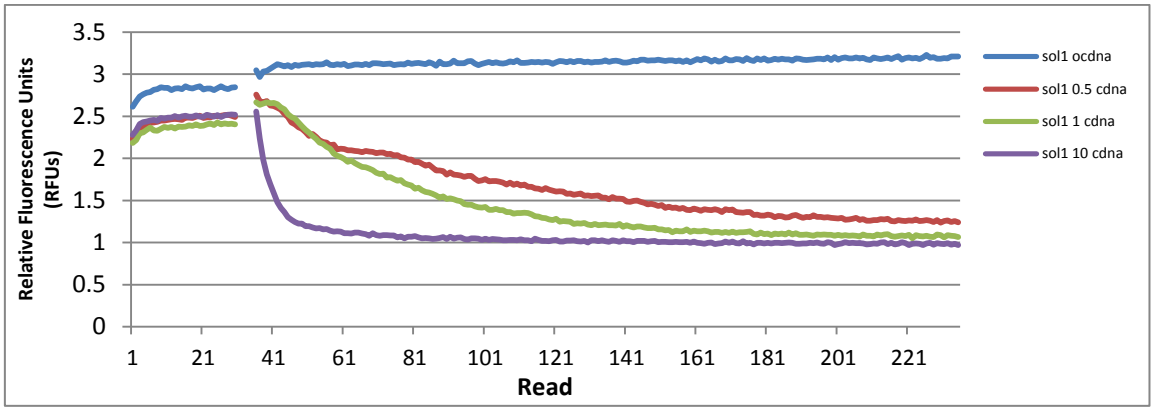

Figure 2-9: FRET values observing the effect of cDNA concentration.

In figure 2-9, a reduction in fluorescence is clearly observable in the three samples in which complementary oligonucleotide was added, with a steep descent in proportion to the concentration of complement, whereas the sample without complementary oligonucleotide added maintains an approximately constant fluorescence. Another noticeable change is that all four samples, at the start of the second section appear to have increased fluorescence in the first readings, with the sample without complementary oligonucleotide maintaining a higher fluorescence than observed in the first section.

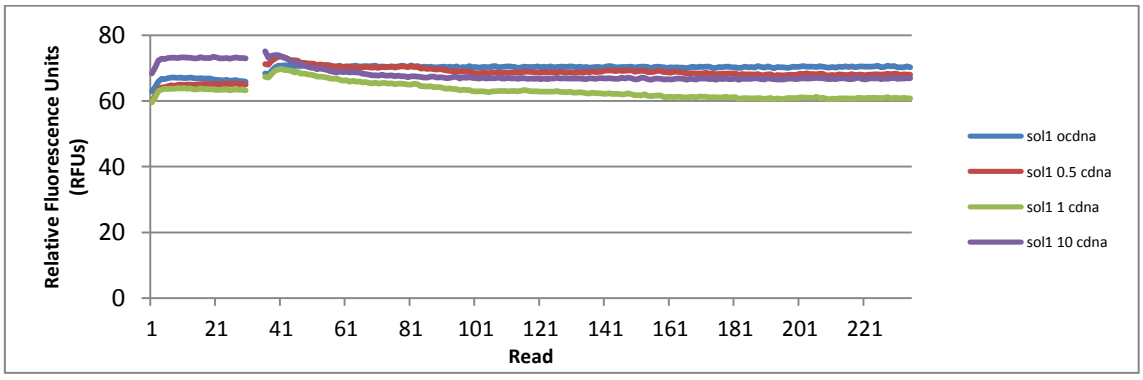

Figure 2-10: Cy5 fluorescence values observing the effect of cDNA concentration.

As can be observed from figure 2-10, fluorescence is relatively constant throughout the experiment, with the exception of an increase in the first few cycles of each section, this lead to fluorescence in the second section being increased in all four samples when compared to the first section.

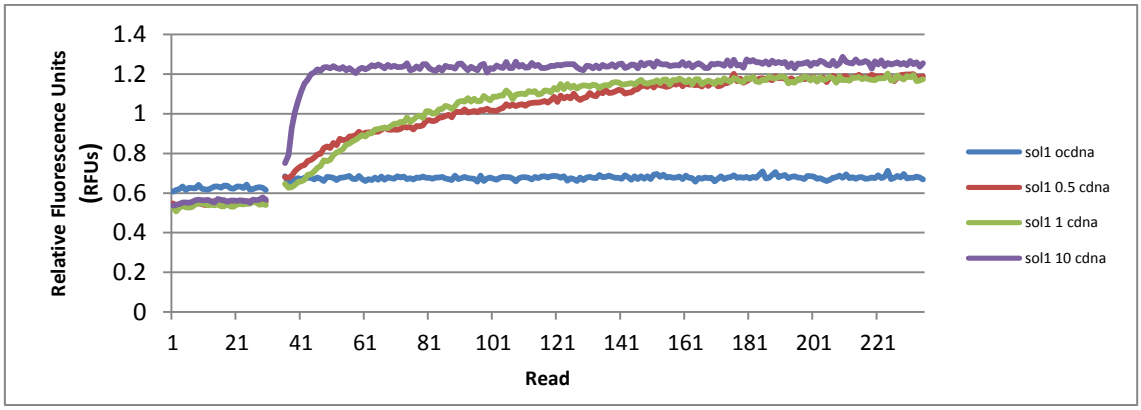

Figure 2-11: Cy3 fluorescence values observing the effect of cDNA concentration.

Figure 2-11 describes an almost mirror image of values of Cy3 fluorescence when compared to FRET; fluorescence increases in a proportionate manner to the concentration of complementary sequence. However, whereas in the logarithmic descent in FRET samples 2 (0.5 µM cDNA) and 3 (1 µM cDNA) are clearly differentiated, in the exponential raise in Cy3 fluorescence of these two samples appear to be less differentiated. It should also be noted that the fluorescence levels of sample 1 (0 cDNA) are raised in the second section, when compared to the initial section.

To eliminate the contribution of an increase in optical density to the decrease in FRET when complementary oligonucleotide is added, control sequences were utilised. These sequences were selected to not form stable duplexes with the c-Kit GQM used or form dimers or any stable secondary structure (figure 2-12).

Ctrl8	5'- CTATCCTATTTGATTCTTATCGTTCGCTCGGTTGTTT -3'
Ctrl12	5'- TCGCCTATCCTATTTGATTCTTATCGTTCGCTCGGTTGTTTGTAT -3'

Figure 2-12: Sequence of the synthesized c-Kit control complementary sequences used in the FRET Assays.

Furthermore, to increase the precision of the experiment, the concentration of GQM used was doubled to 2 µM. This was intended to reduce the influence of very small variations in concentrations that are inevitable in a method that involves manual pipetting. Concentrations of the complementary sequence and Hydrogen peroxide were adjusted accordingly.

2.1.4 Optimised Experiments

2.1.1a First optimised Experiment

The preliminary optimised method was performed on the samples presented in table 2-2. All experiments were performed in 100 mM KCl and all tubes were filled with 10 mM Sodium Cacodylate (pH 7.4) to a total volume of 20 µl at 37°C.

Table 2-2: Samples used in optimised protocol.

	GQM used	GQM concentration	H_2O_2 concentration	Complementary sequence used
C1	Ckit8	2 µM	0	-
C2	Ckit12	2 µM	0	-
1	Ckit8	2 µM	0	Ctrl8 (2 µM)
2	Ckit8	2 µM	0	Comp8 (2 µM)
3	Ckit8	2 µM	2 µM	Ctrl8 (2 µM)
4	Ckit8	2 µM	2 µM	Comp8 (2 µM)
5	Ckit8	2 µM	20 µM	Ctrl8 (2 µM)
6	Ckit8	2 µM	20 µM	Comp8 (2 µM)
7	Ckit8	2 µM	2000 µM (2 mM)	Ctrl8 (2 µM)
8	Ckit8	2 µM	2000 µM (2 mM)	Comp8 (2 µM)
9	Ckit8	2 µM	200.000 µM (0.2 M)	Ctrl8 (2 µM)
10	Ckit8	2 µM	200.000 µM (0.2 M)	Comp8 (2 µM)
11	Ckit8	2 µM	2.000.000 µM (0.2 M)	Ctrl8 (2 µM)
12	Ckit8	2 µM	2.000.000 µM (0.2 M)	Comp8 (2 µM)
13	Ckit12	2 µM	0	Ctrl12 (2 µM)
14	Ckit12	2 µM	0	Comp12 (2 µM)
15	Ckit12	2 µM	2 µM	Ctrl12 (2 µM)
16	Ckit12	2 µM	2 µM	Comp12 (2 µM)
17	Ckit12	2 µM	20 µM	Ctrl12 (2 µM)
18	Ckit12	2 µM	20 µM	Comp12 (2 µM)
19	Ckit12	2 µM	2000 µM (2 mM)	Ctrl12 (2 µM)
20	Ckit12	2 µM	2000 µM (2 mM)	Comp12 (2 µM)
21	Ckit12	2 µM	200.000 µM (0.2 M)	Ctrl12 (2 µM)
22	Ckit12	2 µM	200.000 µM (0.2 M)	Comp12 (2 µM)
23	Ckit12	2 µM	2.000.000 µM (0.2 M)	Ctrl12 (2 µM)
24	Ckit12	2 µM	2.000.000 µM (0.2 M)	Comp12 (2 µM)

To directly visualise the effect of the addition of both the complementary sequence and Hydrogen peroxide, 3 separate runs were performed: the first serving to establish the baseline fluorescence in the presence of the G-

quadruplex structure; the second to observe the effect of H_2O_2; and the third to observe the effect of complementary sequence.

The protocol was as followed:

1. Samples were prepared by adding to each tube:
    a. 4 µl of 10 µM solutions of the ckit8 and ckit12 solutions
    b. 2 µl of a 1 M solution of KCl
    c. 6 µl of 10 mM Sodium Cacodylate (pH 7.4)
2. Tubes were placed in the Rotor-gene Q PCR machine and were run under the following conditions:
    a. 100 cycles of 1 second at 37°C with fluorescence detection after each cycle in the following channels:

Channel	Excitation Wavelength	Emission Wavelength
Cy3	530nm	555nm
Cy5	625nm	660nm
Cy3-Cy5 FRET	530nm	660nm

3. The tubes were removed from the instrument.
4. Hydrogen peroxide was added to the tubes to achieve the concentrations described in table 2-2. All concentrations were achieved using 4 µl of H_2O_2 solutions. To maintain equal volume in all samples, 4 µl of water was added to the samples to which H_2O_2 was not added. Care was taken to pipette the solutions onto the side of the tubes, without being in contact with the rest of the solution; the rotation of the samples within the instrument will mix the solution before the first reading.
5. The tubes were placed in the instrument, which was set to same conditions as in step 2.

6. Tubes were again removed from the instrument.

7. The complementary sequences were added to achieve the concentrations described in table 2-2. Concentrations were achieved using 4ul of 10 µM solutions. To maintain equal volume in all samples, 4 µl of Sodium Cacodylate was added to the samples to which complementary oligonucleotide was not added. The same care was taken when pipetting as in step 4.

8. The tubes were placed in the instrument, which was set to same conditions as in steps 2 and 5.

9. Fluorescence data from the instrument was then analysed, as described throughout this section.

Figures 2-13 to 2-16 represent the results achieved using the above protocol. However, samples using the 12 flanked GQM are not shown as they proved to be less accurate, these can be consulted in the Appendix DVD. It should be noted that the first 10 data points in the first section were omitted, as mixing occurred during this period and FRET values were not representative.

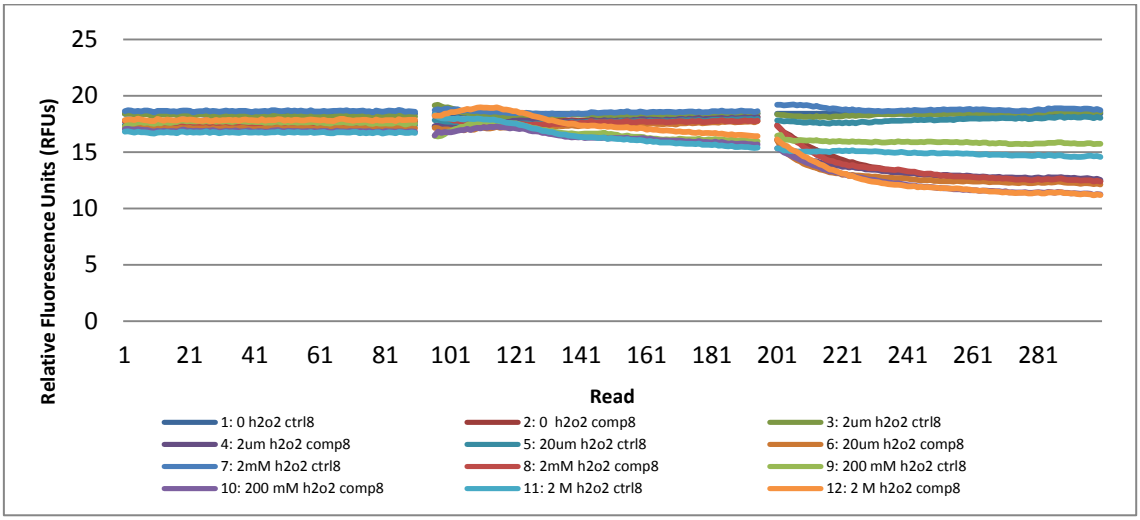

Figure 2-13: FRET values observing effect of Hydrogen peroxide and complementary oligonucleotide concentration.

As can be observed from figure 2-13, this method showed a relatively good precision during the first section, as all samples are within approximately 2 RFUs of each other. At the start of the second and third sections, as observed in other experiments, fluorescence in all samples experiences a shift; in this case, the shift is toward a lower fluorescence value. During the second section (when H_2O_2 was added at different concentrations), samples with 200mM or 2 M H_2O_2 decreased (samples 9-12), whereas samples with 0, 2 µM, 20 µM and 2 mM (samples 1-8) stayed constant throughout the section.

During the third section (in which complementary sequence and control sequence was added), those samples in which complementary oligonucleotide was added decreased sharply, whereas samples to which control sequence was added remained far more stable in general. The exception being the samples to which control sequence was added as well as 200 mM and 2 M H_2O_2 (samples 9 and 11) was added, which decreased a moderate amount, although not as steeply as those samples to which complementary sequence was added.

To eliminate other intervening factors, such as temperature fluctuations, the controls were normalised to a straight line at their average point and all other samples were adjusted, point-by-point by the same values as those required to produce a constant fluorescence for the control sample.

As can be observed from figure 2-14, before normalisation, fluorescence values are disparate from the start and throughout the experiment. In the second section, samples to which 200 mM or 2 M H_2O_2 was added experienced a reduction in fluorescence, whereas the other samples remained constant. In the third section these same samples also decreased substantially whereas the remaining samples remained fairly constant.

After the samples were normalised to the control (sample 1: 0 H_2O_2 with control sequence), results were largely similar, with better differentiation between samples with 200 mM and 2 M H_2O_2 and the remaining samples.

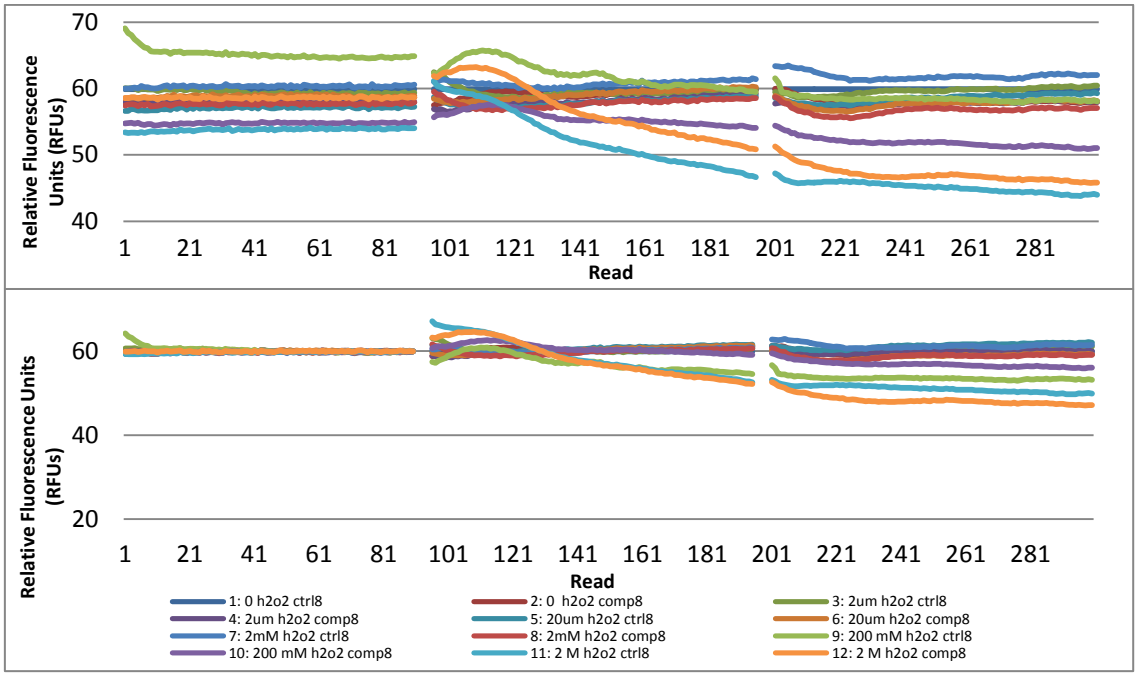

Figure 2-14: Cy5 fluorescence values observing effect of Hydrogen peroxide and complementary oligonucleotide concentration raw data (above) and data normalised to control values (sample 1) (below).

As can be observed from figure 2-15, results reflect an almost mirror image to FRET results, with section one presenting similar, stable values for all the samples (with the exception of sample 9, which although stable, presented higher fluorescence values). During section two, fluorescence levels of samples 9, 10, 11 and 12 (samples with 200 mM and 2 M H_2O_2 added) showed significant increases. In the third section, samples with complementary sequence showed increased fluorescence as did samples 9 and 11.

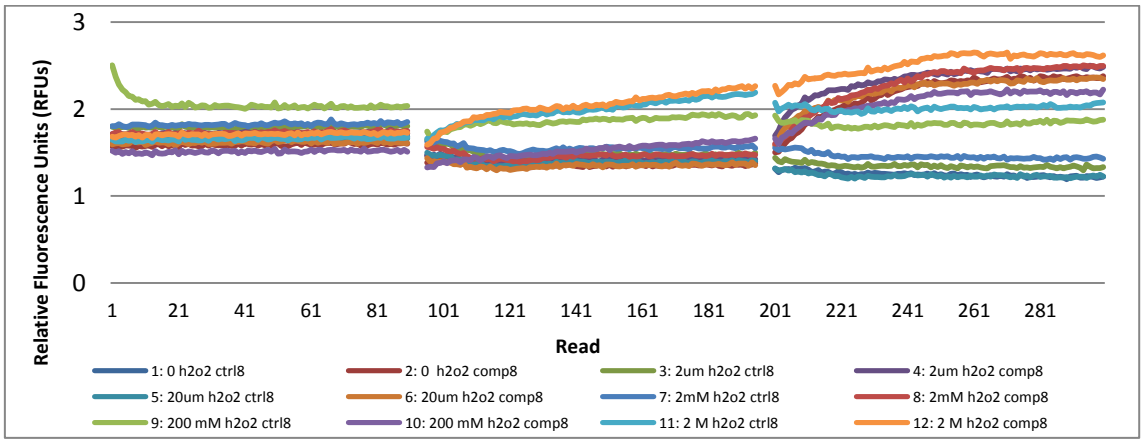

Figure 2-15: Cy3 fluorescence values observing effect of Hydrogen peroxide and complementary oligonucleotide concentration raw data.

Figure 2-16 displays the treated FRET data. Data treatment involves normalisation with respect to the control (as described previously) and, to discount interfering factors that affect FRET by decreasing the fluorescence detected by Cy5, any change in Cy5 fluorescence was taken into account by increasing the FRET value by the same percentage value as the Cy5 fluorescence decreased, and vice-versa.

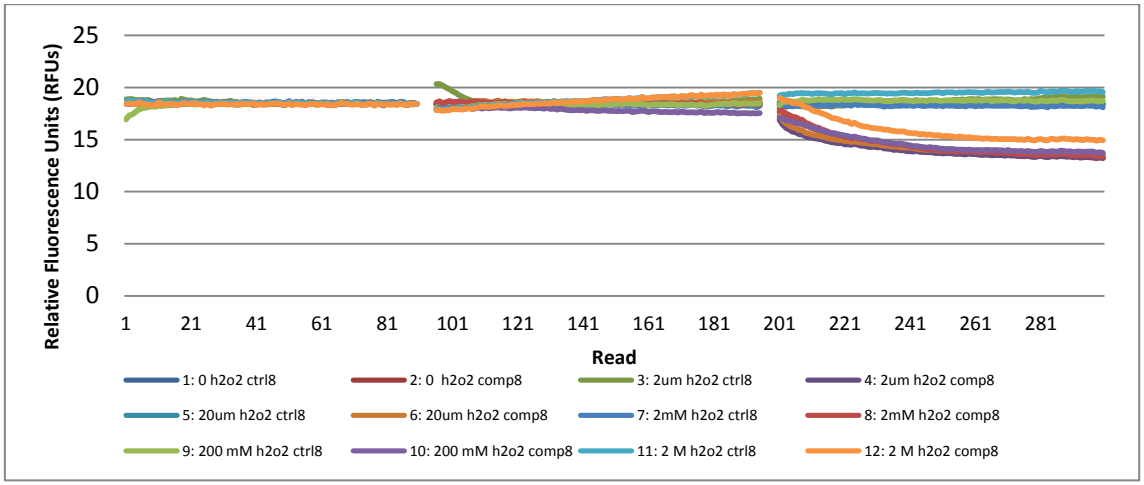

Figure 2-16: FRET values observing effect of Hydrogen peroxide and complementary oligonucleotide concentration, normalised to control and Cy5 fluorescence.

Figure 2-16 represents the same results as shown in figure 2-15, with the data normalised with respect to the control sample (1) and discounting the change in Cy5 fluorescence from the values detected.

Using this data analysis, results can be seen to be constant throughout sections 1 and 2; whilst in the third section; samples with complementary sequence added experienced a decrease in fluorescence.

2.1.1b Second optimised experiment

The fully optimised method, including the specific induction of Fenton's reaction, by adding not only H_2O_2, but also $FeCl_2$ and ascorbic acid, was then performed. To ensure none of the components of this reaction would interfere with fluorescence, controls were performed, isolating each of the components to observe their effect on FRET values observed, as well as "production" samples,

in accordance with table 2-3, observing the effect of each of the components of Fenton's reaction on fluorescence (Figures 2-17 and 2-18).

Table 2-3: Samples used in experiment evaluating effect of Fenton's reaction components.

	GQM conc.	KCl conc.	H_2O_2 conc.	$FeCl_2$ conc.	Ascorbic Acid conc.	Comp. sequence conc.
1	-	100 mM	300 µM	20 µM	50 µM	0
2	-	100 mM	300 µM	20 µM	50 µM	2 µM
3	2 µM	0	300 µM	20 µM	50 µM	0
4	2 µM	0	300 µM	20 µM	50 µM	2 µM
5	2 µM	0	0	0	0	0
6	2 µM	0	0	0	0	2 µM
7	2 µM	100 mM	300 µM	0	0	0
8	2 µM	100 mM	300 µM	0	0	2 µM
9	2 µM	100 mM	0	20 µM	0	0
10	2 µM	100 mM	0	20 µM	0	2 µM
11	2 µM	100 mM	0	0	50 µM	0
12	2 µM	100 mM	0	0	50 µM	2 µM
13	2 µM	100 mM	300 µM	20 µM	0	0
14	2 µM	100 mM	300 µM	20 µM	0	2 µM
15	2 µM	100 mM	0	20 µM	50 µM	0
16	2 µM	100 mM	0	20 µM	50 µM	2 µM
17	2 µM	100 mM	300 µM	20 µM	50 µM	0
18	2 µM	100 mM	300 µM	20 µM	50 µM	2 µM
19	2 µM	100 mM	300 µM	20 µM	50 µM	0
20	2 µM	100 mM	300 µM	20 µM	50 µM	2 µM
Ctrl	2 µM	100 mM	0	0	0	0
Ctrl2	2 µM	100 mM	0	0	0	2 µM

The protocol used was identical to the previously described optimised method, differing only in the constituents of each tube. $FeCl_2$ and Ascorbic acid were added prior to the first reading.

Utilising the samples indicated in table 2-3, yielded fluorescence results presented in Figures 2-17 and 2-18. As with previous experiments, only the most relevant data is represented here, to consult the full data sets, please refer to the Appendix DVD.

As can be observed from figure 2-17, neither method of inducing oxidation (performed after read 100) affected fluorescence values at these concentrations, and there is no apparent difference between the use of Fenton's reaction (used in samples 17 and18) and H_2O_2 alone (samples 7 and 8). However, when observing section 3, both oxidation methods appeared to hinder

binding to the complementary oligonucleotide, producing more tenuous hybridisation curves than the control. The decrease in fluorescence observed at the beginning of each new section is a characteristic that is noticed in every experiment performed with this method including controls, caused by variations in temperature and the inevitable loss of fluorescence that occurs over time.

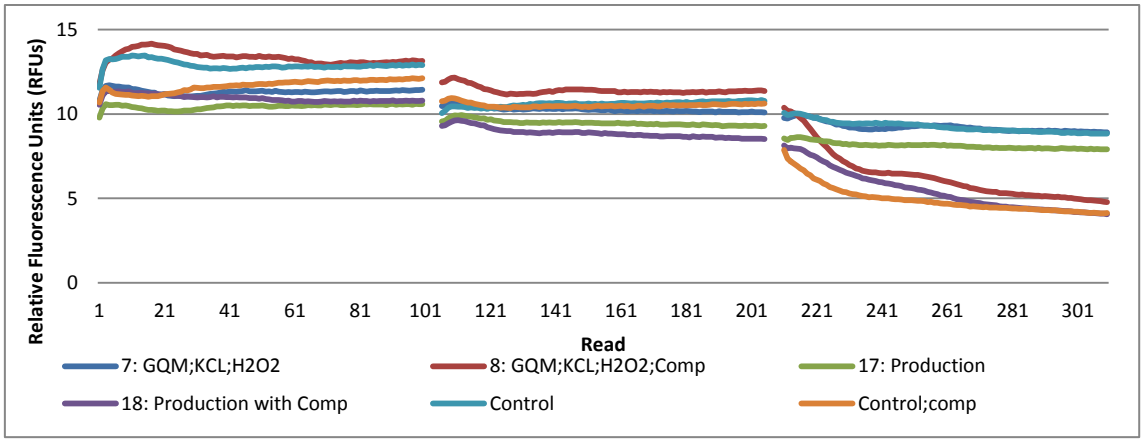

Figure 2-17: FRET values comparing the effect of H_2O_2 and Fenton's reaction.

Figure 2-18 represents the FRET values observing the effect of $FeCl_2$ and Ascorbic acid on Fluorescence, when compared to a control. Figure 2-18 shows that neither $FeCl_2$ nor ascorbic acid (added at read 100) have an observable effect on fluorescence values. However, as in the samples represented in figure 2-18, the samples with $FeCl_2$ or ascorbic acid show an attenuated hybridisation curve.

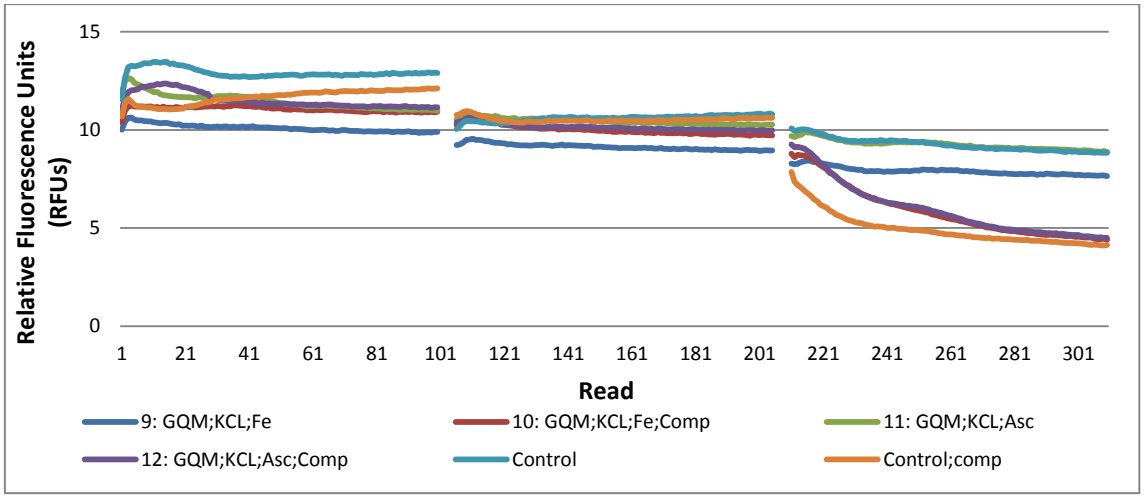

Figure 2-18: FRET values observing the effect of ascorbic acid and $FeCl_2$ on fluorescence.

The components used in this experiment were therefore deemed to be suitable to observe the effect of oxidation. However the concentrations used in this experiment were deemed to be insufficient to induce oxidation under the conditions of the developed method.

2.1.1c Third optimised experiment

Figures 2-19 to 2-22 represent an experiment in which a series of higher concentrations were used, based on a study on Guanine oxidation by White and co-workers (2003) [283], this experiment was aimed at revealing the threshold for oxidation affecting fluorescence, as well as to establish whether a correlation could be made between H_2O_2 concentration and the hybridisation curve incline. This experiment was devised using the samples described in table 2-5, yielding results presented in figures 2-19 to 2-22.

Table 2-4: Samples used in oxidation threshold experiment.

	GQM conc.	KCl conc.	H_2O_2 conc.	$FeCl_2$ conc.	Ascorbic Acid conc.	Comp. sequence conc.
1	2 µM	100 mM	0	0	0	0
2	2 µM	100 mM	0	0	0	2 µM
3	2 µM	100 mM	0	150 µM	50 µM	0
4	2 µM	100 mM	0	150 µM	50 µM	2 µM
5	2 µM	100 mM	10 mM	150 µM	50 µM	0
6	2 µM	100 mM	10 mM	150 µM	50 µM	2 µM
7	2 µM	100 mM	20 mM	150 µM	50 µM	0
8	2 µM	100 mM	20 mM	150 µM	50 µM	2 µM
9	2 µM	100 mM	30 mM	150 µM	50 µM	0
10	2 µM	100 mM	30 mM	150 µM	50 µM	2 µM
11	2 µM	100 mM	40 mM	150 µM	50 µM	0
12	2 µM	100 mM	40 mM	150 µM	50 µM	2 µM
13	2 µM	100 mM	50 mM	150 µM	50 µM	0
14	2 µM	100 mM	50 mM	150 µM	50 µM	2 µM
15	2 µM	100 mM	60 mM	150 µM	50 µM	0
16	2 µM	100 mM	60 mM	150 µM	50 µM	2 µM
17	2 µM	100 mM	70 mM	150 µM	50 µM	0
18	2 µM	100 mM	70 mM	150 µM	50 µM	2 µM
19	2 µM	100 mM	80 mM	150 µM	50 µM	0
20	2 µM	100 mM	80 mM	150 µM	50 µM	2 µM
21	2 µM	100 mM	90 mM	150 µM	50 µM	0
22	2 µM	100 mM	90 mM	150 µM	50 µM	2 µM
23	2 µM	100 mM	100 mM	150 µM	50 µM	0
24	2 µM	100 mM	100 mM	150 µM	50 µM	2 µM

Using these samples, the experiment was run under the same conditions as previously described. Figure 2-19 shows the FRET values for the experiment, with the data treated in the same way as in the first optimised experiment (pp. 74). Data represented in figure 2-19 has been fitted as described on page 74. The first 20 data points of section 1 have also been omitted as mixing occurred during this time and fluorescence values are not representative. Within the legend, the concentrations shown are those of H_2O_2 in the sample.

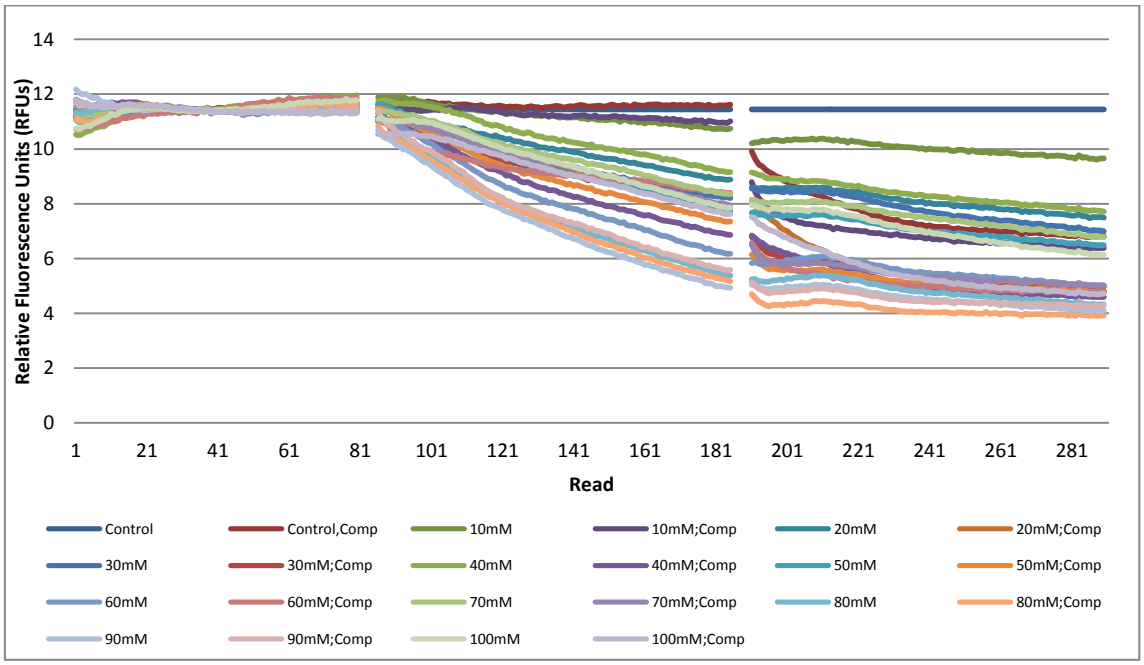

Figure 2-19: FRET values observing effect Fenton's reagent using different concentrations of Hydrogen peroxide.

As can be observed from figure 2-19, the FRET values observed decrease in direct proportion to the concentration of Hydrogen peroxide used. This hypothesis was tested by using Pearson's product moment correlation coefficient to study the effect of the independent variable "concentration of Hydrogen Peroxide" with the dependent variable "decrease in fluorescence over section 2"; results revealed a strong negative correlation (-0.786) between the concentration of H_2O_2 and the variation in FRET values over section 2.

Observing section 3, it is apparent that all samples to which Hydrogen peroxide was added (before section 2) continued to decrease, albeit at a slower rate.

Samples to which equimolar concentrations of complementary DNA was added, showed the characteristic hybridisation curve however due to the lower starting point for samples containing Hydrogen peroxide, these showed a more attenuated incline.

As with previous experiments both Cy5 (figure 2-20) and Cy3 (figure 2-21) fluorescence values were collected and analysed.

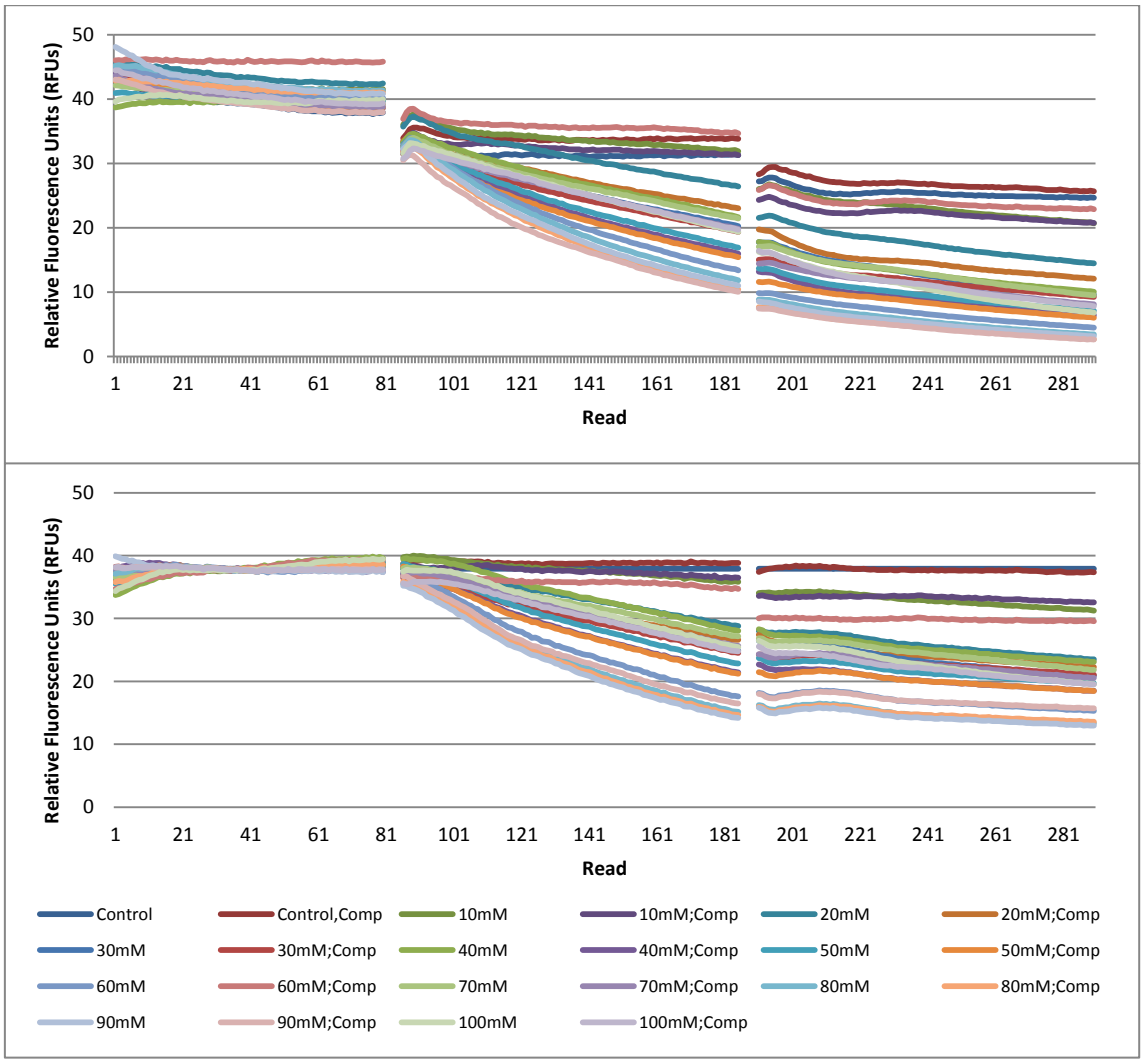

Figure 2-20: Cy5 fluorescence values for oxidation threshold experiment raw data (above) and data normalised to control values (below).

Figure 2-20 denotes the variation in Cy5 fluorescence over the course of the experiment; as can be observed the trend in sections 1 and 2 (up to read 180) are remarkably similar to those observed in the FRET values. In section 3, any

decrease noticed was in proportion to the H_2O_2 concentration, the presence of complementary oligonucleotide had no effect.

Figure 2-21 denotes the Cy3 raw data values, these were, as with previous experiments, an approximate mirror image of the FRET values observed, with fluorescence increasing with increasing concentrations of H_2O_2.

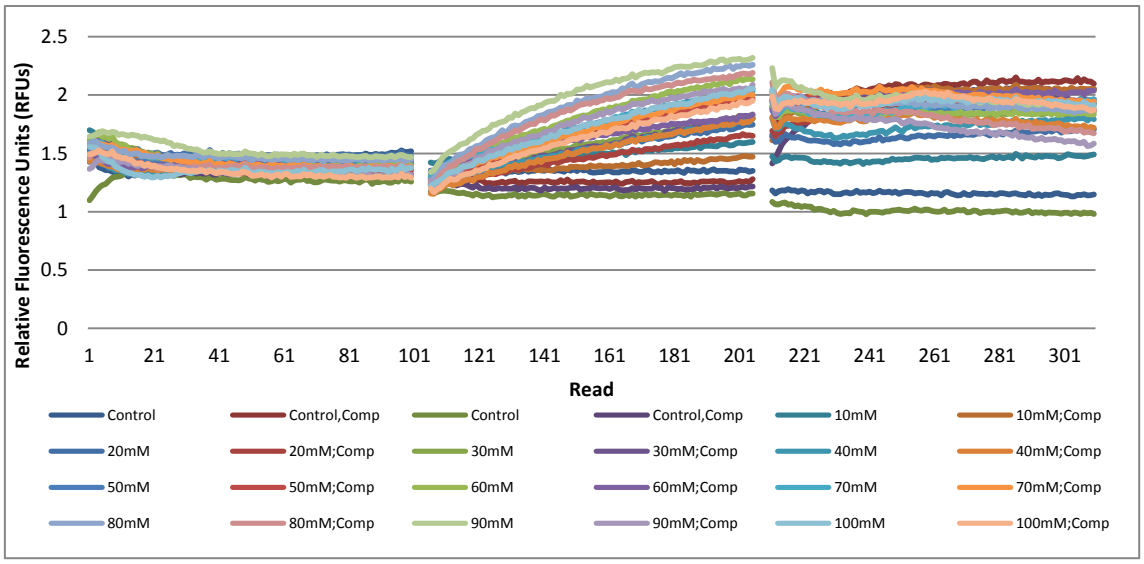

Figure 2-21: Cy3 raw data values from oxidation threshold experiment.

As with the first optimised method (pp. 77), figure 2-22 denotes a far different picture of the FRET values, when the variation in Cy5 fluorescence is discounted; In section 1, all samples coincide, in section 2 however, whereas with the raw data all fluorescence values decreased to some extent, when treated as described, all samples except one (60 mM with complement) ether remain stable, or increase over time. In section 3, the samples to which complementary oligonucleotide was added decrease, and the samples to which no complementary oligonucleotide was added remain stable.

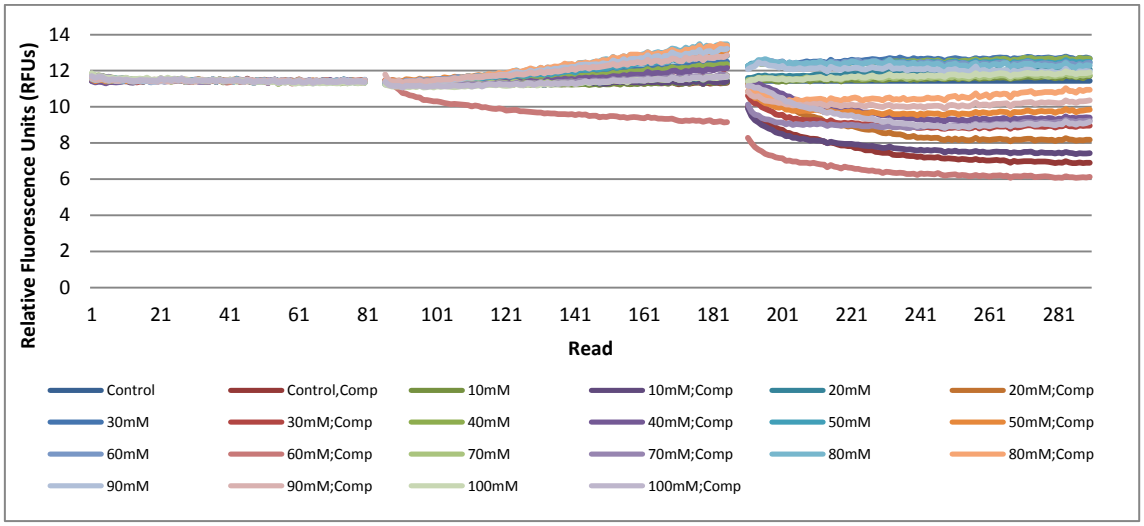

Figure 2-22: FRET values from the oxidation threshold experiment, normalised to control and Cy5 fluorescence.

2.2 Molecular Modelling

All production Molecular Dynamics simulations were performed on a Dell/Alienware workstation with 2.66Ghz Intel core i7 processor (4 core HT), 12Gb RAM and 2Tb HDD running CentOS Linux 5.4 and an HP XC-3000 supercomputer: Infiniband interconnect; 2 login nodes each with 3.2Ghz Intel Xeon 8 core processors, shared 64Gb RAM and 3Tb file-store; plus 872 processor cores across 218 HP DL140 G3 compute nodes each with dual Intel "Woodcrest" 5160 Xeon dual core processors, 8Gb of shared RAM per node and dedicated 33Tb file-store (Astral). Optimisation was performed on the Dell/Alienware workstation, a Dell laptop with 1.7 GHz Intel core i7 processor (4 core HT), 12Gb RAM and 750Gb HDD running Ubuntu Linux 10.10 and 4 x RM workstations clustered with Rocks Linux over 100Mbps Ethernet each with 3.2Ghz Intel Pentium 4HT processors, 2Gb RAM and 320Gb HDD.

GROMACS Molecular Dynamics Suite 4.5.3 [277] single precision was used, in both the optimisation and production runs.

To perform the Molecular modelling, a PERL script was used to automatically perform all the separate commands in a specific sequence, to execute molecule

preparation, equilibration, and production MD simulations and, to a lesser extent, analysis of results.

To perform adequate simulations of the G-quadruplexes, an extensive optimisation was undertaken. This optimisation can be generally divided into two different, yet equally important parts; optimisation of the script, which dictated the different programs used and the general flow of the equilibration and production simulations, and optimisation of the simulation parameters: a fine tuning of the parameters of all of the MD and EM simulations to ensure valid results.

2.2.1 Optimisation of Molecular Modelling Protocol and Parameters

A script was used to open and use all the different programs that constitute the GROMACS package, in a precise order. The script was originally written by Dr Lee Larcombe and used for protein modelling. The initial optimisation of this script was based on a thorough review of the literature, which revealed the general steps recommended for large-scale modelling of biomolecules, presented in the following scheme.

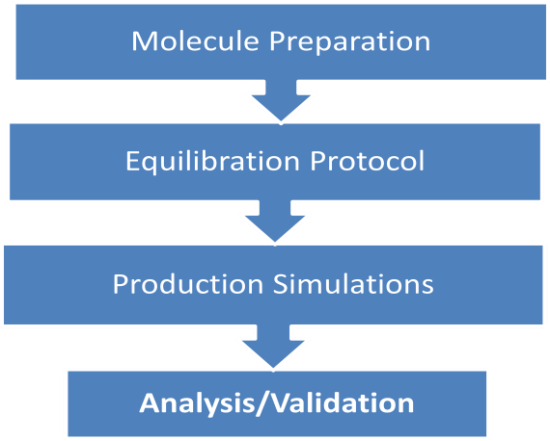

Figure 2-23: Molecular Modelling Overview Flow Chart.

A consistent problem encountered during the first weeks of optimisation was file errors due to inappropriately named files, different programs of the GROMACS package not receiving the correct input files and syntax errors. This was

overcome by implementing a consistent and simple naming system, facilitating the transition between programs and assuring the correct files are used and/or updated on each step.

Hereafter, the optimisation of each of the three steps: Molecule Preparation, Equilibration Protocol and Production Simulation will be described.

2.2.1a Molecule Preparation

For reasons described in section 2.1, the G-quadruplex structure files PDB ID 1XAV [67] and PDB ID 2O3M [125] were taken from the Research Collaboratory for Structural Bioinformatics (RSCB) Protein Data Bank (PDB). These are the structures of G-quadruplexes in the NHE III region of c-Myc, and in the promoter region of c-Kit, respectively. These are both intramolecular quadruplexes with three guanine tetrads. They differ in the length and sequence of the loops, and will be discussed in detail in later chapters. These structures have several features which make them attractive choices for this project: they are biologically relevant; they have both been well studied; and their disparity in loop length and sequence indicate a possible difference in stability between the two.

The B-DNA analogues were created using the same sequence as the PDB G-quadruplex structures (Table 2-5), using AMBER Tools NAB module.

Table 2-5: Sequences for the c-Myc and c-Kit B-DNA molecules.

c-Myc	5'-TGA GGG TGG TAG GGT GGG TAA-3'
	3'-ACT CCC ACC ATC CCA CCC ATT-5'
c-Kit	5'-AGG GAG GGC GCT GGG AGG AGG G-3'
	3'-TCC CTC CCG CGA CCC TCC TCC C-5'

Molecule preparation consisted of structure editing, counterionisation, and solvation. Due to the easy-to-use GUI of SYBYL and our familiarity with its

structure editing features, this platform was planned to be used to perform this step.

However initial tests proved that this platform is inappropriate when working in the .pdb and .gro file formats (used by GROMACS), as alterations made within SYBYL and saved in the .pdb file format, were distorted when opened in any other molecular viewing software.

For this reason, UCSF CHIMERA 1.5.3 [284] and the GROMACS package were selected to perform molecule preparation. The use of flanking sequences was revealed to be flawed as the computational time would be increased exponentially and the benefits to biological accuracy would have been negligible due to the inherent shortcomings of MD force fields.

To simulate the effect of oxidation, all structures were duplicated and to these duplicate structures, one 8-oxo-dG nucleotide was incorporated into the structures. This was performed on a Guanine in the central tetrad of both G-quadruplexes: G7 in c-Myc and G9 in c-Kit; and on the corresponding guanine in the B-DNA analogues. Details of the incorporation and force field modifications performed to accommodate the oxidised nucleotides are described in section 2.2.2b.

Two aspects determined to be crucial to a stable system were counterions and mixed solvent. The counterions would be potassium cations that would neutralise the negatively charged backbone and the mixed solvent (NaCl of KCl salt) would approximate the system to real life situations and also restricts the movement of molecules caused by electrostatic forces, when compared to pure water.

Mixed solvation was initially achieved using CHIMERA to generate a .gro file with NaCl which was added to the structure file using genbox. Water was then added to the system. Although this achieved the desired result in terms of quantity, there were several disadvantages: the structure file added the atoms randomly, without consideration for the other ions that are added – leading to an

unstable environment that requires rigorous equilibration. The other disadvantage was practical, using genbox required the use of a series of extra files and commands in the script which is awkward when performing optimisation and dealing with several “draft-runs” a day.

Counterionisation is a requirement for accurate molecular dynamics simulations of nucleic acids. This was first attempted, using CHIMERA, by manually placing the positive counterions (potassium) 6 Å away from each phosphorous group. However this strategy produced an unstable system, likely due to the repulsive electrostatic forces between the counterions as well as the central potassium ions.

Counterionisation and mixed solvation were then achieved using genion. This program is part of the GROMACS package and is used after solvation in water and replaces random water molecules within the system and has the advantage of enabling the user to define the minimum distance between ions as well as the desired final net charge and the ionic concentration.

The structures were then solvated in approximately 11000 TIP3P water molecules using genbox with 70 Å by 70 Å by 70 Å periodic boundary conditions (PBC) for the G-quadruplex structures and in a Triclinic box with PBC at least 10 Å away from the solute for the B-DNA molecules; and then counterionised with 19 K^+ ions to neutralise the molecule and another 20 K^+ and Cl^- atoms were added to the system to achieve a final Salt concentration of 100 mM and a net neutral charge.

2.2.1b Equilibration Protocol and Production Simulations

The equilibration protocol is an important process, consisting of several EM and MD simulations designed to assure that any unnatural restraints resulting either from the original solute molecule or from the solvation and addition of ions, are resolved before the production molecular dynamics simulations are performed.

There is no standard equilibration protocol; it depends largely on the system under study, the software used and the parameters used in the production simulation.

The initial equilibration protocol consisted of simply an EM followed by a short MD simulation. This was used to test the effectiveness of the script and the transition between separate simulations. Once this script was shown to function, it was expanded in accordance with the equilibration protocol presented by Haider and Neidle (2010) [285], which consisted of a total of five energy minimisations and three molecular dynamics simulations. Although this is a valid protocol, it is designed for use with the AMBER software package with several impracticalities when using GROMACS. As such a protocol was devised taking into consideration several papers on G-quadruplex MD simulations with GROMACS, consisting of two energy minimisations and three molecular dynamics simulations, in accordance with the parameters stated in the digital Appendix. The parameters used in the equilibration protocol were also subjected to a rigorous optimisation process. Going from an initial, crude equilibration, to one with positional restraints, accurate indexing groups (for temperature and pressure coupling), positional restraints and appropriate centre of mass motion removal.

This was subsequently validated to be an adequate equilibration. An overview of the equilibration protocol is presented in Fig. 2-24.

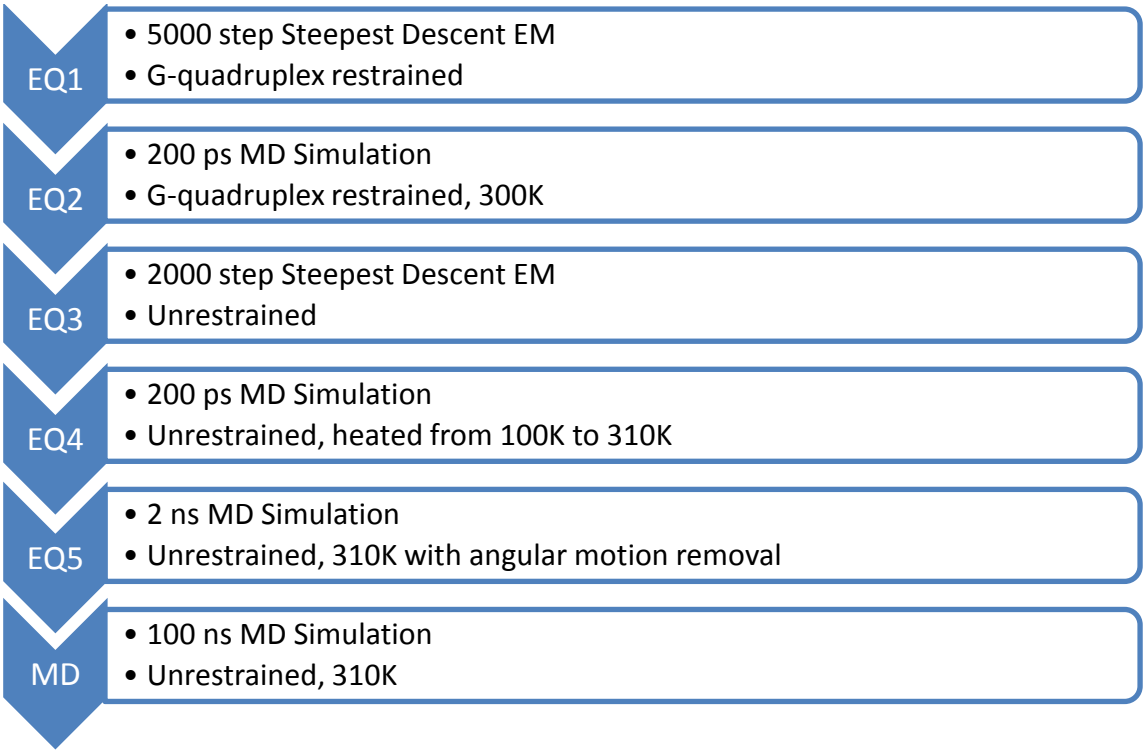

Figure 2-24: Overview Flow Chart of the Equilibration Protocol

The first two steps consist of an initial 5000 step minimisation and 200 ps MD simulation with the G-quadruplex restrained. These are both intended to allow the water and counterions to equilibrate, while restraining the G-quadruplex to prevent structural alterations, which may occur due to the brusque movements of the solvent atoms. The EM will resolve the volatile electrostatic interactions and the MD will then allow the movement of solvent and ions to resolve the other, weaker interactions.

The system was then subjected to a further 2000 step minimisation (EQ3) without restraints and 300 ps MD simulation without restraints with temperature rising from 100 K to 310 K (EQ4), augmenting 75 K every 50 ps. These steps are intended to equilibrate the entire system by resolving the electrostatic interactions between the G-quadruplex and the surrounding solvent and ions and then slowly heating the system from 100 K to the desired production run temperature of 310 K to slowly increase the movement of the atoms in the

system, to better resolve the constraints that may be present in either the original PDB structure or in the solvated and counterionised system.

The final equilibration step was run with the same parameters as the production run (below), differing only in the length of the simulation (2 ns) and centre of mass motion removal which, in this case, was angular. This step has the dual purpose of attenuating the angular momentum of the G-quadruplex and resolving any residual constraints remaining in the system.

The parameters involved in particle interactions were constant throughout the equilibration and test production runs, with the exception of Fourier grid spacing which was increased to 0.12 Å in the production simulations to improve efficiency as described in the following section (Table 2-6).

Table 2-6: Particle Interaction Parameters Used.

ns_type =	grid
Pbc =	xyz
coulombtype =	PME
Rlist =	1
Rcoulomb =	1
rvdw =	1
Fourierspacing =	0.1178
pme_order =	4
ewald_rtol =	1e-5
Vdwtype =	Cut-off

The run output control was adapted for each step, to give approximately 100 snapshots for each of the equilibration steps and the production simulation was set to write out every 10 ps. Table 2-7 describes the principal parameters used in the production MD simulations.

Table 2-7: Summary of the parameters used in the production simulation runs. Descriptions taken from GROMACS online manual (http://manual.GROMACS.org/)

Parameter	Value	Description
dt	0.002	Time step (2 fs)
Nsteps	250.000.000	Length of simulation (500 ns)
comm_mode	Angular	Mode for centre of mass motion removal (translational and rotational)
comm_grps	DNA Water_and_Ions	Groups to remove centre of mass motion
nstcomm	10	Frequency to remove centre of mass motion
tcoupl	Berendsen	Temperature coupling (weak coupling to a Berendsen Thermostat)
tc_grps	DNA Water_and_Ions	groups to couple separately to thermostat
tau_t	0.1 0.1	Time constant for coupling each group to the thermostat
ref_t	310 310	Reference temperature for each group
pcoupl	Berendsen	Pressure coupling (weak coupling to a Berendsen Barostat)
pcoupltype	Isotropic	Barostat to function uniformly in all directions
tau_p	1	Time constant for coupling (1 ps)
compressibility	4.5e-5	Rate of volume change in response to pressure (bar)
ref_p	1	Reference pressure for coupling (1 bar)
refcoord_scaling	All	The reference coordinates are scaled with the scaling matrix of the pressure coupling
gen_vel	No	Do not generate velocities at start-up
constraints	All-bonds	Convert all bonds to constraints
constraint_algorithm	LINCS	LINear Constraint Solver. An algorithm which is highly parallelisable
lincs_order	4	Highest order in the expansion of the constraint coupling matrix
lincs_iter	1	Number of iterations to correct for rotational lengthening in LINCS

The script used for the results presented in this thesis, as well as the full list of parameters for molecule preparation, equilibration protocol and production simulations can be found in the digital Appendix.

2.2.1c Parallelisation and Processing Performance Optimisation

Molecule preparation and equilibration was performed on the Dell/Alienware machine previously described. This protocol took approximately 8 hours to complete, using all 8 available threads. The largest step was EQ5, a 2 ns MD simulation which generally took approximately 3 hours at 16.5 GFlops. The processing performance values for each step in molecule preparation and equilibration for normal c-Myc can be viewed in the digital Appendix, as can the values for the simulations described below.

To perform the production molecular dynamics simulation on the Astral supercomputer, the script was first tested on the RM cluster described previously. Although the simulation was successful, the processing performance was only 4.3 GFlops, which yielded 1.86 ns/day.

Dr Lee Larcombe and Dr Les Oswald then proceeded to compile GROMACS 4.5.3 in double precision on the Astral supercomputer. An initial 10 ns test was performed on the c-Myc G-quadruplex using 32 cores. This test revealed identical results to those performed on the Dell/Alienware machine. However, the simulation had a processing performance of 70 GFlops, and was able to simulate 30 ns/day. However, although this processing performance is over 4 times better than achieved on the Dell/Alienware machine, it was still considered and was only able to simulate 30 ns/day.

To improve speed, the decision was made to perform the simulation in GROMACS 4.5.3 single precision on the Astral supercomputer. This could potentially reduce accuracy, however a study by Gruber and co-workers (2011) [286] describes how any benefit from the use of double precision GROMACS is outweighed by the "noise" produced in simulations at temperatures over 300 K.

The script was then adapted to improve the parallelisation of the simulation. This involved an increase in the PME grid size to 0.12 nm^2 and a reduction of the centre of mass motion removal frequency to every 10 steps.

These alterations produced an significant increase to the processing performance of the MD simulations. By performing this optimised simulation on 96 cores, 140 GFlops, which produces approximately 60 ns/day, was the minimum performance achieved in any simulation.

2.2.2 Force Field Optimisation

2.2.2a Porting AMBER parm99 BSC0 parameters

In 2007, Perez and colleagues [287] modified the standard parm99 force field to correct the misrepresented alpha (O5' - P) and gamma (C4'-C5') conformers, which led to artefacts in the DNA backbone in long (>10 ns) MD simulations. This parm99 BSC0 force field algorithm (available from http://mmb.pcb.ub.es/PARMBSC0/) proved to be very successful and is currently the principal force-field used for simulation DNA in the AMBER MD software package.

Although GROMACS 4.5.3 does not support this modification, these parameters were ported by the author, as they were considered indispensable for accurate representation of the G-quadruplex structure during the intended simulation time.

Reparameterising the alpha and gamma torsions involved modifying the dihedral angles and charge thresholds on the C4' to C5' and O5' and P bonds, as well as creating a new atom type for the C5' atom to avoid the unintentional reparameterisation of protein molecules.

The fact that the AMBER software utilises different units from the GROMACS software required the conversion of these units, as well as taking into account several peculiarities in the porting process from AMBER to GROMACS, especially the fact that AMBER force fields within GROMACS are unable to differentiate between proper and improper dihedral angles – treating all dihedral angles as improper.

An explicit list of the alterations form the AMBER99sb force field can be found in the digital Appendix.

To validate the port, the altered force field files have been sent to Dr. Eric Sorin, the main contributor of previous AMBER force-field ports to GROMACS.

As an internal validation method, 400 ns MD simulations were performed on both G-quadruplex DNA (PDB ID: 1XAV [67]) and the corresponding duplex DNA molecule, with the unmodified AMBER99sb force field parameters, which were then compared to the trajectories obtained with the AMBER99 ParmBSC0 parameters.

All .pdb files and data analysis graphs are contained within the Digital Appendix.

The analysis of these simulations showed that with the 99sb force field, after approximately 100 ns the B-DNA molecule showed substantial distortions, with increased average root mean square fluctuation (RMSF) (2.87 Å) (a measure of the fluctuation of each residue over the course of the simulation) and loss of Hydrogen bonding. However, when the same molecule was simulated with the BSC0 force field, results were vastly improved; the distortions observed in the backbone in the previous simulation (average RMSF of 2.47 Å) were not noticed as the molecule retained a near perfect double helical structure. End fraying and modest DNA bending were observed, however to a far less extent than in the previous simulation. No substantial difference was noticed in Hydrogen bonding nor all-atom RMSD between the two B-DNA molecules, however the results obtained from RMSF and the macroscopic evaluations performed indicated that the parmBSC0 parameters were superior.

With respect to the simulations performed on the G-quadruplex structure, results were less conclusive. Both showed good stability, as confirmed by the low average RMSD values, however the structure simulated under 99sb parameters appeared to fluctuate substantially more than the structure simulated under BSC0 parameters, which was confirmed by a higher RMSF value in almost every residue (averages of 1.38 and 1.10 for the 99sb and BSC0 parameters, respectively), with more pronounced differences in the loops, for the structure simulated under 99sb parameters.

2.2.2b Modifications performed for Simulation of Oxidised Guanine

To simulate the effect of oxidation on G-quadruplex structures, a series of simulations were performed on structures containing one or more 8-oxo-2'-deoxyguanosine residues in the place of Deoxyguanine residues. This was first achieved by simply altering the H8 atom, bonded to C8 to an oxygen atom (O8). This was useful in that it was a simple modification and was used to familiarise the author with the intricacies of adding a modified residue to the GROMACS force fields.

However this proved to be an oversimplified alteration which did not account for the change in electronegativity and charge threshold caused when deoxyguanosine is oxidised to 8-oxo-2'-deoxyguanosine. Appropriate parameters were found from the Bryce Group at the University of Manchester (http://www.pharmacy.manchester.ac.uk/bryce/AMBER) and ported to GROMACS (an explicit list of the modifications can be found in the digital Appendix).

The alterations to the pdb structure file were performed using UCSF CHIMERA 1.5.2 [(284)], by replacing the Hydrogen atom (H8) found at C8 and by adding a Hydrogen atom H7 to the Nitrogen atom at position 7 (N7) (Fig. 2-25).

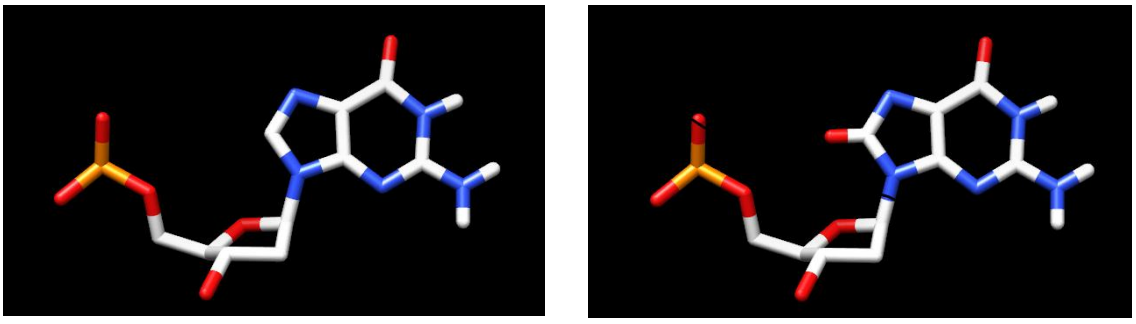

Figure 2-25: Comparison of unmodified Deoxyguanosine (left) and 8-oxo-dG (right).

These alterations were, in conjunction with the modified force field parameters, considered sufficient to accurately represent an 8-oxo-2'deoxyguanosine residue and the B-DNA duplexes and G-quadruplexes in which they are incorporated. As both the representation of the DNA as a whole and the 8-oxo-

2’deoxyguanosine residue would be represented using the most current and industry standard parameters.

2.2.3 Analysis of Results

Analysis of the MD simulations consisted of macroscopic visualisation of the structures over the course of the simulation using UCSF CHIMERA 1.5.3 [284] and analysis of the data produced by several GROMACS tools.

The GROMACS tools used to generate data and a summary of their use is described in table 2-8.

Table 2-8: GROMACS tools used to generate data on the MD simulations.

Tool	Description
g_rms	Calculates the Root Mean Squared Deviation (RMSD) of the solute (DNA) over the course of the simulation
g_energy	Calculates several parameters to describe the energy of the system over the course of the simulation, such as potential and kinetic energy, density and volume.
g_rmsf	Calculates the RMSD of each atom at the end of the simulation
g_hbond	Calculates the number and properties of the Hydrogen bonds present in the system over the course of the simulation
g_sas	Calculates the solvent accessible surface area

To specifically study the stability of the tetrads, without taking into account movement of the loops, index files were created using make_ndx with groups specifying the guanines involved in tetrad formation. This enabled the observation of the specific RMSD of the tetrads, as well as the number of Hoogsteen Hydrogen bonds present in the tetrads. This same process was used to calculate the RMSD of the B-DNA without the influence the end-fraying artefact.

To observe any change in the angles in the DNA backbone, g_angle was used on a group consisting of the atoms of the DNA backbone, created using make_ndx.

Graphs were plotted and basic descriptive statistics were calculated using Microsoft® Excel 2010.

To further explore the results obtained from the MD simulations, Principal component analysis (PCA), was used. This is a multivariate analysis method that identifies large scale collective motions of atoms and therefore separates significant motions from background thermodynamic fluctuations [288]. This analysis provides eigenvectors (principal components) corresponding to directions of motion and often, a large proportion of the motility of the analysed molecule can be explained by a few eigenvectors with high eigenvalues.

The displayed results relate to the optimised MD simulations performed on the Astral supercomputer. GROMACS input files for all simulations performed, including equilibration and molecule preparation can be found in the Appendix DVD, as can .pdb structure files for all the images shown in this section.

2.2.4 Simulations Performed on c-Kit G-quadruplex Structures

As described previously, four 500 nanosecond molecular dynamics simulations were performed on the c-Kit motif; comparing the normal sequence with a sequence with an oxidised guanine incorporated, in both the duplex and quadruplex structures.

Macroscopic analysis was performed using Chimera® 1.5.3 and consisted of viewing the entire trajectory of each simulation, as well as selecting representative snapshots of the simulation. To select these, the clustering module in Chimera was used. This program is based on the NMRCLUST clustering algorithm developed by Kelley and co-workers (1996) [289]. This method of clustering does not involve setting an RMSD threshold for clusters, but utilises its algorithm to define clusters of conformationally related subfamilies, based on the entire data set.

Due to the inherent difficulties in visualising three dimensional structures in two dimensions, the reader is directed to consult the Appendix DVD to visualise all relevant .pdb structure files of the snapshots figures presented here.

Figure 2-26 describes the trajectory over the course of the simulation, presenting the most prevalent cluster for every 100 ns of the 500 ns simulation; direct visualisation of these clusters and the animation of the trajectory, revealed very little change over the course of the simulation. In this all further representations of the Molecular Dynamics simulations, the oxidised nucleotide is presented as white and blue, with all atoms represented, as opposed to the non-oxidised nucleotides, all of which are presented as blue rectangular objects.

Tetrad stacking remained stable throughout, as Guanine orientation of those involved in tetrad formation and the distances between these nucleotides.

With respect to the loops, the nucleotides in single nucleotide loops (A5 and C9), the double nucleotide loop (C11 T12) and the long five nucleotide loop (A16 G17 G18 A19 G20) showed a high amount of movement; twisting around the sugar base and rotating around the backbone. It should be noted, however, that although these nucleotides displayed a lot of movement, the backbone of these loops remained in a similar position throughout the simulation and did not cause the quadruplex structure to distort in any way.

Due to the unprecedented conformation this motif forms, the 3’ tail is involved in tetrad formation, decreasing the flexibility of this tail. However the 5’ tail, consisting of only 1 nucleotide (A1), appeared to stack, through π-π interactions, on top of the tetrads.

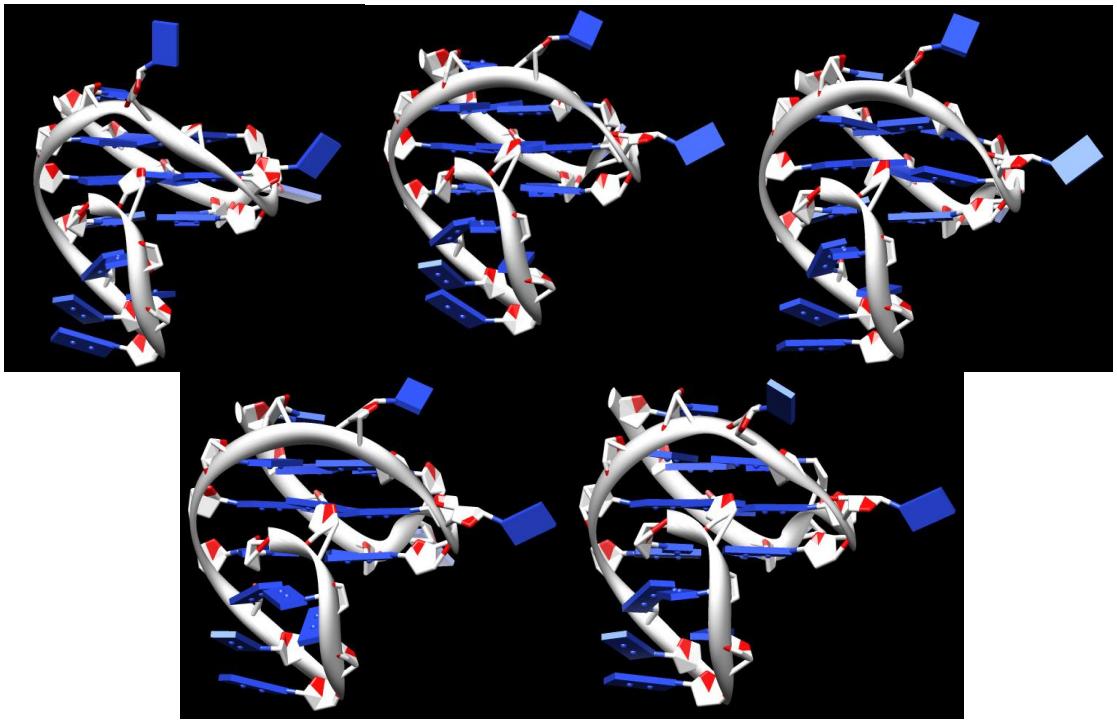

Figure 2-26: Representative snapshots of the cluster analysis of the non-oxidised c-Kit G-quadruplex. Top left: Cluster of 194 frames representative of the first 100ns; Top centre: Cluster of 99 frames representative of the second 100ns; Top right: Cluster of 256 frames representative of the third 100ns; Bottom left Cluster of 119 frames representative of the fourth 100ns; Bottom right: Cluster of 291 frames representative of the last 100ns.

Figures 2-27 describes the representative snapshots of the 500 ns simulation of the oxidised c-Kit G-quadruplex structure, that is to say, the structure to which 8-oxo-dG was incorporated in the central tetrad at the G9 position.

In terms of the loops and tails of the oxidised c-Kit G-quadruplex structure, the only discernible difference from the non-oxidised form was that over the course of the simulation, the long loop showed greater fluctuation, tending to move away from the oxidised nucleotide.

However in terms of the central tetrads, there were discernible changes over the course of the simulation; at around 200 ns the oxidised Guanine, within the central tetrad, started to twist in relation to the other guanines and this distortion spread to the Guanines directly above and below the oxidised Guanine (G6 and G8). At the same time, these Guanines were moving away from the remaining tetrads.

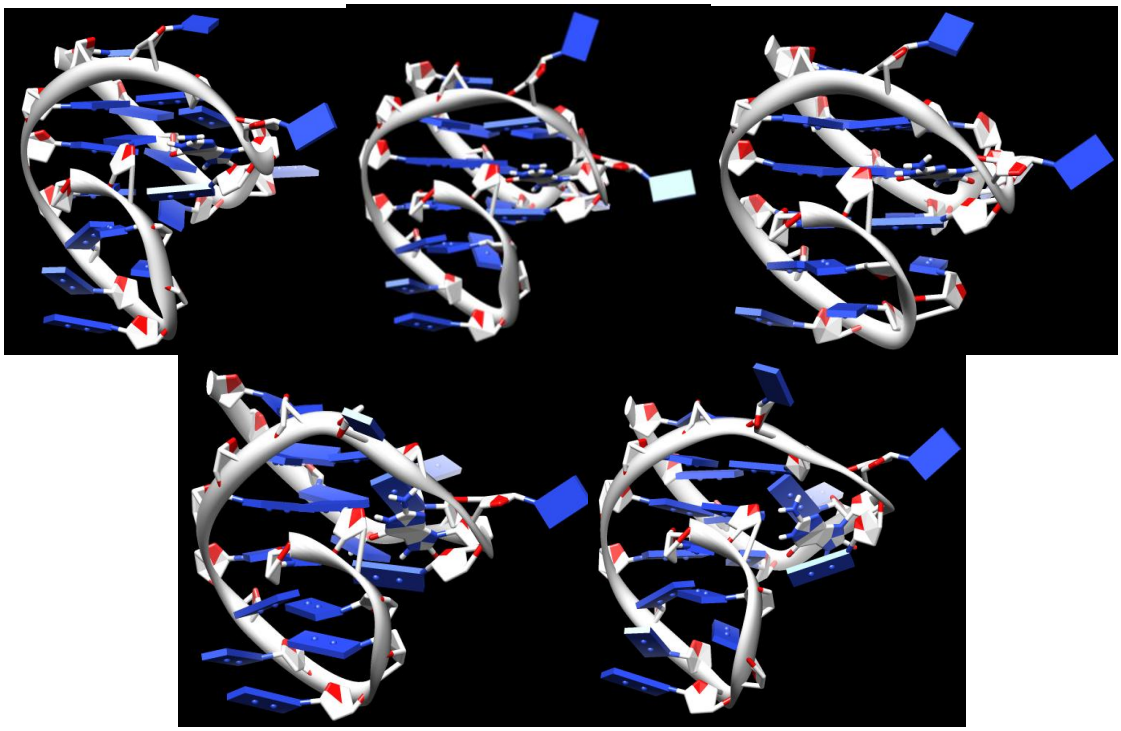

Figure 2-27: Representative snapshots of the cluster analysis of the oxidised c-Kit G-quadruplex. Top left: Cluster of 223 frames representative of the first 100ns; Top centre: Cluster of 108 frames representative of the second 100ns; Top right: Cluster of 448 frames representative of the third 100ns; Bottom left Cluster of 119 frames representative of the fourth 100ns; Bottom right: Cluster of 139 frames representative of the last 100ns..

This distortion became more apparent over the course of the simulation and, as can be observed in figure 2-28. This image represents a focused snapshot of just the Guanines involved in tetrad formation, with the 8-oxo-dG residue represented as a predominantly white nucleotide. The distortion appears to affect both the stacking of the tetrads and the likelihood of Hoogsteen Hydrogen bonds forming between the distorted Guanines and the remaining Guanines.

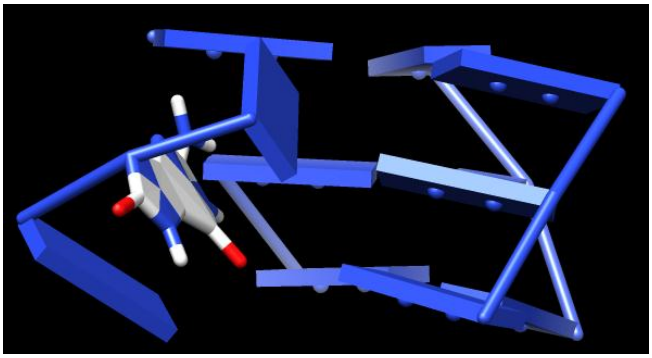

Figure 2-28: Focused snapshot of the Guanines involved in tetrad formation in the final conformation of the oxidised c-Kit G-quadruplex.

The following figures represent the data analysis preformed to enable the comparison between the oxidised and non-oxidised forms of the c-Kit G-quadruplex, to help interpret the full effect of oxidation on this structure.

As can be observed in the RMSD plot in figure 2-29, the results are in general agreement with the macroscopic evaluations performed; the non-oxidised structure remains relatively stable, with an average RMSD of 0.178nm. However, during the final part of the simulation, the structure appears to be destabilising to some extent, evidenced by the rise in RMSD to an average of 0.218 nm in the final 100 ns.

Also in agreement with the macroscopic evaluations, the RMSD values of the oxidised c-Kit structure were on average higher (0.207 nm) and an increase in RMSD was noted from 200 ns and became more pronounced throughout the simulation, with an RMSD for the final 100 ns of the simulation of 0.268 nm.

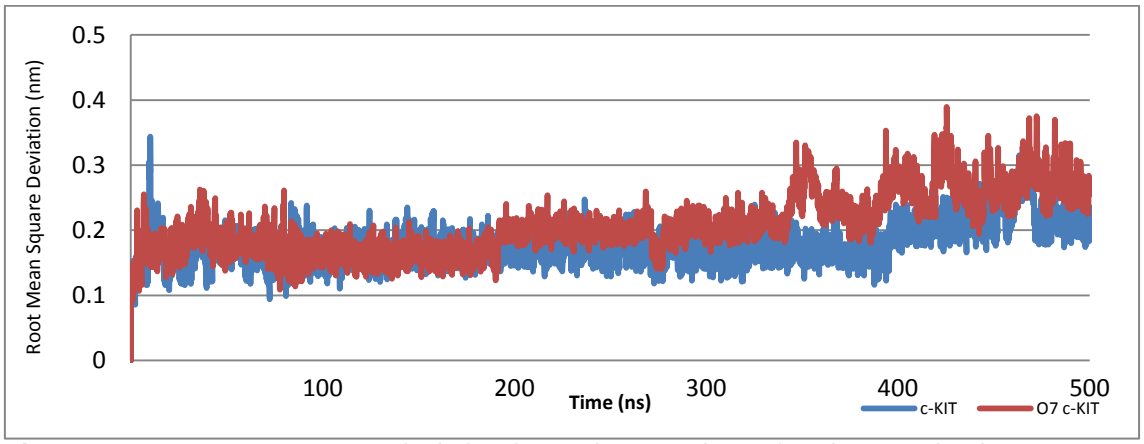

Figure 2-29: Root mean square deviation (RMSD) comparison of c-Kit G-quadruplex structures with and without 8-oxo-dG incorporated.

Observing the RMSD of only the guanines involved in tetrad formation, presented in figure 2-30, the difference between the oxidised and non-oxidised structures is more apparent than in figure 2-29.

For the first 100 ns, the oxidised structure presents a higher RMSD than the non-oxidised form. For the next 200 ns the structures are indistinguishable and at around 300 ns, the RMSD of the oxidised form increases and continues to

increase up to approximately 0.2 nm, whereas the non-oxidised structure remains at a constant RMSD of approximately 0.1 nm.

The average RMSD values for the central tetrads (0.105 nm and 0.149 nm for the non-oxidised and oxidised structures, respectively) are lower than those observed for the entire structure.

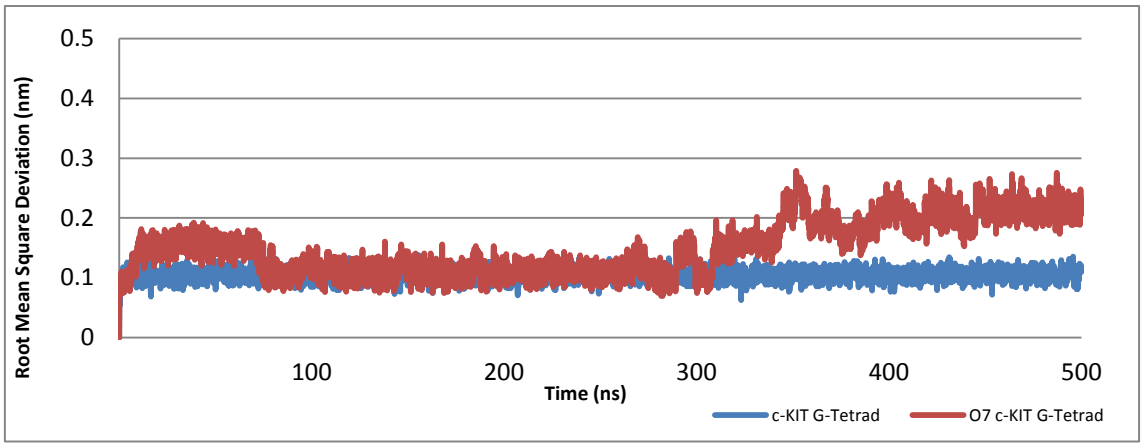

Figure 2-30: Root mean square deviation (RMSD) comparison of the Guanines involved in tetrad formation in the c-Kit G-quadruplex structures with and without 8-oxo-dG incorporated.

Figure 2-31 presents the root mean square fluctuation (RMSF) values of the two c-Kit G-quadruplex structures. RMSF is a measure of the movement of the structure, plotted per residue. This allows the determination of the origin of movements detected using RMSD. Observing this graph, and relating the residue numbers to nucleotides in the G-quadruplex, inferences can be made on the stability of the different parts of the structure.

In general terms, the RMSF values agree with the previous data accrued; the highest values are observed in the residues that correspond to the loops, however the long loop (A16 G17 G18 A19 G20) shows a more pronounced fluctuation in RMSD in the oxidised quadruplex than in the non-oxidised form. With respect to the tetrads, the guanines involved in their formation (G2 G3 G4 G6 G7 G8 G10 G13 G14 G15 G21 G22) presented lower values than the loops, and the oxidised G-quadruplex presented generally higher values than the non-oxidised form, especially in the oxidised guanine and those directly above and below it (G6 G7 G8).

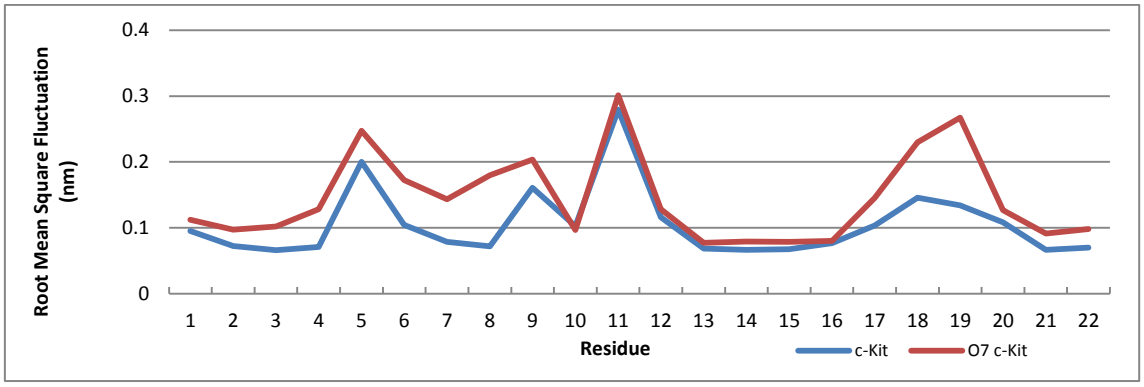

Figure 2-31: Root mean square fluctuation (RMSF) comparison of c-Kit G-quadruplex structures with and without 8-oxo-dG incorporated.

Figure 2-32 denotes the calculation of the number of Hydrogen bonds present between the Guanines involved in tetrad formation over the course of the simulations of both the oxidised and non-oxidised forms of the c-Kit G-quadruplex. Observing this graph, there are some immediate observations that can be made; the average number of hydrogen bonds in the non-oxidised structure is approximately 24, whereas the oxidised version has an average of approximately 20 hydrogen bonds. Furthermore the number of hydrogen bonds in the oxidised structure varies to form an almost mirror image of the focused tetrad RMSD graph, with a decrease in the first 100 ns, then coinciding with the non-oxidised structure for the next 200 ns, at which point the number of Hydrogen bonds decreases again, to approximately 16 and stabilises at this level until the end of the simulation.

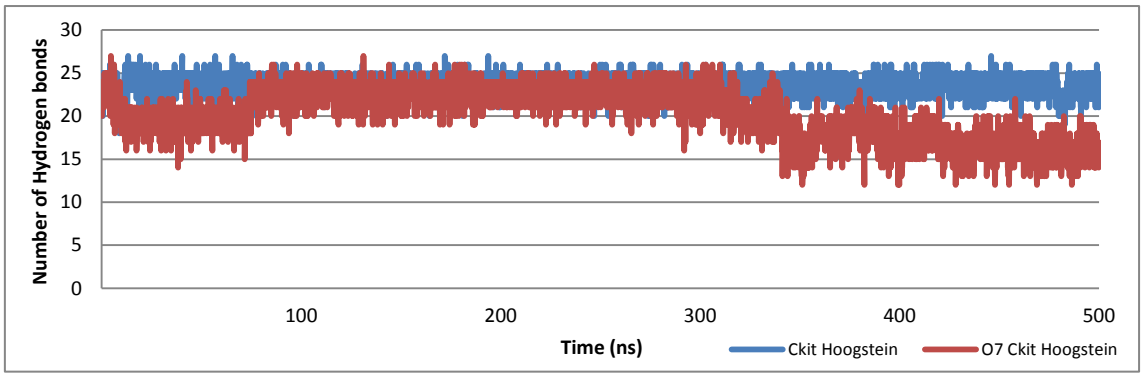

Figure 2-32: Comparison of the number of Hoogsteen Hydrogen bonds in the c-Kit G-quadruplex structures with and without 8-oxo-dG.

To evaluate changes in the topology, the solvent accessible surface area (SASA) of each residue was calculated for both structures (Fig. 2-33).

As can be observed from figure 2-41, the loop regions (A5, C9, C11 and around G18) present a higher SASA than the tetrads. In terms of the differences between the structures, they appear largely similar, with only slight differences in residues G2, G3 and G4 (involved in tetrad formation) and a larger difference at the oxidised Guanine G7.

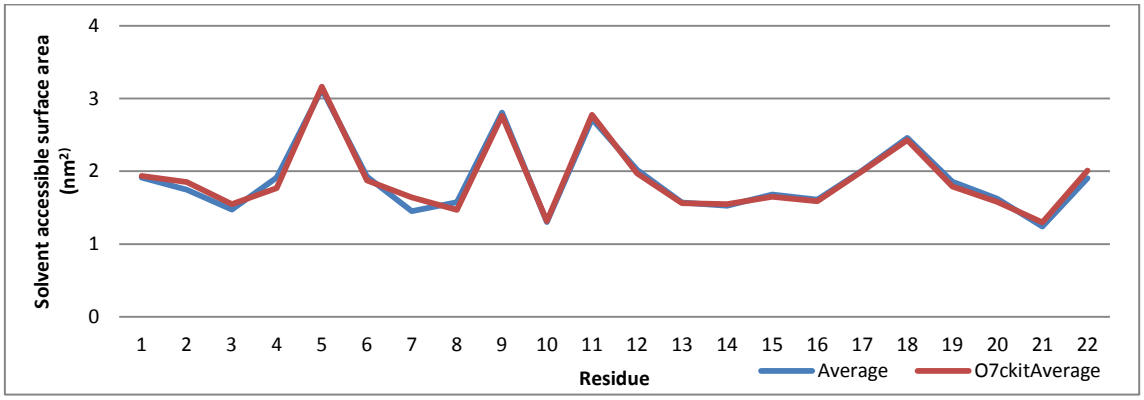

Figure 2-33: Solvent accessible surface area of each residue in the non-oxidised and oxidised c-Kit G-quadruplex Structures.

Principal component analysis (PCA) was also performed to identify any significant motions within the molecules. Figures 2-34, 2-35 and 2-36 represent the results of PCA on both the oxidised and non-oxidised c-Kit G-quadruplex structures.

Figure 2-34 shows the percentage of motility explained by eigenvalues of the eigenvectors; it plots the eigenvalues corresponding to the first (and, therefore, most significant) 50 eigenvectors presented as a percentage of the eigenvalues for the total eigenvectors and as a cumulative percentage. This analysis describes that the first 3 eigenvectors account for a more than half of the motility of the oxidised c-Kit G-quadruplex structure (approximately 53%) but account for a smaller proportion of the non-oxidised c-Kit G-quadruplex structure (approximately 42%).

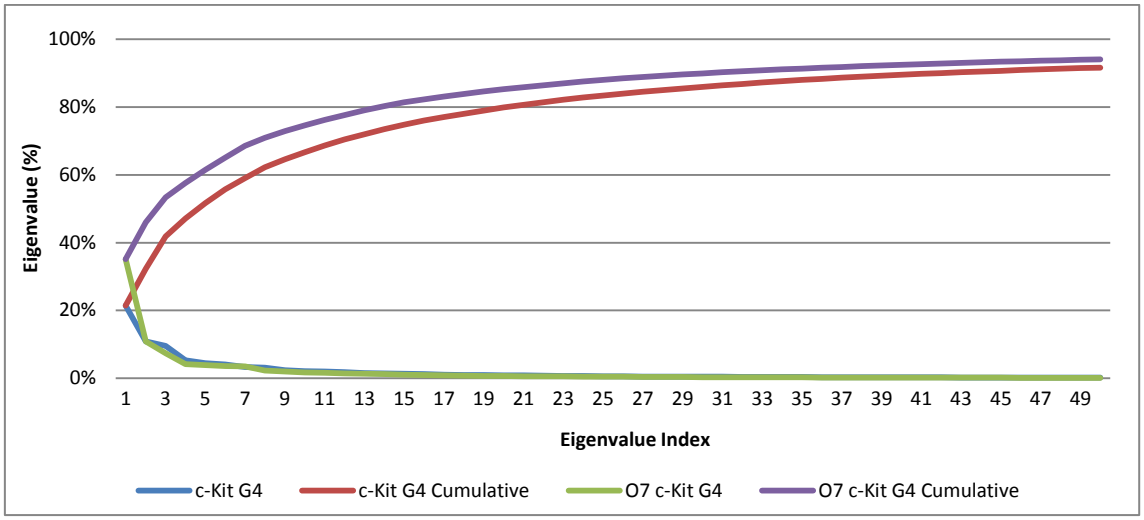

Figure 2-34: Graph plotting percentage of motility explained by eigenvalues against eigenvectors (principal components) of non-oxidised and oxidised c-Kit G-quadruplex structure and cumulative percentage.

Figures 2-35 and 2-36 represent the extreme structures and RMSF values of the first two eigenvectors of the non-oxidised and oxidised c-Kit G-quadruplex structures. The extreme structures are the most distant projections of the structure along the eigenvector. In Figure 2-35, the first eigenvector, representing the most significant motion of the non-oxidised structure corresponds to slight fluctuations in the loops with the most significant motion being in the Cytosine in the double nucleotide loop (C11) which rotated around the backbone, towards the tetrads. The second eigenvector of the same structure shows represents the movement of the remaining loop nucleotides, which were to a smaller extent (maximum 0.193 nm, compared to 0.499 nm in the first eigenvector).

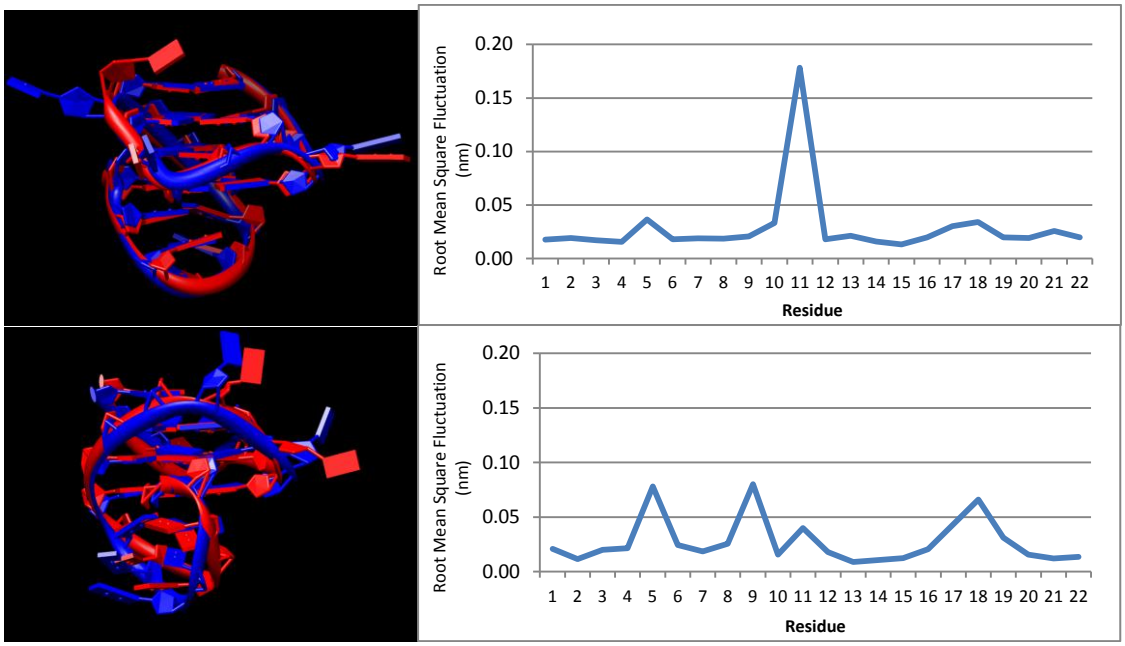

Figure 2-35: Extreme structures and RMSF values of the first (top) and second (bottom) eigenvectors of the non-oxidised c-Kit G-quadruplex.

When observing the oxidised c-Kit G-quadruplex structure (Figure 2-36), the structures pertaining to the first two eigenvectors support the previous observations made; both eigenvectors corresponded to the same effects (although to a greater extent in the first eigenvector). The main distortion was observed in the G-tetrads, with the oxidised lesion and the Guanines directly above and below it twisting and moving away from the remaining Guanines in the tetrads. Smaller movements of the loop nucleotides were also observed, in directions away from the oxidised lesion.

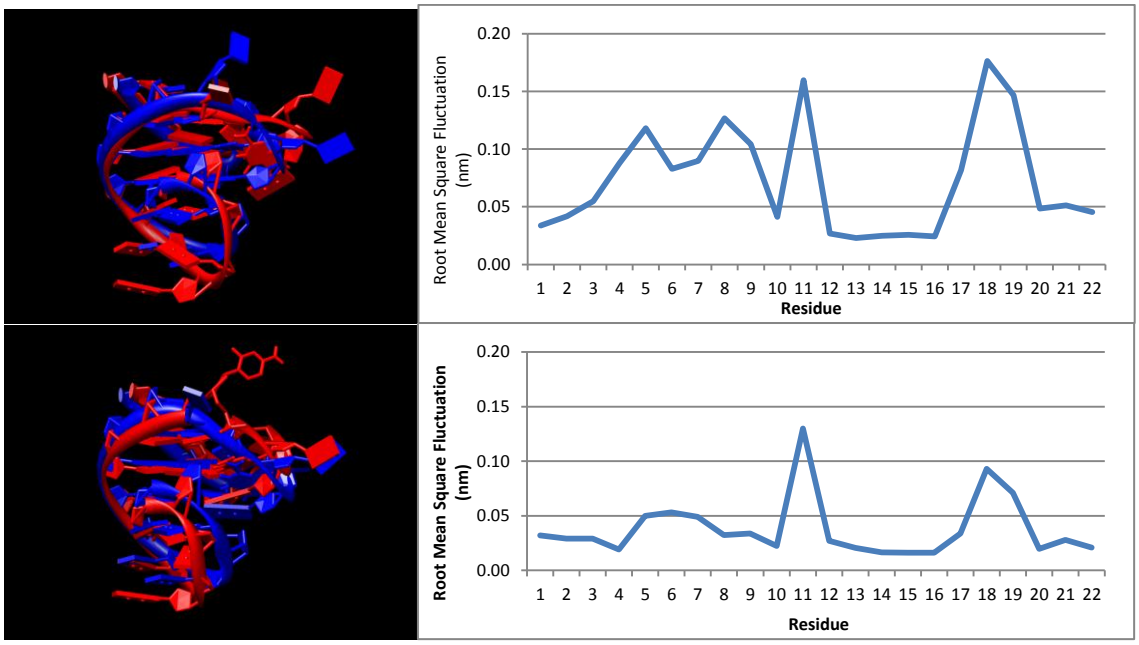

Figure 2-36: Extreme structures and RMSF values of the first (top) and second (bottom) eigenvectors of the oxidised c-Kit G-quadruplex.

2.2.5 Simulations Performed on c-Kit B-DNA Structures

To complement the simulations performed on the c-Kit G-quadruplexes, MD simulations under the same conditions were performed on the same sequence in the B-DNA form, with and without 8-oxo-dG incorporated into the 9 position.

Figure 2-37 represents snapshots of most prevalent clusters in each 100 ns of the entire simulation of the non-oxidised c-Kit B-DNA molecule. Throughout the simulation performed on this molecule, fluctuations were observed in the backbone (bending) and fraying of the 5' and 3' ends of the molecule. However, in general, the molecule remained remarkably stable and there were no discernible differences between the snapshots representing each 100 ns of the simulation Base stacking and the orientation of the nucleotides also remained stable throughout the simulation.

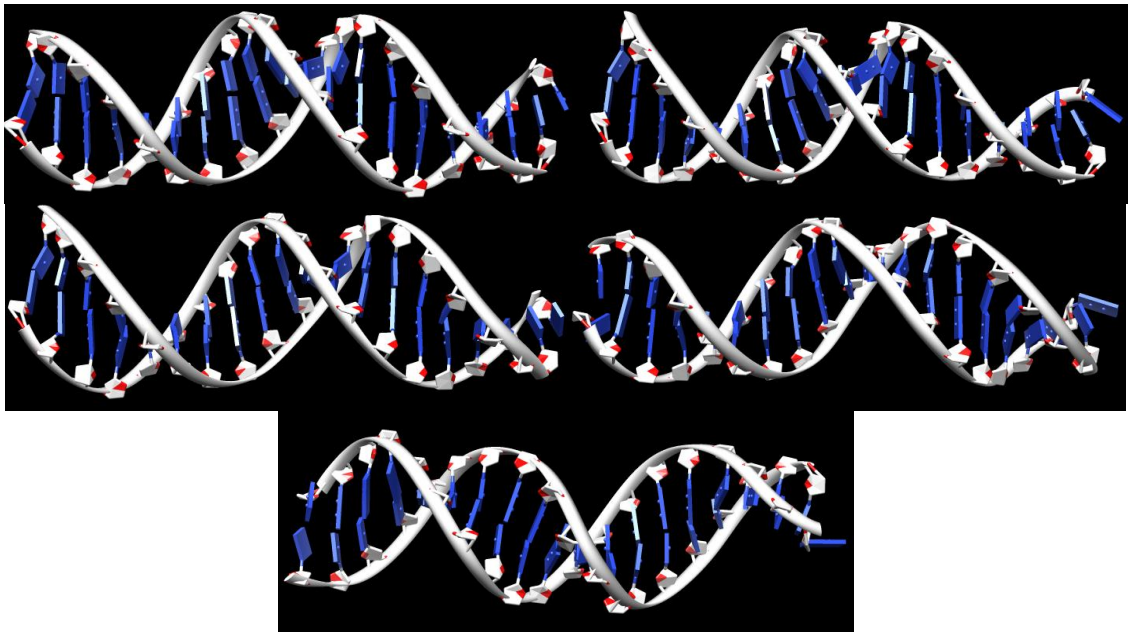

Figure 2-37: Representative snapshots of the cluster analysis of the non-oxidised c-Kit B-DNA molecule. Top left: Cluster of 81 frames representative of the first 100ns; Top right: Cluster of 136 frames representative of the second 100ns; Centre left: Cluster of 92 frames representative of the third 100ns; Centre right: Cluster of 49 frames representative of the fourth 100ns; Bottom: Cluster of 75 frames representative of the last 100ns.

Figure 2-38 represents snapshots of most prevalent clusters in each 100 ns of the entire simulation of the oxidised c-Kit B-DNA molecule and figure 2-39 represents a focused snapshot of the base pairing between the oxidised nucleotide and its complementary base. Throughout the 500ns simulation, the same observations were made in terms of the backbone bending, stable base stacking and pairing and end fraying. The only discernible difference was an increase in flexibility of the molecule, leading to an increase in bending from the oxidised nucleotide (G7).

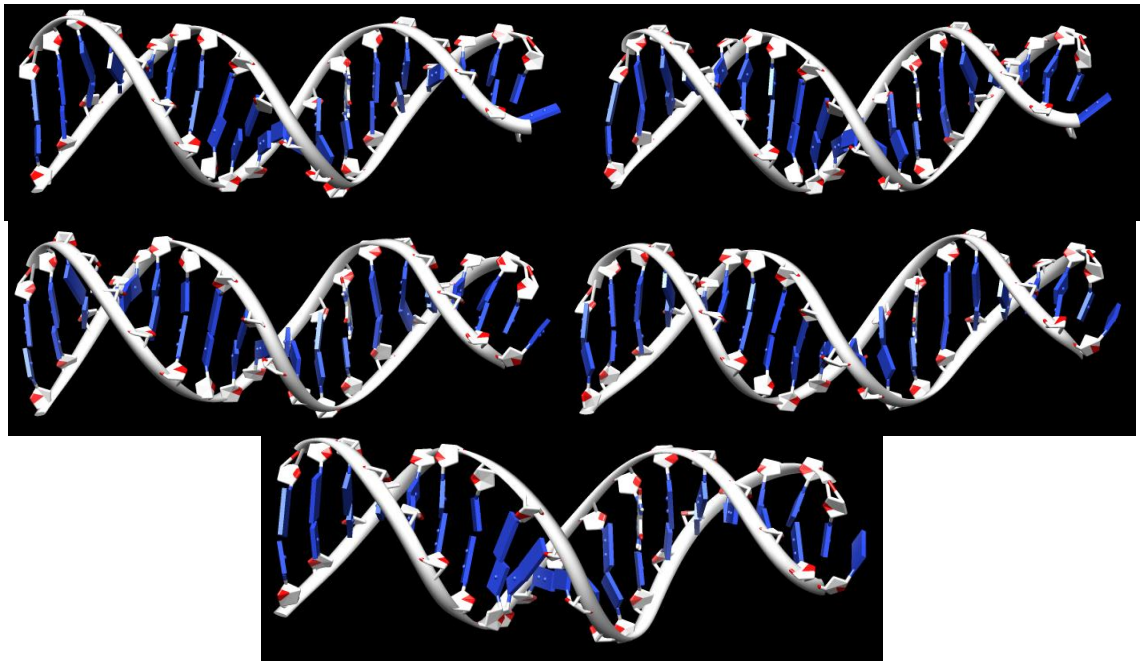

Figure 2-38: Representative snapshots of the cluster analysis of the oxidised c-Kit B-DNA molecule. Top left: Cluster of 178 frames representative of the first 100ns; Top right: Cluster of 98 frames representative of the second 100ns; Centre left: Cluster of 85 frames representative of the third 100ns; Centre right: Cluster of 87 frames representative of the fourth 100ns; Bottom: Cluster of 146 frames representative of the last 100ns.

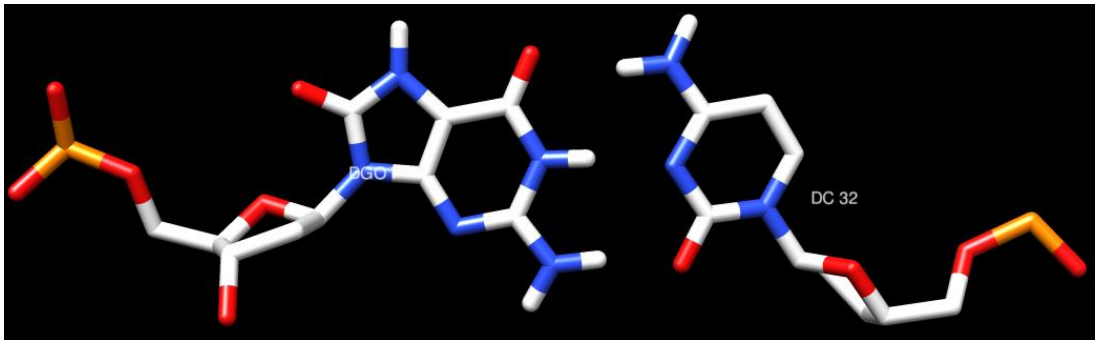

Figure 2-39: Focused snapshot of the base pairing/orientation of the oxidised Guanine and its complementary Cytosine at the end of the simulation.

The following figures represent the data analysis performed and comparisons performed on the oxidised and non-oxidised c-Kit B-DNA molecules. Observing figure 2-40, results are concordant with the macroscopic evaluations, RMSD values between oxidised and non-oxidised B-DNA molecules are very similar, with the oxidised structure presenting only slightly higher values (averages of 0.298 nm and 0.315 nm for non-oxidised and oxidised B-DNA molecules, respectively) and dispersion (standard deviations of 0.067 nm and 0.077 nm for non-oxidised and oxidised B-DNA molecules, respectively). Both molecules

appear to be equilibrated and stable, with no discernible shifts in RMSD values, nor any apparent incline.

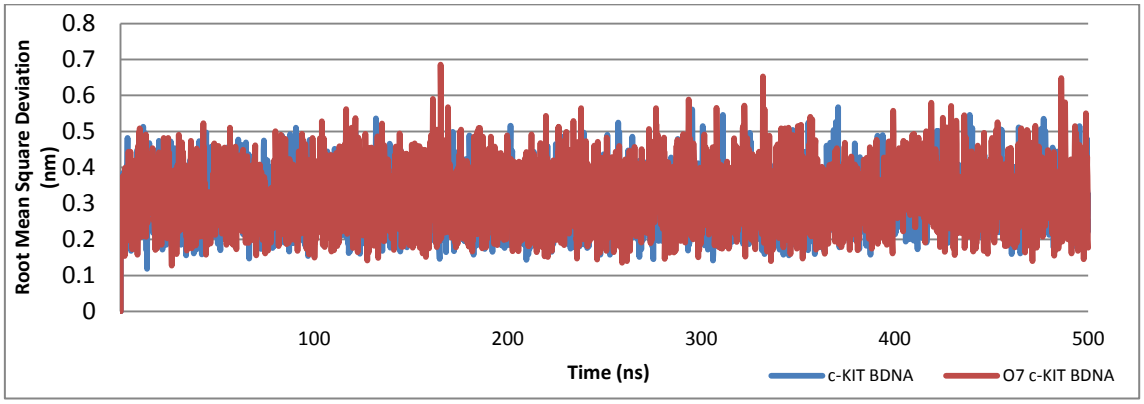

Figure 2-40: Root mean square deviation (RMSD) comparison of c-Kit B-DNA structures with and without 8-oxo-dG incorporated.

Observing the RMSF values plotted in figure 2-41, the results appear concordant with the accrued data; the terminal ends, present a higher RMSF value than the rest of the molecule. Between the two molecules, the most discernible differences are also observed at the terminal ends; at the 3' end of the sense strand and 5' end of the antisense strand of the molecules, where the non-oxidised structure presents a higher value, indicating more pronounced fraying. The reverse is apparent at the 5' end of the sense strand and 3'end of the antisense strands of the two molecules, with the oxidised structure presenting higher RMSD values than the non-oxidised structure.

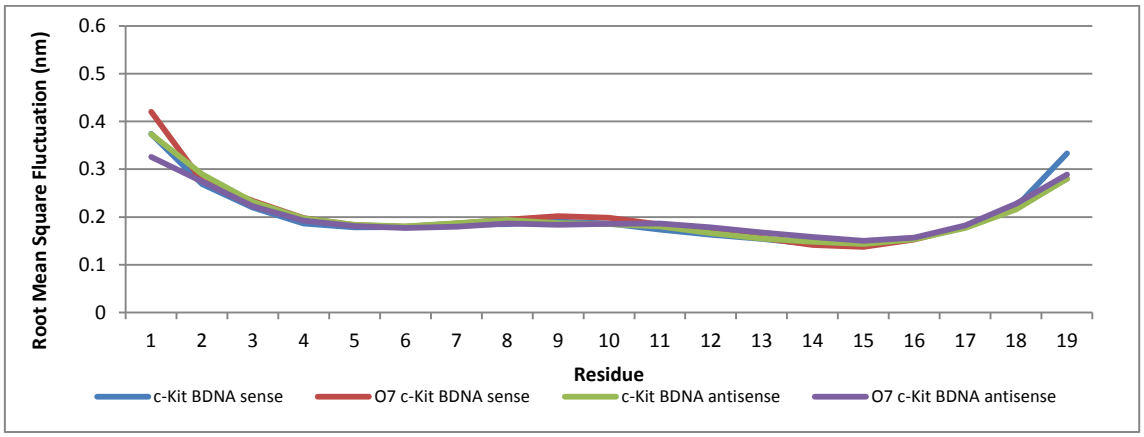

Figure 2-41: Root mean square fluctuation (RMSF) comparison of c-Kit B-DNA structures with and without 8-oxo-dG incorporated.

The Hydrogen bonding data presented in figure 2-42 supports previous data in that the number of Hydrogen bonds remains stable over time in both molecules (standard deviations of 0.184 nm and 0.181 nm for the non-oxidised and oxidised molecules) and structures appear to have very similar stabilities, with average Hydrogen bond number of 54 for both the non-oxidised and oxidised molecules.

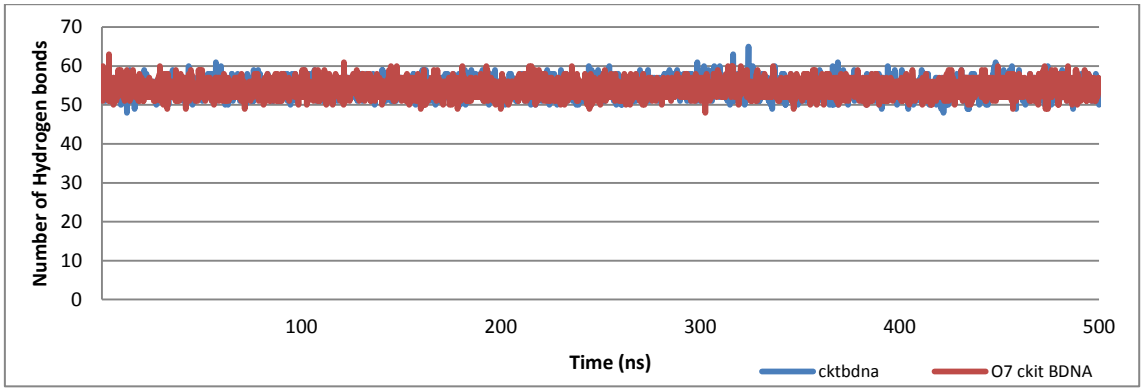

Figure 2-42: Comparison of the number of Hydrogen bonds in the c-Kit B-DNA structures with and without 8-oxo-dG.

In terms of the solvent accessible surface area (figure 2-43), there was no visible difference between the two molecules; both were generally low (around 2 nm^2) with spikes at the 5' and 3' ends.

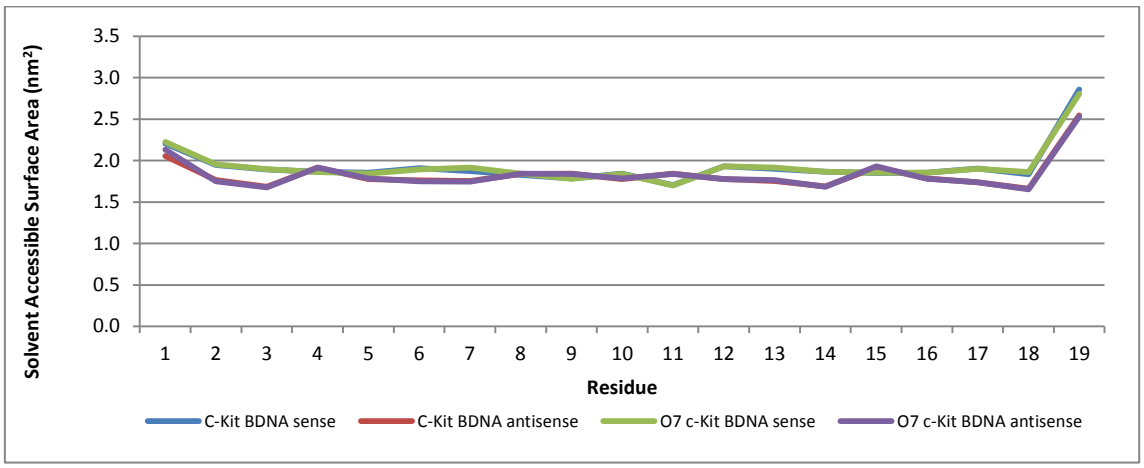

Figure 2-43: Solvent accessible surface area of the non-oxidised and oxidised c-Kit B-DNA molecules.

Principal component analysis (PCA) was also performed to identify any significant motions within the molecules. Figures 2-44, 2-45 and 2-46 represent

the results of PCA on both the oxidised and non-oxidised c-Kit B-DNA molecules.

The analysis of the percentage of motility explained by eigenvalues of the eigenvectors (Fig. 2-44) describes that the non-oxidised and oxidised structures showed very similar results, with the first 3 eigenvectors accounting for a large proportion of the motility of both B-DNA molecules (approximately 58% in the non-oxidised molecule and 60% in the oxidised molecule).

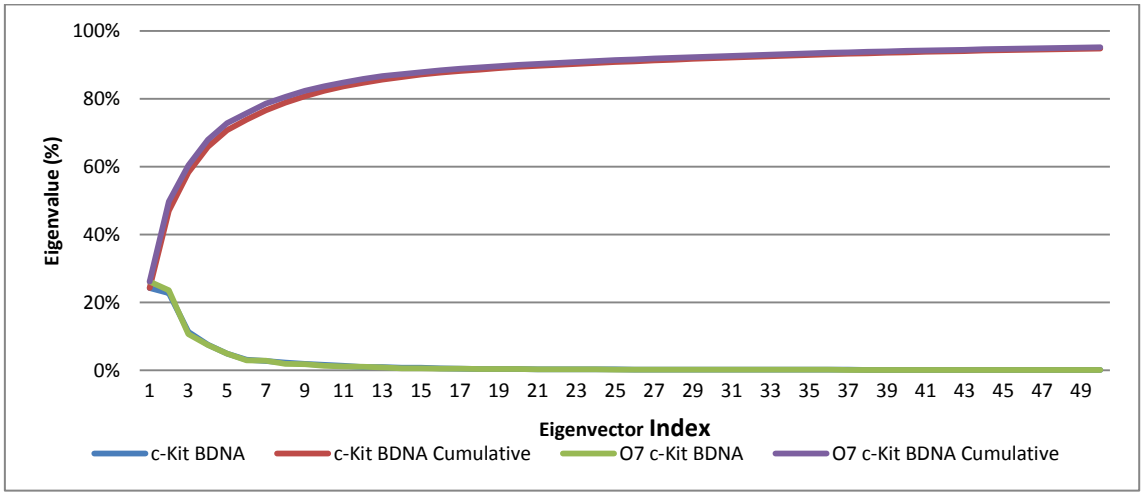

Figure 2-44: Graph plotting percentage of motility explained by eigenvalues against eigenvectors (principal components) of non-oxidised and oxidised c-Kit B-DNA molecule and cumulative percentage.

Figures 2-45 and 2-46 represent the extreme structures and RMSF values of the first two eigenvectors of the non-oxidised and oxidised c-Kit B-DNA molecules. In Figure 2-45, the first two eigenvectors of the non-oxidised c-Kit B-DNA molecule are almost perfect mirror images of each other; both show high fluctuation at the 5' and 3' ends of the double stranded molecule however in the first eigenvector, the antisense strand appears to twist further around the sense strand, producing a more "bent" molecule, and also increasing the RMSD in the antisense strand. Whereas in the second eigenvector, the opposite is true, the sense strand appears to twist and present higher RMSD values.

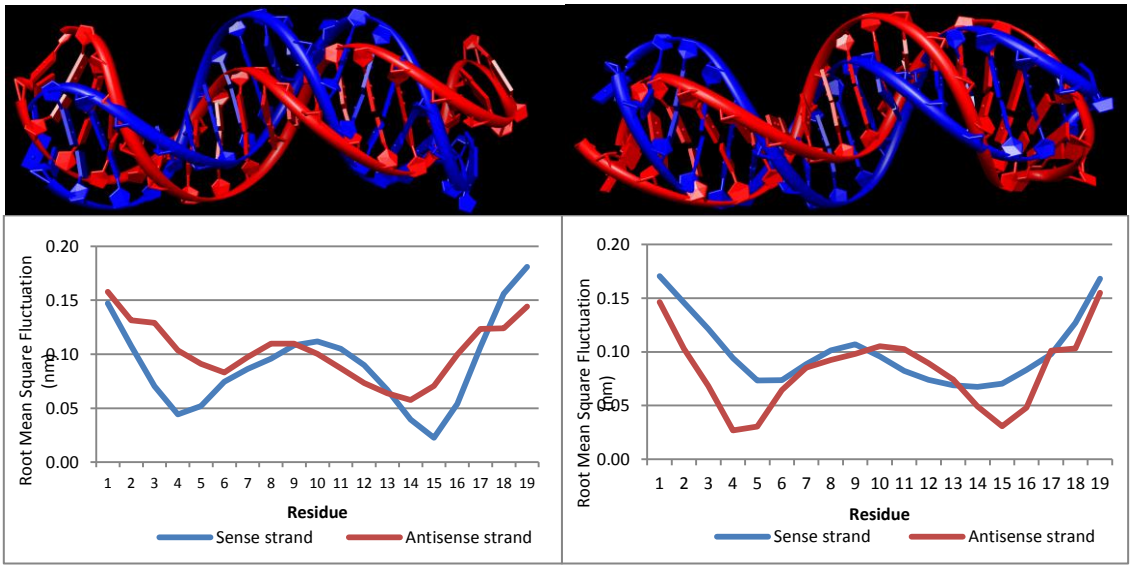

Figure 2-45: Extreme structures and RMSF values of the first (left) and second (right) eigenvectors of the non-oxidised c-Kit B-DNA molecule.

When observing the oxidised c-Kit B-DNA molecule (Figure 2-46), the structures pertaining to the first two eigenvectors are similar to those observed for the non-oxidised structures. Both show the same phenomena macroscopically and no discernible difference was noted. In terms of the RMSF data, the differences noted were attributed to fluctuations occurring at different points in the molecule, with the same overall result for the structure.

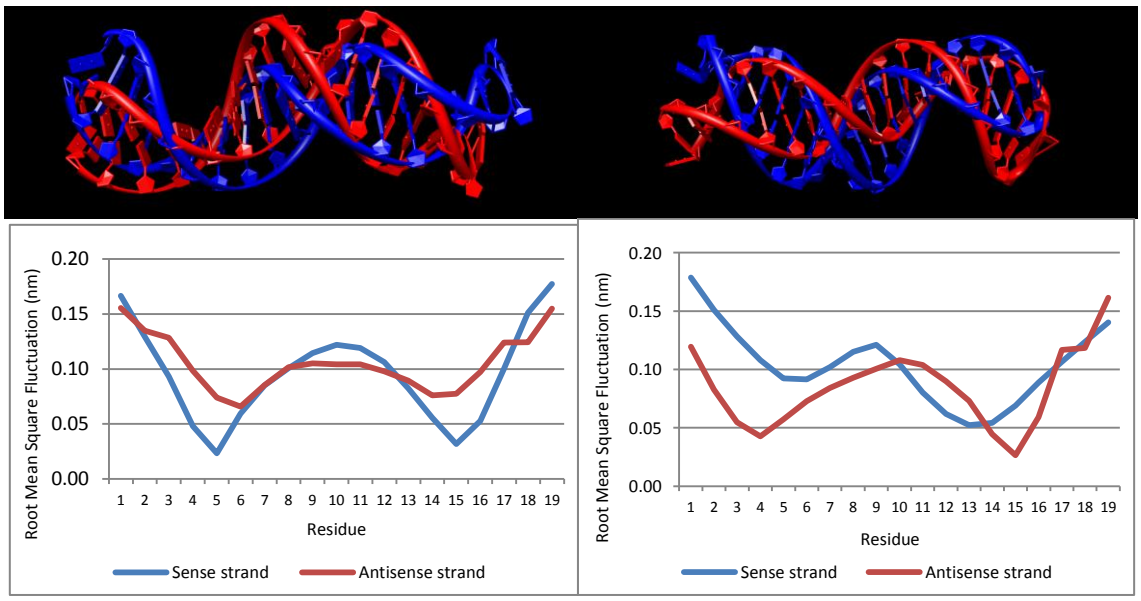

Figure 2-46: Extreme structures and RMSF values of the first (left) and second (right) eigenvectors of the oxidised c-Kit B-DNA Molecule.

2.2.6 Simulations Performed on c-Myc G-Quadruplex Structures

To observe the effect of oxidation on the c-Myc G-quadruplex and to perform comparisons between this structure and that of the c-Kit G-quadruplex, the same simulations and analyses were performed as described in section 2.2.4.

Observing figure 2-47, in terms of the overall structure, the non-oxidised c-Myc G-quadruplex remains relatively stable. However there appears to be a “compression” of the structure, in a perpendicular orientation to the tetrads, leading to a more “propeller” type quadruplex structure.

In terms of the loops, these remained very stable, with little fluctuation over the course of the simulation. The 5’ (T1 G2 A3) and 3’ (T20 A21 A22) tails fluctuated during the initial 100 nanoseconds of the simulation, however at this point they appeared to be forming π- π interactions with both the upper and lower tetrads, through T1 and A22 and stabilised at these positions, reducing the appearance of the overall size and making the structure appear more “compressed”.

In terms of the G-tetrads, these remained stable throughout the simulation, appearing to retain the π- π stacking with the adjacent tetrad(s) and each guanine, although fluctuating, did not appear to lose its orientation in relation to the remaining guanines in the G-tetrad.

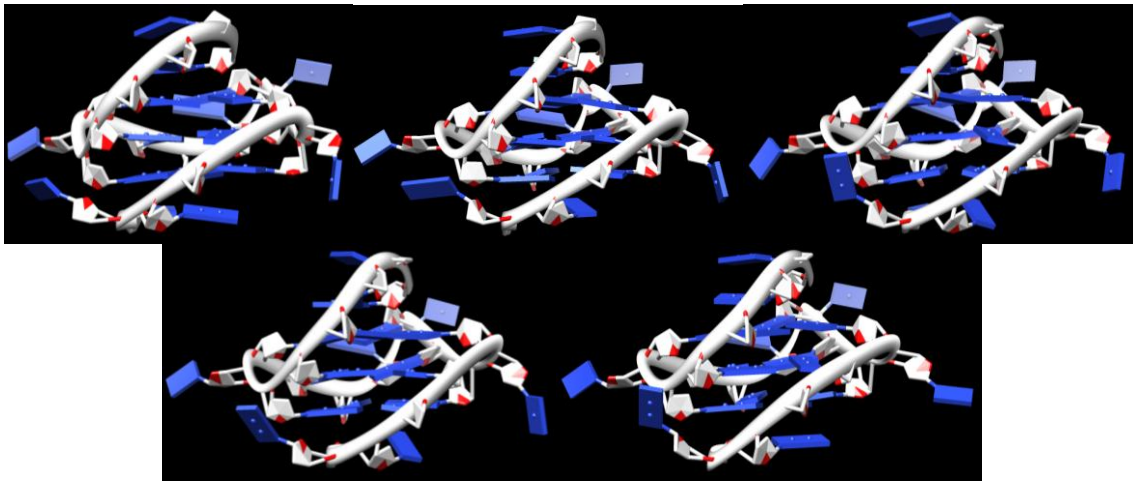

Figure 2-47: Representative snapshots of the cluster analysis of the non-oxidised c-Myc G-quadruplex. Top left: Cluster of 166 frames representative of the first 100ns; Top centre: Cluster of 312 frames representative of the second 100ns; Top right: Cluster of 243 frames representative of the third 100ns; Bottom left Cluster of 249 frames representative of the fourth 100ns; Bottom right: Cluster of 158 frames representative of the last 100ns.

Overall, the oxidised c-Myc G-quadruplex structure presented in figure 2-48, remains stable throughout the 500 ns simulation. However there are some subtle differences with the non-oxidised structure.

In terms of the three loops, the first single nucleotide loop (T7) appears to greatly fluctuate and throughout the simulation appears to rotate outwards, increasing the distance from the oxidised guanine (G9). This same phenomenon is observed in the double nucleotide loop (T11 and A12). The remaining single nucleotide loop (T16) does not suffer prom this phenomenon.

And, the 5’ (T1 G2 A3) and 3’ (T20 A21 A22) tails were not able to form the stacking to the same extent observed in the previous simulation, resulting in continuously fluctuating tails throughout the course of the simulation although it should be noted that by the end of the simulation, A22 appeared to have formed a stable stacking interaction with the upper tetrad.

In terms of the G-tetrads, these remained relatively stable, maintaining orientation and stacking. However, the all the guanines involved in tetrad formation appeared to twist, subtly, although maintaining stacking and likely Hoogsteen hydrogen bonding

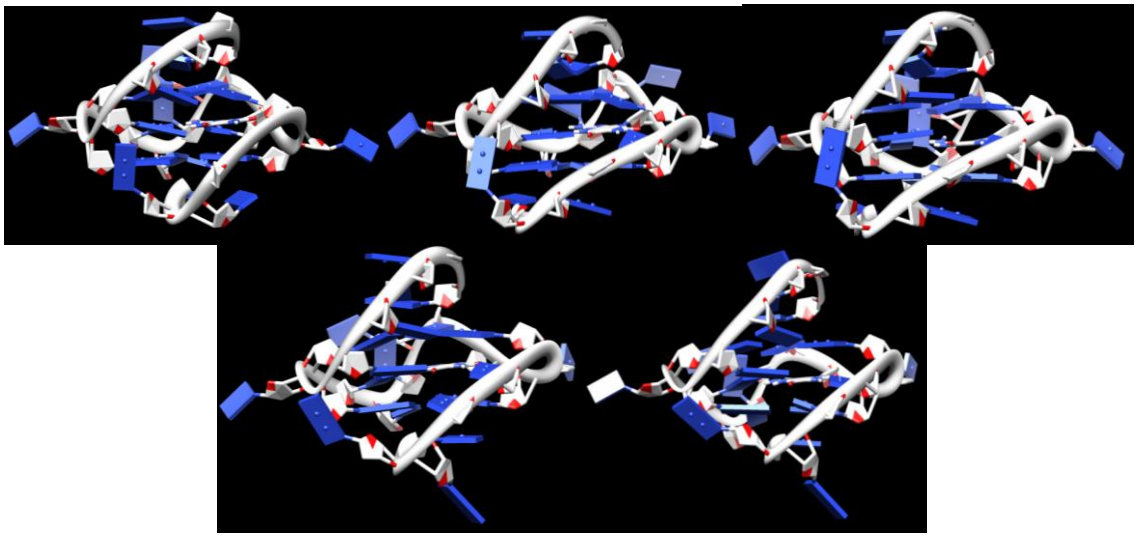

Figure 2-48: Representative snapshots of the cluster analysis of the oxidised c-Myc G-quadruplex. Top left: Cluster of 133 frames representative of the first 100ns; Top centre: Cluster of 216 frames representative of the second 100ns; Top right: Cluster of 168 frames representative of the third 100ns; Bottom left Cluster of 158 frames representative of the fourth 100ns; Bottom right: Cluster of 134 frames representative of the last 100ns.

Figures 2-49 to 2-56 present the summary of the comparisons performed between the non-oxidised and oxidised forms of the c-Myc G-quadruplex structure.

As can be observed in figure 2-49, the RMSD of the two c-Myc G-quadruplex structures over time, reveals a more substantial difference; whereas the non-oxidised form remains fairly constant at an average of 0.282 nm (standard deviation of 0.023 nm) the RMSD values of the oxidised structure vary far more, fluctuating between values similar to those of the non-oxidised structure and lower within the first 100 ns (with an average of 0.236 nm) for this part of the simulation). After which point the RMSD values continually rise and stay above the non-oxidised structure with an average of 0.329 nm for the final 100 nanoseconds.

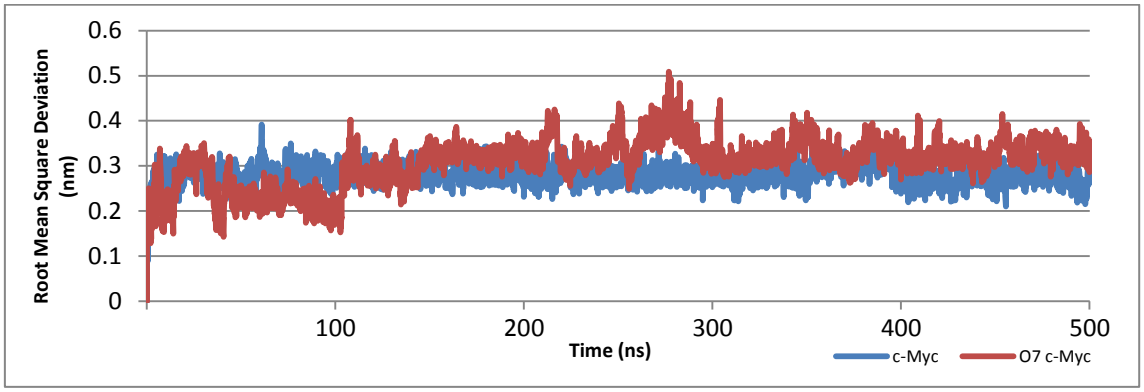

Figure 2-49: Root mean square deviation (RMSD) comparison of c-Myc G-quadruplex structures with and without 8-oxo-dG incorporated.

The focused RMSD of the G-tetrads (figure 2-50) shows that these remained far more stable that the rest of the molecule, and similarly to the general RMSD, the oxidised structures presented higher values (with averages of 0.122 nm and 0.142 nm for the non-oxidised and oxidised forms, respectively). Furthermore, the graph denotes similar tendencies for both structures; with the non-oxidised structure remaining stable throughout the simulation (with a standard deviation of only 0.014 nm) and the oxidised version increasing to higher RMSD than the non-oxidised structure after around 100ns, to an average of 0.148 nm for the final 400 ns. However, the focused RMSD of the oxidised structure showed to be far more stable throughout the simulation after 100ns with a standard deviation of only 0.017 nm for this portion of the simulation.

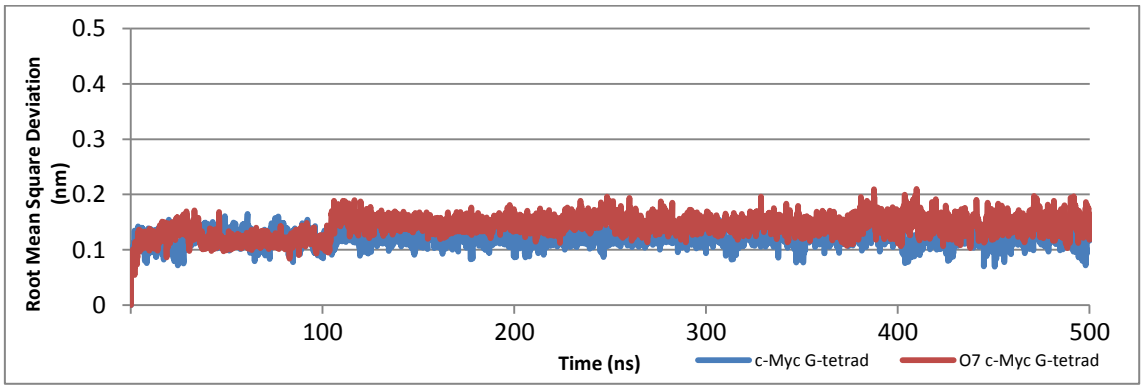

Figure 2-50: Root mean square deviation (RMSD) comparison of the Guanines involved in tetrad formation in the c-Myc G-quadruplex structures with and without 8-oxo-dG incorporated.

The RMSF values presented in figure 2-51 are in agreement with the previous results; both structures present higher values at the tails and loops, and the differences between the structures is more pronounced in these loops and the 5' tail.. The oxidised structure presents a massive increase in the 5' tail and in the T7 single nucleotide loop and the double (T11 A12) nucleotide loop. The RMSF values of both structures for the oxidised Guanine (G9) are remarkably similar.

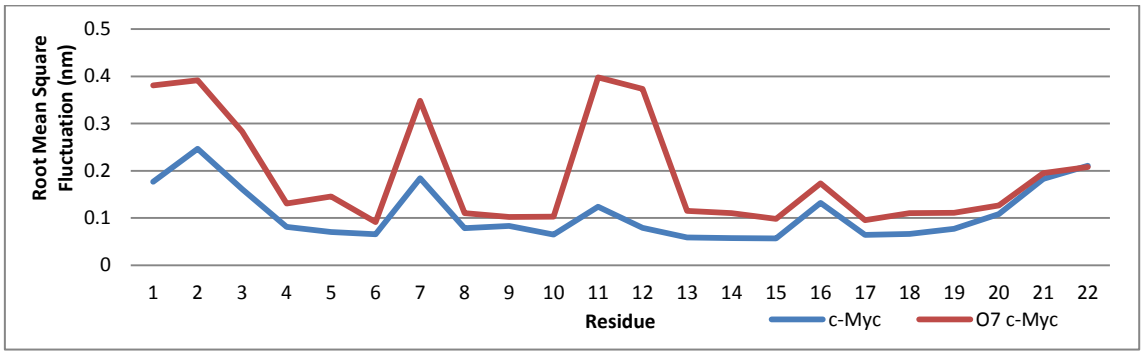

Figure 2-51: Root mean square fluctuation (RMSF) comparison of the c-Myc G-quadruplex structures with and without 8-oxo-dG incorporated.

Hydrogen bonding data of the G-tetrads of the two c-Myc G-quadruplex structures, presented in figure 2-52 is concordant with the focused G-tetrad RMSD; both structures presenting an average Hydrogen bond number of 24 for the first 100 ns, after which point the oxidised guanine appears to lose a Hydrogen bond, with an average over the last 400 ns of the simulation of 23, whereas the non-oxidised structure remains at a constant average of 24.

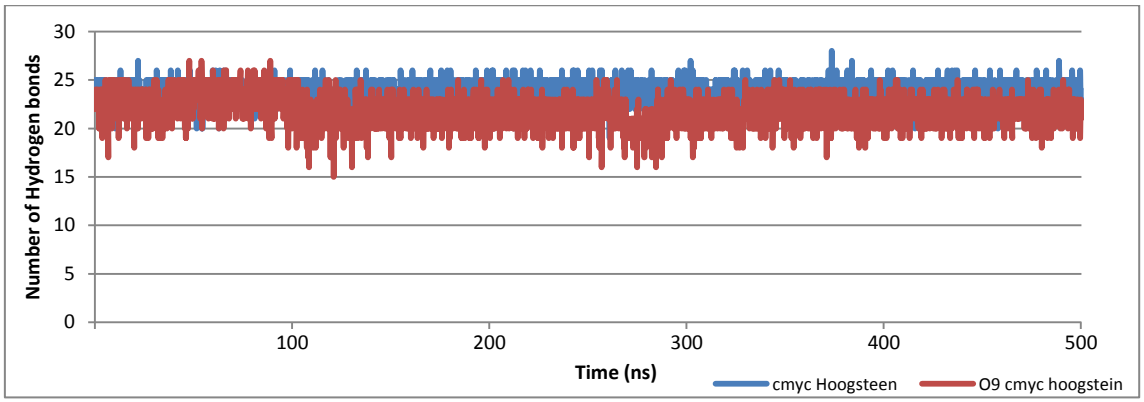

Figure 2-52: Comparison of the number of Hydrogen bonds in the c-Myc G-quadruplex structures with and without 8-oxo-dG.

Observing the SASA plots for the two structures (figure 2-53), it is apparent that there are discernible differences between the two structures. The G-tetrads present low values (approximately 1.5 nm^2) and the loops present higher values (up to 3 nm^2). The differences between the two structures are only apparent in the Guanines involved in tetrad formation (G4, G5, G6, G8, G9, G10, G13, G14, G15, G17, G18 and G19); where the oxidised structure presents generally higher values, indicating a greater surface area.

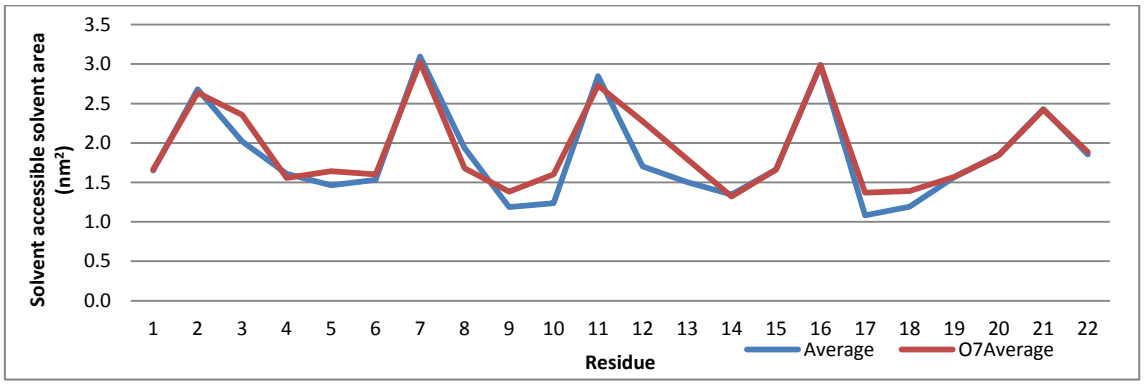

Figure 2-53: Solvent accessible surface area of the non-oxidised and oxidised c-Myc G-quadruplex structures.

PCA was also performed to identify any significant motions within the molecules. Figures 2-54, 2-55 and 2-56 represent the results of PCA on both the oxidised and non-oxidised c-Myc G-quadruplex structures.

Figure 2-54 shows the percentage of motility explained by the eigenvalues of the eigenvectors. This analysis describes that the first 3 eigenvectors account for a larger proportion of the motility of the oxidised and c-Myc G-quadruplex structure (approximately 52%) than of the non-oxidised c-Myc G-quadruplex structure (approximately 40%).

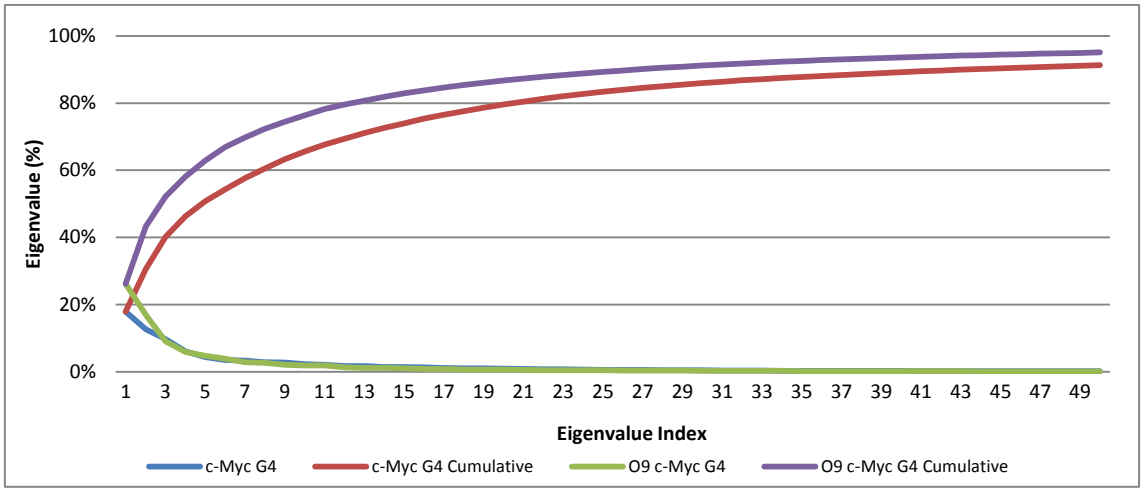

Figure 2-54: Graph plotting percentage of motility explained by eigenvalues against eigenvectors (principal components) of the non-oxidised and oxidised c-Myc G-quadruplex structure and cumulative percentage.

Figures 2-55 and 2-56 represent the extreme structures and RMSF values of the first two eigenvectors of the non-oxidised and oxidised c-Myc G-quadruplex structures. Within figure 2-55, the first eigenvector, representing the most significant motion of the non-oxidised structure corresponds to the fluctuation of the 3' tail of the structure, towards the final conformation previously observed, with T20 and A22 stacking on top of the upper tetrad. G2 was the only other residue to suffer significant motion, twisting slightly, without significantly affecting the backbone. The second eigenvector of the same structure shows represents the movement of the 5' tail, which fluctuated to a conformation which should enable the establishment of stacking interactions with the lower tetrad.

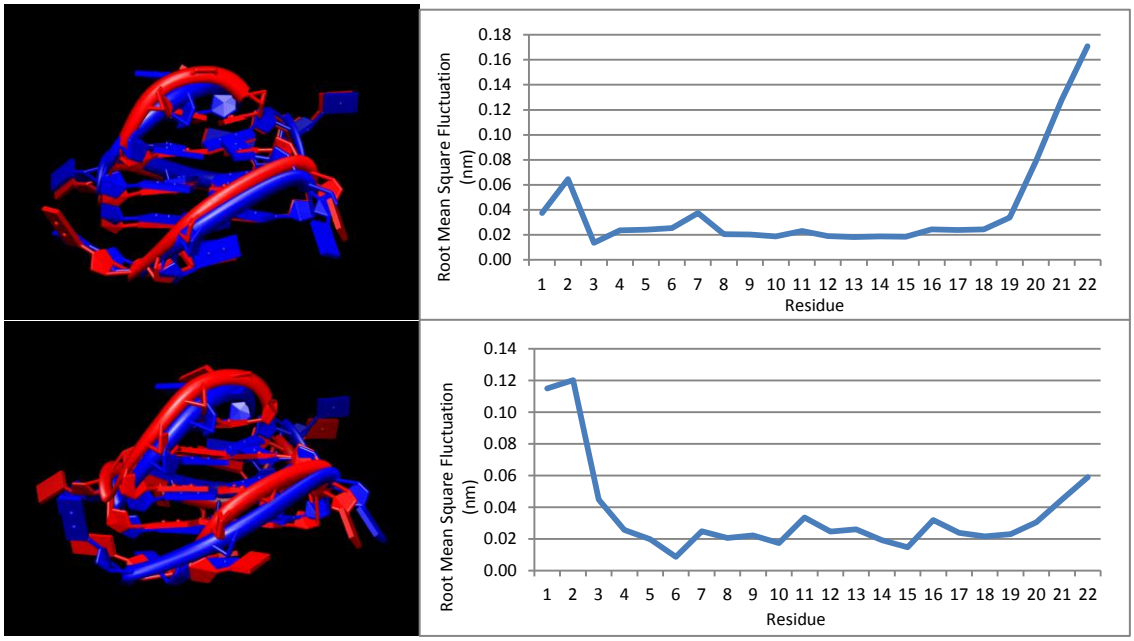

Figure 2-55: Extreme structures and RMSF values of the first (top) and second (bottom) eigenvectors of the non-oxidised c-Myc G-quadruplex.

When observing the oxidised c-Myc G-quadruplex structure (Figure 2-56), the structures pertaining to the first eigenvector corresponds to distortions in the 5' tail and loop sequences, with high amounts of movement observed in T11 and A12 in the double nucleotide loop, which twisted outward, away from the oxidised lesion. The second eigenvector represents similar alterations in the 5' tail (although to a lesser extent) and to the T7 loop nucleotide, which again twisted outward, away from the oxidised Guanine.

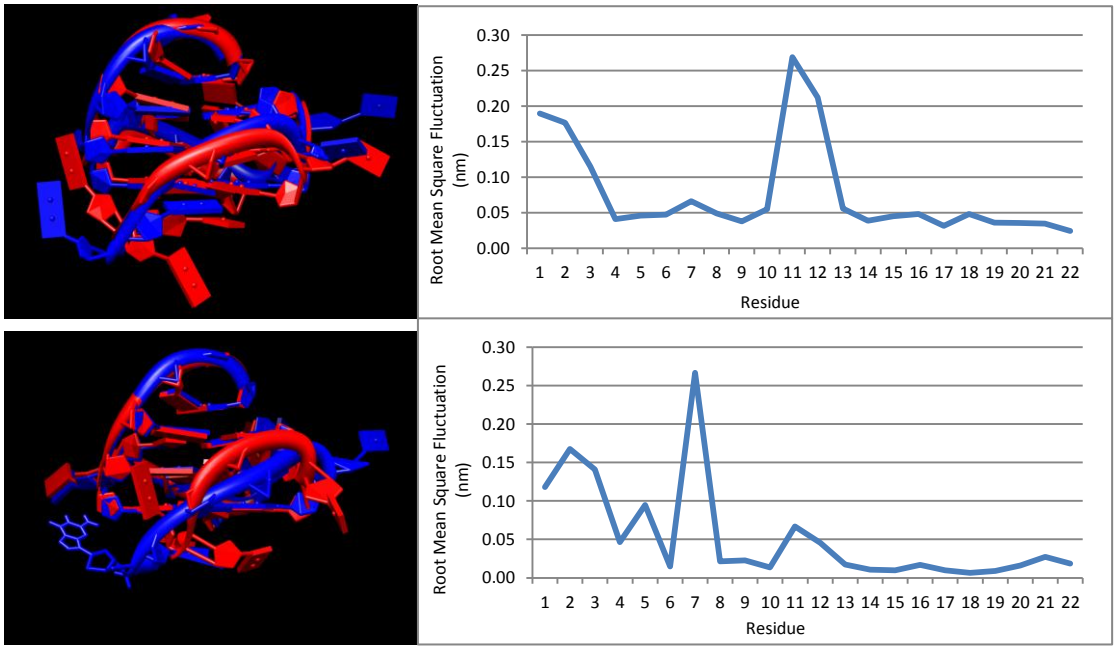

Figure 2-56: Extreme structures and RMSF values of the first (top) and second (bottom) eigenvectors of the oxidised c-Myc G-quadruplex.

2.2.7 Simulations Performed on c-Myc B-DNA Structures

As with the c-Kit G-quadruplex motif, the B-DNA form of the c-Myc G-quadruplex motif was also simulated with and without incorporation of 8-oxo-dG.

Figures 2-57 presents the most prevalent clusters observed for every 100 ns over the entire simulation of the non-oxidised c-Myc G-quadruplex motif. As can be observed from these figures, the structure remained largely stable; no discernible changes were observed throughout the course of the simulation, apart from end fraying and bending.

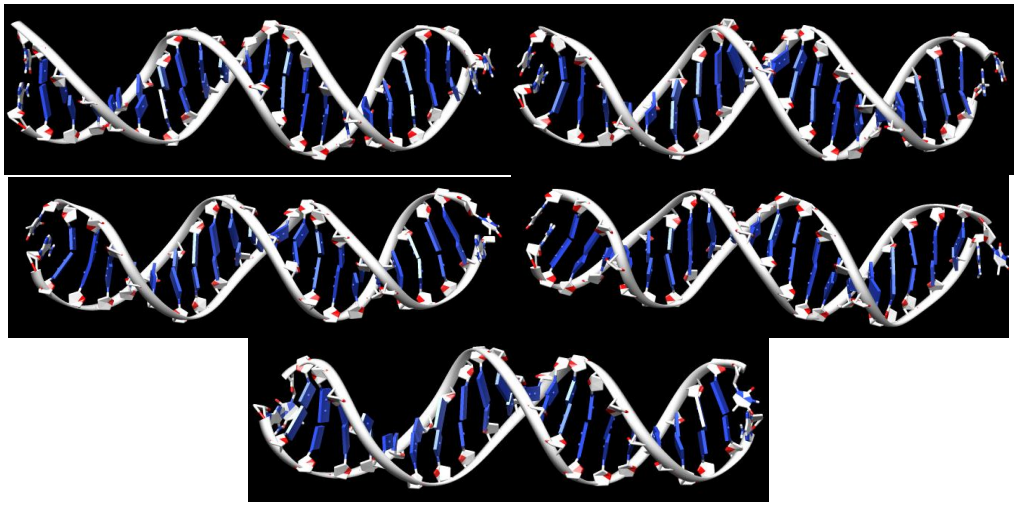

Figure 2-57: Representative snapshots of the cluster analysis of the non-oxidised c-Myc B-DNA molecule. Top left: Cluster of 59 frames representative of the first 100ns; Top right: Cluster of 114 frames representative of the second 100ns; Centre left: Cluster of 77 frames representative of the third 100ns; Centre right: Cluster of 130 frames representative of the fourth 100ns; Bottom: Cluster of 95 frames representative of the last 100ns.

Figures 2-58 represents the same cluster analysis, performed on the oxidised c-Myc B-DNA molecule. Throughout the 500ns simulation, the same observations were made in terms of the backbone, base stacking and pairing and end fraying, however the bending observed was more pronounced.

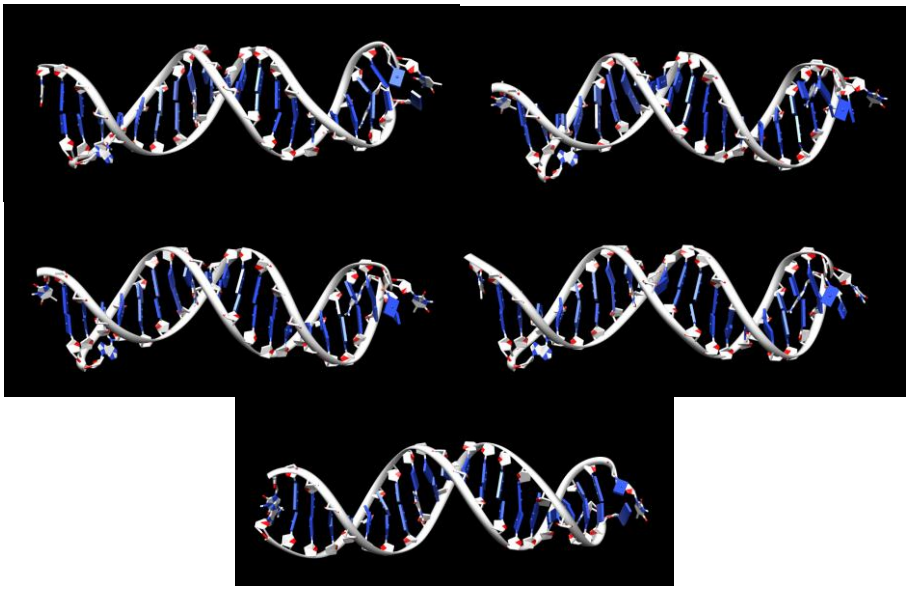

Figure 2-58: Representative snapshots of the cluster analysis of the non-oxidised c-Myc B-DNA molecule. Top left: Cluster of 164 frames representative of the first 100ns; Top right: Cluster of 97 frames representative of the second 100ns; Centre left: Cluster of 73 frames representative of the third 100ns; Centre right: Cluster of 96 frames representative of the fourth 100ns; Bottom: Cluster of 133 frames representative of the last 100ns.

Figures 2-59 to 2-65 represent the data analysis performed and comparisons performed on the oxidised and non-oxidised c-Myc B-DNA molecules.

Observing figure 2-59, results corroborate the macroscopic evaluations, RMSD values of the oxidised molecule where higher than those of the non-oxidised structure throughout the simulation (averages of 0.338 nm and 0.433 nm for non-oxidised and oxidised B-DNA molecules, respectively). Both molecules appear to be equilibrated and stable, with no discernible shifts in RMSD values after the first few nanoseconds, nor any apparent incline.

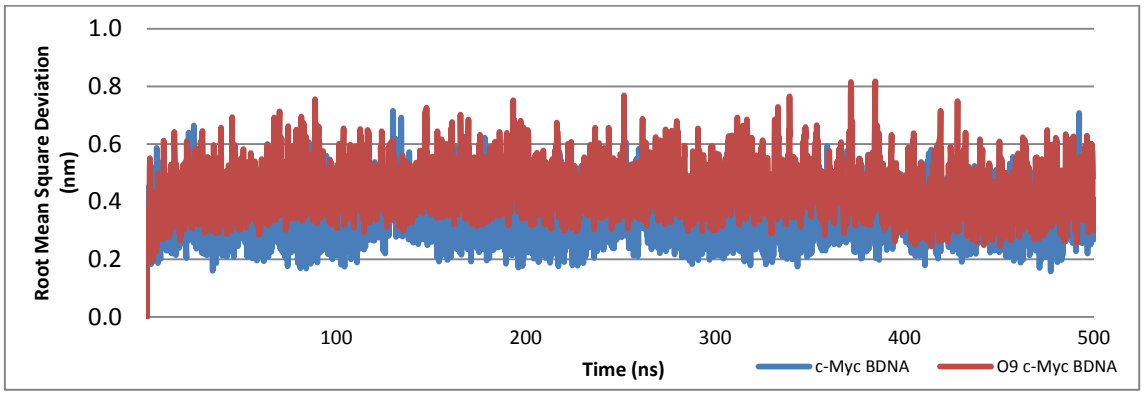

Figure 2-59: Root mean square deviation (RMSD) comparison of c-Myc B-DNA structures with and without 8-oxo-dG incorporated.

Observing the RMSF values plotted in figure 2-60, the results appear concordant with the accrued data; the terminal ends present a higher RMSF value than the rest of the molecule. Between the two molecules, the most discernible differences are observed at the 3' terminal ends of the sense strands and the 5' terminal ends of the antisense strands, where the non-oxidised structure presents a higher value, indicating more pronounced fraying. This situation is reversed at the other terminal end (5' terminal ends of the sense strands and 3' terminal ends of the antisense strands) which present higher RMSD values for the oxidised structure.

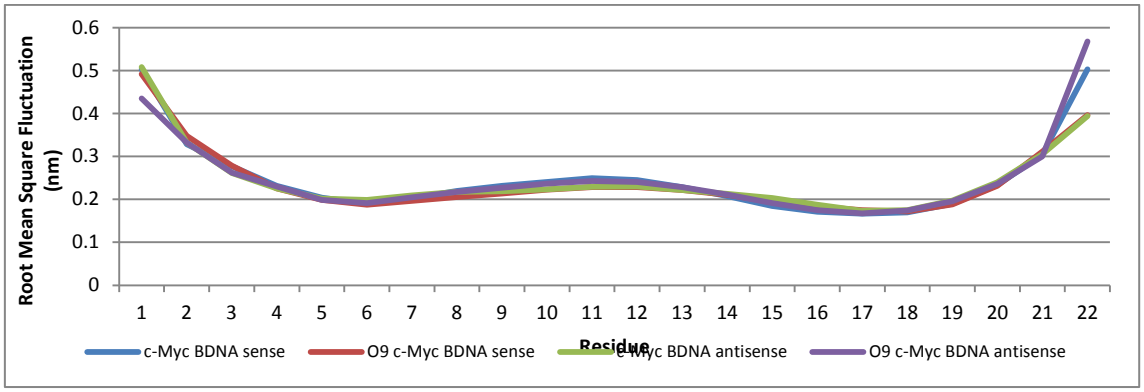

Figure 2-60: Root mean square fluctuation (RMSF) comparison of c-Myc B-DNA structures with and without 8-oxo-dG incorporated.

The Hydrogen bonding data presented in figure 2-61 supports previous data in that the number of Hydrogen bonds remains stable over time in both molecules (standard deviations of approximately 2 for both molecules) and structures appear to have very similar stabilities, with average Hydrogen bond number of 56 for both the non-oxidised and oxidised molecules.

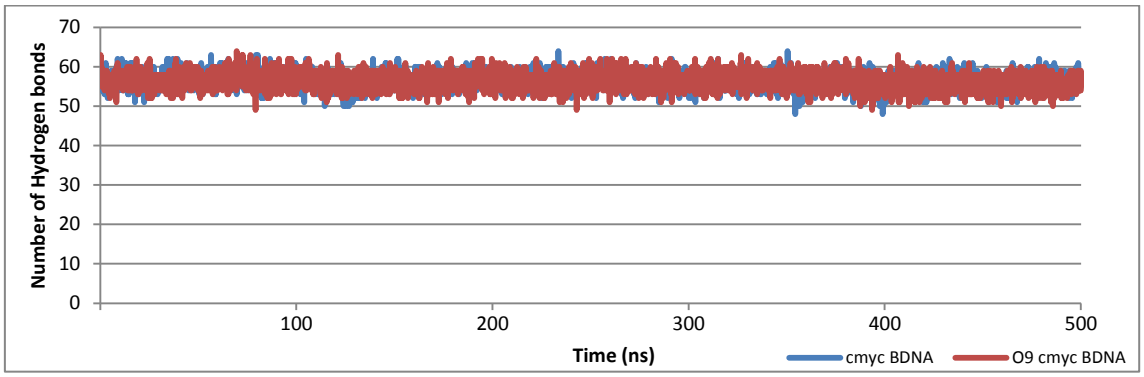

Figure 2-61: Comparison of the number of Hydrogen bonds in the c-Myc B-DNA structures with and without 8-oxo-dG.

In terms of the solvent accessible surface area (figure 2-62), this analysis showed the most pronounced differences between the non-oxidised and oxidised structures of the c-Myc B-DNA molecule. As expected by the previous data, the highest levels of SASA are observed at the 5' and 3' ends. Substantial differences were observed in several stretches along the molecule; in general, the oxidised molecule showed a lower SASA at these nucleotides than the non-oxidised molecule.

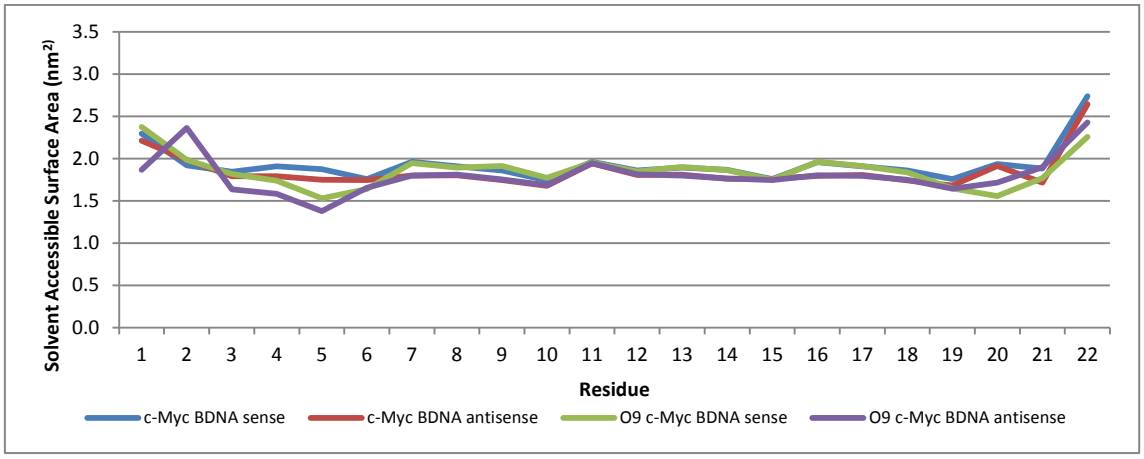

Figure 2-62: Solvent accessible surface area of the non-oxidised and oxidised c-Myc B-DNA molecules.

Principal component analysis (PCA) was also performed to identify any significant motions within the molecules. Figures 2-63, 2-64 and 2-65 represent the results of PCA on both the oxidised and non-oxidised c-Myc B-DNA molecules.

The analysis of the percentage of motility explained by eigenvalues of the eigenvectors (Fig. 2-63) describes that the non-oxidised and oxidised structures showed very similar results, with the first 3 eigenvectors accounting for approximately the same proportion of the motility of both B-DNA molecules (approximately 57%).

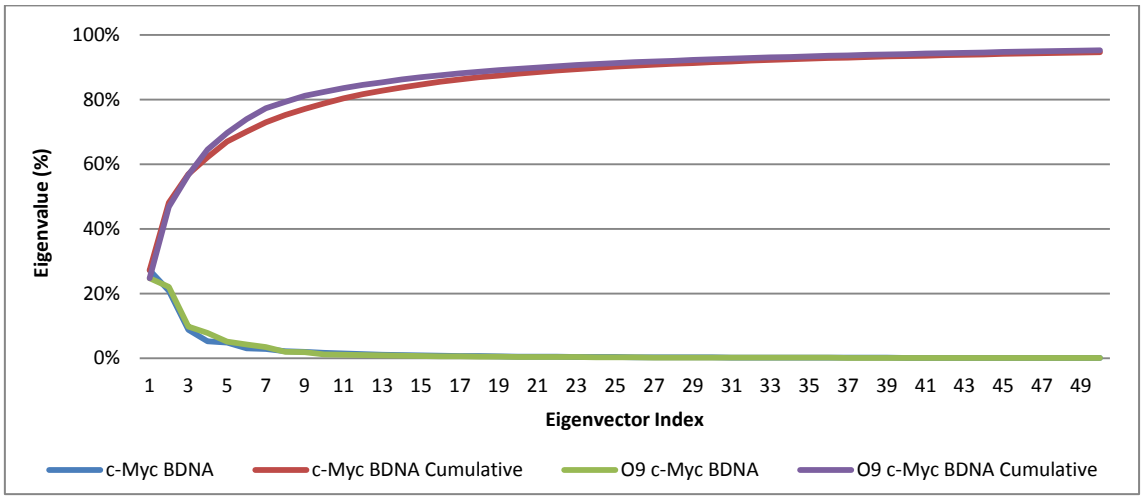

Figure 2-63: Graph plotting percentage of motility explained by eigenvalues against eigenvectors (principal components) of non-oxidised and oxidised c-Myc B-DNA molecule and cumulative percentage.

Figures 2-64 and 2-65 represent the extreme structures and RMSF values of the first two eigenvectors of the non-oxidised and oxidised c-Myc B-DNA molecules. In figure 2-64, the first two eigenvectors of the non-oxidised c-Myc B-DNA molecule are almost perfect mirror images of each other; both show high fluctuation at the 5' and 3' ends of the double stranded molecule however in the first eigenvector, the sense strand appears to twist further around the sense strand, producing a more "bent" molecule, and also increasing the RMSD in the sense strand. Whereas in the second eigenvector, the opposite is true, the antisense strand appears to twist and present higher RMSD values.

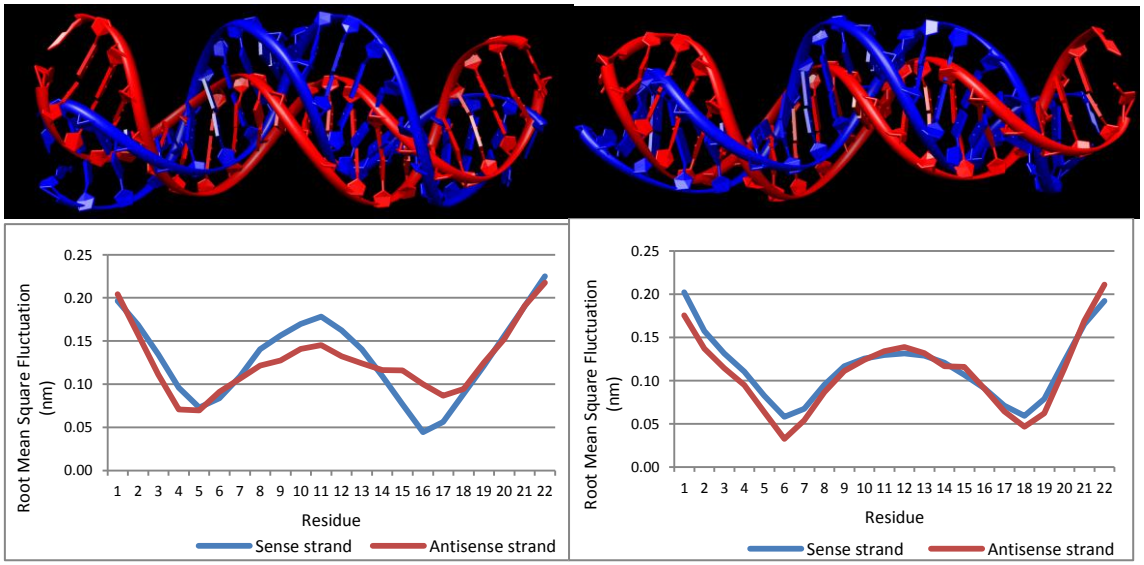

Figure 2-64: Extreme structures and RMSF values of the first (left) and second (right) eigenvectors of the non-oxidised c-Myc B-DNA molecule.

When observing the oxidised c-Myc B-DNA molecule (Figure 2-73), the structures pertaining to the first two eigenvectors show similar results to those observed for the non-oxidised structures. The only difference being that in the first eigenvector represents a high fluctuation in the 3' end of the antisense strand.

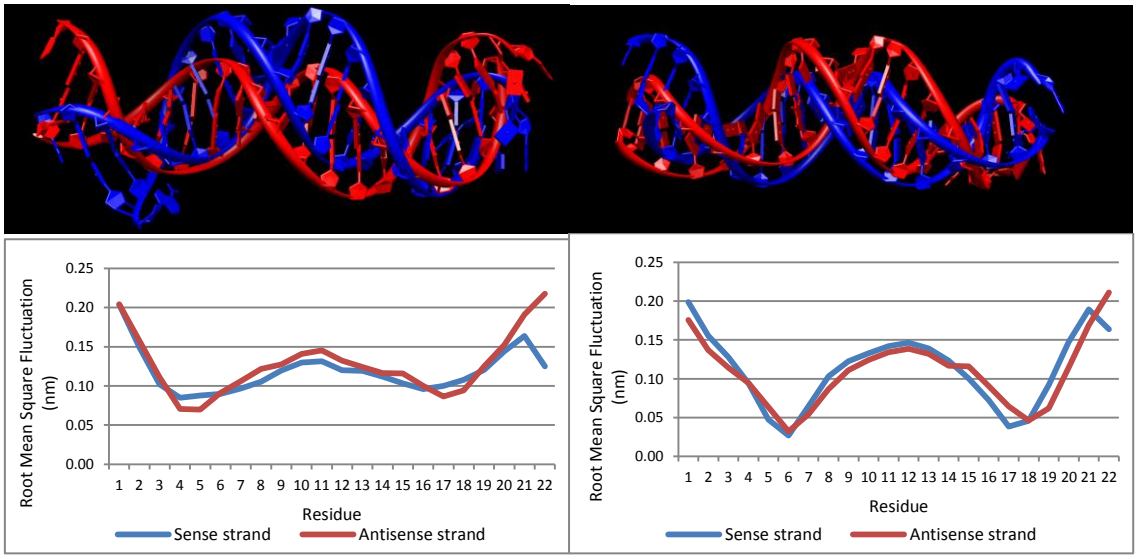

Figure 2-65: Extreme structures and RMSF values of the first (left) and second (right) eigenvectors of the oxidised c-Myc B-DNA Molecule.

2.3 Gene Expression Data Analyses

To shed light on the possible extent and implications of our findings in c-Kit and c-Myc, several bioinformatics analyses were designed and interpreted by the author and the code was written by Dr Lee Larcombe. These experiments aimed to establish:

1) The prevalence of c-Kit and c-Myc similar motifs in upstream promoter regions, UTR and ORF of human genes;

2) Whether the type of G-quadruplex motif has an influence in the extent of the effect on gene expression in helicase deficient cells;

3) whether genes whose expression is significantly altered by oxidation are more likely to contain G-quadruplex motifs;

4) whether the type and/or location of the G-quadruplex motif present has an influence in the extent of the effect on gene expression when oxidised;

To establish these points, three separate analyses were devised. The following is a summary of the methods used and results obtained, including statistical

treatment used. All work was undertaken on a Dell Alienware Aurora workstation; 2.66 GHz Intel core i7, 12 GB RAM, 2Tb HDD running x86_64 CentOS 5 GNU/Linux OS 2.6.18. All PERL and R scripts are available in the Digital Appendix.

Genome data was retrieved from the Ensembl resource [290] (GRCh37 release 63 – www.ensembl.org). The BioMart [291] custom data selection interface was used to generate PERL code, which was subsequently incorporated into scripts (PERL 5.8.8 – www.cpan.org) to obtain specific parts of the total resource required for the in-house database. 2000bp of the 5' flank and the gene sequence (including 5' & 3' UTR) were retrieved and exported to FASTA text files for subsequent processing.

A simple database schema was designed using MySQL 5.0.77 (www.mysql.com), and the FASTA-formatted genome data imported into the relevant tables using PERL scripts. The initial evaluation of the genome for the presence of G4 motifs: c-Kit1, c-Kit2, c-Kit87up, c-Myc, and the general motif (table 2-9); was performed using the Regular Expression-based search features of SQL. Genes including these motifs in the 5' UTR, flank and gene sequences were counted.

The general motif was determined by Huppert and Balasubramanian [24] and is the one currently used to assess the prevalence of GQMs in gene expression data analyses in the G-quadruplex field. To assess the prevalence of c-Kit like structures, three motifs were used, termed c-Kit1, c-Kit2 and c-Kit87up. The c-Kit1 and the c-Kit87up motifs are essentially the same sequence, however c-Kit1 defines the guanines "traditionally" expected to participate in tetrad formation and the c-Kit87up motif defines the guanines determined to be involved in tetrad formation by Phan and co-workers [125]. The c-Kit2 motif refers to the G-quadruplex in the c-Kit promoter region whose structure was elucidated by Kuryavyi and co-workers [69]. With respect to the c-Myc G-quadruplex, only one motif was investigated: the "wild type" c-Myc motif used to elucidate the

structure by Ambrus and co-workers, although it should be noted that this NMR structure was determined from a mutated sequence to avoid polymorphisms [67].

These motifs were then incorporated into a PERL script to automate the retrieval of genes from the database containing these motifs in their 5' UTR, flank and gene regions. The script accepts a list of Human Genome Nomenclature Committee (www.genenames.org) standard gene symbols as input. These lists were obtained from experimental datasets reported in the literature [55] [202] as described in subsequent sections.

Table 2-9: G-quadruplex motifs used in bioinformatics studies.

Motif definition	Motif sequence
General motif	$G_{(3-5)}N_{(1-7)}G_{(3-5)}N_{(1-7)}G_{(3-5)}N_{(1-7)}G_{(3-5)}$ $C_{(3-5)}N_{(1-7)}C_{(3-5)}N_{(1-7)}C_{(3-5)}N_{(1-7)}C_{(3-5)}$
c-Kit1	$G_3N_1G_3N_4G_3N_4G_3$ $C_3N_1C_3N_4C_3N_4C_3$
c-Kit2	$G_3N_1G_3N_5G_3N_2G_3$ $C_3N_1C_3N_5C_3N_2C_3$
c-Kit87up	$G_3N_1G_3N_1G_1N_2G_3N_5G_2$ $C_3N_1C_3N_1C_1N_2C_3N_5C_2$
c-Myc	$G_3N_1G_4N_1G_3N_1G_4$ $C_3N_1C_4N_1C_3N_1C_4$

2.3.1 Analysis of the Prevalence of c-Myc and c-Kit Similar Motifs in Genic Regions of the Human Genome

To ascertain the prevalence of these motifs in the Human genome, as well as to establish the base line prevalence for use in subsequent analyses, the motifs under investigation were checked against the ORF and upstream promoter region (2000 bases from TSS) of the Human Genome.

Table 2-10 displays the results of this analysis. The values presented are the total number of genes fulfilling the specified conditions, with the percentages of the total number of genes (24,717) in parentheses.

Table 2-10: Number and percentage of the different GQMs found in the different locations in the Gene.

	General motif	c-Kit1	c-Kit2	c-Kit87up	c-Myc
Upstream promoter region	10,451 (42.28%)	975 (3.94%)	995 (4.03%)	962 (3.89%)	206 (0.83%)
ORF	17,396 (70.38%)	4,600 (18.61%)	3,328 (13.46%)	4,288 (17.35%)	726 (2.94%)
Genic region	19.029 (76.99%)	5295 (21.42%)	4069 (16.46%)	4,978 (20.14%)	920 (3.72%)

The above table describes a high number of genes with the general motif in the genic region (76.99%), as calculated by adding those found in the upstream promoter region and in the ORF, removing any duplicate entries found. The majority of these were found in the ORF with a large proportion of genes found to have a GQM in the upstream promoter region, also having one or more present in the ORF.

In terms of the c-Kit and c-Myc motifs, in general, the prevalence of c-Kit > c-Myc motifs, and, within the different c-Kit motifs, c-kit87up > c-Kit1 > c-Kit2 motif. All four of these motifs followed the general motif in terms of proportion in the different locations, with the most prevalent location being the ORF.

2.3.2 Analysis of Expression Profiles of Cells Subjected to Oxidation for the Presence of GQMs

To ascertain more information about the effect of oxidation on G-quadruplex regulated genes, a gene expression analysis of Cells under oxidative stress was compared with the Helicase deficient cell results.

This data set was obtained from Briede and co-worker's (2010) [202] global gene expression analysis of Caco-2 cells, studying the response to Hydroxyl and superoxide induced oxidative stress. The authors used both H_2O_2 and Menadione to study the effect of oxidation. Three lists of genes were obtained from the supplementary material (available at http://toxsci.oxfordjournals.org /content/114/2/193/suppl/DC1), relating to genes significantly affected by H_2O_2 treatment, Menadione treatment and genes significantly affected by both oxidising agents were used as input for the PERL script described previously,

with these lists being cross-referenced with the lists of genes with the different types of G-quadruplexes present.

To assess the significance of these results, the total number of genes observed to have each motif in each location were compared through a Chi^2 test with one degree of freedom to the totals expected to be observed, considering the results obtained using the entire genome (Table 2-10).

To analyse the data the null hypothesis (H_0) was determined as "there is no significant difference between the observed and expected results" and the alternate hypothesis (H_a) was determined as "there is a significant difference between the observed and expected results".

The Chi-squared test with one degree of freedom is a variant of the Chi-squared test which uses Yates' correction to reduce the $\chi^2_{calculated}$ value. This variant was used to ensure the significance of the results obtained, as these can be overestimated when sample sizes are small with Chi-squared test. Equation 2-1 shows the formula used to obtain the $\chi^2_{calculated}$ value:

Equation 2-1: Equation used to obtain $\chi^2_{calculated}$ value for all gene expression data analysis results.

$$\chi^2\text{calculated} = \sum \left[\frac{(|observed - expected| - 0.5)^2}{expected} \right]$$

Utilising the results obtained from this equation, $\chi^2_{calculated}$ values were compared to the $\chi^2_{critical}$ values, which with one degree of freedom and a significance of 0.05 is 3.841.

If $\chi^2_{calculated} > \chi^2_{critical}$ then the null hypothesis is rejected indicating that there is a significant difference between the expected and observed results.

Tables 2-11 to 2-13 represent the results of the analysis performed on the gene expression data of cells subjected to oxidation. Results labelled in green denote a significantly higher number of GQMs observed than expected (χ^2_{calc}>3.841

and observed number of genes positive for GQM is greater than expected number of genes positive for GQM); results labelled in red denote a significantly lower number of GQMs observed than expected (χ^2_{calc}>3.841 and observed number of genes positive for GQM is less than expected number of genes positive for GQM) and results labelled in grey indicate that there was no significant difference between the observed and expected results (χ^2_{calc}<3.841). This significance test is not appropriate if the expected value is less than 5, as such, occasions in which this occurs are labelled NA. All calculations were performed using Microsoft® Excel®.

Results represented in Tables 2-11, 2-12 and 2-13, overall, describe no significant enrichment observed in any of the gene lists. The exceptions being enrichment in the general motif in the ORF in cells treated by both Menadione and H_2O_2 and ckit87up in the Flank region in cells treated with Menadione. Furthermore, there were significantly less GQMs observed than expected in the H_2O_2 dataset (general motif in the entire genic region) and in the Menadione dataset (general motif in the entire genic region and c-Kit1 in the ORF).

Table 2-11: Gene expression data analysis of genes significantly affected by both H_2O_2 and Menadione treatment.

		General motif			c-Kit1			c-Kit2			c-Myc			c-Kit87up		
		Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}
Flank	Observed	137	160		13	284		12	285		4	293		17	280	
	Expected	125.5	171.4	1.65	11.7	285.2	0.05	11.9	285.0	0.02	2.4	294.52	NA	11.55	285.4	2.20
ORF	Observed	226	71		55	242		42	255		8	289		54	243	
	Expected	209.0	87.96	4.38	55.2	241.7	0.00	39.9	257.0	0.07	8.7	288.27	0.01	51.52	245.4	0.09
Gene	Observed	234	63		64	233		51	246		12	285		67	230	
	Expected	228.6	68.3	0.45	63.62	233.3	0.00	48.8	248.1	0.06	11.04	285.95	0.02	59.8	237.1	0.94

Table 2-12: Gene expression data analysis of genes significantly affected by H_2O_2 treatment.

		General motif			c-Kit1			c-Kit2			c-Myc			c-Kit87up		
		Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}
Flank	Observed	547.00	798.0		52.00	1293.0		60.00	1285.0		11.00	1334.0		53.0	1292	
	Expected	568.70	776.3	1.37	53.06	1291.9	0.01	54.14	1290.8	0.55	11.21	1333.7	0.01	52.3	1292	0.00
ORF	Observed	928.00	417.0		251.00	1094.0		196.00	1149.0		36.00	1309.0		241.0	1104	
	Expected	946.62	398.4	1.17	250.31	1094.6	0.00	181.10	1163.9	1.32	39.51	1305.4	0.24	233.3	1111.6	0.27
Gene	Observed	973.00	372.0		286.00	1059.0		241.00	1104.0		45.00	1300.0		279.0	1066	
	Expected	1035.5	309.5	16.12	288.13	1056.8	0.01	221.36	1123.6	1.98	50.01	1294.9	0.42	270.8	1074	0.27

Table 2-13: Gene expression data analysis of genes significantly affected by Menadione treatment.

		General motif			c-Kit1			c-Kit2			c-Myc			c-Kit87up		
		Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}
Flank	Observed	383.00	547.0		46.00	884.00		39.00	891.00		9.00	921.00		49.00	881.0	
	Expected	393.23	536.7	0.42	36.69	893.31	2.20	37.44	892.56	0.03	7.75	922.25	0.07	36.20	893.8	4.35
ORF	Observed	648.00	282.0		149.00	781.00		122.00	808.00		29.00	901.00		150.0	780.0	
	Expected	654.54	275.4	0.19	173.08	756.92	3.95	125.22	804.78	0.07	27.32	902.68	0.05	161.3	768.6	0.88
Gene	Observed	678.00	252.0		182.00	748.00		148.00	782.00		38.00	892.00		185.0	745.0	
	Expected	715.98	214.0	8.53	199.23	730.77	1.79	153.06	776.94	0.16	34.58	895.42	0.26	187.3	742.7	0.02

This was the complete antithesis of what was expected and led to the formation of the hypothesis that it would make sense, in a biological context, for G-quadruplexes to not be involved in the regulation of genes involved in DNA damage repair as, if G-quadruplexes were destabilised through DNA damage, cells would be unable to repair themselves.

To investigate this hypothesis, the same analysis was performed on the most prevalent clusters in terms of ontology as defined by the authors, excluding any genes involved in DNA damage repair. This produced two lists of genes, one for each of the clusters found to be significantly affected by H_2O_2 (with 558 genes) and Menadione treatment (with 277 genes). These lists were then cross-referenced against the lists of genes with the different GQMs in the Flank, ORF and in the entire genic region.

Significance was analysed and tables constructed using the same methods previously described.

Tables 2-14 and 2-15 describe the results for the genes affected by H_2O_2 and Menadione treatment, respectively. As can be observed, this analysis generally showed more enrichment of GQMs. The general motif was enriched in all locations in both datasets. The c-Kit motifs showed different results for the H_2O_2 and Menadione datasets; in the H_2O_2 dataset, both c-Kit1 and c-Kit2 were enriched in the ORF but not in the Flank and the c-kit87up motif was enriched in both. However, in the Menadione dataset, the c-Kit1 motif and the c-kit87up motif were enriched in all locations and the c-Kit2 motif was not enriched in any location.

In terms of the c-Myc motif, the prevalence was insufficient to attribute statistical significance to the results in the upstream promoter region, but was sufficient in the ORF and entire genic region datasets; with no enrichment detected in the H_2O_2 dataset but detected in the Menadione dataset.

Table 2-14: Gene expression data analysis of gene clusters not involved in DNA damage repair significantly affected by H2O2 treatment.

		General motif			c-Kit1			c-Kit2			c-Myc			c-Kit87up		
		Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}
Flank	Observed	302.00	256		30.00	528.00		31.00	527.00		9.00	549.00		33.00	525.0	
	Expected	235.94	322.1	31.57	22.01	535.99	2.65	22.46	535.54	3.00	4.65	553.35	NA	21.72	536.2	5.57
ORF	Observed	496.00	62.00		140.00	418.00		116.00	442.00		17.00	541.00		139.0	419.0	
	Expected	392.72	165.2	90.81	103.85	454.15	15.04	75.13	482.87	25.07	16.39	541.61	0.00	96.80	461.2	21.73
Gene	Observed	520.00	38.00		163.00	395.00		141.00	417.00		25.00	533.00		161.0	397.0	
	Expected	429.59	128.4	81.77	119.54	438.46	19.65	91.84	466.16	30.87	20.75	537.25	0.71	112.3	445.6	25.80

Table 2-15: Gene expression data analysis of gene clusters not involved in DNA damage repair significantly affected by Menadione treatment.

		General motif			c-Kit1			c-Kit2			c-Myc			c-Kit87up		
		Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}	Positive	Negative	χ^2_{calc}
Flank	Observed	160.00	117.0		21.00	256.00		15.00	262.00		5.00	272.00		25.00	252	
	Expected	117.12	159.9	26.57	10.93	266.07	8.73	11.15	265.85	1.05	2.31	274.69	NA	10.78	266.2	18.16
ORF	Observed	251.00	26.00		66.00	211.00		47.00	230.00		15.00	262.00		67.00	210	
	Expected	194.95	82.05	53.43	51.55	225.45	4.64	37.30	239.70	2.62	8.14	268.86	5.13	48.06	228.9	8.57
Gene	Observed	259.00	18.00		81.00	196.00		58.00	219.00		20.00	257.00		84.00	193	
	Expected	213.26	63.74	41.71	59.34	217.66	9.60	45.59	231.41	3.72	10.30	266.70	8.54	55.79	221.2	17.24

3 Discussion

This chapter is divided into sections corresponding to each of the techniques used which aim to discuss the results and to critically evaluate the methodology used. A further chapter attempts to present the implications of the work produced in this thesis.

3.1 Fluorescence Resonance Energy Transfer Experiments

3.1.1 Overview of results

Overall, the FRET experiments performed over the course of this project fell short of fully achieving the objective of definitively describing the effect of oxidation on the conformation of G-quadruplex structures *in vitro*. However, the FRET experiments did reveal some important information, and a new methodology for the verification of G-quadruplexes was developed.

The experiments performed on the Varioskan spectral scanner showed poor results, due to a lack of sensitivity at the desired concentrations of fluorophores. Therefore this present section will focus on discussing and contextualising the results obtained from the Rotor-gene Q PCR machine.

These experiments showed a clear difference between c-Myc and c-Kit samples run with and without KCl, this KCl dependent FRET is a hallmark of G-quadruplex formation and was considered indicative of the presence of these structures [249].

However, with regards to observing hybridisation dynamics with the complementary strands, experiments performed on the c-Myc structure were not able to observe a hybridisation curve, indicating that the FRET observed was either not due to the presence of a G-quadruplex, or the excess complementary strand concentrations used were unable to force the quadruplex/duplex equilibrium towards the duplex state. Due to the KCl and pH dependent FRET observed (data not shown); it was considered that the method

was detecting a G-quadruplex structure. Further reading of the literature revealed that although hybridisation curves had been observed in the c-Myc G-quadruplex structure in smFRET with 200x excess complementary sequence [292] which is far higher than the 5x excess used, other studies had shown that in the conditions used, the c-Myc G-quadruplex structure was remarkably stable, even at very high temperatures (>70 °C) [293]. Furthermore, the smFRET experiments used a long 3' tail, to which a complementary strand was already attached, which would favour the duplex formation [251] [292].

The experiments performed on the c-Kit G-quadruplex structure successfully observed KCl dependent FRET and hybridisation with the complementary sequence, in a manner proportionate to the complementary oligonucleotide concentration.

Furthermore, studies into the sequence of this structure showed no stable secondary structure or self-dimer using Primer3 v.0.4.0 (available from http://frodo.wi.mit.edu/primer3/). It was concluded, therefore, that the only possible explanation for the FRET values observed was a G-quadruplex structure.

From these results we can conclude that under near physiological conditions, the c-Myc G-quadruplex structure is more stable than the c-Kit G-quadruplex structure. This is to be expected as the c-Kit structure shows significantly more loop nucleotides and therefore longer loops than the c-Myc structure, which is known to be a determinant to stability [81].

However these results were only noted in the c-Kit oligonucleotide containing 8 flanking bases at the 5' and 3' ends. The results pertaining to the c-Kit GQM with 12 flanking bases were not presented in this thesis as they showed high fluctuations in FRET values, as evidenced by high standard deviation values during various experiments. This was thought to be due to the fluctuations in the single stranded ends that would of course be more pronounced in an oligonucleotide with longer flanks.

With regards to the oxidation experiments, as will be discussed in section 3.1.2, the selection of fluorophores limited the usefulness of the information gained from these experiments at high (> 1 mM) H_2O_2 concentrations. However, experiments performed at concentrations of Fenton's reaction at physiologically relevant concentrations (approximately 300 µM H_2O_2, 20 µM $FeCl_2$ and 50 µM Ascorbic acid) showed no significant variation in the FRET values of the c-Kit G-quadruplex motif. Furthermore, when observing Hybridisation curves, no substantial differences were observed that could be attributed to the oxidation levels.

This could be for one of three reasons, which will be discussed individually: either the G-quadruplex and its competition with the duplex structure are not affected by oxidation; or the method was unable to accurately simulate the effects of DNA oxidative stress, i.e. Fenton's reaction was unable to transform Guanine to 8-oxo-dG; or the method was not sensitive enough to observe the subtle effects caused by oxidation of G-quadruplexes.

It appears unlikely that the duplex-quadruplex equilibrium is not affected by oxidation both through previous evidence accumulated by Gros and co-workers [(180)] on 8' substitutions on Guanines and through the evidence observed in the *in silico* experiments.

It is equally unlikely that the experiment was unable to produce 8-oxo-dG as the concentrations used were, in some cases, far superior to those considered to be physiologically relevant. Even in cases where as much as much as 100 mM H_2O_2 was used, once the data had been normalised to discount the effect of oxidation on the fluorophores, no correlation could be made between H_2O_2 concentration and decrease in fluorescence or changes in the hybridisation curve. Furthermore, studies have shown that concentrations far lower than those used are sufficient to induce the transformation of Guanine to 8-oxo-dG [(283)].

The most likely explanation of the results is that the method was not sensitive enough to detect the effects of oxidation on the G-quadruplex and its competition with the duplex form. This is possible as these effects would likely be subtle, and their effect may not be destructive enough to impede the formation of G-quadruplexes. However, in a sensitive and complex environment such as in a cell, this subtle effect could be sufficient to favour the duplex form, to the detriment the G-quadruplex structure. Furthermore, the technique developed used relatively high (1 – 2 μM) concentrations of the oligonucleotide containing a GQM. This factor could imitate the effect of molecular crowding, increasing the stability of the G-quadruplex structure [76] [77], counteracting any destabilising effect of oxidation.

3.1.2 Critical Appraisal of Materials and Methods

This project aimed to investigate the effect of G-quadruplex oxidation using *in vitro* techniques, as it was desirable to obtain *in vitro* results to support data obtained using the *in silico* techniques.

To achieve this, a variety of biophysical methods are available, including CD and UV spectroscopy and FRET and structural analyses such as NMR and AFM.

Amongst these, FRET was selected as the most appropriate method, for various reasons; it had been widely used in this field (including on the c-Myc and c-Kit G-quadruplexes [251] [292]); the method could be performed using instruments present at the facilities in Cranfield Health; was relatively simple and inexpensive; and would produce all the information required to achieve our objectives.

FRET, as stated before, has been widely used in the investigation of the biophysical properties of G-quadruplexes, however the majority of these studies utilise the more sensitive single molecule FRET technique (smFRET) as opposed to ensemble FRET, as was used in this project.

Despite the clear advantages of smFRET, the instrumentation required was not available at the facilities at Cranfield University. Additionally, it was intended that the current ensemble FRET methods could be adapted to produce a novel technique for investigating G-quadruplex stability under different conditions, using instrumentation present in all molecular biology laboratories.

The most widely used FRET method for studying G-quadruplex stability is described by Balasubramanian and co-workers [251] [292], involving the use of three ssDNA stands, two of which are labelled (described in Chapter 1, pp. 56).

However, this method was not considered appropriate for this project, for several reasons, firstly, it involved the use of three strands using smFRET and, when using ensemble FRET, this would add to the complexity and likelihood of intra- or inter-strand dimers, especially at the relatively high concentrations (1 – 2 μM), influencing the results.

Furthermore, it was intended to obtain information on the relative stability of normal G-quadruplexes vs. G-quadruplexes under oxidative stress, with the simplest way of achieving this being to use a melting curve. With the three strand method, this would be complicated as an increase in temperature would denature the dsDNA, reducing FRET, without observing the effect on the G-quadruplex structure.

For the above reasons, the G-quadruplex motifs used were dual end labelled with both the donor fluorophore (at the 5’ end) and the acceptor fluorophore (at the 3’ end).

The solution used to mimic near physiological conditions was identical to that used in the aforementioned smFRET experiment performed on the c-Myc G4 [292]. It consisted of 10 mM Sodium cacodylate which would act as a buffer, maintaining a constant, physiological pH of 7.4, and 100mM KCl to ensure near physiological levels of K+, necessary for the formation of G-quadruplexes.

As stated previously, a temperature step was used to observe the unfolding kinetics of the G-quadruplex structure and to compare it to the oxidised structure. However, the fluorophores selected, Cy3 and Cy5, were found to be thermolabile, with their fluorescence reducing significantly above 40°C . Reviewing the literature, it was observed that Cyanine dyes, such as Cy3 and Cy5 are extremely thermo-labile, showing a 70% reduction in their quantum yield at 65°C [294]. This was unexpected as there are various, current examples of temperature melting studies using these dyes [295], and even one study using Cy3 and Cy5 dyes to observe the effect of temperature on G-quadruplexes [252] which are, in fact, flawed.

Both instruments utilised to assess FRET had the required specifications to do so, however, after the results obtained from the Varioskan Flash Spectral Scanner (see digital Appendix), this instrument was determined to be inadequate at accurately measuring fluorescence at the desired concentrations.

The Rotor-Gene Q PCR machine, on the other hand, was able to overcome this flaw but did not have the advantage of automatic dispensers; as such the method had to be adapted to enable the addition of reagents without affecting fluorescence levels.

This, ultimately, led to the use of a 3 part method in which the base-line or normal fluorescence values were taken 100 times, then the tubes were removed instantly and, protecting the tubes from light, the oxidation reagents were added to the walls of the tubes, without contacting the initial solution. After which the tubes were place back into the instrument and another 100 readings were taken. The same procedure was then applied to adding the complementary sequences.

This was intended to ensure that any alteration in the fluorescence levels could be attributed to changes to the G-quadruplex structure, and not to prolonged exposure to light or air. This method also ensured that the reagents would only

be mixed when the instrument started to spin, at which point measurements were taken.

The greatest inconvenience to the use of this equipment is the use of relative fluorescence units (RFUs), as these are individual to each experiment and are hard to correlate to energetic values, and therefore to physical parameters such as the distance between the fluorophores, without the use of a calibration curve, which would have been prohibitively expensive.

Nevertheless, although the absolute values observed in different experiments could not be compared, results within the same experiment can. For this reason each experiment had up to 24 samples, corresponding to duplicate runs of samples with varying concentrations of components.

With respect to the oxidation experiments performed using Fenton's reaction to create ROSs to induce the transformation of one or more Guanines within the GQM to 8-oxo-dG, experiments were performed at a wide range of concentrations and with various controls, to be certain that the methodology would detect any change in the G-quadruplex structure due to oxidation, if possible.

To this end, control samples were analysed, containing, in most cases, all but one reagent, to assure that said reagent was not incorrectly influencing the results. This analysis that neither $FeCl_2$ nor Ascorbic Acid affected fluorescence levels in the absence of H_2O_2. Additionally the effect of H_2O_2 is exacerbated in the presence of $FeCl_2$ and Ascorbic Acid, indicative of Fenton Chemistry.

A wide range of concentrations, from 30 µM to 1 M H_2O_2 with varying concentrations of the other components was used. This range was used as, at concentrations considered to be physiologically relevant, no alteration to the FRET values was observed. As such higher concentrations were used to attempt to establish the levels at which FRET values were altered through changes in the G-quadruplex structure caused by oxidation. However when concentrations were increased to the milimolar range, the individual

fluorescence of Cy3 and Cy5 were reduced significantly and this was factored in, no changes to the FRET values was observed.

In summary, the intended ensemble FRET methodology developed was an ineffective tool relative to the study of the effect of chemical modifications on the G-quadruplex structure.

The results obtained do not however, invalidate the methodology for future use in this field; it was shown to be a simple and relatively inexpensive method of detecting the presence of G-quadruplexes and monitoring the hybridisation to its complementary strand using instruments present in all molecular biology laboratories.

Furthermore, the fact that the method does not appear to be suitable to detect the presumably subtle alterations caused by chemical modifications does not preclude the method from being used in other investigations, such as the use of stabilising or destabilising G-quadruplex ligands, as these have been shown to work *in cellulo* and as such would have a substantial enough effect to be detected by the FRET method developed.

The only major inconvenience of this is the expense of dual labelled oligonucleotides. However, in laboratories without sophisticated instruments aimed at structural/biophysical elucidation of biomolecules, this method is a viable option for the detection or verification of G-quadruplexes.

3.2 Molecular Dynamics Simulations

3.2.1 Overview of Results

The results obtained from the MD simulations achieved the objective of modelling the effect of 8-oxo-dG over a 500 ns MD simulation and formed the theoretical basis on which to base our hypothesis of the association of oxidation, quadruplexes and carcinogenesis. This section aims to contextualise these results, paying specific attention to the biological relevance.

In relation to the stabilities of the non-oxidised G-quadruplexes, these appear remarkably stable, with little or no fluctuation in the central tetrads, although the RMSD values for the whole structure are relatively high, when considering the size of the structure. The loops of both the c-Myc and c-Kit structures did fluctuate, but this was not considered to be a sign of low stability and their contribution to the overall RMSD was negligible. When comparing the loops in the two structures, their fluctuation appeared to be directly proportional to the length of the loop in question, as expected. The largest contributions to RMSD were, by far, from the 5' and 3' tails and, as these tails are not representative of flanking regions of G4s within the chromosome, their contribution can be disregarded. These results, aside from indicating stable structures, under the conditions used, indicate that the equilibration protocol was correctly performed, and this was corroborated by observing data relating to the energetic properties of the structures (data shown in Appendix DVD).

When observing the oxidised c-Kit G-quadruplex structure, there are obvious differences with the non-oxidised structure. Again, the contributions of the 5' and 3' tails were not considered relevant in a biological context.

The loops, in general, behave similarly to the non-oxidised structure; however the loop nucleotides closest to the oxidised lesion twisted outward, away from the 8-oxo-dG nucleotide. This is likely due to a localised pocket of increased potential energy around the oxidised nucleotide, caused by the unfavourable electrostatic interactions between the oxidised sugar base and the surrounding bases. To resolve this increased potential energy, the distance between the oxidised nucleotide and those around it, is increased. This movement is, of course, more easily performed in nucleotides with less attractive forces maintaining them in place, hence why the loop nucleotides twist outwards.

When observing the Guanines involved in tetrad formation, these also suffered substantial distortions; the oxidised lesion, as well as the guanines above and below it (contiguous nucleotides on the same strand), appeared to twist and move outwards from the remainder of the tetrad, as if the backbone was being

pulled away. These distortions are likely to be caused by the same mechanism of resolving the increased potential energy in this area; the electrostatic repulsion between the 8-oxo-dG and the remainder of the Guanines was sufficient to counteract the stacking and hydrogen bonds which act as attractive/stabilising forces. The repulsion was likely exacerbated by the presence of the central, positively charged cations. Furthermore the oxidised guanine twisted into a non-planar arrangement, this caused unfavourable electrostatic interactions between the oxidised lesion and the upper and lower Guanine. This was compensated for by these guanines also twisting, enabling stacking between the three guanines.

Overall, the effect of oxidation clearly destabilised the c-Kit G-quadruplex structure, both Hoogsteen Hydrogen bonding and tetrad stacking are affected which would severely affect stability.

When considering the macroscopic evaluation and data analysis of the c-Myc G-quadruplex, this structure also suffered distortions due to oxidation, although, overall, these were not to the same extent as observed in the c-Kit G4.

The loop nucleotides suffered similar distortions as observed in the c-Kit G4, twisting away from the oxidised lesion. However, in the case of the c-Myc G4, the repulsive forces induced by unfavourable electrostatic interactions around the oxidised Guanine were insufficient to cause any guanine, or group of contiguous Guanines, to be “repulsed” from their original position, which allowed Hoogsteen Hydrogen bonding to not be greatly affected. However, there was some distortion, which took the form of slight twisting of the oxidised guanine into a non-planar arrangement. As the Guanine is still involved in tetrad formation, all the other Guanines involved in tetrad formation also appeared to twist, to resolve the unfavourable electrostatic interactions.

Furthermore, it was interesting to note that although the solvent accessible surface area (SASA) was not altered substantially in the c-Kit G4, the c-Myc G4 showed a substantial increase in SASA when oxidised. This would have

implications in facilitating the interaction with both naturally occurring G4 interacting proteins and artificial G4 ligands.

Taken altogether, these results show that oxidation destabilises the G-quadruplex structure, however the extent of the destabilisation is not uniform amongst different G4 structures. It seems likely that known factors such as loop length play a role in attenuating the effect of oxidation, as longer loop lengths do not restrict the movement of the Guanines to the same extent as shorter loop lengths do, thereby maintaining Hoogsteen Hydrogen bonding. However it does appear that the stabilising forces contributed by the stacking interactions could be weakened by the twisting of nucleotides, independent of loop length.

In a biological context, as these structures exist in an equilibrium with the duplex form, even the less extensive destabilisation observed in the c-Myc structure could be sufficient to shift the equilibrium towards the duplex structure.

To shed further light on the effect of oxidation on the duplex/quadruplex equilibrium, 500 ns simulations were also performed on duplex B-DNA molecules with the same sequences as the quadruplex structures as well as these same molecules with 8-oxo-dG incorporated into the same positions as in the G4 structure.

The non-oxidised c-Kit and c-Myc B-DNA molecules were both considered to be stable, when considering the trajectories and data analysis results, with the most significant movements being end fraying and DNA bending, two phenomena known to occur in both *in vivo* and *in vitro* investigations of duplex DNA.

When observing the oxidised B-DNA molecules, there was no substantial difference with the non-oxidised molecules attributable to oxidation. This implies that oxidation destabilises G4 structures and does not destabilise B-DNA, further confirming the Hypothesis that DNA oxidative stress shifts the duplex/quadruplex equilibrium towards the duplex state.

3.2.2 Critical Appraisal of Materials and Methods

The GROMACS molecular modelling software package used was, as expected, appropriate for the purposes of this project. The flexibility of the software, allowing not only to fine tune every parameter of the simulation and even to modify force field parameters as well as the speed of simulations and ability to scale up when using the Astral supercomputer enabled this project to produce some of the longest simulations on DNA molecules performed to date and, at the time of writing, the longest simulations ever performed on the c-Myc and c-Kit G-quadruplex structures.

The flexibility and transparency of the system allowed the author to, with relatively little experience with the GROMACS package, implement force field parameters not available within the distributed package.

This refers to the AMBER Parm99BSC0 parameters which, as stated before, are the standard parameters used to simulate nucleic acid simulations, since their development in 2007 [287] and implementation in the AMBER software package. As GROMACS is traditionally used to simulate protein dynamics, the force field has not been, thus far, ported to the GROMACS system. However, aside from the desire to perform the most accurate simulations possible, as with the FRET experiments, we sought to be able to perform our simulations using tools available to most, if not all, research groups interested in investigating G-quadruplexes. As such, we thought it an important objective to implement these parameters in the open source GROMACS system.

For these parameters to be used with confidence, and to become a force field distributed with the GROMACS package, they must be extensively validated, which usually involves the use of customised scripts to perform numerous simulations in both the AMBER and GROMACS software suites to compare the results. However AMBER package was not purchased and so it was assumed that, for the purposes of this project, performing simulations of the c-Myc G-quadruplex structure and corresponding B-DNA molecule using the

Parm99BSC0 parameters and comparing them to the previous standard force field, Parm99SB, would sufficiently validate the force field.

This analysis showed that the new force field was superior at simulating the DNA backbone of both B-DNA and G-quadruplexes, when compared to the previous standard, Parm99sb parameters. This led to pronounced differences in the B-DNA structure after approximately 100 ns, with the 99sb parameters inducing artificial distortions over the entire molecule. This was expected as the parmBSC0 parameters were chosen due to the improvement in the simulation of alpha and gamma conformers, enabling the more accurate simulation of DNA.

However, in terms of the G-quadruplex structure, the differences were less pronounced. The only substantial differences noted were in the loop regions, which fluctuated far more in the simulation performed under 99sb parameters. Although this may not be hugely detrimental to stability, it is important to perform simulations as accurately as possible, and it is likely that the unusual c-Kit conformation which includes long loops, shown in this project to be less stable and more susceptible to destabilisation, would be affected by the less accurate 99sb parameters to a greater extent than the c-Myc G-quadruplex.

The use of these force field parameters, combined with the implementation of a new nucleotide residue specifically designed to simulate 8-oxo-dG and the inbuilt algorithms used to calculate atomic interactions, of which Particle Mesh Ewald stands out, enabled this project to produce the most accurate representations of both oxidised and non-oxidised G-quadruplex structures feasible.

The G-quadruplex structures used as proof of concept were the oncogenic promoter G-quadruplexes found in c-Myc and c-Kit. The former was selected as it is the most extensively studied promoter G-quadruplex and has been shown to affect expression and has been used in FRET experiments. The latter was chosen to contrast with the c-Myc G4, as the *in vitro* experiments had

suggested it would be too stable to be affected by oxidation. It was apt as a contrast as it had longer loops, a proven determinant of G4 stability, it was also chosen as it had also been widely studied, showed an effect on expression and was such an unusual structure, when compared to the more traditional c-Myc structure. These structures were considered, therefore, apt as models for proof of concept of the hypothesis.

To investigate the hypothesis, an oxidised G-quadruplex had to be constructed; to do this, the author chose to replace one of the Guanines involved in the formation of the central tetrad. This method was selected as the work performed by Gros and co-workers [180] had demonstrated that 8' Carbon modifications were more deleterious to G-quadruplex stability when they occurred in these Guanines.

As stated previously, this was, indeed, shown as deleterious to stability in our simulations. However to accurately characterise the effect of oxidation, a complete study, examining the effect of 8-oxo-dG incorporation at every Guanine position in each of the G-quadruplexes would have been desirable. Furthermore, although 8-oxo-dG is the standard nucleotide modification used to simulate oxidative DNA damage, the comparison with 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapydG) would also have been desirable. However, time constraints made this approach impractical and the method used was designed to ensure that if oxidation damages a Guanine involved in central tetrad formation, the structure is affected.

Accurate simulations of complex biomolecules require careful consideration of both the environment (i.e. solvent) and equilibration before proceeding to the "production" simulations.

Simulation systems were, in all instances, set up so that the solute was never closer than 1 nm from the periodic boundaries. Although this increased computational time as it introduced a large number of solvent molecules, it was

thought necessary to impede periodic images affecting the dynamics of the solute.

G-quadruplexes require cations for stability, as discussed in Chapter 1, therefore a mixed solvent consisting of water and an appropriate ion was required. The ion that has been determined to be more favourable to G4 stability and the most widely used in *in vitro* studies of these structures is potassium. For this reason, potassium was selected for use in both the FRET and molecular modelling experiments. However, when introducing potassium cations into the simulation system, literature pertaining to novel parameters of this ion was not taken into consideration. These parameters generally use reduced ion radii [263] as the standard AMBER/GROMACS potassium radius is too large for the central cavity of G-quadruplex structures, leading sometimes to instability and even to expulsion of the ion. These observations were not observed in the simulations of non-oxidised G-quadruplexes and were therefore, in hindsight, not considered to be vital to produce accurate simulations.

To enable more direct comparisons between the results of the molecular modelling studies and the *in vitro* experiments, it was decided that the solvent used in the simulations should depict the solvent used in the FRET experiments. For this reason, solvation was performed in an ionic environment of 100 mM KCl, to a net neutral total charge, in water. This was considered to be an appropriate solvent environment for the study of these structures.

The equilibration protocol was designed to release any artificial constraints and resolve all the unfavourable interactions prior to performing the production simulations that would be used to analyse the effect of oxidation. The equilibration had five steps, the first two were designed to resolve the major electrostatic interactions (using energy minimisation) and the less powerful interactions (short MD simulation) in the solvent, by keeping the G-quadruplex restrained. The second two steps aimed to resolve the major electrostatic interactions and less powerful interactions between the solvent and the G-

quadruplex as well as to start resolving any unfavourable interactions within the solute. The final step was aimed at resolving any residual unfavourable interactions, principally within the G-quadruplex by a relatively long (2 ns) MD simulation.

This methodology was considered sound, but, in hindsight, a longer MD simulation at the last step would have been desirable, to ensure equilibration prior to the production simulations, as some simulations showed evidence of possibly not being equilibrated until after 100 ns of the production simulation.

With respect to the simulation parameters used, as stated in the experimental section, these were based on a standard protocol designed for the study of G-quadruplexes using the AMBER software package, and adapting it for use in the GROMACS system, taking into consideration the available literature on the use of GROMACS to study these structures.

The parameters used to perform the simulations can be broadly divided into 3 groups: processing and output control, control of physical conditions and control of particle interactions.

The processing and output control parameters were standard and not thought to influence the results in any way, and the parameters used are those considered to be the standard ones to be used in MD simulations. The only parameter that requires justification is the length of the simulation. As stated before, this project produced extremely long simulations, in comparison to current literature; however it would have been desirable to achieve simulations on the microsecond timescale as the kinetics of G-quadruplexes, like most bio-macromolecules, is on this timescale. Simulations on this timescale were, however, impractical, owing to the computational time required, combined with the numerous technical faults that occurred in the supercomputing facilities at Cranfield. Furthermore, the longer the simulation time, the more inevitable it is that the accumulation of artefacts implicit in MD simulations will affect the results. The simulation time of 500 ns was thus considered to be a good

compromise, whilst still pushing the boundaries of what has been done in this field. Furthermore, this time scale proved to be sufficient to visualise a difference between oxidised and non-oxidised structures. However it is possible that results would have been different, given a longer simulation time.

The control of physical conditions was also considered to be fit for purpose, with most of these being standard parameters used in simulations of bio-macromolecules. The one deviation from standard parameters was the temperature, which was set to 310 K (37°C) to simulate physiological conditions and those of the FRET experiments. This was important as temperature is a determinant of stability in these structures and it was important to be able to, as much as possible, assess the results in a biological context.

The control of particle interactions is by far, the most intricate and complex group of parameters and where several judgements had to be made, that may or may not have detrimentally affected the results. These parameters included the constraint algorithm used and associated conditions, the electrostatic and long range interaction types and associated conditions and the van der Waals conditions. This last one was kept at the default values, with no evident reason to alter them.

The selection of the constraint algorithm however, was not straight forward. The LINear Constraint Solver (LINCS) constraint algorithm was used. This algorithm is more stable and faster than the alternative SHAKE algorithm; however it does not work with angle constraints. To compensate for this, the maximum rotation allowed for a bond in any step was set to 30°, if any bond rotated more than this, a warning would be printed in the log. As this was not noted in any production simulation, the constraint algorithm was considered appropriate. All other parameters associated with constraints were maintained at the default values.

The electrostatic interactions were set to use PME to calculate these forces. PME is a computationally efficient method that enables the simulation of large

molecules with long range interactions by creating a grid or mesh in the simulation and separating the short range interactions from the long range ones and calculating the former in real space and the latter in Fourier space, with little loss of accuracy. The grid dimensions were set to 1.2 Å, this could have been shortened to improve accuracy, however this would have been impractical in the long (500 ns) simulations and the dimensions used were considered to be apt. All other parameters associated with these interactions were maintained at the default values.

Although care was taken to select the most appropriate parameters for our system, current opinions of the exact parameter set to be used are unclear and sometimes contradictory. As such we cannot guarantee that the simulations performed are perfect representations of the G-quadruplex structures.

Furthermore, recent investigations in quantum biology [296] [297] highlight the role of non-trivial quantum effects in complex biological systems, such as DNA. The papers demonstrate that electrostatic and dispersion forces are equally important to maintain the structure of DNA molecules. This implies that accurate molecular models, with the physical, structural and functional properties of DNA are at present unattainable. This highlights the fact that molecular models can only ever be, even at the best of times, an approximation to reality and should be treated as such.

However, taking this into account, the methodology of performing simulations on both oxidised and non-oxidised structures allows us to disregard any artefacts introduced to the system, as all structures were treated with the exact same parameters. It should also be made clear that perfectly simulating these structures to extract biophysical information was not objective of this work; the objective was to be able to relate the simulations performed to a biological context, which was, as much as possible, successful.

The analysis performed on the production simulations was comprehensive and designed to extract any and all information relevant to the stability and/or function of these structures in a biological context.

However, it was a significant oversight by the authors not to include the calculation of free energy values. This would have been extremely useful in the comparison of stabilities but, as it was not included in the parameters of the production MD simulations, it was impossible to calculate free energy values, without repeating the entire simulation.

3.3 Gene Expression Data Analyses

3.3.1 Overview of Results

The Gene expression data analyses performed were not designed to be conclusive in their own right, but shed light on the association of G-quadruplexes and oxidation, in cells. Broadly speaking, these results achieved the desired objective of elucidating the association between oxidation of G-quadruplexes and gene expression.

The initial estimation of the numbers of GQMs in the genome revealed far greater numbers than expected. In the case of the general motif, this was previously estimated to be present in the promoter region of approximately 33% of genes [(31)], however our estimation revealed this motif to be present in the upstream promoter region of 42.28% of genes. This discrepancy is due to the previous study having considered a promoter region of only 1000 base pairs (bp) upstream of the TSS, whereas we considered 2000 base pairs upstream of the TSS to be relevant to gene regulation.

There was also a large discrepancy between the values obtained for the c-Kit87up motif and available literature. A study by Todd and co-workers (2007) [(127)], demonstrated that sequences similar to this motif were exceptionally rare and not found within 100 bp of a TSS within the genome. However, our results

showed this motif to be present in the 2000 bp upstream of a TSS in approximately 4% of genes, a substantial number.

It is debatable as to how large the region upstream of the TSS is relevant to regulation of transcription, however, the fact that G-quadruplexes are suspected to be involved in a variety of processes that can affect transcription, such as histone occupancy, transcription factor binding and interference with DNA dependent RNA polymerase progress [32] [97] [106] [177] [298] [299], it is probable that G-quadruplexes in regions up to 2000 bp upstream of the TSS might affect gene expression. Additionally, the presence of GQMs in the open reading frame is suspected to effect transcription and splicing [66] [300] [301] [302].

It is interesting to note the difference in prevalence of the different motifs analysed in the upstream promoter region. These differences are, in the most part, due to more rigid constraints on some of the motifs, i.e. some motifs, have a greater proportion of their nucleotides defined to be a G, which would decrease the possible permutations, and therefore, decrease the occurrence of these motifs. This is very apparent went observing the difference between c-Myc and c-Kit motifs, as the former has less loop nucleotides, reducing prevalence.

There is also a clear increase in the prevalence of every motif in the ORF, when compared to the upstream promoter region. This is, at least in part, due to the greater length of the ORF, which includes 5’ and 3’ UTRs introns and exons.

Although drawing definitive conclusions as to the distribution of prevalence of the different motifs from these estimations is unviable, the principal reason for performing this analysis is to establish the prevalence of each motif in each location, to then access the significance of the prevalence observed in subsequent analyses.

With respect to the analyses of gene expression in cells subjected to oxidation, the initial analysis for the presence of GQMs revealed, with only 2 exceptions no statistically significant enrichment of any motif in any location. In fact, some

motifs were significantly under-represented in some locations. This was not expected as, even if there was no correlation between GQMs and susceptibility to oxidation, under-representation would not be expected. Upon reflection, it became apparent that, in a biological context, it is intuitive that genes involved in DNA damage repair should not be as susceptible to DNA damage themselves, as this would undermine the cell's repair mechanisms, impeding survival. This logic led to the belief that genes involved in DNA damage repair would be less likely than other genes to be regulated by G-quadruplexes, which we theorise to be susceptible to DNA damage.

To investigate this, the significant ontology clusters, calculated by the authors of the paper were used to select all genes that were significantly altered and whose ontology is not involved in DNA damage repair. The analysis of these genes revealed that GQMs were enriched.

Results for both Menadione and H_2O_2 treatment were broadly similar, with the only major difference being that the c-Myc motif was enriched in the ORF and entire genic region in the Menadione dataset, and not in the H_2O_2 dataset. This could imply that, as Menadione tends to produce superoxide anions ($O_2^{\cdot-}$) and H_2O_2 tends to produce hydroxyl radicals (HO·), that the type of ROS is a determinant of the effect on G-quadruplex structure and consequent expression.

These results, taken as a whole, suggest that the presence of a GQM in any part of a gene not involved in DNA damage repair, increase the susceptibility of this gene to being affected by oxidation. This supports our hypothesis that G-quadruplexes are destabilised by oxidation, which leads to altered gene expression levels.

Furthermore, results could also imply that, due to the lower stability of the c-Kit G-quadruplex, this structure is more susceptible to oxidation and is destabilised in the presence of either superoxide or hydroxyl radicals, whereas the more stable c-Myc G4 is only susceptible to the presence of superoxide anions.

3.3.2 Critical Appraisal of Materials and Methods

The gene expression data analyses (GEDA) performed revealed potentially interesting and somewhat controversial results, requiring a thorough inspection of the methodology used.

The motifs utilised in these analysis were apt for purpose as, taken together, they can be used to establish differences and similarities between the prevalence of any GQM (using the general motif) the c-Myc motif (using both the wild type and mutated motif) and the c-Kit motif (using the previously described 2 motifs corresponding to the c-Kit87up G-quadruplex and the c-Kit2 G-quadruplex).

As no study had been published at the time of performing these analyses on the classification of G-quadruplex structures, this was the most accurate way of estimating the prevalence of motifs similar to the c-Kit and c-Myc motifs.

However, a recent study by Todd and Neidle (2011) [33] use *in silico* techniques to perform this classification and would have been a superlatively useful tool for our study. Any future study attempting to establish a differential effect of oxidation on G-quadruplexes should use the results described in this paper.

The paper studying the effect of oxidation was an apt choice as it performed a global gene expression evaluation, analysing two different ROS and so was a good starting point from which to attempt to extract data on the effect of oxidation on G-quadruplexes.

The statistical significance test selected, Chi-squared test with one degree of freedom is one of many tests which could have been used in this situation. However this one was selected as it is perfectly tailored for datasets with only two outcomes (GQM present/not present) in which you know the expected numbers by chance (using the prevalence across the genome, presented in table 2-17, pp. 150) and where the datasets are relatively small. This last variable can influence other tests such as fisher's exact test and other Chi-

squared tests by inducing an overestimation of the significance of a difference detected in small groups; Yates' correction counteracts this overestimation, leading to a greater certainty of when a parameter presents a significant difference, it is, in reality, a truly significant difference.

3.4 General Overview of Biological Implications

Taken together, the results presented in this project point towards oxidative DNA damage destabilising G-quadruplexes, with direct effects on gene expression.

The molecular dynamics simulations clearly show an effect of oxidation in both the c-Kit and c-Myc structures, and the gene expression data analysis allow us to speculate on the effect this destabilisation would have on gene expression.

We propose that the destabilisation of G-quadruplexes in the upstream promoter regions (and possibly in the ORF) alters gene expression. The two structures used as proof of concept (c-Myc and c-Kit) both support this hypothesis by being destabilised by oxidation and their motifs being enriched, under certain conditions, in datasets of genes affected by oxidation.

This project presents a novel route for the association of oxidative DNA damage and carcinogenesis. As so many known proto-oncogenes present G-quadruplexes in their upstream promoter regions, including the two studied in this work, it appears likely that the increase in expression levels caused by destabilisation of G-quadruplexes would increase the likelihood of developing cancer.

This association is further strengthened by the presence of G-quadruplexes in the upstream promoter regions of proto-oncogenes associated with all six hallmarks of cancer.

Furthermore, it appears that the extent of the effect of oxidation is dependent on G4 structure, as c-Myc and c-Kit motifs present different sensitivities to

oxidation, both from the MD simulations and the analysis of expression data in cells subjected to oxidative stress.

This would also have implications in oncogene expression and could explain the strong association between the expression of some oncogenes and oxidative stress and not others.

A thorough review of literature has also revealed several studies which support the findings presented in this work.

Benzoyl peroxide, a ROS generator, for instance, has been shown to oxidise Guanine double and triple repeats sequences site specifically in dsDNA, with effects on regulation of tumour suppression genes and oncogene expression [(303)].

Additionally, the results explain the observation that the G-quadruplex selective RecQ helicases WRN and BLM preferentially unwind quadruplex substrates containing 8-oxo-dG lesions [(304)], as a G-quadruplex structure destabilised by oxidative stress is likely to be more susceptible to unwinding.

In terms of the implications in telomeric G-quadruplexes, one must take into account that these results may be specific only to promoter G4s, as telomeric G-quadruplexes have a different structure and are subjected to different forces.

However, it is possible that these G4s are also affected by oxidation. This would have the opposite effect of the G-quadruplex stabilising ligands targeted at the telomeres. These ligands, of which telomestatin is a prime example, bind to telomeric G-quadruplexes and impede telomerase activity, eventually leading to senescence in cancer cells. In the case of oxidation, as G4s would be destabilised, this would facilitate the binding of telomerase and the ability of cancer cells to maintain telomere lengths, delaying senescence.

Furthermore a study by Cysewski and Czelen (2010) [(305)] which investigated the effect of oxidation on the complexation of telomeric repeat binding factor 1 (TRF1) with a GQM telomeric repeat unit (although, no mention of G-quadruplexes was made in the study) observed that oxidation of this repeat

sensitised it to TRF1 binding. As TRF1 binding is associated with telomerase inhibition, this would enhance the link between oxidation destabilising the G-quadruplex structure, leading to telomere shortening.

Another study has shown that stabilisation of telomeric G-quadruplexes induces the dissociation of TRF2 from telomeres, leading to telomere shortening in the absence of telomerase [306]. With this knowledge, analysing the study by Cysewski and Czelen, suggests the possibility that oxidation may destabilise the telomeric G-quadruplex structure, which would facilitate the formation of protein-DNA complexes. The consequence could be that in cancer cells, where TRF1 and TRF2 are often downregulated and telomerase is often upregulated, oxidation could exacerbate carcinogenesis, or play a role in establishing immortality.

The destabilisation of G-quadruplexes by oxidation, as stated before, would shift the quadruplex/duplex equilibrium. However this may be only a transient phenomenon or even a phenomenon that only mildly affects the equilibrium and consequent gene regulation. On the other hand, when taking into account the composite DNA damage repair mechanism, of which base excision repair (BER) is the principal component, it seems likely that the consequences of quadruplex oxidation may be more permanent, as the BER mechanism would remove the oxidised guanine which would, most likely impede the reformation of the quadruplex structure. This would lead to more permanent changes in expression profiles in cells subjected to oxidation and might also lead to more permanent effects in telomeres.

4 Conclusions & Further Work

This project aimed to investigate the hypothesis that oxidative stress destabilises G-quadruplex structures, leading to changes in expression levels of proto-oncogenes, which may be associated with carcinogenesis, through the use of FRET bases experiments, MD simulations and Gene expression data analyses.

The FRET based methodology developed was successful at verifying the presence of G-quadruplex structures and their unfolding in the presence of their complementary strand. However the methodology was unsuccessful at detecting any change in FRET under conditions of oxidative stress that could be associated with changes in G-quadruplex structures, and thus failed to achieve the intended objective of providing *in vitro* evidence of a change in G4 conformation due to oxidation.

The MD simulations proved to be successful at identifying substantial alterations to the G-quadruplex structure when simulating oxidised structures vs. non-oxidised structures that were interpreted as destabilising effects, therefore fulfilling the desired objective of observing the effects of oxidation during these simulations.

The molecular modelling methodology used was also successful, producing both an automated script with all simulations necessary for studying large DNA molecules as well as the Parm99BSC0 GROMACS port, currently undergoing validation for distribution with the GROMACS package.

The gene expression data analyses, although by no means conclusive in itself, achieved the objective of shedding light on the effect of G4 oxidation on gene expression and supports the hypothesis. This was implied by the enrichment of G-quadruplexes in the upstream promoter region and/or open reading frame in genes affected by oxidation that are not involved in DNA damage repair.

We therefore conclude that, overall, the aim of this project, to investigate a novel route for the association of oxidative stress and carcinogenesis was achieved. However, we were not able to support the hypothesis with conclusive *in vitro* work. We propose that the mechanism by which oxidation of G-quadruplexes is involved in carcinogenesis is as described below:

- As the most common effect of DNA oxidative damage is the oxidation of deoxyguanosine to 8-oxo-dG [307], this nucleotide would be expected to occur in the G-rich sequences that form G-quadruplexes, indeed guanines forming part of G-quadruplexes have been shown to be more susceptible to oxidation than guanines in the B-DNA duplex;

- Our results have shown that the incorporation of only one 8-oxo-dG into the central tetrad of two oncogenic promoter G4s has a detrimental effect on stability, through MD simulations;

- In a biological context, even slight changes to stability would likely shift the duplex/quadruplex equilibrium, towards the formation of the duplex structure. Furthermore, it appears likely that, at some point, BER would remove the oxidised Guanine, impeding the reformation of the quadruplex;

- Studies [55] [308] [309] have shown that numerous GQMs act as repressors of gene expression when in the folded form and that the unfolding of these structures causes upregulation of the corresponding genes;

- The overexpression of these oncogenes would lead to, or predispose a cell to become cancerous.

Although the results suggest this mechanism to be true, further work should be performed not only to prove this hypothesis in a biological context, but also to further examine the relationship between G-quadruplexes, oxidation and cancer.

Further work in the short term should concentrate on further proving the hypothesis using *in vitro* and *in cellulo* methods, as well as improving and furthering the *in silico* methodologies to yield more information about this phenomenon.

As an initial verification of the effect of oxidation, it would be relevant to perform native PAGE experiments, comparing the migration of quadruplexes subjected to oxidation with quadruplexes not subjected to oxidation. This method has been previously used to purify and to verify the presence of quadruplexes, and a detailed protocol is available [310].

However the most important objective to be fulfilled is to investigate the effects of oxidation on the G-quadruplex structure and consequent expression in human cells.

The simplest way of performing this study would be to select a non-cancerous human cell line and utilise Fenton chemistry to simulate oxidative stress in a subset of these cells and then evaluate the expression levels of several proto-oncogenes (including c-Myc and c-Kit) and tumour suppressor genes in these cells and compare it to a control subset of cells.

A more comprehensive way to perform this study would be complemented by calibration experiments involving the use unspecific G-quadruplex stabilising and destabilising agents and cross referencing the genes (and, in particular, proto-oncogenes) affected by these agents with the list of genes with a GQM in their promoter region or ORF. The results of this could be used to identify key genes to study for the effect of oxidation, as described above, as well as facilitate the inference of the effect of oxidation on G-quadruplex structures.

The MD simulations should be improved upon by performing these same studies and including free energy calculations and, depending on the availability of computing facilities, examining the effect of both 8-oxo-dG and FapydG incorporation in each Guanine involved in tetrad or loop formation. It would also be interesting to perform these simulations over, at least, one µs.

There is also scope to utilise some of the methodologies developed to study other chemical modifications, such as DNA methylation.

The MD methodology, including simulations performed, the equilibration protocol and the force field port developed could be used to examine the effect of methylated Guanine residues on the stability of G-quadruplexes.

The FRET methodology developed could also be a viable option to investigate the effect of methylation on GQMs *in vitro*. This would likely be more successful than the use of this method to investigate the effect of methylation as there is no reason to believe that the fluorophores would be affected by methylation. Additionally, this chemical modification has been shown to substantially affect G4 stability [172] [173] and be associated with G-quadruplex function [178].

It is logical to assume, and indeed a study by Kawai and co-workers (2005) [222] has made the link between oxidation at the telomeres and G-quadruplexes. The aforementioned mechanism of favouring the binding of telomerase in cancerous cells would be an interesting hypothesis to investigate. This would require, however, a different set of experiments to prove this hypothesis.

5 Bibliography

1. **Neidle, Stephen and Balasubramanian, Shankar.** *Quadruplex Nucleic Acids.* Cambridge : Royal Society of Chemistry , 2006.

2. *G-Quartets 40 Years Later: From 5'-GMP to Molecular Biology and Suplamolecular Chemistry.* **Davis, Jeffery T.** 2004, Angewandte Chemie International Edition, Vol. 43, pp. 668-698.

3. *Helix Formation by Guanylic Acid.* **Gellert, Martin, Lipsett, Marie N and Davies, David R.** 1962, Proceedings of National Academy of Science, Vol. 48, pp. 2013-2018.

4. *Four-Stranded Nucleic Acid Structures 25 Years Later: From Guanosine Gels to Telomer DNA.* **Guschlbauer, Wilhelm, Chantot, Jean-Francoes and Thiele, Danielle.** 1990, Journal of Biomolecular Structure & Dynamics, Vol. 8, pp. 491-511.

5. **Wong, Han-Min, Rodgers, Simon and Huppert, Julian.** About Quadruplexes. *Quadruplex.org.* [Online] University of Cambridge. [Cited: 28 10 2010.] http://www.quadruplex.org/?view=about.

6. *Physicochemical Properties of Nucleosides 3. Gel Formation by 8-Bromoguanosine.* **Chantot, Jean-Francois and Guschlbauer, Wilhelm.** 1969, FEBS Letters, Vol. 4, pp. 173-176.

7. *Evolution and Ecology of Spider Coloration.* **Oxford, G S and Gillespie, R G.** 1998, Annual Reveiws Entomology, Vol. 43, pp. 619-643.

8. *The Eyes of Deep-Sea Fish I: Lens Pigmentation, Tapeta and Visual Pigments.* **Douglas, R H, Partridge, J C and Marshall, N J.** 1998, Progress in Retinal and Eye Research, Vol. 17, pp. 597-636.

9. **Kaucher, Mark S, Harrell, William A and Davis, Jeffery T.** The G-Quartet in Supramolecular Chemistry and Nanoscience. [book auth.] Stephen Neidle

and Shankar Balasubramanian. *Quadruplex Nucleic Acids.* Cambridge : Royal Society of Chemistry, 2006, pp. 253-296.

10. **Parkinson, Gary Nigel.** Fundamentals of Quadruplex Structures. [book auth.] Stephen Neidle and Shankar Balasubramanian. *Quadruplex Nucleic Acids.* Cambridge : RSC Publishing, 2006, 1, pp. 1-30.

11. *G-Quadruplex DNA Structures and Organic Chemistry: More Than One Connection.* **Franceschin, Marco.** 2009, European Journal of Organic Chemistry, Vol. 14, pp. 2225-2238.

12. *Hoogsteen G-G Base Pairing is Dispensable for Telomere Healing in Yeast.* **Lustig, Arthur J.** 1992, Nucleic Acids Research, Vol. 20, pp. 3021-3028.

13. *Human DNA Polymerase I Incorporates dCTP Opposite Template G via a G.C+ Hoogsteen Base Pair.* **Nair, Deepak T, *et al.*** 2005, Structure, Vol. 13, pp. 1569-1577.

14. *The Crystal and Molecular Structure of a Hydrogen-Bonded Complex Between 1-Methylthymine and 9-Methyladenine.* **Hoogsteen, Karst.** 1963, Acta Crystallographica, Vol. 16, pp. 907-916.

15. *Four-Stranded Nucleic Acids: Structure, Function and Targeting of G-Quadruplexes.* **Huppert, Julian Leon.** 2008, Chemical Society Reviews, Vol. 37, pp. 1375-1384.

16. *Structures, Folding Patterns, and Functions of Intramolecular DNA G-quadruplexes found in Eukaryotic Promoter Regions.* **Qin, Yong and Hurley, Laurence H.** 2008, Biochimie, Vol. 90, pp. 1149-1171.

17. *Biological Aspects of DNA/RNA Quadruplexes.* **Shafer, Richard H and Smirnov, Ivan.** 2001, Nucleic Acid Sci, Vol. 56, pp. 209-227.

18. *Guanine Quartet Networks Stabilized by Cooperative Hydrogen Bonds.* **Otero, Roberto, *et al.*** 2005, Angewandte Chemie International Edition, Vol. 44, pp. 2270-2275.

19. *G-quadruplex DNA Structures - Variations on a Theme.* **Simonsson, Tomas.** 2001, Biological Chemistry, Vol. 382, pp. 621-628.

20. *Structures for Polyionosinic Acid and Polyguanylic Acid.* **Arnott, Struther, Chandrasekaran, Rengaswami and Marttila, Constance M.** 1974, Biochemical Journal, Vol. 141, pp. 537-543.

21. *Kinetics of Tetramolecular Quadruplexes.* **Mergny, Jean-Louis, *et al.*** 2005, Nucleic Acids Research, Vol. 33, pp. 81-94.

22. *A Sodium-Potassium Switch in the Formation of Four-Stranded G4-DNA.* **Sen, Dipankar and Gilbert, Walter.** 1990, Nature, Vol. 344, pp. 410-414.

23. *Highly Prevalent Putative Quadruplex Sequence Motifs in Human DNA.* **Todd, Alan K, Johnston, Matthew and Neidle, Stephen.** 2005, Nucleic Acids Research, Vol. 33, pp. 2901-2907.

24. *Prevalence of Quadruplexes in the Human Genome.* **Huppert, Julian L and Balasubramanian, Shankar.** 2005, Nucleic Acids Research, Vol. 33, pp. 2908-2916.

25. *Hunting G-Quadruplexes.* **Huppert, J L.** 2008, Biochimie 90, Vol. 90, pp. 1140-1148.

26. *G-Quadruplexes in Promoters Throughout the Human Genome.* **Huppert, Julian L and Shankar, Balasubramanian.** 2007, Nucleic Acids Research, Vol. 35, pp. 406-413.

27. *Human Telomere, Oncogenic Promoter and 5'-UTR G-Quadruplexes: Diverse Higher Order DNA and RNA Targets for Cancer Therapeutics.* **Patel,**

Dinshaw J, Phan, Anh Tuan and Kuryavyi, Vitaly. 2007, Nucleic Acids Research, Vol. 35, pp. 7429-7455.

28. *Biophysical and Biological Properties of Quadruplex Oligodeoxyribonucleotides.* **Dapic, Virna, *et al.*** 2003, Nucleic Acids Research, Vol. 31, pp. 2097-2107.

29. *Genome-Wide Computational and Expression Analyses Reveal G-Quadruplex DNA Motifs as Conserved cis-Regulatory Elements in Human and Related Species.* **Verma, Anjali, *et al.*** 2008, Journal of Medicinal Chemistry, Vol. 51, pp. 5641-5649.

30. *Intracellular Treanscription of G-rich DNAs Induces Formation of G-loops, Novel Structures Containing G4 DNA.* **Duquette, Michelle L, Handa, Priya and Vincent, Jack A.** 2004, Genes & Development, Vol. 18, pp. 1618-1629.

31. *Greglist: a Database Listing Potential G-Quadruplex Regulated Genes.* **Zhang, Ren, Lin, Yan and Zhang, Chun-Ting.** 2008, Nucleic Acids Research, Vol. 36, pp. 372-376.

32. *Making Sense of G-Quadruplex and i-Motif Functions in Oncogene Promoters.* **Brooks, Tracy A, Kendrick, Samantha and Hurley, Laurence.** 2010, FEBS Journal, Vol. 277, pp. 3459-3469.

33. *Mapping the Sequences of Potential Guanine Quadruplex Motifs.* **Todd, Alan K and Neidle, Stephen.** 2011, Nucleic Acids Research, Vol. 39, pp. 4917-4927.

34. *Influence of Loop Size on the Stability of Intramolecular DNA Quadruplexes.* **Risitano, Antonina and Fox, Keith R.** 2004, Nucleic Acids Research, Vol. 32, pp. 2598-2606.

35. *Hairpin and Parallel Quartet Structures for Telomeric Sequences.* **Balagurumoorthy, P, *et al.*** 1992, Nucleic Acids Research, Vol. 20, pp. 4061-4067.

36. *Loop-Length-Dependent Folding of G-Quadruplexes.* **Hazel, Pascale, *et al.*** 2004, Journal of the American Chemical Society, Vol. 126, pp. 16405-16415.

37. *Analysis of Thermal Melting Curves.* **Mergny, Jean-Louis and Lacroix, Laurent.** 2003, Oligonucleotides, Vol. 13, pp. 515-537.

38. *Thermodynamics of G-Tetraplex Formation by Telomeric DNAs.* **Lu, Min, Guo, Qiu and Kallenbach, Neville R.** 1993, Biochemistry, Vol. 32, pp. 498-601.

39. *Stability and Structure of Telomeric DNA Sequences Forming Quadruplexes Containing Four G-Tetrads with Different Topological Arrangements.* **Petraccone, Luigi, *et al.*** 2004, Biochemistry, Vol. 43, pp. 4877-4884.

40. *Parallel and Antiparallel G-DNA Structures from a Complex Telomeric Sequence.* **Venczel, Eduard A and Sen, Dipankar.** 1993, Biochemistry, Vol. 32, pp. 6220-6228.

41. *DNA Comes in Many Forms.* **Rich, Alexander.** 1993, Gene, Vol. 135, pp. 99-109.

42. *Following G-Quartet Formation by UV-Spectroscopy.* **Mergny, Jean-Louis, Phanb, Anh-Tuan and Lacroix, Laurent.** 1998, FEBS Letters, Vol. 435, pp. 74-78.

43. *Human Telomeric DNA: G-Quadruplex, i-Motif and Watson-Crick Double Helix.* **Phan, Anh Tuan and Mergny, Jean-Louis.** 2002, Nucleic Acids Research, Vol. 30, pp. 4618-4625.

44. *Human Telomeric G-Quadruplex: Thermodynamic and Kinetic Studies of Telomeric Quadruplex Stability.* **Chaires, Jonathan B.** 2010, FEBS Journal, Vol. 277, pp. 1098-1106.

45. *Evaluation of Human Telomeric G-Quadruplexes: The Influence of Overhanging Sequences on Quadruplex Stability and Folding.* **Viglasky, Viktor, *et al.*** 2010, Journal of Nucleic Acids, Vol. 1.

46. *G-Quadruplex Structures: in vivo Evidence and Function.* **Lipps, Hans J and Rhodes, Daniela.** 2007, Trends in Cell Biology, Vol. 19, pp. 414-422.

47. *Human Telomeric G-Quadruplex: Structures of DNA and RNA Sequences.* **Phan, Anh Tuan.** 2010, FEBS Journal, Vol. 277, pp. 1107-1117.

48. *Human Telomeric G-Quadruplex.* **Maiti, Souvik.** 2010, FEBS Journal, Vol. 277, p. 1097.

49. *Four-Stranded DNA: Cancer, Gene Regulation and Drug Development.* **Huppert, Jonathan L.** 2007, Philosophical Transactions of Mathematical Physical & Engineering Sciences, Vol. 365, pp. 2969-2984.

50. *Clinical Utility of Telomerase in Cancer.* **Hiyama, Eiso and Hiyama, Keiko.** 2002, Oncogene, Vol. 21, pp. 643-649.

51. *Human Telomeric G-Quadruplex: The Current Status of Telomeric G-Quadruplexes as Therapeutic Targets in Human Cancer.* **Neidle, Stephen.** 2010, FEBS Journal, Vol. 277, pp. 1118-1125.

52. *Physiological Relevance of Telomeric G-Quadruplex Formation: A Potential Drug Target.* **Oganesian, Liana and Bryan, Tracy M.** 2007, Bioessays, Vol. 29, pp. 155-165.

53. *Targeting the Limitless Replicative Potential of Cancer: The Telomerase/Telomere Pathway.* **Kelland, Lloyd.** 2007, Clinical Cancer Research, Vol. 13, pp. 4960-4963.

54. *Targeting Telomeres and Telomerase.* **De Cian, Anne, *et al.*** 2007, Biochimie, Vol. 90, pp. 131-155.

55. *Altered Gene Expression in the Werner and Bloom Syndromes is Associated with Sequences Having G-Quadruplex Forming Potential.* **Johnson, Jay E, *et al.*** 2010, Nucleic Acids Research, Vol. 38, pp. 1114-1112.

56. *DNA Helicase Deficiencies Associated with Cancer Predisposition and Premature Ageing Disorders.* **Mohaghegh, Payam and Hickson, Ian D.** 2001, Human Molecular Genetics, Vol. 10, pp. 741-746.

57. *WRN at Telomeres: Implications for Aging and Cancer.* **Multani, Asha S and Chang, Sandy.** 2007, Journal of Cell Science, Vol. 120, pp. 713-721.

58. *Functions of RecQ Family Helicases: Possible Involvement of Bloom's and Werner's Syndrome Gene Products in Guarding Genome Integrity During DNA Replication.* **Enomoto, T.** 2001, Journal of Biochemistry, Vol. 129, pp. 501-507.

59. *Werner Syndrome, Aging and Cancer.* **Ozgenc, A and Loeb, L A.** 2006, Genome Dynamics, Vol. 1, pp. 206-217.

60. *Gene Expression Profiling in Werner Syndrome Closely Resembles that of Normal Aging.* **Kyng, K J, *et al.*** 2003, Proceeding of the National Academy of Sciences of the United States of America, Vol. 100, pp. 12259-12264.

61. *Telomere Replication: Poised but Puzzling.* **Sampathi, Shilpa and Chai, Weihang.** 2011, Journal of Cellular and Molecular Medicine, Vol. 15, pp. 3-13.

62. *Molecular Genetics of Bloom's Syndrome.* **Ellis, N A and German, J.** 1996, Human Molecular Genetics, Vol. 5, pp. 1457-1463.

63. *Central Dogma of Molecular Biology.* **Crick, Francis.** 1970, Nature, Vol. 227, pp. 561-563.

64. *Systematic Discovery of Regulatory Motifs in Human Promoters and 3' UTRs by Comparison of Several Mammals.* **Xie, Xiaohui, *et al.*** 2005, Nature, Vol. 434, pp. 338-345.

65. *The Relationship of Potential G-Quadruplex Sequences in cis-Upstream Regions of the Human Genome to SP1-Binding Elements.* **Todd, Alan K and Neidle, Stephen.** 2008, Nucleic Acids Research, Vol. 36, pp. 2700-2704.

66. *Conserved Elements with Potential to form Polymorphic G-Quadruplex Structures in the First Intron of Human Genes.* **Eddy, Johanna and Maizels, Nancy.** 2008, Nucleic Acids Research, Vol. 36, pp. 1321-1333.

67. *Solution Structure of the Biologically Relevant G-Quadruplex Element in the Human c-MYC Promoter. Implications for G-Quadruplex Stabilization.* **Ambrus, Attila, *et al.*** 2005, Biochemistry, Vol. 44, pp. 2048-2058.

68. *NMR Solution Structure of the Major G-Quadruplex Structure Formed in the Human BCL2 Promoter Region.* **Dai, Jixun, *et al.*** 2006, Nucleic Acids Research, Vol. 34, pp. 5133-5144.

69. *Solution Structures of all Parallel-Stranded Monomeric and Dimeric G-Quadruplex Scaffolds of the Human c-kit2 Promoter.* **Kuryavyi, Vitaly, Phan, Anh Tuan and Patel, Dinshaw J.** 2010, Nucleic Acids Research, Vol. Advance Access, pp. 1-17.

70. *Effect of Flanking Bases on Quadrplex Stability and Watson-Crick Duplex Competition.* **Arora, Amit, Nair, Divya and Maiti, Souvik.** 2009, FEBS Journal, Vol. 276, pp. 3628-3640.

71. *Dynamics Approach to DNA Breathing.* **Metzler, Ralf and Ambjornsson, Tobias.** 2005, Journal of Biological Physics, Vol. 31, pp. 339-350.

72. *Breathing Dynamics in Heteropolymer DNA.* **Ambjornsson, Tobias, *et al.*** 2007, Biophysical Journal, Vol. 92, pp. 2674-2684.

73. *Spectroscopic Studies of Position-Specific DNA "Breathing" Fluctuations at Replication Forks and Primer-Template Junctions.* **Jose, Davis, *et al.*** 2009, Proceedings of the National Academy of Science, Vol. 106, pp. 4231-4236.

74. *The i-Motif in Nucleic Acids.* **Gueron, Maurice and Leroy, Jean-Louis.** 2000, Current Opinion in Structural Biology, Vol. 10, pp. 326-331.

75. *Human Telomeric DNA: G-Quadruplex, i-Motif and Watson-Crick Double Helix.* **Phan, Anh Tuan and Mergny, Jean-Louis.** 2002, Nucleic Acids Research, Vol. 30, pp. 4618-4625.

76. *Molecular Crowding Regulates the Structural Switch of the DNA G-Quadruplex.* **Miyoshi, Daisuke, Nakao, Akihiro and Sugimoto, Naoki.** 2002, Biochemistry, Vol. 41, pp. 15017-15024.

77. *Human Telomeric G-Quadruplexes Undergo Dynamic Conversion in a Molecular Crowding Environment.* **Xu, Liang, Feng, Shuo and Zhou, Xiang.** 2011, Chemical Communications, Vol. 47, pp. 3517-3519.

78. *Role of Molecular Crowding in Perturbing Quadruplex-Watson Crick Duplex Equilibrium.* **Kumar, Niti and Maiti, Souvik.** 2008, Nucleic Acids Symposium Series, Vol. 52, pp. 157-158.

79. *Duplex Dissociation of Telomere DNAs Induced by Molecular Crowding.* **Miyoshi, Daisuke, *et al.*** 2004, Journal of the American Chemical Society, Vol. 126, pp. 165-169.

80. *Molecular Crowding Induces Telomere G-Quadruplex Formation under Salt-Deficient Conditions and Enhances its Competition with Duplex Formation.* **Kan, Zhong-yuan, *et al.*** 2006, Angewandte Chemie International Edition, Vol. 45, pp. 1629-1632.

81. *Effect of Loop Length Variation on Quadruplex-Watson Crick Duplex Competition.* **Kumar, Niti, *et al.*** 2008, Nucleic Acids Research, Vol. 36, pp. 4433-4442.

82. *Stability of Intramolecular Quadruplexes: Sequence Effects in the Central Loop.* **Guedin, Aurore, Alberti, Patrizia and Mergny, Jean-Louis.** 2009, Nucleic Acids Research, Vol. 37, pp. 5559-5567.

83. *Sequence Effects in Single-Base Loops for Quadruplexes.* **Guedin, Aurore, *et al.*** 2008, Biochimie, Vol. 90, pp. 686-696.

84. *Loop Residues of Thrombin-Binding DNA Aptamer Impact G-Quadruplex Stability and Thrombin Binding.* **Nagatoishi, Satoru, *et al.*** 2011, Biochimie, Vol. 93, pp. 1231-1238.

85. *Welcome the Family of FANCJ-like Helicases to the Block of Genome Stability Maintenace Proteins.* **Wu, Y, Suhasini, A N and Brosh, R M.** 2009, Cellular and Molecular Life Sciences, Vol. 66, pp. 1209-1222.

86. *FANCJ Helicase Defective in Fanconia Anemia and Breast Cancer Unwinds G-Quadruplex DNA to Defend Genomic Stability.* **Wu, Yuliang, Shin-ya, Kazuo and Brosh, Robert M.** 2008, Molecular and Cellular Biology, Vol. 28, pp. 4116-4128.

87. *Disruption of Dog-1 in Caenorhabditis elegans Triggers Deletions Upstream of Guanine-Rich DNA.* **Cheung, Iris, *et al.*** 2002, Nature Genetics, Vol. 31, pp. 405-409.

88. *Spectrum of Mutational Events in the Absence of DOG-1/FANCJ in Caenorhabditis elegans.* **Zhao, Yang, *et al.*** 2008, DNA Repair, Vol. 7, pp. 1846-1854.

89. *Mutagenic Capacity of Endogenous G4 DNA Underlies Genome Instability in FANCJ-Defective C. elegans.* **Kruisselbrink, Evelien, *et al.*** 2008, Current Biology, Vol. 18, pp. 900-905.

90. *The Emerging Genetic and Molecular Basis of Fanconi Anaemia.* **Joenje, Hans and Patel, J.** 2001, Nature Reviews Genetics, Vol. 2, pp. 446-457.

91. *FANCJ is a Structure-Specific DNA Helicase Associated with the Maintenance of Genomic G/C Tracts.* **London, Timothy B C, *et al.*** 2008, The Journal of Biological Chemistry, Vol. 283, pp. 36132-36139.

92. *Structural Polymorphism Within a Regulatory Element of the Human KRAS Promoter: Formation of G4-DNA Recognized by Nuclear Proteins.* **Cogoi, Susanna, *et al.*** 2008, Nucleic Acids Research, Vol. 36, pp. 3765-3780.

93. *Protein hnRNP A1 and its Derivative Up1 Unfold Quadruplex DNA in the Human KRAS Promoter: Implications for Transcription.* **Paramasivam, M, *et al.*** 2009, Nucleic Acids Research, Vol. 37, pp. 2841-2853.

94. *DNA Replication Through G-Quadruplex Motifs is Promoted by the Saccharomyces cerevisiae Pif1 DNA Helicase.* **Paeschke, Katrin, Capra, John A and Zakian, Virginia A.** 2011, Cell, Vol. 145, pp. 678-691.

95. *Human Pif1 Helicase is a G-Quadruplex DNA-Binding Protein With G-Quadruplex DNA-Unwinding Activity.* **Sanders, C M.** 2010, Biochemical Journal, Vol. 15, pp. 119-128.

96. *The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences In Vivo.* **Ribeyre, Cyril, *et al.*** 2009, PLoS Genetics, Vol. 5, p. e1000475.

97. *Genomic Distribution and Functional Analyses of Potential G-Quadruplex-Forming Sequences in Saccharomyces cerevisiae.* **Hershman, Stebe G, *et al.*** 2008, Nucleic Acids Research, Vol. 36, pp. 144-156.

98. *Retinoblastome Susceptibility Genes Contain 5' Sequences with a High Propensity to Form Guanine-Tetrad Structures.* **Murchie, Alastair I H and Lilley, David M.** 1992, Nucleic Acids Research, Vol. 20, pp. 49-53.

99. *Unusual DNA Structure of the Diabetes Susceptibility Locus IDDM2 and Its Effect on Transcription by the Insulin Promoter Factor Pur-1/MAZ.* **Lew, Amy, Rutter, William J and Kennedy, Giulia C.** 2000, Proceedings of National Academy of the United States of America, Vol. 97, pp. 12508-12512.

100. *Facilitation of a Structural Transition in the Polypurine/Polypyrimidine Tract Within the Proximal Promoter Region of the Human VEGF Gene by the*

Presence of Potassium and G-Quadruplex-Interactive Agents. **Sun, Daekyu, *et al.*** 2005, Nucleic Acids Research, Vol. 33, pp. 6070-6080.

101. *The Fragile X Syndrome d(CGG)n Nucleotide Repeats Form a Stable Tetrahelical Structure.* **Fry, Michael and Loeb, Lawrence A.** 1994, Proceedings of the National Academy of Sciences of the United States fo America, Vol. 91, pp. 4950-4954.

102. *Fragile X Mental Retardation Protein Targets G Quartet mRNAs Important for Neuronal Function.* **Darnell, Jennifer C, *et al.*** 2001, Cell, Vol. 107, pp. 489-499.

103. *The Tetraplex (CGG)n Destabilizing Proteins hnRNP A2 and CBF-A Enhance the in vivo Translation of Fragile X Premutation mRNA.* **Khateb, Samer, *et al.*** 2007, Nucleic Acids Research, Vol. 35, pp. 5775-5788.

104. *Gene Function Correlates with Potential for G4 DNA Formation in the Human Genome.* **Eddy, Johanna and Maizels, Nancy.** 2006, Nucleic Acids Research, Vol. 34, pp. 3887-3896.

105. *Formation of Pseudosymmetrical G-Quadruplex and i-Motif Structures in the Proximal Promoter Region of the RET Oncogene.* **Guo, Kexiao, *et al.*** 2007, Journal of the American Chemical Society, Vol. 129, pp. 10220-10228.

106. *G-Quadruplex Formation Within the Promoter of the KRAS Proto-Oncogene and its Effect on Transcription.* **Cogoi, Susanna and Xodo, Luigi E.** 2006, Nucleic Acids Research, Vol. 34, pp. 2536-2549.

107. *The Hallmarks of Cancer.* **Hanahan, Douglas and Weinberg, Robert A.** 2000, Cell, Vol. 100, pp. 57-70.

108. *G-quadruplex Nucleic Acids as Therapeutic Targets.* **Balasubramanian, Shankar and Neidle, Stephen.** 2009, Current Opinion in Chemical Biology, Vol. 13, pp. 345-353.

109. *The Role of Supercoiling in Transcriptional Control of MYC and its Importance in Molecular Therapeutics.* **Brooks, Tracy A and Hurley, Laurence H.** 2009, Nature Reviews Cancer, Vol. 9, pp. 849-861.

110. *myc Function and Regulation.* **Marcu, Kenneth B, Bossone, Steven A and Patel, Amanda J.** 1992, Annual Reviews Biochemistry, Vol. 61, pp. 809-858.

111. *Action of Myc in vivo - Proliferation and Apoptosis.* **Pelengaris, Stella, Rudolph, Bettina and Littlewood, Trevor.** 2000, Current Opinion in Genetics & Development, Vol. 10, pp. 100-105.

112. *Control of c-myc Regulation in Normal and Neoplastic Cells.* **Spencer, Charlotte A and Groudine, Mark.** 1991, Advances in Cancer Research, Vol. 56, pp. 1-48.

113. **Dexheimer, Thomas S, Fry, Michael and Hurley, Laurence H.** DNA Quadruplexes and Gene Regulation. [book auth.] Stephen Neidle and Shankar Balasubramanian. *Quadruplex Nucleic Acids.* Cambridge : Royal Society of Chemistry, 2006, pp. 180-207.

114. *Chromatin Structure and Protein Binding in the Putative Regulatory Region of the c-myc Gene in Burkitt Lymphoma.* **Siebenlist, Ulrich, *et al.*** 1984, Cell, Vol. 37, pp. 381-391.

115. *PuF/NM23-H2/NDPK-B Transactivates a Human c-myc Promoter-CAT Gene Via a Functional Nuclease Hypersensitive Element.* **Berberich, S J and Postel, E H.** 1995, Oncogene, Vol. 10, pp. 2343-2347.

116. *Ribonucleoprotein and Protein Factors Bind to an H-DNA-Forming c-myc DNA Element: Possible Regulators of the c-myc Gene.* **Davis, Terri L, Firulli, Anthony B and Kinniburgh, Alan J.** 1989, Proceedings of the National Academy of Sciences of the United States of America, Vol. 86, pp. 9682-9686.

117. *DNA Tetraplex Formation in the Control Region of c-myc.* **Simonsson, Tomas, Kubista, Mikael and Pecinka, Petr.** 1998, Nucleic Acids Research, Vol. 26, pp. 1167-1172.

118. *Direct Evidence for a G-Quadruplex in a Promoter Region and its Targeting with a Small Molecule to repress c-MYC Transcription.* **Siddiqui-Jain, Adam, *et al.*** 2002, Proceedings of the National Academy of Sciences of the United States of America, Vol. 99, pp. 11593-11598.

119. *The Dynamic Character of the G-Quadruplex Element in the c-MYC Promoter and Modification by TMPyP4.* **Seenisamy, Jeyaprakashnarayanan, *et al.*** 2004, Journal of the American Chemical Society, Vol. 126, pp. 8702-8709.

120. *Propeller-Type Parallel-Stranded G-Quadruplexes in the Human c-myc Promoter.* **Phan, Anh Tuan, Modi, Yasha S and Patel, Dinshaw J.** 2004, Journal of the American Chemical Society, Vol. 126, pp. 8710-8716.

121. *Multiple Single-Stranded cis Elements are Associated with Activated Chromatin of the Human c-myc Gene In Vivo.* **Michelotti, Gregory A, *et al.*** 1996, Molecular and Cellular Biology, Vol. 16, pp. 2656-2669.

122. *The Cationic Porphyrin TMPyP4 Down-Regulates c-MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth in Vivo.* **Grand, Cory L, *et al.*** 2002, Molecular Cancer Therapeutics, Vol. 1, pp. 565-573.

123. *Metastases Supressor NM32-H2 Interaction with G-Quadruplex DNA within c-MYC Promoter Nuclease Hypersensitive Element Induces c-MYC Expression.* **Thakur, Ram Krishna, *et al.*** 2009, Nucleic Acids Research, Vol. 37, pp. 172-183.

124. *Antitumor Activity of G-Quadruplex-Interactive Agent TMPyP4 in K562 Leukemic Cells.* **Mikami-Terao, Yoko, *et al.*** 2008, Cancer Letters, Vol. 261, pp. 226-234.

125. *Structure of an Unprecidented G-Quadruplex Scaffold in the Human c-kit Promoter.* **Phan, Anh Tuan, *et al.*** 2007, Journal of the American Chemical Society, Vol. 129, pp. 4386-4392.

126. *G-Quadruplex-Binding Benzo[a]phenoxazines Down-Regulate c-Kit Expression in Human Gastric Carcinoma Cells.* **McLuckie, Keith I.E., *et al.*** 2011, Journal of the American Chemical Society, Vol. 133, pp. 2658-2663.

127. *Sequence Occurrence and Structural Uniqueness of a G-Quadruplex in the Human c-Kit Promoter.* **Todd, Alan K, *et al.*** 2007, Nucleic Acids Research, Vol. 35, pp. 5799-5808.

128. *G-Quadruplex-Binding Benzo[a]phenoxazines Down-Regulate c-Kit Expression in Human Gastric Carcinoma Cells.* **Mc-Luckie, Keith I.E., *et al.*** 2011, Journal of the American Chemical Society, Vol. 133, pp. 2658-2663.

129. *A Comparison of DNA and RNA Quadruplex Structures and Stabilities.* **Joachimi, Astrid, Benz, Armin and Hartig, Jorg S.** 2009, Bioorganic & Medicinal Chemistry, Vol. 17, pp. 6811-6815.

130. *RNA Quadruplex-Based Modulation of Gene Expression.* **Wieland, Markus and Hartig, Jorg S.** 2007, Chemistry and Biology, Vol. 14, pp. 757-763.

131. *An RNA G-Quadruplex in the 5' UTR of the NRAS Proto-Oncogene Modulates Translation.* **Kumari, Sunita, *et al.*** 2007, Nature Chemical Biology, pp. 218-221.

132. *Inhibition of Translation in Living Eukaryotic Cells by an RNA G-Quadruplex Motif.* **Arora, Amit, *et al.*** 2008, RNA, Vol. 14, pp. 1290-1296.

133. *Predictable Suppression of Gene Expression by 5'-UTR-Based RNA Quadruplexes.* **Halder, Kangkan, Wieland, Markus and Hartig, Jorg S.** 2009, Nucleic Acids Research, Vol. 37, pp. 6811-6817.

134. *Small Molecule-Mediated Inhibition of Translation by Targeting a Native RNA G-Quadruplex.* **Bugaut, Anthony, *et al.*** 2010, Organic & Biomolecular Chemistry, Vol. 8, pp. 2771-2776.

135. *Telomeric Repeat-Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends.* **Azzalin, Claus M, *et al.*** 2007, Science, Vol. 318, pp. 798-801.

136. *G-Quadruplexes: The Beginning and End of UTRs.* **Huppert, Julian Leon, *et al.*** 2008, Nucleic Acids Research, Vol. 36, pp. 6260-6268.

137. *The BCL-2 5' Untranslated Region Contains an RNA G-Quadruplex-Forming Motif that Modulates Protein Expression.* **Shahid, Ramla, Bugaut, Anthony and Balasubramanian, Shankar.** 2010, Biochemistry, Vol. 49, pp. 8300-8306.

138. *G-Quadruplex Structures in TP53 Intron 3: Role in Alternative Splicing and in Production of p53 mRNA Isoforms.* **Marcel, Virginie, *et al.*** 2010, Carcinogenesis, Vol. 32, pp. 271-278.

139. *Telomerase Downregulation Induced by the G-Quadruplex Ligand 12459 in A549 Cells is Mediated by hTERT RNA Alternative Splicing.* **Gomez, Dennis, *et al.*** 2004, Nucleic Acids Research, Vol. 32, pp. 371-379.

140. *Structure Monomorphism of RNA G-Quadruplex that is Independent of Surrounding Condition.* **Zhang, Dong-Hao and Zhi, Gao-Ying.** 2010, Journal of Biotechnology, Vol. 150, pp. 6-10.

141. *A Hitchhiker's Guide to G-Quadruplex Ligands.* **Monchaud, David and Teulade-Fichou, Marie-Paule.** 2008, Organic & Biomolecular Chemistry, Vol. 6, pp. 627-636.

142. **Riou, Jean-Francois, *et al.*** Quadruplex Ligand Recognition: Biological Aspects. [book auth.] Stephen Neidle and Shankar Balasubramanian.

Quadruplex Nucleic Acids. Cambridge : Royal Society of Chemistry, 2006, pp. 154-179.

143. *Targeting G-Quadruplex Structure in the Human c-Kit Promoter with Short PNA Sequences.* **Amato, Jussara, *et al.*** 2011, Bioconjugate Chemistry, Vol. 22, pp. 654-663.

144. *Promoting the Formation and stabilization of Human Telomeric G-Quadruplex DNA, Inhibition of Telomerase and Cytotoxicity by Phenanthroline Derivatives.* **Wang, Lihua, *et al.*** 2011, Organic & Biomolecular Chemistry, Vol. 9, pp. 2648-2653.

145. *Activity of a Novel G-Quadruplex-Interactive Telomerase Inhibitor, telomestatin (SOT-095), Against Human Leukemia Cells: Involvement of ATM-dependent DNA Damage Response Pathways.* **Tauchi, Tetsuzo, *et al.*** 2003, Oncogene, Vol. 22, pp. 5338-5347.

146. *Telomere Damage Induced by the G-Quadruplex Ligand RHPS4 has an Antitumor Effect.* **Salvati, Erica, *et al.*** 2007, The Journal of Clinical Investigation, Vol. 117, pp. 3236-3247.

147. *Senescence and Telomere Shortening Induced by Novel Potent G-Quadruplex Interactive Agents, Quindoline Derivatives, in Human Cancer Cell Lines.* **Zhou, J-M, *et al.*** 2006, Oncogene, Vol. 25, pp. 503-511.

148. *Telomerase Inhibition with a Novel G-Quadruplex-Interactive Agent, Telomestatin: in vitro and in vivo Studies in Acute Leukemia.* **Tauchi, T, *et al.*** 2006, Oncogene, Vol. 25, pp. 5719-5725.

149. *Quinolino-benzo-[5,6]-dihydroisoquindolium Compounds Derived from Berbine: A New Class of Highly Selective Ligands for G-Quadruplex DNA in c-Myc Oncogene.* **Ma, Yan, *et al.*** 2011, European Journal of Medicinal Chemistry, Vol. 46, pp. 1906-1913.

150. *Trisubstituter Isoalloxazines as a New Class of G-Quadruplex Binding Ligands: Small Molecule Regulation of c-kit Oncogene Expression.* **Bejugam, Mallesham, *et al.*** 2007, Journal of the American Chemical Society, Vol. 129, p. 12926.

151. **Drygin, Denis, *et al.*** Quarfloxin (CX-3543) Disrupts the Nucleolin/rDNA Quadruplex Complexes, Inhibits the Elongation by RNA Polimerase I and Exhibits Potent Antitumor Activity in Models of Cancer. [Online] April 2008. [Cited: 03 12 2010.] http://www.cylenepharma.com/pdf/Poster3-QuarfloxinNucleolinQuadruplexAACR4-08.pdf.

152. *Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis.* **Drygin, Denis, *et al.*** 2009, Cancer Research, Vol. 69, pp. 7653-7661.

153. *DNA and RNA Quadruplex Ligands.* **De Cian, Anne, *et al.*** 2008, Nucleic Acids Symposium Series, Vol. 52, pp. 7-8.

154. *V-Shaped Scaffold: A New Architectural Motif Identified in an A*(G*G*G*G) Pentad-Containing Dimeric DNA Quadruplex Involving Stacked G(anti)*G(anti)*G(anti)*G(syn) Tetrads.* **Zhang, Na, *et al.*** 2001, Journal of Molecular Biology, Vol. 311, pp. 1063-1079.

155. *Dimeric DNA Quadruplex Containing Major Groove-Aligned A*T*A*T and G*C*G*C Tetrads Stabilized by Inter-Subunit Watson-Crick A*T and G*C Pairs.* **Zhang, Na, *et al.*** 2001, Journal of Molecular Biology, Vol. 312, pp. 1073-1088.

156. *Structure of the Parallel-Stranded DNA Quadruplex d(TTAGGGT)4 Containing the Human Telomeric Repeat: Evidence for A-Tetrad Formation from NMR and Molecular Dynamics Simulations.* **Gavathiotis, Evripidis and Searle, Mark S.** 2003, Organic & Biomolecular Chemistry, Vol. 1, pp. 1650-1656.

157. *Association of DNA Quadruplexes Through G:C:G:C Tetrads. Solution Structure of d(GCGGTGGAT).* **da Silva, Mateus Webba.** 2003, Biochemistry, Vol. 42, pp. 14356-14365.

158. *Experimental Demonstration of T:(G:G:G:G):T Hexad and T:A:A:T Tetrad Alignments within a DNA Quadruplex Stem.* **da Silva, Mateus Webba.** 2005, Biochemistry, Vol. 44, pp. 3754-3764.

159. *NMR Observation of T-Tetrads in a Parallel Stranded DNA Quadruplex Formed by Saccharomyces Cerevisiae Telomere Repeats.* **Patel, P K and Hosur, R V.** 1999, Nucleic Acids Research, Vol. 27, pp. 2457-2464.

160. *NMR Observation of a Novel C-Tetrad in the Structure of the SV40 Repeat Sequence GGGCGG.* **Patel, P K, Bhavesh, Neel S and Hosur, R V.** 2000, Biochemical and Biophysical Research Communications, Vol. 270, pp. 967-971.

161. *Crystal Structure of an RNA Purine-Rich Tetraplex Containing Adenine Tetrads: Implications for Specific Binding in RNA Tetraplexes.* **Pan, Baocheng, *et al.*** 2003, Structure, Vol. 11, pp. 815-823.

162. *Solution Structure of an Unusually Stable RNA Tetraplex Containing G- and U-Quartet Structures.* **Cheong, Chaejoon and Moore, Peter B.** 1992, Biochemistry, Vol. 31, pp. 8406-8414.

163. *Solution Structure of a Na Cation Stabilized DNA Quadruplex Containing G*G*G*G and G*C*G*C Tetrads Formed by G*G*G*C Repeats Observed in Adeno-Associated Viral DNA.* **Kettani, Abdelali, *et al.*** 1998, Journal of Molecular Biology, Vol. 282, pp. 619-636.

164. **Mergny, Jean-Louis, *et al.*** Energetics, Kinetics and Dynamics of Quadruplex Folding. [book auth.] Stephen Neidle and Shankar Balasubramanian. *Quadruplex Nucleic Acids.* Cambridge : Royal Society of Chemistry, 2006, pp. 31-80.

165. *Dramatic Effect of Single-Base Mutation on the Conformational Dynamics of Human Telomeric G-Quadruplex.* **Lee, Ja Yil and Kim, D S.** 2009, Nucleic Acids Research, Vol. 37, pp. 3625-3634.

166. *Structure-Function Correlations of the Insulin-Linked Polymorphic Region.* **Catasti, Paolo, *et al.*** 1996, Journal of Molecular Biology, Vol. 264, pp. 534-545.

167. *Mutations in the G-Quadruplex Silencer Element and Their Relationship to c-MYC Overexpression, NM23 Repression, and Therapeutic Rescue.* **Grand, Cory L, *et al.*** 2004, Proceedings of the National Academy of Sciences of the United States of America, Vol. 101, pp. 6140-6145.

168. *Quadruplex-duplex Competition in the Nuclease Hypersensitive Element of Human c-Myc Promoter: C to T Mutation in C-rich Strand Enhances Duplex Association.* **Halder, Kangkan, *et al.*** 2005, Biochemical and Biophysical Research Communications, Vol. 327, pp. 49-56.

169. *A PNA4 Quadruplex.* **Krishnan-Ghosh, Yamuna, Stephens, Elaine and Balasubramanian, Shankar.** 2004, Journal of The American Chemical Society, Vol. 126, pp. 5944-5945.

170. *NMR Solution Structure of a Parallel LNA Quadruplex.* **Randazzo, Antonio, *et al.*** 2004, Nucleic Acids Research, Vol. 32, pp. 3083-3092.

171. *A Conformationally Constrained Nucleotide Analogue Controls the Folding Topology of a DNA G-Quadruplex.* **Dominick, Pamela K and Jarstfer, Michael B.** 2004, Journal of the American Chemical Society, Vol. 126, pp. 5050-5051.

172. *Effect of O6-Methylguanine on the Stability of G-Quadruplex DNA.* **Mekmaysy, Chongkham S, *et al.*** 2008, Journal of the American Chemical Society, Vol. 130, pp. 6710-6711.

173. *Effects of 8-Methlguanine on Structure, Stability and Kinetics of Formation of Tetramolecular Quadruplexes.* **Tran, Phong Lan Thao, *et al.*** 2010, Biochimie, Vol. Article In Press, pp. 1-10.

174. *Effect of a Modified Thymine on the Structure and Stability of [d(TGGGT)]4 Quadruplex.* **Petraccone, Luigi, *et al.*** 2003, International Journal of Biological Macromolecules, Vol. 31, pp. 131-137.

175. *The Epigenetic Progenitor Origin of Human Cancer.* **Feinberg, Andrew P, Ohlsson, Rolf and Henikoff, Steven.** 2006, Nature Reviews Genetics, Vol. 7, pp. 21-33.

176. *Epigenetic Changes in Cancer.* **Gronbaek, Kirsten, Hother, Christoffer and Jones, Peter A.** 2007, APMIS, Vol. 115, pp. 1039-1059.

177. *Genome-Wide Analysis Predicts DNA Structural Motifs as Nucleosome Exclusion Signals.* **Halder, K, Halder, R and Chowdhury, S.** 2009, Molecular Biosystems, Vol. 5, pp. 1703-1712.

178. *DNA Secondary Structures and Epigenetic Determinants of Cancer Genome Evolution.* **De, Subhajyoti and Michor, Franziska.** 2011, Nature Structural and Molecular Biology, p. Advanced online publication.

179. *The Effect of Chemical Modifications on the Thermal Stability of Different G-Quadruplex-Forming Oligonucleotides.* **Sacca, Barbara, Lacroix, Laurent and Mergny, Jean-Louis.** 2005, Nucleic Acids Research, Vol. 33, pp. 1182-1192.

180. *Guanines are a Quartet's Best Friend: Impact of Base Substitutions on the Kinetics and Stability of Tetramolecular Quadruplexes.* **Gros, Julien, *et al.*** 2007, Nucleic Acids Research, Vol. 35, pp. 3064-3075.

181. *Hydroperoxide Metabolism in Mammalian Organs.* **Chance, Britton, Sies, Helmut and Boveris, Alberto.** 1979, Physiological Reviews, Vol. 59, pp. 527-605.

182. *DNA Damage by Oxygen-Derived Species. Its Mechanism and Measurement in Mammalian Systems.* **Halliwell, B and Aruoma, O I.** 1991, FEBS Lett, Vol. 281, pp. 9-19.

183. *Oxidative DNA Damage: Mechanisms, Mutation, and Disease.* **Cooke, Marcus S, *et al.*** 2003, Journal of the Federation of American Societies for Experimental Biology, Vol. 17, pp. 1195-1214.

184. *Molecular Bases of Progeroid Syndromes.* **Navarro, Claire L, Cau, Pierre and Levy, Nicolas.** 2006, Human Molecular Genetics, Vol. 15, pp. 151-161.

185. *Base Excision Repair of Oxidative DNA Damage and Association with Cancer and Aging.* **Maynard, Scott, *et al.*** 2009, Carcinogenesis, Vol. 30, pp. 2-10.

186. *Aging and Defense Against Generation fo 8-oxo-7,8-dihydro-2'-deoxyguanosine in DNA.* **Mikkelsen, Lone, *et al.*** 2009, Free Radical Biology & Medicine, Vol. 47, pp. 608-615.

187. *The Transcriptosomal Response of Human A549 Lung Cells to a Hydrogen Peroxide-Generating System: Relationship to DNA Damage, Cell Cycle Arrest, and Caspase Activation.* **Dandrea, Tiziana, *et al.*** 2004, Free Radical Biology & Medicine, Vol. 36, pp. 881-896.

188. *Mechanism of Oxidative DNA Damage Repair and Relevance to Human Pathology.* **D'Errico, Mariarosaria, Parlanti, Eleonora and Dogliotti, Eugenia.** 2008, Mutation Research, Vol. 659, pp. 4-14.

189. *Accumulation of 8-oxo-deoxyguanosine in Cardiovascular Tissues with the Development of Hypertension.* **Ohtsubo, Toshio, *et al.*** 2007, DNA Repair, Vol. 6, pp. 760-769.

190. *Oxidative Nucleotide Damage: Consequences and Prevention.* **Sekiguchi, Mutsuo and Tsuzuki, Teruhisa.** 2002, Oncogene, Vol. 21, pp. 8895-8904.

191. *8-Oxoguanine Enhances Bending of DNA that Favors Binding to Glycosylases.* **Miller, John H, *et al.*** 2003, Journal of the American Chemical Society, Vol. 125, pp. 6331-6336.

192. *Effects of Base Excision Repair Proteins on Mutagenesis by 8-oxo-7,8-dihydroguanine (8-hydroxyguanine) Paired with Cytosine and Adenine.* **Suzuki, Tetsuya, Harashima, Hideyoshi and Kamiya, Hiroyuki.** 2010, DNA Repair, Vol. 9, pp. 542-550.

193. *Different DNA Repair Strategies to Combat the Threat from 8-Oxoguanine.* **Russo, Maria Teresa, *et al.*** 2007, Mutation Research, Vol. 614, pp. 69-76.

194. *Oxidation in the Nucleotide Pool, the DNA Damage Response and Cellular Senescence: Defective Bricks Build a Defective House.* **Rai, Priyamvada.** 2010, Mutation Research, Vol. 703, pp. 71-81.

195. *Cellular Response to Oxidative Stress: Signaling for Suicide and Survival.* **Martindale, Jennifer L and Holbrook, Nikki J.** 2002, Journal of Cellular Physiology, Vol. 192, pp. 1-15.

196. *Repression of Gene Expression by Oxidative Stress.* **Morel, Yannick and Barouki, Robert.** 1999, Biochemistry Journal, Vol. 342, pp. 481-496.

197. *Oxidative Stress and Apoptosis in Metal Ion-Induced Carcinogenesis.* **Shi, Honglian, Hudson, Laurie G and Liu, Ke Jian.** 2004, Free Radical Biology & Medicine, Vol. 37, pp. 582-593.

198. *Oxidation of DNA Bases by Tumor Promoter-Activated Processes.* **Frenkel, Krystyna.** 1989, Environmental Health Perspectives, Vol. 81, pp. 45-54.

199. *Repression of Gene Expression by Oxidative Stress.* **Morel, Yannick and Barouki, Robert.** 1999, Biochemistry Journal, Vol. 342, pp. 481-496.

200. *Hydrogen Peroxide-Induced Gene Expression Across Kingdoms: A Comparative Analysis.* **Vandenbroucke, Korneel, *et al.*** 2008, Molecular Biology and Evolution, Vol. 25, pp. 507-516.

201. *Diverse and Specific Gene Expression Responses to Stresses in Cultured Human Cells.* **Murray, John Isaac, *et al.*** 2004, Molecular Biology of the Cell, Vol. 15, pp. 2361-2374.

202. *Global Gene Expression Analysis Reveals Differences in Cellular Responses to Hydroxyl- and Superoxide Anion Radical-Induced Oxidative Stress in Caco-2 Cells.* **Briede, Jacob J, *et al.*** 2010, Toxicological Sciences, Vol. 114, pp. 193-203.

203. *Oxidative Stress and Gene Regulation.* **Allen, R G and Tresini, Maria.** 2000, Free Radical Biology & Medicine, Vol. 28, pp. 463-499.

204. *Biological Consequences of Oxidative Stress-Induced DNA Damage in Saccharomyces cerevisiae.* **Salmon, Tiffany B, *et al.*** 2004, Nucleic Acids Research, Vol. 32, pp. 3712-3723.

205. *Gene Expression After Treatment with Hydrogen Peroxide, Menadione, or t-Butyl Hydroperoxide in Breat Cancer Cells.* **Chuang, Yao-Yu Eric, *et al.*** 2002, Cancer Research, Vol. 62, pp. 6246-6254.

206. *Transient Excess of MYC Activity Can Elicit Genomic Instability and Tumorigenesis.* **Felsher, D W and Bishop, J M.** 1999, Proceedings of the National Academy of Sciences of the United States of America, Vol. 96, pp. 809-860.

207. *c-Myc Can Induce DNA Damage, Increase Reactive Oxygen Specied, and Mitigate p53 Function: A Mechanism for Oncogene-Induced Genetic Instability.* **Vafa, Omid, *et al.*** 2002, Molecular Cell, Vol. 9, pp. 1031-1044.

208. *High Glucose and Hydrogen Peroxide Increase c-Myc and Haeme-Oxygenase 1 mRNA Levels in Rat Pancreatic Islets Without Activating NF kB.* **Elouil, H, *et al.*** 2005, Diabetologia, Vol. 48, pp. 496-505.

209. *Hydrogen Peroxide-Induced Expression of the Proto-Oncogenes, c-jun, c-fos and c-myc in Rabbit Lens Epithelial Cells.* **Li, David Wan-Cheng and Spector, Abraham.** 1997, Molecular and Cellular Biochemistry, Vol. 173, pp. 59-69.

210. *Evidence for Oxidative Activation of c-Myc-Dependent Nuclear Signaling in Human Coronary Smooth Muscle Cells and in Early Lesions of Watanabe Heritable Hyperlipidemic Rabbits: Protective Effects of Vitamin E.* **de Nigris, Filomena, *et al.*** 2000, Circulation, Vol. 102, pp. 2111-2117.

211. *Role of [Ca2+]i in Induction of c-fos, c-jun, and c-myc mRNA in Rat PTE after Oxidative Stress.* **Maki, Atsuhiko, *et al.*** 1992, The Federation of American Societies for Experimental Biology Journal, Vol. 6, pp. 920-924.

212. *Hydrogen Peroxide-Induced Arachidonic Acid Release in L929 Cells: Roles of Src, Protein Kinase C and Cytosolic Phospholipase A2a.* **Taniguchi, T, *et al.*** 2006, European Journal of Pharmacology, Vol. 546, pp. 1-10.

213. *E2F1 and c-Myc Potentiate Apoptosis through Inhibition of NF-kB Activity that Facilitates MnSOD-Mediated ROS Elimination.* **Tanaka, Hirokazu, *et al.*** 2002, Molecular Cell, Vol. 9, pp. 1017-1029.

214. *Antioxidants Prevent Oxidative DNA Damage and Cellular Transformation Elicited by the Over-Expression of c-MYC.* **KC, Sagun, Carcamo, Juan M and Golde, David W.** 2006, Mutation Research, Vol. 593, pp. 64-79.

215. *Sodium Arsenite Induces ROS Generation, DNA Oxidative Damage, HO-1 and c-Myc Proteins, NF-kB Activation and Cell Proliferation in Human Breast Cancer MCF-7 Cells.* **Ruiz-Ramos, Ruben, *et al.*** 2009, Mutation Research, Vol. 674, pp. 109-115.

216. *c-Myc Target Genes Involved in Cell Growth, Apoptosis and Metabolism.* **Dang, Chi V.** 1999, Molecular and Cellular Biology, Vol. 19, pp. 1-11.

217. *c-Myc Phosphorylation is Required for Cellular Response to Oxidative Stress.* **Benassi, Barbara, *et al.*** 2006, Molecular Cell, Vol. 21, pp. 509-519.

218. *Decreased Ability of Cells Overexpressing MYC Proteins to Reduce Peroxide and Hydroperoxides.* **Tuttle, S, *et al.*** 1996, British Journal of Cancer, Vol. 74, pp. 140-144.

219. *Charge Conduction Properties of a Parallel-Stranded DNA G-Quadruplex: Implications for Chromosomal Oxidative Damage.* **Huang, Yu Chuan, *et al.*** 2009, Biochemistry, Vol. 48, pp. 6794-6804.

220. *Electron Transfer in Tetrads: Adjacent Guanines Are Note Hole Traps in G-Quartets.* **Szalai, Veronika A and Thorp, H Holden.** 2000, Journal of The American Chemical Society, Vol. 122, pp. 4524-4525.

221. *Charge Transport in DNA Duplex/Quadruplex Conjugates.* **Delaney, Sarah and Barton, Jacqueline K.** 2003, Biochemistry, Vol. 42, pp. 14159-14165.

222. *Selective Guanine Oxidation by UVB-Irradiation in Telomeric DNA.* **Kawai, Kiyohiko, Fujitsuka, Mamoru and Majima, Tetsuro.** 2005, Chemical Communications, pp. 1476-1477.

223. *Long-Distance Radical Cation Transport in DNA: Horizontal Charge Hopping in a Dimeric Quadruplex.* **Ndlebe, Thabisile and Schuster, Gary B.** 2006, Organic & Biomolecular Chemistry, Vol. 4, pp. 4015-4021.

224. *Effect of Quadruplex Conformation on Radiation-Induced Formation of 8-Hydroxyguanine and Unaltered Base Release in Polyguanylic Acid.* **Grygoryev, D and Zimbrick, J D.** 2010, Radiation Research, Vol. 173, pp. 110-118.

225. **Haugland, Richard P.** *The Handbook: A guide to Fluorescent Probes and Labeling Technologies.* 10th. California : Invitrogen Corp, 2005.

226. *Comparison of the Formation of 8-Hydroxy-2'-deoxyguanosine and Single- and Double-Strand Breaks in DNA Mediated by Fenton Reactions.* **Lloyd, Daniel R, Carmichael, Paul L and Phillips, David H.** 1998, Chemistry Research in Toxicology, Vol. 11, pp. 420-427.

227. *Quantification of Oxidative Single-Base and Intrastrand Cross-Link Lesions in Unmethylated and CpG-Methylated DNA Induced by Fenton-Type Reagents.* **Cao, Huachuan and Wang, Yinsheng.** 2007, Nucleic Acids Research, Vol. 35, pp. 4833-4844.

228. *Molecular self-assembly and Nanochemistry: A Chemical Strategy for the Synthesis of Nanostructures.* **Whitesides, G M, Mathias, J P and Seto, C T.** 1991, Science, Vol. 29, pp. 1312-1319.

229. *Self-Assembling Organic Nanotubes Based on a Cyclic Peptide Architecture.* **Ghadiri, M Reza, *et al.*** 1993, Nature, Vol. 366, pp. 324-327.

230. *A New DNA Nanostructure, the G-Qire, Imaged by Scanning Probe Microscopy.* **Marsh, Thomas C, Vesenka, James and Henderson, Eric.** 1995, Nucleic Acids Research, Vol. 23, pp. 696-700.

231. *Auto-orientation of G-Wire DNA on Mica.* **Vesenka, James, *et al.*** 2007, Colloids and Surfaces B: Biointerfaces, Vol. 58, pp. 256-263.

232. *Functionalization of DNA G-Wires for Patterning and Nanofabrication.* **Lyonnais, Sebastien, *et al.*** 2008, Nucleic Acids Symposium Series, Vol. 52, pp. 689-690.

233. *Artificial G-Qire Switch with 2,2'-Bipyridine Units Responsive to Divalent Metal Ions.* **Miyoshi, Daisuke, *et al.*** 2007, Journal of the American Chemical Society, Vol. 129, pp. 5919-5925.

234. *Thrombin-Binding DNA Aptamer Forms a Unimolecular Quadruplex Structure in Solution.* **Macaya, R F, *et al.*** 1993, Proceedings of the National Academy of Sciences of the United States of America, Vol. 90, pp. 3745-3749.

235. *Thrombin-Binding Aptamer Quadruplex Formation: AFM and Voltammetric Characterization.* **Diculescu, Victor Constantin, *et al.*** 2010, Journal of Nucleic Acids, p. Article In Press.

236. *A Biomimetic Potassium Responsive Nanochannel: G-Quadruplex DNA Conformational Switching in a Synthetic Nanopore.* **Hou, Xu, *et al.*** 2009, Journal of the American Chemical Society, Vol. 131, pp. 7800-7805.

237. *G-Quadruplex-Based DNAzyme for Colorimetric Detection of Cocaine: Using Magnetic Nanoparticles as the Separation and Amplification Element.* **Du, Yan, *et al.*** 2011, Analyst, Vol. Article in Press.

238. *Protein Recognition by an Ensemble of Fluorescent DNA G-Quadruplexes.* **Margulies, David and Hamilton, Andrew D.** 2009, Angewandte Chemie International Edition, Vol. 48, pp. 1771-1774.

239. *Guest Editor's Introduction: Methods for Studying Quadruplex Nucleic Acids.* **Neidle, Stephen.** 2007, Methods, Vol. 43, p. 245.

240. *Quadruplex Melting.* **Rachwal, Phillip A and Fox, Keith R.** 2007, Methods, Vol. 43, pp. 291-301.

241. *Structure and Stability of Sodium and Potassium Complexes of dT4G4 and dT4G4T.* **Lu, Min, Guo, Qiu and Kallenbach, Neville R.** 1992, Biochemistry, Vol. 31, pp. 2455-2463.

242. *Circular Dichroism of Quadruplex DNAs: Applications to Structure, Cation Effects and Ligand Binding.* **Paramasivan, Sattanathan, Rujan, Iulian and Bolton, Philip H.** 2007, Methods, Vol. 43, pp. 324-331.

243. *NMR Methods for Studying Quadruplex Nucleic Acids.* **da Silva, Matues Webba.** 2007, Methods, Vol. 43, pp. 264-277.

244. *Crystallographic Studies of Quadruplex Nucleic Acids.* **Campbell, Nancy H and Parkinson, Gary N.** 2007, Methods, Vol. 43, pp. 252-263.

245. *Crystal Structure of Parallel Quadruplexes From Human Telomeric DNA.* **Gary, N Parkinson, Lee, M P and Neidle, Stephen.** 2002, Nature, Vol. 417, pp. 876-880.

246. *Structure of a Two-G-Tetrad Intramolecular G-Quadruplex Formed by a Variant Human Telomeric Sequence in K+ Solution: Insights Into the Interconversion of Human Telomeric G-Quadruplex Structures.* **Zhang, Zhenjiang, *et al.*** 2010, Nucleic Acids Research, Vol. 38, pp. 1009-1021.

247. *A Practical Guide to Single-Molecule FRET.* **Roy, Rahul, Hohng, Sungchul and Ha, Taekjip.** 2008, Nature Methods, Vol. 5, pp. 507-516.

248. *Probing the Interaction Between two Single Molecules: Fluorescence Resonance Energy Transfer Between a Single Donor and a Single Acceptor.* **Ha, T, *et al.*** 1996, Proceedings of the National Academy of Science, Vol. 93, pp. 6264-6268.

249. *Single Molecule Conformational Analysis of DNA G-quadruplexes.* **Shirude, Pravin S and Balasubramanian, Shankar.** 2008, Biochimie, Vol. 90, pp. 1197-1206.

250. *Protein Folding Studied by Single-Molecule FRET.* **Schuler, Benjamin and Eaton, William A.** 2008, Current Opinion in Structural Biology, Vol. 18, pp. 16-26.

251. *Single-Molecule Conformational Analysis of G-Quadruplex Formation in the Promoter DNA Duplex of the Proto-Oncogene C-Kit.* **Shirude, Pravin S, *et al.*** 2007, Journal of the American Chemical Society, Vol. 129, pp. 7484-7485.

252. *Ensemble and Single Molecule FRET Analysis of the Structure and Unfolding Kinetics of the c-Kit Promoter Quadruplexes.* **Fegan, Adrian, *et al.*** 2010, Chemical Communications, Vol. 46, pp. 946-948.

253. **Leach, Andrew R.** *Molecular Modelling: Principals and Applications.* 2nd. Harlow : Prentice Hall, 2001.

254. **Schlick, Tamar.** *Molecular Modeling and Simulation: An Interdisciplinary Guide.* New York : Springer-Verlag, 2000.

255. **Moore, Elaine.** *Molecular Modelling and Bonding.* Milton Keynes : The Open University, 2002.

256. **Tripos International.** Tripos Bookshelf 7.3. *1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.*

257. *An All Atom Force Field for Simulations of Proteins and Nucleic Acids.* **Weiner, Scott J, *et al.*** 1986, Journal of Computational Chemistry, Vol. 7, pp. 230-252.

258. *A New Force Field for Molecular Mechanical Simulation of Nucleic Acids.* **Weiner, Scott J, *et al.*** 1984, Journal of the American Chemical Society, Vol. 106, pp. 765-784.

259. *How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules?* **Wang, Junmei, Cieplak, Piotr and Kollman, Peter A.** 2000, Journal of Computational Chemistry, Vol. 21, pp. 1049-1074.

260. *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.* **Cornell, Wendy D, *et al.*** 1995, Journal of the American Chemical Society, Vol. 117, pp. 5179-5197.

261. *Molecular Mechanical Models for Organic and Biological Systems Going Beyond the Atom Centered Two Body Additive Approximation.* **Cieplak, Piotr, Caldwell, James and Kollman, Peter.** 2001, Journal of Computational Chemistry, Vol. 22, pp. 1048-1057.

262. **Westhof, E, Rubin-Carrez, C and Fritsch, V.** The Use of Molecular Dynamics Simulations for Modelling Nucleic Acids. [book auth.] Julia M Goodfellow. *Computer Modelling in Molecular Biology.* Weinheim : VCH, 1995, pp. 104-131.

263. *Single Stranded Loops of Quadruplex DNA as Key Benchmark for Testing Nucleic Acids Force Fields.* **Fadrna, Eva, *et al.*** 2009, Journal of Chemical Theory and Computation, Vol. 5, pp. 2514-2530.

264. *Stability and Cations Coordination of DNA and RNA 14-Mer G-Quadruplexes: A Multiscale Computational Approach.* **Pagano, Bruno, *et al.*** 2008, Journal of Physical Chemistry B, Vol. 112, pp. 12115-12123.

265. *Molecular Dynamics and Principal Components Analysis of Human Telomeric Quadruplex Multimers.* **Haider, Shozeb, Parkinson, Gary N and Neidle, Stephen.** 2008, Biophysical Journal, Vol. 95, pp. 296-311.

266. *Molecular Modeling Study of Binding Site Selectivity of TQMP to G-Quadruplex DNA.* **Aixiao, Li, *et al.*** 2010, European Journal of Medicinal Chemistry, Vol. 45, pp. 983-991.

267. *A Molecular Model for Drug Binding to Tandem Repeats of Telomeric G-Quadruplexes.* **Haider, Shozeb M and Neidle, Stephen.** 2009, Biochemical Society Transactions, Vol. 37, pp. 583-588.

268. *Molecular Dynamics Study on the Interactions of Porphyrin with Two Antiparallel Human Telomeric Quadruplexes.* **Li, Ming-Hui, Luo, Quan and Li, Ze-Sheng.** 2010, Journal of Physical Chemistry B, Vol. 114, pp. 6216-6224.

269. *Energetics of the Human Tel-22 Quadruplex-Telomestatin Interaction: A Molecular Dynamics Study.* **Agrawal, Saurabh, Ojha, Rajendra Prasad and Maiti, Souvik.** 2008, Journal of Physical Chemistry B, Vol. 112, pp. 6828-6836.

270. *G-Quadruplex Structure Can Be Stable with Only Some Coordination Sites Being Occupied by Cations: A Six: Nanosecond Molecular Dynamics Study.* **Chowdhury, Shibasish and Bansal, Manju.** 2001, Journal of Physical Chemistry B, Vol. 105, pp. 7572-7578.

271. *A Combined QM and MM Investigation Into Quanine Quadruplexes.* **Clay, Emma H and Gould, Ian R.** 2005, Journal of Molecular Graphics and Modelling, Vol. 24, pp. 138-146.

272. *Molecular Dynamics Simulations and Their Application to Four-Stranded DNA.* **Sponer, Jiri and Spackova, Nad'a.** 2007, Methods, Vol. 43, pp. 278-290.

273. *Force-Induced Unfolding of Human Telomeric G-Quadruplex: A Steered Molecular Dynamics Simularion Study.* **Li, Hui, Cao, En-hua and Gisler, Thomas.** 2009, Biochemical and Biophysical Research Communications, Vol. 379, pp. 70-75.

274. **Case, D A, *et al.*** *AMBER 11.* s.l. : University of California, San Francisco, 2010.

275. **Tripos International.** SYBYL 7.3. *1699 South Hanley Rd., St Louis, Missouri, 63144, USA.*

276. *MacroModel - an Integrated Software System for Modelling Organic and Bioorganic Molecules Using Molecular Mechanics.* **Mohamadi, F, *et al.*** 1990, Journal of Computational Chemistry, Vol. 11, p. 440.

277. *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation.* **Hess, B, *et al.*** 2008, Journal of Chemical Theory and Computation, Vol. 4, pp. 435-447.

278. **van der Spoel, D, *et al.*** *Gromacs User Manual version 4.5.* 2010.

279. *A Smooth Particle Mesh Ewald Method.* **Essmann, Ulrich, *et al.*** 1995, Journal of Chemical Physics, Vol. 103.

280. *Optimized Particle-Mesh Ewald/Multiple-time Step Integration for Molecular Dynamics Simulations.* **Batcho, Paul F, Case, David A and Schlick, Tamar.** 2001, Journal of Chemical Physics, Vol. 115, pp. 4003-4016.

281. *Tetraplex DNA and its Interacting Proteins.* **Fry, M.** 2007, Frontiers in Bioscience, Vol. 12, pp. 4336-4351.

282. *Effect of Loop Length Variation on Quadruplex-Watson Crick Duplex Competition.* **Kumar, Niti, *et al.*** 2008, Nucleic Acids Research, Vol. 36, pp. 4433-4442.

283. *Oscillating Formation of 8-Oxoguanine during DNA Oxidation.* **White, Blanaid, *et al.*** 2003, Journal of the American Chemical Society, Vol. 125, pp. 6604-6605.

284. *UCSF Chimera - A Visualization System for Exploratory Research and Analysis.* **Pettersen, E F, *et al.*** 2004, Journal of Computational Chemistry, Vol. 25, pp. 1605-1612.

285. **Haider, Shozeb and Neidle, Stephen.** Molecular Modeling and Simulation of G-Quadruplexes and Quadruplex-Ligand Complexes. [book auth.] P Bauman. *G-Quadruplex DNA: Methods and Protocols.* s.l. : Human Press, 2010, Vol. 608, pp. 17-37.

286. *Systematic Benchmarking of Large Molecular Dynamics Simulations Employing GROMACS on Massive Multiprocessing Facilities.* **Gruber, Christian C and Pleiss, Jurgen.** 2011, Journal of Computational Chemistry, Vol. 32, pp. 600-606.

287. *Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of alpha/gamma Conformers.* **Perez, Alberto, *et al.*** 2007, Biophysical Journal, Vol. 92, pp. 3817-3829.

288. *Molecular Dynamics Studies of DNA A-Tract Structure and Flexibility.* **Sherer, Edward C, *et al.*** 1999, Journal of the American Chemical Society, Vol. 121, pp. 5981-5991.

289. *An Automated Approach for Clustering an Ensemble of NMR-Derived Protein Structures into Conformationally Related Subfamilies.* **Kelley,**

Lawrence A, Gardner, Stephen P and Sutcliffe, Michael J. 1996, Protein Engineering, Vol. 9, pp. 1063-1065.

290. *Ensembl 2008.* **Flicek, P, *et al.*** 2008, Nucleic Acids Research, Vol. 36, pp. 707-714.

291. *Bio-Mart - Biological Queries Made Easy.* **Smedley, D, *et al.*** 2009, BMC Genomics, Vol. 10, pp. 10-22.

292. *Single Molecule Conformational Analysis of the Biologically Relevant DNA G-Quadruplex in the Promoter of the Proto-Oncogene c-Myc.* **Shirude, Pravin S, Ying, Liming and Balasubramanian, Shankar.** 2008, Chemical Communications, pp. 2007-2009.

293. *Stability of Intramolecular DNA Quadruplexes: Comparison with DNA Duplexes.* **Risitano, Antonina and Fox, Keith R.** 2003, Biochemistry, Vol. 42, pp. 6507-6513.

294. **Marras, Salvatore A.** Selection of Fluorophore and Quencher Pairs for Fluorescence Nucleic Acid Hybridization Probes. [book auth.] V V Didenko. *Methods in Molecular Biology: Fluorescent Energy Transfer Nucleic Acid Probes: Designs and Protocols.* Totowa : Human Press Inc, 2003.

295. *Effects of Temperature and ATP on the Kinetic Mechanism and Kinetic Step-size for E. coli RecBCD Helicase-catalyzed DNA Unwinding.* **Lucius, Aaron L and Lohman, Timothy M.** 2004, Journal of Molecular Biology, Vol. 339, pp. 751-771.

296. *Double-Helical - Ladder Structural Transition in the B-DNA is Induced by a Loss of Dispersion Energy.* **Cerny, Jiri, Kabelac, Martin and Hobza, Pavel.** 2008, Journal of the American Chemical Society, Vol. 130, pp. 16055-16059.

297. *The Relevance of Continuous Variable Entanglement in DNA.* **Rieper, Elisabeth, Anders, Janet and Vedral, Vlatko.** 2010, arXiv:1006.4053.

298. *DNA G-Quadruplex: Structure, Function and Human Disease.* **Gomez-Marquez, Jaime.** 2010, The FEBS Journal, Vol. 277, p. 3451.

299. *Functional and Dysfunctional Roles of Quadruplex DNA in Cells.* **Arthanari, Haribabu and Bolton, Philip H.** 2001, Chemistry & Biology, Vol. 8, pp. 221-230.

300. *Essential Role for the Interaction Between hnRNP H/F and a G Quadruplex in Maintaining p53 pre-mRNA 3'-end Processing and Function During DNA Damage.* **Decorsiere, Adrien, *et al.*** 2011, Genes & Development, Vol. 25, pp. 220-225.

301. *Occurence of a Quadruplex Motif in a Unique Insert Within Exon C of the Bovine Estrogen Receptor a Gene (ESR1).* **Derecka, Kamila, *et al.*** 2010, Biochemistry, Vol. 49, pp. 7625-7633.

302. *Selection for the G4 DNA Motif at the 5' End of Human Genes.* **Eddy, Johanna and Maizels, Nancy.** 2009, Molecular Carcinogenesis, Vol. 48, pp. 319-325.

303. *Site Specific Oxidation at GG and GGG Sequences in Double-Stranded DNA by Benzoyl Peroxide as a Tumor Promoter.* **Kawanishi, Shosuke, *et al.*** 1999, Biochemistry, Vol. 38, pp. 16733-16739.

304. *Telomeric D-Loops Containing 8-Oxo-2'-Deoxyguanosine are Preferred Substrates for Werner and Bloom Syndrome Helicases and are Bound by POT1.* **Ghosh, Avik, *et al.*** 2009, Journal of Biological Chemistry, Vol. 284, pp. 31074-31084.

305. *Structural and Energetic Consequences of Oxidation of d(ApGpGpGpTpT) Telomere Repeat Unit in Complex with TRF1 Protein.* **Cysewski, Piotr and Czelen, Przemyslaw.** 2010, Journal of Molecular Modelling, Vol. 16, pp. 1797-1807.

306. *G-Quadruplex Stabilization be Telomestatin Induces TRF2 protein Dissociation from Telomeres and Anaphase Bridge Formation Accompanied by Loss of the 3' Telomeric Overhang in Cancer Cells.* **Tahara, H, *et al.*** 2006, Oncogene, Vol. 25, pp. 1955-1966.

307. *Repair of 8-oxodeoxyguanosine Lesions in Mitochondrial DNA Depends on the Oxoguanine DNA Glycosylase (OGG1) and 8-Oxoguanine Accumulates in the Mitochondrial DNA of OGG1-defective Mice.* **de Souza-Pinto, Nadja C, *et al.*** 2001, Cancer Research, Vol. 61, pp. 5378-5381.

308. *Genome-wide Analysis of a G-Quadruplex-Specific Single-Chain Antibody that Regulates Gene Expression.* **Fernando, Himesh, *et al.*** 2009, Nucleic Acids Research, Vol. 37, pp. 6716-6722.

309. *Evidence of Genome-Wide G4 DNA-Mediated Gene Expression in Human Cancer Cell Lines.* **Verma, Anjali, *et al.*** 2009, Nucleic Acids Research, Vol. 37, pp. 4194-4204.

310. **Moon, Ian K and Jarstfer, Michael B.** Preparation of G-Quartet Structures and Detection by Native Gel Electrophoresis. [ed.] P. Bauman. *G-Quadruplex DNA: Methods and Protocols.* s.l. : Human Press, 2010, Vol. 608, pp. 51-63.

311. *ArrayExpress update - from an archive of functional genomics experiments to the atlas of gene expression.* **Parkinson, H, *et al.*** 2009, Nucleic Acids Research, Vol. 37, pp. 686-872.

6 APPENDICES

Due to the large amount of data produced by the different techniques used, a digital appendix was created on a companion DVD which includes all relevant results not disclosed in this thesis, as well as .pdb structure files and abbreviated trajectory files to enable 3D visualisation of all the MD results referred to in this work, as well as all PERL scripts used. For the readers' convenience, the appendices are ordered into folders corresponding to the sections and sub-sections present in this thesis.