UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Centre for Biological Sciences

Volume 1 of 1

The Effects of DNA Supercoiling and G-quadruplex Formation

by

Doreen Asiya T. Sekibo

Thesis for the degree of Doctor of Philosophy

December 2013

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Doctor of Philosophy

Thesis for the degree of Doctor of Philosophy

THE EFFECTS OF DNA SUPERCOILING AND G-QUADRUPLEX FORMATION

Doreen Asiya T. Sekibo

The self-association of guanine bases in a tetrameric square planar arrangement was first determined in the early 1960s. The tetramer, commonly termed the G-quartet, can stack upon other G-quartets to form four-stranded helices termed G-quadruplexes. Bioinformatics studies have revealed that guanine-rich sequences with the propensity to adopt these structures are found in telomeric DNA and throughout the human genome, particularly in gene promoter regions. It is thought that the location of these sequences is not a coincidence and that the folding potential of guanine-rich DNA *in vivo* may play an important role in biological events such as gene regulation. Repetitive guanine tracts of G-quadruplex-forming DNAs form highly polymorphic structures with parallel or antiparallel strand orientations, depending on the ionic condition and the length of the connecting loops, and can be assembled as inter- or intra-molecular complexes. While extensive research has demonstrated their formation *in vitro*, there is little direct evidence to support their formation *in vivo*. With the exception of the single-stranded telomeric DNA, all genomic guanine-rich sequences are always present in the duplex configuration. Therefore, these structures will need to compete with the duplex that is normally generated with the complementary cytosine-rich strand. In order for this to happen would first require the local dissociation of the strands. Negative supercoiling results from the unwinding of the DNA helix and is known to provide energy to facilitate the formation of a number of alternative DNA structures. The work described in this thesis therefore aims to investigate the formation of G-quadruplexes under negatively supercoiled conditions. This was examined by preparing plasmids that contained multiple copies of G-rich oligonucleotides, based on the sequences $(G_3T)_n$ and $(G_3T_4)_n$, cloned into the pUC19 vector.

The formation of G-quadruplexes within these repeats has been assessed using the chemical probes dimethyl sulphate (DMS) and potassium permanganate, and the single-strand specific endonuclease S1. DMS probing revealed some evidence for G-quadruplex formation in $(G_3T)_n$ sequences, though this was not affected by DNA supercoiling. However, probing with $KMnO_4$ failed to detect exposed thymines in the loop regions, though there was some supercoil-dependent reactivity in the surrounding sequences, suggesting that this had been affected by the G-rich region. In contrast, the $(G_3T_4)_n$ sequences did not demonstrate protection from DMS, suggesting that G-quadruplex formation had not taken place. Surprisingly, the $KMnO_4$ reactions identified structural alterations around, but not within, the inserted G-rich fragments. S1 nuclease digestions did not detect any structural perturbations in any of the sequences apart from a mutant plasmid containing an inverted quadruplex repeat at the 3'-end.

Two-dimensional gel electrophoresis of DNA topoisomers was also conducted to detect any supercoil-dependent B-DNA to quadruplex transitions. Neither the $(G_3T)_n$ nor $(G_3T_4)_n$ plasmids showed any such structural changes. However, the mutant plasmid did demonstrate some supercoil-dependent changes, though these may correspond to cruciform rather than G-quadruplex formation. These results do not support the suggestion that negative supercoiling can induce the formation of G-quadruplex structures.

Contents

ABSTRACT ..... **i**

Contents ..... **iii**

List of tables ..... **vii**

List of figures ..... **ix**

Acknowledgements ..... **xvii**

Definitions and Abbreviations ..... **xix**

Chapter 1: Introduction **1**

1.1 Background 1
1.2 Fundamentals of G-quadruplexes 2
1.2.1 General Features of the G-quadruplex Structure 2
1.2.2 Stoichiometry 4
1.2.3 Quadruplex Loops 7
1.2.4 Cation Binding and Stability 9
1.2.5 Sequence Effects 11
1.3 i-motif 12
1.4 Duplex-Quadruplex Competition 13
1.5 Telomeric G-quadruplexes 14
1.5.1 G-Quadruplex Folding Topologies of the Human Telomere 15
1.5.2 Folding Topologies of Other Telomeric Sequences 23
1.6 Telomere End-Binding Proteins (TEBPs) 24
1.7 G-quadruplexes and Gene Regulation 25
1.7.1 Chicken-β-globin Gene 27
1.7.2 c-myc 28
1.7.3 c-kit 31
1.7.4 Bcl-2 33
1.8 RNA Quadruplexes 34
1.8.1 Stability of RNA G-quadruplexes RNA Quadruplexes in Telomeres 35
1.8.2 RNA Quadruplexes in Telomeres 36
1.8.3 mRNA Quadruplexes 36
1.9 Bioinformatics 37
1.9.1 Prevalence of G-quadruplexes in the Human Genome 37
1.9.2 Loop Length and Sequence Distribution Patterns 38
1.9.3 G-quadruplex Motif Distribution Patterns 42

1.10 Quadruplex Stabilising Ligands as Therapeutic Targets 44
1.10.1 BRACO-19 47
1.10.2 TMPyP4 49
1.10.3 Telomestatin 50
1.11 Negative Supercoiling of DNA and G-quadruplexes 52
1.11.1 DNA G-quadruplexes *in vivo* 54
1.12 Aims 56
Chapter 2: Materials and Methods 59
2.1 Instrumentation 59
2.2 Materials 59
2.2.1 Chemicals and Reagents 59
2.2.2 Buffers and Solutions 60
2.2.3 Oligonucleotide Sequences 61
2.3 Methods 63
2.3.1 General Techniques 63
2.3.2 Cloning 65
2.3.3 Sequencing 67
2.4 Probing of G-quadruplex Formation 69
2.4.1 Chemical Probing: Chemical Footprinting of G-rich Strand 69
2.4.2 Piperidine Cleavage of Modified Bases 71
2.4.3 S1 Mapping 72
2.5 Two-Dimensional Gel Electrophoresis 74
2.5.1 Preparation of Topoisomer Distributions 74
2.5.2 Experimental Procedure for Two-Dimensional Gel Electrophoresis 75
2.6 Circular Dichroism 76
2.6.1 CD Measurements 76
Chapter 3: The Cloning and Characterisation of pUC19 G-quadruplex-Forming Plasmids 79
3.1 Introduction 79
3.1.1 Circular Dichroism 80
3.1.2 Cloning: Blue-White Selection 81
3.1.3 Dideoxy Sequencing – The Sanger Method 82
3.2 Experimental Theory and Design 83
3.2.1 Oligonucleotide Design Considerations 84
3.2.2 Circular Dichroism 86
3.2.3 Cloning and Sequencing 86

3.3 Results 87
3.3.1 Circular Dichroism 87
3.3.2 Cloning of Potential G-Quadruplex Forming Oligonucleotide Sequences 98
3.3.3 Generating Plasmids with Extended G-Quadruplex Forming Sequences 102
3.3.4 Investigations into Whether the Sequencing Reaction Induces G-Quadruplex Self-Cleavage 108
3.4 Discussion 113
3.4.1 Circular Dichroism 113
3.5 Conclusion 117
Chapter 4: Chemical and Enzymatic Probing of G-Quadruplex Formation within Supercoiled DNA 119
4.1 Introduction 119
4.1.1 Investigating G-Quadruplex Structures in Supercoiled DNA using Chemical and Enzyme Probes 119
4.2 Experimental Theory and Design 120
4.2.1 Chemical Probing: The use of Dimethyl Sulphate and Potassium Permanganate as Chemical Probes to Investigate the Formation of G-Quadruplexes 121
4.2.2 S1 Nuclease 123
4.3 Results 125
4.3.1 *In Vitro* Chemical Footprinting using DMS 125
4.3.2 *In Vitro* Chemical Footprinting using $KMnO_4$ 140
4.3.3 The use of S1 Nuclease to Map the Sites of G-Quadruplex Formation Within Supercoiled DNA 149
4.3.4 Probing an Extended G-rich Sequence Within Supercoiled DNA: $pG_n,4,X_n$ 156
4.4 Discussion 162
4.4.1 Investigating G-quadruplex Formation Using DMS and KMnO4 Chemical Probes and S1 Nuclease as an Enzymatic Probe 163
4.5 Conclusion 171
Chapter 5: Two-Dimensional Gel Electrophoresis of G-Quadruplex Forming Supercoiled DNA Topoisomers 173
5.1 Introduction 173
5.1.1 Topology of Supercoiled DNA and Topoisomers 174
5.1.2 Topoisomerases 178
5.1.3 Two-Dimensional Gel Electrophoresis 179
5.1.4 The Energetics of Supercoiled DNA 184
5.2 Experimental Design 185

5.3 Results 186
5.3.1 Two-Dimensional Gel Electrophoresis of Plasmid pUC19 186
5.3.2 Two-Dimensional Gel Electrophoresis of plasmid pG_4,1,1 and pG_4,1,2. 188
5.3.3 Two-Dimensional Gel Electrophoresis of plasmid pG_4,4,1 and pG_4,4,2. 190
5.3.4 Two-Dimensional Gel Electrophoresis of plasmid pG_n,4,X_n 192
5.4 Discussion 197
5.5 Conclusion 202
Chapter 6: Final Discussion and Conclusions 205
6.1 Final Discussion 205
6.1.1 Analysis of the $p(G_3T)_n$ Sequences 206
6.1.2 Analysis of the $p(G_3T_4)_n$ sequences 210
6.1.3 Analysis of the $p(G_3T_8)$ sequences 211
6.1.4 Analysis of the plasmid mutant pG_n,4,X_n 212
6.2 The Effect of Negative Supercoiling and G-Quadruplex Formation 214
6.3 Final Conclusions 217
6.4 Future Work 217
Appendices 219
Appendix 1 Commercial sequencing of pUC19 G-quadruplex clones 221
List of References 231

List of tables

Table 1.1 Ionic radii of quadruplex stabilising cations. 10

Table 1.2 Some potential G-quadruplex-forming motifs identified in oncogenic promoter sequences. 26

Table 1.3 DNA sequences of the NHE III_1 G-rich strand from the c-myc promoter and its truncated and modified derivatives: 31

Table 1.4 Promoter sequences of the *bcl-2* gene and its modifications. 34

Table 1.5 PQS loop lengths: 41

Table 1.6 Loop sequence frequency found in PQS. 42

Table 2.1 List of G-rich oligonucleotides sequences with complimentary C-rich strand. 62

Table 2.2 Volumes (μl) required for the preparation of topoisomer distributions .. 75

Table 3.1 Summary of Cloned G-quadruplex forming plasmids (see appendices for sequencing). 112

Table 5.1 List of G-quadruplex forming supercoiled DNA used in the two-dimensional gel electrophoresis experiments. 185

List of figures

Figure 1.1 Assembly of a G-quadruplex structure. 3

Figure 1.2 Glycosidic bond of guanine bases in G-quadruplex structures 4

Figure 1.3 Examples of quadruplex formation. 6

Figure 1.4 Schematic representation of quadruplex loop. a) Edgewise, b) Diagonal and c) Double-chain-reversal. 7

Figure 1.5 The association of C-rich strands to form a four-stranded *i-motif* structure. 13

Figure 1.6 Schematic diagrams of the 22 nt human telomeric sequence. 17

Figure 1.7 Schematic representation of NMR analysis of the folding topologies of the human telomeric sequence in the presence of potassium. 18

Figure 1.8 A schematic model of G-quadruplex formation on the human telomeric DNA sequence Tel26. 19

Figure 1.9 Schematic diagram of the Tel26 Hybrid-1 folding topology of the intramolecular human telomeric G-quadruplex in K^+ solution. 20

Figure 1.10 Schematic diagram of the wtTel26 Hybrid-2 folding topology of the intramolecular human telomeric G-quadruplex in K^+ solution. 21

Figure 1.11 A schematic diagram of the Hybrid-3 folding topology of the intramolecular human telomeric G-quadruplex in K^+ solution. 22

Figure 1.12 Schematic diagram *htel27[Br22]* antiparallel (2+2) G-quadruplex. 22

Figure 1.13 Intramolecular “cinched” G-quadruplex structure as proposed by Howell *et al.* 28

Figure 1.14 Schematic structures of human c-myc promoter: 30

Figure 1.15 Topology of *c-kit1*scaffold. 32

Figure 1.16 Topology of *c-kit2* G-quadruplex fold. 33

Figure 1.17 A graph to show that the probability of finding PQS is directly related to its proximity with the TSS..... 43

Figure 1.18 Molecular structure of a) BRACO19, b) TMPyP4 c) TMPyP2 and d) Telomestatin 47

Figure 1.19 The biological unit in the crystal (PDB id 3CE5) with BRACO-19. 49

Figure 1.20 Chemical structure of the macrocyclic hexaoxazole L2H2-6M(2)OTD with proton naming indicated. 52

Figure 1.21 Immunofluorescence visualisation of G-quadruplex structures for the BG4 antibody on chromosomes isolated from HeLa cells. 56

Figure 2.1 pUC19 cloning vector. 66

Figure 2.2 Diagram depicting the potential region of S1 cleavage of the G-rich strand of the DNA plasmid. 73

Figure 3.1 CD spectra of intramolecular G-quadruplexes. 81

Figure 3.2 Diagram of a dideoxynucleotide. 82

Figure 3.3 CD spectra of the T-loop quadruplex-forming oligonucleotides. 90

Figure 3.4 CD spectra of the T_4-loop quadruplex-forming oligonucleotides 91

Figure 3.5 CD spectra of the T_8-loop quadruplex-forming oligonucleotides. 92

Figure 3.6 CD spectra of the T-loop quadruplex-forming oligonucleotides in the presence of its complementary C-rich strand. 95

Figure 3.7 CD spectra of the T_4-loop quadruplex-forming oligonucleotides in the presence of its complementary C-rich strand. 96

Figure 3.8 CD spectra of the T_8-loop quadruplex-forming oligonucleotides in the presence of its complementary C-rich strand. 97

Figure 3.9 Automated chain-termination DNA sequencing of a $p(G_3T)_n$ clone. 99

Figure 3.10 Automated chain-termination DNA sequencing of a $p(G_3T_4)_n$ clone. . 100

Figure 3.11 Manual dideoxy sequencing of G-rich plasmids. 101

Figure 3.12 Dideoxy sequencing of plasmids containing G-rich oligonucleotides (extended inserts). 103

Figure 3.13 Dideoxy sequencing of plasmids containing the sequence $pG_n,4,X_n$. ... 106

Figure 3.14 Automated DNA sequencing of an extended G-quadruplex forming sequence $p(G_n,4,X_n)$ cloned into pUC19 using a universal reverse primer. 107

Figure 3.15 Determination of plasmid self-cleavage using agarose gel electrophoresis. 109

Figure 3.16 Automated DNA sequencing of $p(G_4,4,X_n)$. 111

Figure 4.1 Schematic reaction of the proposed Maxam and Gilbert guanine base modification. 122

Figure 4.2 A Schematic reaction of the oxidation of thymine by potassium permanganate. 123

Figure 4.3 The proposed model for the formation of a G-quadruplex structure induced by negative supercoiling within a duplex plasmid. 124

Figure 4.4 *In vitro* footprinting of plasmid pG_4,1,1 with DMS. 128

Figure 4.5 *In vitro* footprinting of plasmid pG_4,1,2 with DMS. 129

Figure 4.6 *In vitro* footprinting of plasmid pG_5,1,1 with DMS. 132

Figure 4.7 *In vitro* footprinting of plasmid pG_5,1,2 with DMS. 133

Figure 4.8 *In vitro* footprinting of the plasmid pG_4,4,1 with DMS. 135

Figure 4.9 *In vitro* footprinting of the plasmid pG_4,4,2 with DMS. 136

Figure 4.10 *In vitro* footprinting of the plasmid pG_5,4,1 with DMS. 138

Figure 4.11 *In vitro* footprinting of the plasmid pG_5,4,2 with DMS. 139

Figure 4.12 $KMnO_4$ *in vitro* footprinting on the plasmids: pG_4,1,1 and 2. pG_4,1,2. 142

Figure 4.13 $KMnO_4$ *in vitro* footprinting on the plasmids: pG_5,1,1 and 2. pG_5,1,2. 144

Figure 4.14 $KMnO_4$ *in vitro* footprinting of the plasmids: pG_4,4,1 and 2. pG_4,4,2.. 146

Figure 4.15 $KMnO_4$ *in vitro* footprinting on the plasmids: pG_5,4,1 and 2. pG_5,4,2. 148

Figure 4.16 Mapping of S1 nuclease-sensitive sites in plasmids a) pG_4,1,1 and b) pG_4,1,2:..... 151

Figure 4.17 Mapping of S1 nuclease-sensitive sites in plasmids a) pG_4,4,1 and b) pG_4,4,2:..... 153

Figure 4.18 Mapping of S1 nuclease-sensitive sites in pUC19. 155

Figure 4.19 Formation of a pUC19 cruciform. 155

Figure 4.20 *In vitro* footprinting on the plasmid pG_n,4,X_n with DMS. 157

Figure 4.21 $KMnO_4$ *in vitro* footprinting of plasmid pG_n,4,X_n. 159

Figure 4.22 Mapping of S1 nuclease-sensitive sites in plasmid pG_4,4,X_n..... 161

Figure 4.23 Formation of a possible cruciform structure from the inverted repeat sequence in the plasmid p(G_4,4,X_n)..... 170

Figure 5.1 Electron micrographs of Polyoma virus DNA. 174

Figure 5.2 Relationship between linking number, twist and writhe for a linear DNA molecule. 178

Figure 5.3 DNA intercalators: Ethidium bromide and Chloroquine. 180

Figure 5.4 Schematic representation of two-dimensional gel electrophoresis of a circular DNA molecule that does not undergo supercoil-dependent structural transitions. 182

Figure 5.5 Schematic representation of two-dimensional gel electrophoresis of a circular DNA molecule that undergoes supercoil-dependent structural transitions. 183

Figure 5.6 Separation of pUC19 DNA topoisomers by two-dimensional gel electrophoresis. 187

Figure 5.7 Separation of a) pG_4,1,1 and b) pG_4,1,2 DNA topoisomers by two-dimensional gel electrophoresis. 189

Figure 5.8 Separation of a) pG_4,4,1 and b) pG_4,4,2 DNA topoisomers by two-dimensional gel electrophoresis. 191

Figure 5.9 Separation of pG_n,4,X_n DNA topoisomers by two-dimensional gel electrophoresis. 195

Figure 5.10 a) Separation of a) pUC19 and b) pG_n,4,X_n DNA topoisomers by two-dimensional gel electrophoresis at high chloroquine concentration..... 196

DECLARATION OF AUTHORSHIP

I, Doreen Asiya T. Sekibo declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

The Effects of DNA Supercoiling and G-quadruplex Formation.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

3. Where I have consulted the published work of others, this is always clearly attributed;

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

5. I have acknowledged all main sources of help;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. None of this work has been published before submission.

Signed: .....

Date: .....

Acknowledgements

I would like to thank Prof Keith Fox for supporting this research project and the Fox group for their support and friendship. I would like to say a big thank you to Dr Mark Coldwell and Dr Jo Cowan for their help and advice on DNA sequencing, Mr Neville Wright for helping and assisting me with the CD spectroscopy and chemical probing experiments and our lab technician Mrs Marilyn Jenkins for her invaluable support and encouragement, I am truly grateful. I would like to thank my friends and colleagues in Prof Proud's group, it was fun working with you all. Thank you to Susannah and all of my friends for all your love and the faith you had in me.

To support my studies, I worked in the university's libraries, and I want to express my heartfelt thanks to the library team, especially my managers, Elizabeth, Wendy, Ric, Mark and Jon. They have not only been my colleagues but also my friends who have shown me great support and understanding.

I would like to say a massive thank you to Alexander John Boyland, you believed in me, you encouraged me and you supported me throughout my studies. I will always be truly grateful for this. Thank You!

Lastly, I would like to express my sincerest gratitude to my parents and family. The support and great sacrifices that you have made for me throughout out my life were not in vain and gave me the strength, belief and confidence to achieve my PhD. Lots of love and thank you so much!

Definitions and Abbreviations

A	Adenine
bp	Base pairs
C	Cytosine
cccDNA	Covalently close circular DNA
CD	Circular dichroism
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
FRET	Fluorescence resonance energy transfer
G	Guanine
G4P	Potential quadruplex-formation
G-quadruplex	Guanine-quadruplex
G-quartet	Guanine-quartet
G-tetrad	Guanine-tetrad
kb	Kilobases
kDa	Kilodalton
L	Linear (in respect to DNA)
Lk	Linking number
NHE	Nuclease hypersensitivity element
NMR	Nuclear magnetic resonance
nt	Nucleotides
OD	Optical density
PCR	Polymerase chain reaction
PNA	Peptide nucleic acid
POT1	Protection of telomeres 1
PQS	Putative quadruplex-forming sequence
RNA	Ribonucleic acid
SC	Supercoiled
T	Thymine
TBE	Tris-boric acid-EDTA (gel running buffer)
TEBP	Telomere-end binding protein
TEMED	N,N,N,N’-Tetramethylethylenediamine
T_m	Melting temperature
TSS	Transcription start site
Tw	Twist
UTR	Untranslated region
UV	Ultraviolet
Wr	Writhe

Chapter 1:
Introduction

1.1 Background

The elucidation of the deoxyribonucleic acid (DNA) double helix structure described by Watson and Crick in 1953 (Watson and Crick, 1953) has made an enormous impact in molecular biology. Today this model is still highly significant and important in modern day biochemical research in understanding the sequences and structure of DNA in respect to its biological function. The majority of DNA present in cells is considered to adopt the double helix B-form configuration, which consists of two antiparallel DNA strands that are bonded by the complementarity of Watson-Crick base pairs. However, under certain biological conditions, some sequences in both DNA and RNA (ribonucleic acid) have the potential to form alternative secondary structures, such as cruciforms, hairpins and triplexes. This research will focus on another such secondary structure; a guanine-rich four-stranded nucleic structure termed the "guanine (G)-quadruplex".

The ability of millimolar concentrations of guanylic acid to form viscous gels in water was first reported in 1910 by a German Chemist Ivar Bang (Bang, 1910). However, it was not until 1962 when Gellert *et al*. (Gellert et al., 1962) were able to explain this phenomena through X-ray diffraction studies of guanosine-5'-monophosphate gels. They determined that the reaction was due to the tetrameric arrangement of four guanylic acid molecules in a co-planar square array, which resulted in this structure being described as the guanine quartet/tetrad. Synthesis of polyguanine oligonucleotide strands led to further discoveries with X-ray diffraction studies revealing that guanine residues in extended oligonucleotides self-associate to form G-quartets that were able to stack and form a four-stranded helical structure termed the G-quadruplex or quadruplex (Arnott et al., 1974). For several years following this discovery, very little attention was given to the possible roles of G-quartets and quadruplexes in biology until 1987 when Henderson *et al*. reported that synthetic oligonucleotides of the guanine-rich (G-rich) strand of telomeric DNA exhibited unexpectedly high electrophoretic mobility on non-

denaturing polyacrylamide gels (Bryan and Baumann, 2011, Henderson et al., 1987). *In vitro* structural probing analysis later showed that single-stranded synthetic oligomers based on the G-rich sequences of chromosomal telomeres and the immunoglobulin switch region showed that the guanine bases could self-associate at physiological salt concentrations to form the four stranded G-quadruplexes (Henderson et al., 1987, Sen and Gilbert, 1990, Sundquist and Klug, 1989, Williamson et al., 1989, Sen and Gilbert, 1988). These discoveries led scientists to postulate whether these structures may hold some physiological important roles *in vivo*. Over the years, the possible formation of G-quadruplex structures *in vivo* has been implicated in a number of biologically significant events with initial bioinformatics studies identifying a number of quadruplex-forming motifs in key regions of the human genome with a vast majority being located within promoter regions of oncogenes (Huppert and Balasubramanian, 2005, Todd et al., 2005). This led to further speculation on whether these structures play a vital role in gene regulation, generating huge pharmaceutical interest, particularly in developing targets for anticancer therapeutics (Neidle, 2009, Parkinson, 2006).

1.2 Fundamentals of G-quadruplexes

1.2.1 General Features of the G-quadruplex Structure

The key component to a G-quadruplex structure is the G-quartet. A G-quartet is composed of four guanine bases that self-assemble to form a square planar cyclic array (Figure 1.1a). Unlike the hydrogen-bonding of the classical Watson-Crick G-C base pairing in duplex DNA, the cyclic guanine array is instead held together by eight Hoogsteen hydrogen-bonds. By utilising the N1 and N2 of one face and the O6 and N7 of another, each guanine base becomes a hydrogen bond donor on one face and a hydrogen bond accepter on the other and when two or more G-quartets stack together, they construct the quadruplex scaffold (Figure 1.1b) that may be held together by loops which are a specific feature of unimolecular and bimolecular quadruplexes (discussed below). Thus, the definition of a G-quadruplex refers to any four-stranded DNA containing stacked guanine quartets (Parkinson, 2006, Simonsson, 2001). All G-quadruplex structures possess a central cavity which is a unique characteristic of a G-quadruplex and due to the inwardly facing O6 guanine carbonyls of the tetrameric arrangement of the four guanine bases there is an aggregation of negative charge

resulting in electrostatic repulsion that can be stabilised by a locally placed cation in the central cavity.

G-quadruplex structures are extremely polymorphic and can adopt a number of different inter- or intra-molecular folding topologies depending on; the number of strands, the polarity of the strands (parallel of antiparallel), the glycosidic bond angles of the guanine bases (Figure 1.2), DNA sequence, temperature and most significantly the presence of a monovalent cation. Like duplex DNA, G-quadruplexes have grooves (four) which, depending on the structural topology vary in dimensions. Each groove contains a network of water molecules ordered about hydrogen-bond donors and acceptors of the exocyclic amino groups N2 and heterocyclic N3 atoms. For parallel quadruplexes the grooves are all medium in size whilst antiparallel quadruplexes are a combination of wide, medium and narrow depending on strand polarities.

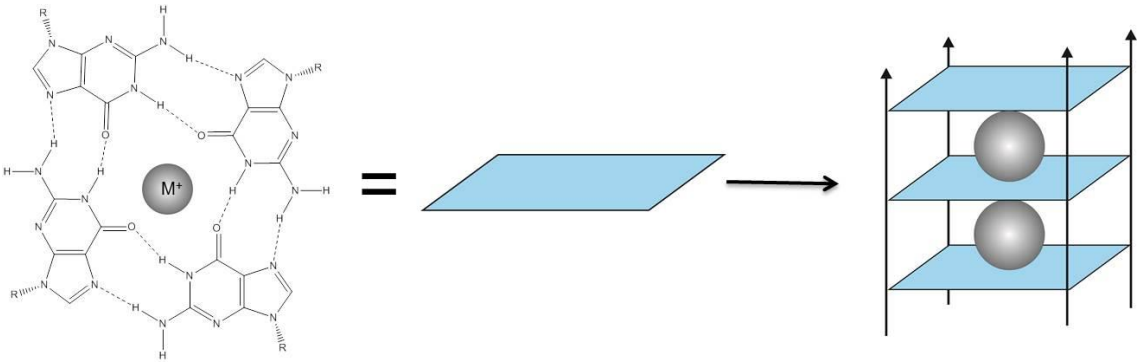


Figure 1.1 Assembly of a G-quadruplex structure. a) Cyclic arrangement of a G-quartet where each guanine base interacts with its neighbour in Hoogsteen hydrogen bonding. G-quartets stack on top of one another which are stabilised by a central cation. b) Schematic representation of three stacked G-quartets to form the parallel G-quadruplex structure.

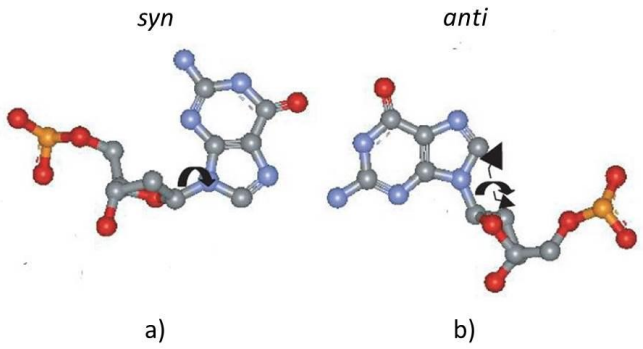


Figure 1.2 Glycosidic bond of guanine bases in G-quadruplex structures Rotation around the glycosidic bond allows guanine to form both a) *syn* and b) *anti* conformations. Adapted from Fundamentals of G-quadruplex Structure by G. Parkinson (Parkinson, 2006).

1.2.2 Stoichiometry

G-Quadruplexes are extremely polymorphic and can be categorised as either inter- or intra-molecularly folded, parallel or antiparallel structures depending on their strand orientation. They associate via one (unimolecular) (Henderson et al., 1987), two (bimolecular) (Sundquist and Klug, 1989, Keniry et al., 1995) or four (tetramolecular) (Sen and Gilbert, 1988, Laughlan et al., 1994) G-rich oligonucleotide strands (Figure 1.3) which are typically characterised by circular dichroism (CD) (discussed in further detail in Chapter 3). CD is a popular technique used for studying the polymorphism of G-quadruplexes. It is an indirect method that is reliant on spectral data of known structural motifs in solution for comparison (Randazzo et al., 2013, Parkinson, 2006). There are a few general rules that can be applied in determining strand orientation: parallel quadruplexes are characterised by a positive ellipticity at around 264 nm and a negative minimum at 240 nm whilst antiparallel complexes usually have a positive maximum at 295 nm and a negative minimum at 265 nm (Gray et al., 2008, Balagurumoorthy et al., 1992, Vorlickova et al., 2005, Lu et al., 1993). However, it is important to note that these are empirical rules and there may be instances in which CD spectroscopy may not be able to correctly indicate strand orientation.

The tetramolecular assembly is the simplest form of the G-quadruplex and occurs via the intermolecular association of four G-rich DNA molecules Figure 1.3a). Quadruplexes adopting this formation have all their strands in a parallel arrangement and each guanine participating in this structure adopts an *anti* glycosidic torsion angle

(Figure 1.2), resulting in each phosphodiester backbone orienting in the same direction. Although, in principle there are several ways in which the four strands could associate, to date only the parallel strand arrangements have been observed for monomeric solutions containing single runs of guanine (Parkinson, 2006). One of the most studied examples is the G-quadruplex formed by the sequence d(TGGGGT), which has been determined by NMR spectroscopy (Aboulela et al., 1992) and high resolution X-ray crystallography (Laughlan et al., 1994). Each strand is orientated in a parallel arrangement and each guanine tract is capped at both ends by a thymine residue. The structure has four identical grooves with each guanine glycosidic torsion angle adopting the *anti* configuration. Tetramolecular complexes can also adopt unusual topologies, for example in the case of the sequence d(GGGT) where the guanine tract is not capped at the 5'-end a more complex arrangement is observed. In this instance eight G-rich strands form an interlocked but extended G-quadruplex by the dimerisation of two staggered or "slipped" quadruplexes (Krishnan-Ghosh et al., 2004).

Bimolecular structures are formed by the dimerisation of two G-rich strands. NMR spectroscopy (Phan et al., 2004b, Smith et al., 1994) and X-ray crystallography (Haider et al., 2002, Haider et al., 2011) have shown that these structures exhibit a varied folding topology compared to tetramolecular quadruplexes. Strand connectivity, the number of guanines in each tract, strand length and loop composition all play a key role in structural topology. Figure 1.3b depicts some examples of bimolecular formation: (i) shows the head-to-head assembly, whilst (ii) illustrates the head-to-tail, both connected by edgewise loops. In both these cases, the strands are antiparallel and the glycosidic bonds of the guanine bases on each strand can alternate either between *anti* and *syn* or one strand exclusively containing all *syn* configurations whilst the other all *anti*. Bimolecular structures may form via diagonal loops where the strands are antiparallel as depicted in Figure 1.3b iii) or in 3+1 hybrid topology as shown in Figure 1.3b iv) with one strand containing three G-tracts and the other just one. In some cases they may also form an all parallel strand configuration where the four strands are linked together by two double-chain reversal loops illustrated Figure 1.3b v) (Zhang et al., 2005).

Unimolecular quadruplexes are the most diverse and complex of all the folding topologies. They are formed by the intramolecular folding of a single G-rich strand containing at least four tracts of guanines that can be linked together by three loop regions. There are four main categories of intramolecular quadruplexes depicted in Figure 1.3c: i) antiparallel "chair" type configuration containing three edgewise loops; ii) antiparallel "basket" type, with one diagonal loop and two edgewise loops; iii) parallel propeller-type with all three loops in the double-chain reversal configuration and iv) the hybrids 3+1 or 1+3 hybrids with three parallel strands and one antiparallel strand.

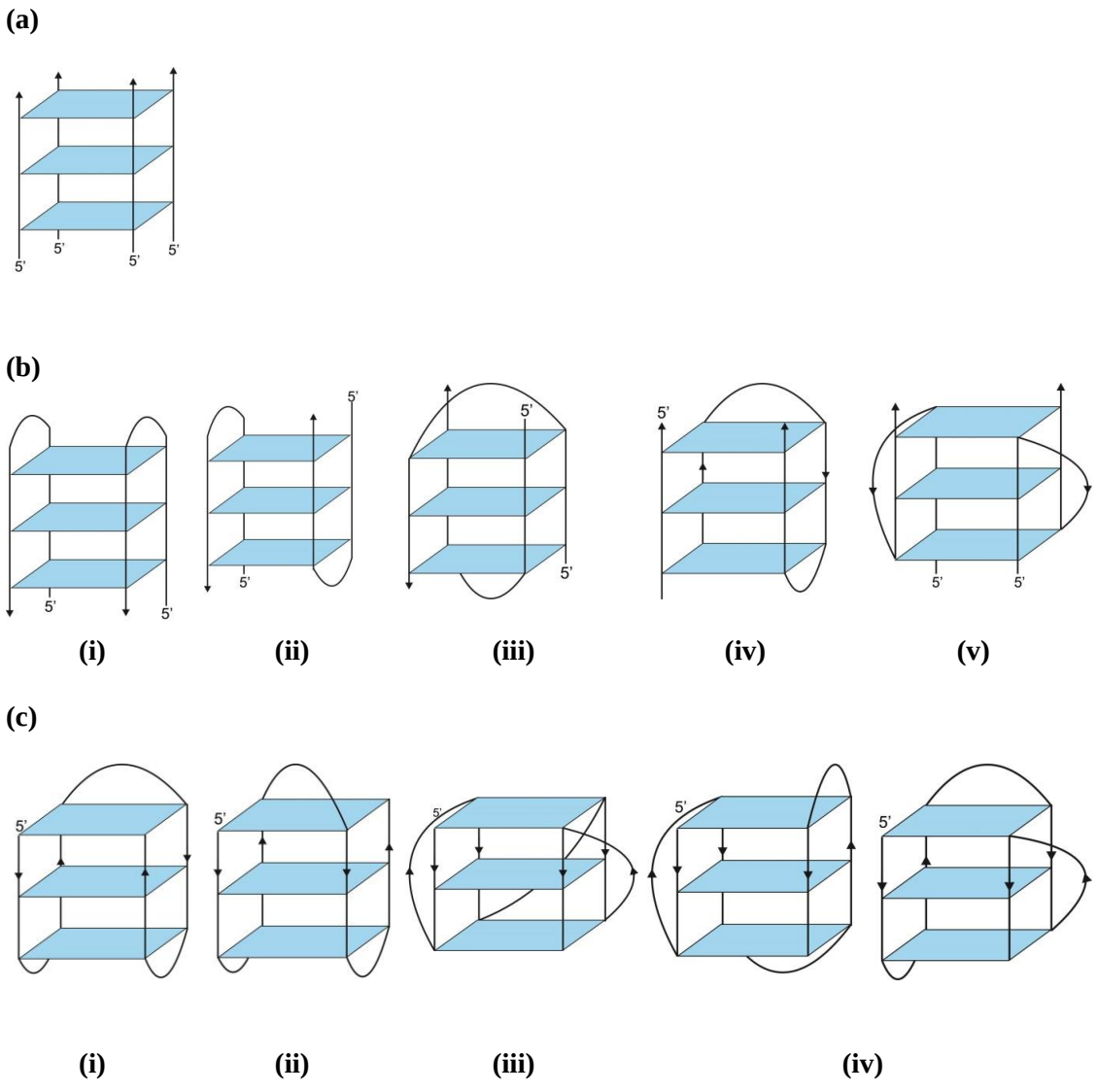


Figure 1.3 Examples of quadruplex formation. a) Four-stranded tetramolecular assembly, b) two stranded bimolecular arrangements of: i) Antiparallel head-to-head, ii) Antiparallel head-to-tail, iii) Antiparallel diagonal, iv) parallel double-chain reversal iv) 3+1 hybrid and c) Single-stranded intramolecular folds i) Antiparallel "chair" type ii) Antiparallel "basket" type iii) parallel double-chain-reversal iv) 3+1 and 1+3 hybrid.

1.2.3 Quadruplex Loops

The G-tracts in bimolecular and unimolecular quadruplexes are linked by loops that connect the G-quartets. Loop length and composition not only play a key role in determining both the topology and stability of the G-quadruplexes, but can influence competition between the Watson-Crick duplex and quadruplex formation (Hazel et al., 2004, Rachwal et al., 2007a, Risitano and Fox, 2003a, Risitano and Fox, 2004, Risitano and Fox, 2003b). Loops are categorised into three main groups: i) edgewise (or lateral) that connect two adjacent antiparallel strands, ii) diagonal which connect two opposing strands and iii) double-chain reversal (or propeller) which connect adjacent parallel strands (Burge et al., 2006, Simonsson, 2001) as depicted in Figure 1.4.

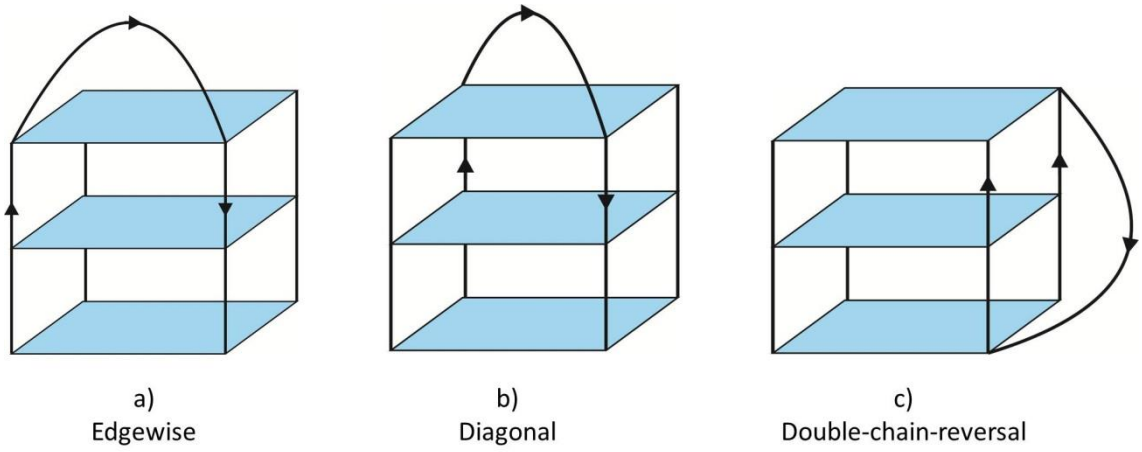


Figure 1.4 Schematic representation of quadruplex loop. a) Edgewise, b) Diagonal and c) Double-chain-reversal.

1.2.3.1 Loop length

In general, loops containing short linkers of a single nucleotide are constrained to the double chain reversal topology for structures that contain up to three stacked quartets, whilst structures with four stacked quartets require loops consisting of at least two nucleotides. Therefore, short linkers impose certain topological constraints on the folding of the G-quadruplex structure with single nucleotide loops typically restricted to the parallel folded topology (Hazel et al., 2004).

CD spectroscopy and gel electrophoresis studies on the sequence $d(GGGGT_nGGGG)_2$ were used to investigate the role of thymine residues on the formation of quadruplexes (Balagurumoorthy et al., 1992). The studies revealed that a single thymine loop formed a parallel strand arrangement, but on addition of a second thymine the quadruplexes resulted in a mixed parallel and antiparallel arrangement. Loops containing three or four thymines exhibited only an antiparallel arrangement. By increasing loop length to two nucleotides the loops are less restricted and can form either edgewise or double-chain reversal conformations. However, both single and double nucleotide loops are unable to form diagonal loops as they are too short to span the distance across the G-tetrad, whereas, longer loops have the freedom to form any loop configuration.

1.2.3.2 Loop stabilising effect

Loop length and composition contribute greatly in the stabilising of G-quadruplex structures. Typically, G-quadruplexes containing short loops are more stable than those with longer loops. Structures in which all loops consist of a single thymine have been shown to be the most stable (Risitano and Fox, 2004, Rachwal et al., 2007a, Rachwal et al., 2007b). DNA melting studies have been widely used to investigate the stability of duplex DNA; their melting transitions are usually detected by measuring the change in absorbance at 260 nm, which increases by about 25% on denaturation. Melting studies can also be used in order to establish the stability of G-quadruplex structures by measuring the change in absorbance at 290-295 nm (Rachwal and Fox, 2007) where the higher the melting temperature the more stable the structure. With the aid of fluorescence melting (Rachwal and Fox, 2007) Rachwal *et al*. (Rachwal et al., 2007b) investigated how changes in loop length affect quadruplex properties. They found that in the presence of either 100 mM potassium or 100 mM sodium G-rich sequences containing four tracts of three guanines and all single thymine loops were the most stable and as the size of the loop increased the stability of the structure decreased. By substituting just one of the single thymine loops with a loop containing four thymines, the stability of the structure decreased by 20 ºC irrespective of whether the replacement was made at either the central or peripheral loops.

1.2.3.3 Loop composition

Loop composition also plays a role in determining the stability of G-quadruplexes. In telomeric sequences G-quadruplex loops commonly consist of thymine and adenine bases with thymine being the most frequent loop nucleotide. In quadruplex research, thymines are typically used instead of adenines in the connecting loops as they not only stabilise the scaffold but are sterically favourable and can fit into a compact structure such as hairpins (Risitano and Fox, 2003b). The imino proton of the thymine base is a potential H-bond donor that may facilitate hydrogen-bond donation in order to provide stabilisation of the loop (Keniry et al., 1997). By comparison, poly-adenine tracts are very rigid due to their bulky purine ring (Nelson and Klug, 1988), meaning that in *in vitro* studies it is rare to find multiple adenines in loop regions. In experiments conducted by Risitano and Fox (Risitano and Fox, 2003b) where the d(TTA) loop of the human telomeric sequence was substituted with a trinucleotide d(AAA) loop, a complete destabilisation of the quadruplex structure was observed. Where possible, cytosines are often avoided in quadruplex loop-forming sequences, as they have a tendency to form Watson-Crick base pairs with neighbouring guanines and disrupt the formation of the G-quartet (Keniry et al., 1997, Parkinson, 2006). Also, at physiological pH, cytosines also lack an imino proton required for hydrogen bond donation. Guanines may also be present in the loop region, especially when the repeating pattern of G-tracts are not equal in length, however, this could lead to structural diversity (Parkinson, 2006) which is particularly observed in the 27 nt sequences of the nuclease hypersensitivity element III_1 (NHE III_1) of *c-myc* oncogene promoter sequence which contains six uneven runs of guanines (Ambrus et al., 2005, Phan et al., 2004a) (see section 1.6).

1.2.4 Cation Binding and Stability

The formation and stability of G quadruplexes is highly dependent on the ionic interactions between the phosphate backbone, hydrogen bonding between the guanine bases, the stacking forces between each G-quartet and most importantly the coordination of a centrally placed monovalent cation (Keniry et al., 1997). The Hoogsteen hydrogen-bonds of the N7:N2H and O6:N1H in the central cavity of the structure may appear to stabilise the G-quadruplex structure, in effect the formation and stability is completely dependent on the metal binding interactions of dehydrated cations (Hud and Plavec, 2006). This is because the central core of the G-quartets is not

electrochemically stable due to the build-up of negative charge from the four inwardly facing O6 carbonyl groups. Hence, a centrally located cation is required to balance the charge and stabilise the structure. Depending on the ionic radius of the cation, they can either interact with the four O6 carbonyl oxygens in the plane of a single G-quartet in a square planner array such as Na^+ with radius 0.97 Å or they can be sandwiched between two tetrads and interact with the eight O6 atoms in an octahedral bipyramid antiprismic arrangement, such as K^+ which has a larger radius of 1.33 Å (Hud and Plavec, 2006). Unlike other nucleic acid structures, the stability of a quadruplex is dependent on the type of cation and a wide variety of monovalent and divalent cations (Table 1.1) are able to induce the formation and stabilisation of quadruplexes (Sen and Gilbert, 1990). In G-quadruplex research the monovalent cations Na^+ and K^+ are considered to be the most efficient in stabilising these structures with the latter being the most effective and most extensively studied due to its physiological importance. In general the most stable to the least stable quadruplex stabilising cations are: $K^+ > Na^+ > Rb^+ > NH4^+ > Cs^+ > Li^+$ (Parkinson, 2006).

Table 1.1 Ionic radii of quadruplex stabilising cations. Taken from (Parkinson, 2006).

Cation	K^+	Na^+	Rb^+	NH_4^+	Cs^+	Li^+
Ionic Radius (Å)	1.33	0.97	1.66	1.43	1.81	0.90

Cations not only facilitate the formation and stabilisation of G-quadruplexes, but they also influence their folding topology. Whilst the formation of tetramolecular structures is not dependent on the nature of the cation, the topology of multi-stranded structures can show cation-dependent polymorphism. The bimolecular four-stack structure derived from the *oxytricha nova* telomeric sequence $d(GGGGTTTTGGGG)_2$ comprising of two symmetrical diagonal loops spanning across the top and bottom terminal tetrads has the same topology in the presence of both K^+ and Na^+ (Shim et al., 2009, Schultze et al., 1999). NMR analysis suggests that the Na^+ ions coordinate within the planes of the quartets allowing the cations to interact with the O2 carbonyl of the third positioned thymine (T7) within the loop. Yet, by substituting Na^+ with either K^+ or NH_4^+ the cations coordinate between the planes of the quartets, obstructing coordination with the loop thymines, therefore allowing the loop more freedom of mobility (Schultze et al., 1999).

Unimolecular structures display a greater degree of cation-dependent polymorphism. Sodium generally stabilises the antiparallel strand arrangement whilst potassium stabilises parallel structures. This is well established for the highly documented human telomeric repeat sequence $d(TTAGGG)_n$, where NMR studies in the presence of sodium show that the sequence $d[AGGG(TTAGGG)_3]$ has an antiparallel strand arrangement (Wang and Patel, 1993b) whilst X-ray crystallography determined in the presence of potassium shows an all parallel strand arrangement (Parkinson et al., 2002) (see section 1.5).

1.2.5 Sequence Effects

The formation of G-quadruplexes does not necessarily derive from neat tandem repeats of guanine tracts that are equal in length. Potential G-quadruplex-forming sequences are distributed throughout the genome in uneven tracts and lengths (Huppert and Balasubramanian, 2005, Todd et al., 2005, Todd and Neidle, 2011) and it is important to highlight that these sequences can also form stable G-quadruplex structures.

Sequences that contain runs of four unequal guanine tracts can produce a number of different folded topologies as some guanines may be forced to participate in the loop region or may be excluded from the scaffold. Similarly, sequences containing five or more runs of guanines can also exhibit a high degree of polymorphism as a number of folding combinations are possible. For example a sequence containing five guanine tracts, may form a structure with the first four G-rich tracts at the 5'-end or with the terminal four tracts at the 3'-end. Alternatively the central G-rich tract may be included in an extended loop region, which may all coexist in dynamic equilibrium. This has been observed in genomic DNA sequences such as the six G-tract sequence of the NHE III_1 G-quadruplex-forming sequence found in the promoter region of the *c-myc* oncogene $d(TG_4AG_3TG_4AG_3TG_4A_2G_2$ (Ambrus et al., 2005, Phan et al., 2004a) which is described in more detail in section 1.7 .

1.3 i-motif

The formation of a G-quadruplex structure within a region of duplex DNA must require the release of the complementary cytosine-rich (C-rich) strand before for it can fold into the quadruplex. NMR (Gehring et al., 1993, Leroy et al., 1994) and crystallographic studies (Chen et al., 1994) have shown that like quadruplexes, C-rich DNA is also capable of adopting an alternative four stranded structure termed the "*i-motif*".

These can also form tetra, bi and unimolecular structures in sequences resembling those of the G-quadruplex, with extended C-rich tracts forming the most thermodynamically stable structures (Mergny et al., 1995). Yet, unlike their G-rich counterparts these structures do not form a cyclic quartet array, they are cation independent and typically form and are stabilised at low pH. Instead the *i-motif* structure requires protonation of one cytosine at N3 position to form C-C^+ base pairs (depicted in Figure 1.5a) and creates two parallel intercalated hemiprotonated duplexes that associate with one another in a head-to-tail orientation (Esmaili and Leroy, 2005) (Figure 1.5b and c). By means of NMR spectroscopy Gehring *et al*. (Gehring et al., 1993) first described the structure of the DNA hexamer d(TCCCCC) in acidic pH. They described a structure consisting of two base-paired parallel-stranded duplexes that are fully intercalated in an antiparallel arrangement (Figure 1.5a and b), which was confirmed by Chen *et al*. (Chen et al., 1994) using crystallographic techniques on the sequence d(CCCC). The results revealed the same four-stranded molecule comprised of two parallel intercalated duplexes held together by hemiprotonated cytosine-cytosine base pairs, with the two duplexes pointing in opposite directions. *i-motifs* of various C-rich DNA sequences can form *in vitro* and are stable under acidic conditions, though there are fewer studies on their assembly at physiological pH. A study conducted by Mergny and co-workers (Mergny et al., 1995) demonstrated that C-rich sequences that form the *i-motif* are much less stable at neutral pH and that the melting temperature of both intra- and intermolecular complexes are extremely pH-dependent. Their findings also revealed that lowering the pH by just one unit typically led to a melting temperature increase of 20 to 25 ºC, which they concluded was due to the low pKa value of cytosine (~ 4.5), reinforcing the requirement for C-C^+ base pairing.

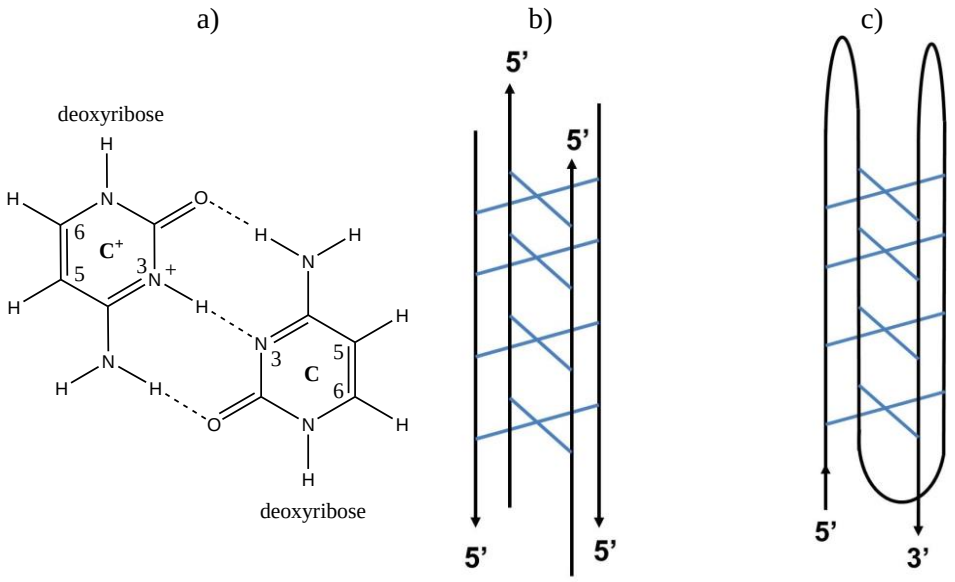


Figure 1.5 The association of C-rich strands to form a four-stranded *i-motif* structure. a) C-C^+ base pair b) intermolecular and c) intramolecular hemiprotonated parallel duplexes bound by the intercalation of C-C^+ base pairs

As C-rich motifs are also prevalent throughout the human genome it may be speculated that their formation and function could have biological implications within cellular DNA. Yet, *in vitro* studies of the *i-motif* suggest that due to their pH-dependent formation (~ pH 5) these structures may not actually exist *in vivo* under physiological conditions. However, in one *in vitro* study, chemical and enzymatic probing of a G-rich clone demonstrated that negative supercoiling can facilitate and stabilise not only the G-quadruplex but also the *i-motif* under physiological-like conditions of pH and ionic strength (Sun and Hurley, 2009). This study helps to provide evidence that transcriptionally generated negative supercoiling *in vivo* may induce these alternative structures although evidence to support this theory is still circumstantial.

1.4 Duplex-Quadruplex Competition

G-quadruplex forming motifs throughout both prokaryotic and eukaryotic genomes exist in the company of their complementary C-rich strands, therefore quadruplex assembly must compete with duplex formation. Since G-quartets contain eight hydrogen bonds (two H-bonds per base) while Watson-Crick base pairs contain two or three

(average 1.25 H-bonds per base), it might be expected that G-quadruplexes will be more stable than duplexes. In fact, the melting temperature of many G-quadruplexes is above 70 ºC under physiological conditions, indicating that they are highly stable (Ambrus et al., 2005, Parkinson et al., 2002) and have the potential to compete with duplex structures *in vivo*.

NMR and UV spectroscopy analysis investigating the conversion of duplex DNA into intramolecular G-quadruplex and *i-motif* structures for an equimolar mixture of the 22-mer telomeric fragments; d[AGGG(TTAGGG)$_3$] and d[(CCCTAA)$_3$CCCT], demonstrated that at near-physiological conditions of pH, temperature and ionic concentration, the duplex DNA was predominantly formed (Phan and Mergny, 2002). However, at lower pH, the G-quadruplex and the *i-motif* were able to compete with the duplex. In contrast a fluorescence melting study revealed, that the G-quadruplex-forming sequence d[(GGGT)$_3$GGG] containing a single thymine loop was able to compete with the duplex conformation and form even in a 50-fold excess of the complementary C-rich strand (Risitano and Fox, 2003b). Studies on the behaviour and thermodynamic properties of the telomeric duplex DNA revealed that the structural polymorphism of telomeric DNA to dissociate into two strands can be induced by molecular crowding thus generating the potential for the DNA to form G-quadruplex structures (Miyoshi et al., 2004). Although these studies have shown that G-quadruplex formation is possible in the presence of its complementary C-strand, it must be noted that these investigations were conducted on short linear fragments. However, *in vivo* it is probable that in the absence of external factors, such as DNA supercoiling (discussed in section 1.11), G-rich sequences existing within the extremely long genomic DNA will remain in the duplex formation.

1.5 Telomeric G-quadruplexes

One of the most abundant sources of G-quadruplex-forming DNA is found in chromosomal telomeres and these sequences have been by far the most extensively studied species in quadruplex research. Telomeres are DNA and protein structures located at the ends of linear chromosomes. They “cap” the end of chromosomes and protect them from exonucleolytic degradation, recombination, end-to-end fusion and

nuclease attack (Jiang et al., 2007, McEachern et al., 2000). Telomeres consist of simple tandem repeats, for example human telomeric sequences consist of the hexanucleotide repeat sequence d(TTAGGG), whilst the *Oxytricha nova* and *Tetrahymena* sequences are repeats of d(GGGGTTTT) and d(GGGGTT), respectively. Telomeric sequences can be up to several kilobases in length and in humans are single-stranded for the last 100 to 200 bases at the extreme 3' termini (Wright et al., 1997). During each round of cell division, telomeres shorten in length due to the "end replication problem" of DNA-polymerase which is unable to fully replicate the extreme 3'overhang. During each round of replication small segments of each telomere are lost until it reaches a critical short length triggering cellular senescence and apoptosis (Laughlan et al., 1994, Miyoshi et al., 2004, Williamson, 1994, Yu et al., 2010, Greider and Blackburn, 1985). In almost all eukaryotes this chromosome end-replication problem is solved by a cellular ribonucleoprotein reverse transcriptase called telomerase, discovered by Greider and Blackburn in 1985 (Greider and Blackburn, 1985). Telomerase has the ability to synthesise telomeric repeats and to extend the telomere length. It is inactive or not present in most differentiated cells, but is active in stem cells and germ cells. It has been observed in approximately 85% of human cancers where the DNAs of their telomeres and lengths are continually maintained making them "immortal". *In vivo*, human telomeres are protected from higher order assembly by interaction with several proteins such as human protection of telomeres 1 (hPOT1). *In vitro* studies have shown that these sequences adopt G-quadruplex structures under physiologically-relevant conditions, and several hypotheses have been proposed for their formation *in vivo* such as the promotion of telomere-telomere interactions, protecting the 3'-end from nucleolytic attack (Maizels, 2006).

1.5.1 G-Quadruplex Folding Topologies of the Human Telomere

The vast majority of G-quadruplex research has used sequences that form the three quartet stacked motif. The 200 nucleotide single-stranded overhang of the human telomeric sequence d(TTAGG) has the potential to form approximately eight tri-quartet stacked quadruplex structures. It has been speculated that if it is possible that all eight structures are folded at the same time, they may form a cylindrical quasi-superhelix, suitable for interaction with telomeric binding proteins or ligands and the inhibition of telomerase extension (Parkinson et al., 2002). However, the full structural arrangement

of this sequence is yet to be established. This hexanucleotide human telomeric repeat sequence can form tetramolecular, bimolecular and unimolecular G-quadruplex structures (Patel et al., 2007). However, the majority of human telomeric research is focused on short linear fragments of single-stranded DNA approximately 24 nucleotides in length as this is considered the minimum length requirement to form a three quartet stacked structure that would fold into an intramolecular structure as this is most likely to be formed by the folding of the 3'- single-stranded overhang (Neidle, 2012, Phan, 2010).

The human telomeric G-quadruplex structure is highly polymorphic particularly with loop geometry, strand orientation and the identity of the monovalent cations (Williamson, 1994). NMR and molecular dynamics modelling (Wang and Patel, 1993b) of the solution structure of the 22-mer sequence d[(AGGG(TTAGGG)$_3$] in sodium revealed an antiparallel stranded G-quadruplex 'basket-type' topology, comprised of three stacked G-quartets that were connected by three d(TTA) loops; one diagonal and two edgewise as depicted in Figure 1.6a. In this structure, the glycosidic torsion angle of the guanine bases in each quartet alternates between *syn* and *anti* in a *syn-syn-anti-anti* configuration. In contrast, X-ray crystallography analysis of the same sequence (Parkinson et al., 2002) revealed an all parallel stranded structure in the presence of potassium (Figure 1.6b). This structure also consisted of three stacked G-quartets, however it was held together with three double-chain-reversal d(TTA) side loops, where the adenine in each TTA trinucleotide loop is swung back so as to intercalate between the two thymines. In this arrangement, all guanines adopt the *anti* conformation and as the loops extend laterally up to 10 Å from the core G-quartets, the scaffold generates three very different types of surfaces; a polarised surface, and a hydrophobic aromatic planar surface which are both ideal for potential drug recognition binding sites and a TTA loop hydrogen bonding interface. The general shape of this quadruplex appears as a flattened disk in contrast to the globular shape of the sodium folded structure.

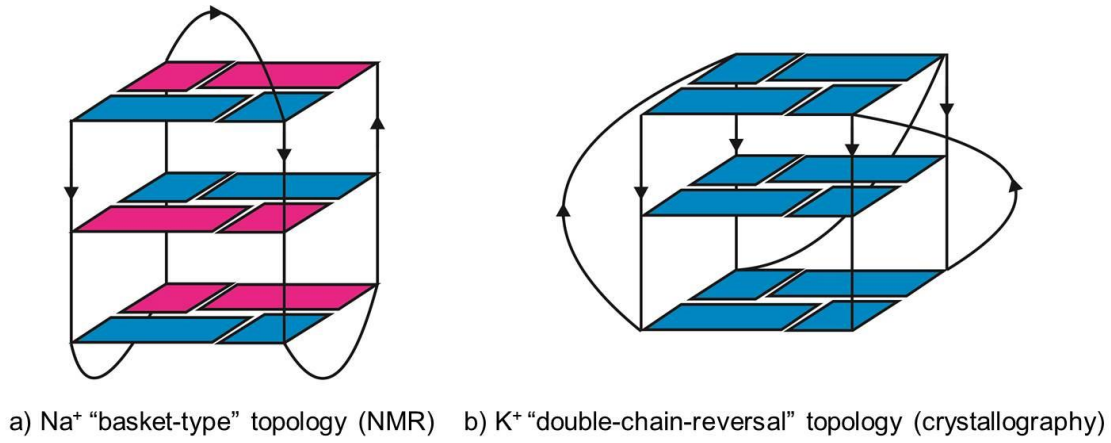


Figure 1.6 Schematic diagrams of the 22 nt human telomeric sequence. d(AGGGTTAGGGTTAGGGTTAGGG) in the presence of different cations by a) NMR analysis in the presence of Na^+ and b) crystallography analysis in the presence of K^+. *Syn* and a*nti* glycosidic bonds are indicated in pink and blue respectively (Wang and Patel, 1993b, Parkinson et al., 2002).

Multiple solution studies of the four repeat sequence in the presence of the potassium cation have revealed various conformations (Phan and Patel, 2003, Redon et al., 2003, Risitano and Fox, 2005). For example; whilst some studies report that in the presence of potassium the four repeat structure interconverts between parallel and antiparallel G-quadruplex structures (Phan and Patel, 2003, Ying et al., 2003), other reports suggest either the propeller-type parallel formation (Risitano and Fox, 2005) or the antiparallel "chair" configuration (He et al., 2004).

The dramatic difference between the NMR and the crystallographic structures in the presence of potassium suggests that the dehydrated conditions of the crystallographic structure may have selected for a particular conformational form and/or alternatively that the conformation was selected by BRACO-19 a known small molecule 'driver' that was included in the crystal mixture (Li et al., 2005, Han et al., 1999b). Therefore the crystal structure may not be the most appropriate model to use for the structure based design and other physiologically more relevant structures may predominate in solution (Li et al., 2005), particularly as it is not uncommon for the topology of G-quadruplex structures obtained by crystallography to exhibit some variation from those obtained in solution (Keniry, 2000, Li et al., 2005).

1.5.1.1 Human telomeric hybrid structures

As potassium is the major intracellular cation (~150 mM), determination of a 'solution' structure of this sequence in K^+ was highly sought. Xu *et al*. (Xu et al., 2006) substituted the guanine bases in the 22 nucleotide sequence with 8-bromoguanine and examined the resultant structure and its thermal stability by CD spectroscopy. They proposed a mixed parallel/antiparallel strand configuration containing two edgewise and one double-chain reversal loop. Following this there have been several NMR studies on the structure of this sequence, within the context of different flanking bases. So far, these studies have elucidated a few types of hybrid structures; Hybrid-1, Hybrid-2 and Hydrid-3 (Figure 1.7) (Ambrus et al., 2006, Dai et al., 2007a, Lim et al., 2009).

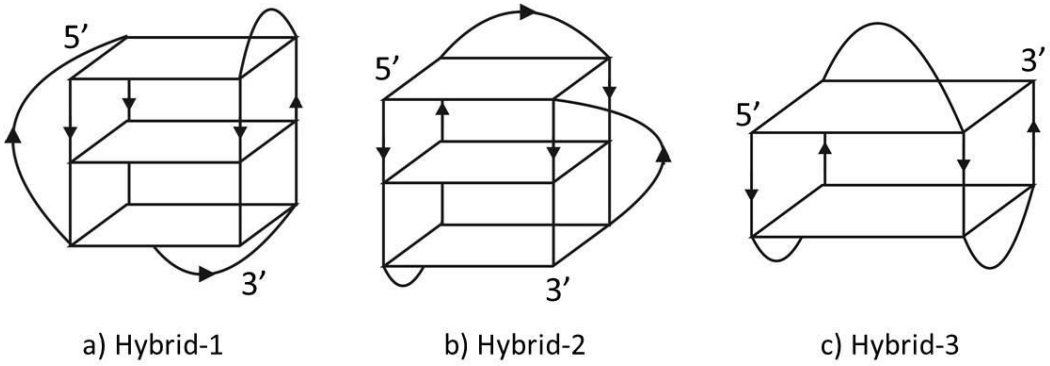


Figure 1.7 Schematic representation of NMR analysis of the folding topologies of the human telomeric sequence in the presence of potassium. a) Hybrid-1, b) Hybrid-2 and c) Hybrid-3(Ambrus et al., 2006, Dai et al., 2007a, Lim et al., 2009).

The 24 nucleotide sequence d[TTGGG(TTAGGG)$_3$A] investigated by Luu *et al*. (Luu et al., 2006) and the extended 26 nucleotide sequence d[AAAGGG(TTAGGG)$_3$AA] known as Tel26, investigated by Ambrus *et al.* (Ambrus et al., 2006) each exhibited a 3+1 hybrid-type G-quadruplex known as Hybrid-1. The Hybrid-1 structure consists of a mixed parallel/antiparallel-stranded conformation with the first, second and fourth strand oriented parallel to one another and the third strand lying antiparallel to the other three (Figure 1.7a). The first and second 5' guanine-strands are linked with a double-chain-reversal TTA loop, whilst the second and third, and the third and fourth strands are linked with two TTA edgewise loops (Figure 1.7). The three G-quartets have mixed guanine-arrangements, with the top quartet adopting a *syn-syn-anti-syn* configuration

whilst the bottom two quartets are in an *anti-anti-syn-anti* arrangement. This unusual folding topology suggests that it might be selectively targeted by G-quadruplex-interactive small molecule drugs. It was also noted that this arrangement allows for a compact-stacking for multiple structures adopting the same conformation. This is because the 5'- and 3'-ends of the Hybrid-1 telomeric G-quadruplex structure point in opposite directions, allowing the Hybrid-type G-quadruplex to readily fold and stack end-to-end in an elongated single-stranded linear sequence (Figure 1.8).

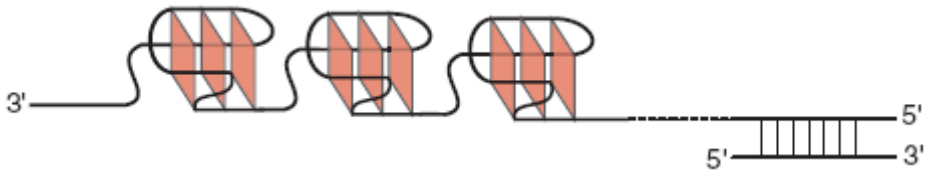


Figure 1.8 A schematic model of G-quadruplex formation on the human telomeric DNA sequence Tel26.This hybrid-type telomeric G-quadruplex structure can readily fold and stack end-to-end to form a compact-stacking structure for multimers of this conformation in an elongated linear manner (Ambrus et al., 2006)

Similarly, the terminal adenines in the flanking sequence of the Tel26 molecule (Figure 1.9) imitate this stacking interaction. The extended aromatic ring system (Ambrus et al., 2006) of adenines provides greater stacking interactions than pyrimidines (Ambrus et al., 2005, Phan et al., 2004a) therefore offering additional stabilising properties to the G-quadruplex structure (Ambrus et al., 2006, Dai et al., 2007b). The basket-type G-quadruplex conformation stabilised by the sodium cation is less favoured for consecutive end-to-end stacking in extended sequences as its 5'- and 3'-ends along with the diagonal loop all point in the same direction. As a result the long length and repetitive nature of the human telomeric sequence suggests the Hybrid-1 structure is naturally more physiologically relevant (Ambrus et al., 2006).

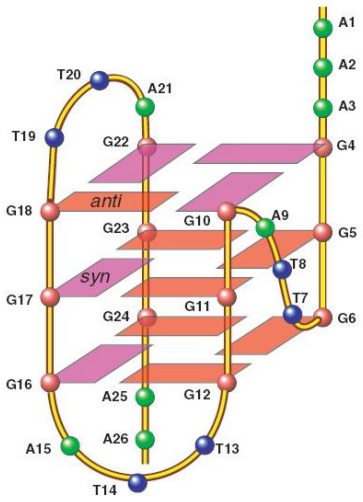
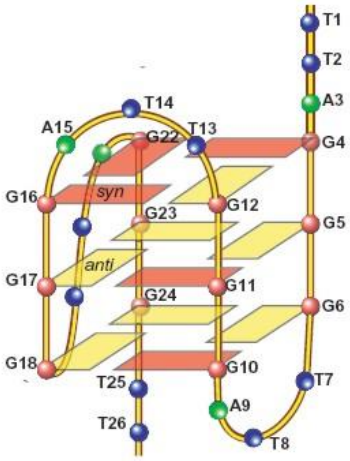


Figure 1.9 Schematic diagram of the Tel26 Hybrid-1 folding topology of the intramolecular human telomeric G-quadruplex in K^+ solution. . Red ball, guanine; red box, (anti) guanine; magenta box, (syn) guanine; green ball, adenine; blue ball, thymine (Ambrus et al., 2006)

In contrast, NMR analysis of the potassium solution structure of Tel26 reveals an alternative conformation known as Hybrid-2 (Figure 1.7b) (Phan et al., 2006, Dai et al., 2007a). This structure still maintains the same 3+1 core, but differs in loop arrangements.

Altering the terminal d(AA) flanking sequences, to d[TAGGG(TTAGGG)$_3$TT] (wtTel26) formed a double-chain-reversal loop on the third linker rather than the first (Figure 1.10) as in the case of the Hybrid-1. The Hybrid-2 assembly contains a novel T:A:T triplet capping structure which covers the bottom end of the G-quadruplex (Figure 1.10). This is formed by the T8 and A9 of the first d(TTA) edgewise loop and the T25 of the 3'-flanking segment which allows for hydrogen-bonds between T8 and T25 which also appears to play an important role in the selective stabilisation of this structure.



Hybrid-2

Figure 1.10 Schematic diagram of the wtTel26 Hybrid-2 folding topology of the intramolecular human telomeric G-quadruplex in K^+ solution. Red ball, guanine; red box, (anti) guanine; magenta box, (syn) guanine; green ball, adenine; blue ball, thymine (Dai et al., 2007a)

In a recent study, NMR analysis of the potassium solution structure of the telomeric sequence d[GGG(TTAGGG)$_3$T] revealed that the sequence forms a Hybrid-3 fold topology (Lim et al., 2009). This structure differs from both the Hybrid-1 and Hybrid-2 structures as it contains an antiparallel intramolecular "basket-type" G-quadruplex comprising of only two stacked G-quartets, where each quartet has the arrangement *syn-syn-anti-anti* (Figure 1.11) (Lim et al., 2009). There have been several other reports of human telomeric sequences starting with 5' guanine for example; d[GGG(TTAGGG)$_3$], d[GGG(TTAGGG)$_3$TT] and d[GGG(TTAGGG)$_3$TTA] which have also been shown to form the Hybrid-3 structure (Lim et al., 2009).

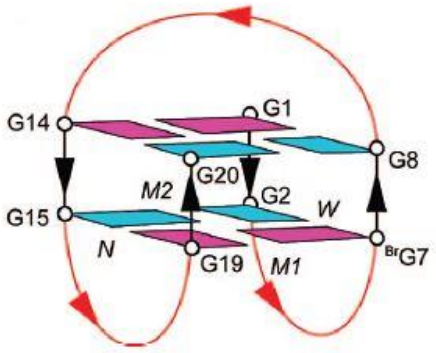


Figure 1.11 A schematic diagram of the Hybrid-3 folding topology of the intramolecular human telomeric G-quadruplex in K^+ solution. The *anti* and *syn* guanines are coloured blue and pink respectively. Adapted from (Lim et al., 2009).

Whilst the basket form is predominantly used as the reference for the G-quadruplex structure in sodium solution, a recent study has reported a new antiparallel (2+2) G-quadruplex topology. The 27 nt human telomeric sequence *htel27[Br22]* d[$(TTAGGG)_3TTA(^{Br}G)GGTTA$] forms a three stacked G-quartet scaffold, with two edgewise loops (first and third), however this conformation differs from the basket-type topology, where the second loop spans diagonally across the top of the terminal tetrad (Figure 1.9) and instead forms a double-chain-reversal loop (Figure 1.12) (Lim et al., 2013) and the glycosidic conformations of the guanines are *anti-anti-syn-syn*.

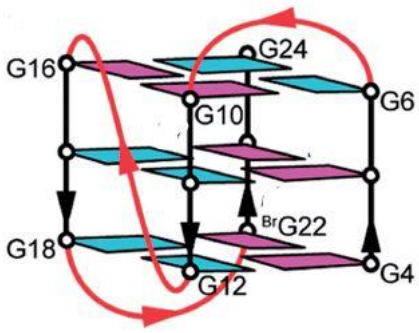


Figure 1.12 Schematic diagram *htel27[Br22]* antiparallel (2+2) G-quadruplex. The *anti* guanines are coloured blue and the *syn* and 8-bromoguanine guanines are coloured pink. Adapted from (Lim et al., 2013).

1.5.2 Folding Topologies of Other Telomeric Sequences

1.5.2.1 Tetrahymena

The telomeric repeat sequence of the ciliate protozoa *Tetrahymena* differs from the human telomeric sequence by one guanine for an adenine $(TTGGGG)_n$ and has been reported to adopt different folding topologies. The sequence d(TTGGGGT) folds intermolecularly in an all parallel-stranded configuration in the presence of K^+ (Wang and Patel, 1993a) with all guanines in an *anti* conformation, whereas the two repeat sequence d(TGGGGTTGGGGT) forms two asymmetric bimolecular G-quadruplex structures in sodium-containing solution (Phan et al., 2004b). The first structure folds in a head-to-head configuration where the two loops are at one end of the G-tetrad core, whilst the second structure folds head-to-tail with the two loops positioned on opposite ends of the G-tetrad core.

In contrast, the four-repeat sequence $(TTGGGG)_4$ forms an intramolecular G-quadruplex in the presence of Na^+ with two edgewise loops (first and second) and one double-chain-reversal (third) (Wang and Patel, 1994). Although this sequence has the potential to form structures consisting of four stacked G-quartets, it instead contains only three with the following loop arrangements: the first loop d(GTTG), the second loop d(TTG) and the third d(TT) (Wang and Patel, 1994). This structure contains a (3+1) G-tetrad core structure in which three strands are oriented parallel to each other and the fourth is antiparallel to the other three.

1.5.2.2 Oxytricha nova

The G-quadruplex-forming sequence of the marine ciliate *oxytricha nova* telomeric motif $d(GGGGTTTT)_n$ is another commonly studied telomeric sequence. Both X-ray crystallography (Haider et al., 2002, Horvath and Schultz, 2001) and NMR (Schultze et al., 1999, Smith and Feigon, 1992) studies have shown that in the presence of K^+ or Na^+, the two repeat sequence d(GGGGTTTTGGGG) forms a bimolecular, antiparallel, four G-quartet, G-quadruplex structure containing two diagonal T_4-loops (Schultze et al., 1994). NMR studies have shown that the modified sequence $d(G_3T_4G_3)_2$ also forms the same antiparallel bimolecular diagonally looped quadruplex, whilst X-ray

crystallography analysis of the d($G_4T_3G_4$) sequence and the brominated analogue d($G_4{}^{Br}UTTG_4$) show that they form head-to-head dimers with two edgewise loops (Hazel et al., 2006), strengthening the hypothesis that T_4-loops form diagonal loops while loop sequences of two or three thymines favour edgewise loops, presumably as a result of conformational constraints.

Reducing the number of guanines in either the 3'-tract; d($G_3T_4G_4$) (Crnugelj et al., 2003) or the 5'-tract; d($G_4T_4G_3$) (Crnugelj et al., 2002) both produce asymmetric folded structures. d($G_3T_4G_4$) produces a three G-quartet structure independent of cation type (Na^+ or K^+), that contains one diagonal and one edgewise loop, three parallel strands and one antiparallel strand with a guanine overhang at the 5'- and 3'-end. The sequence d($G_4T_4G_3$) with the 3' terminal dG residue missing from the 3'-end also forms a structure that consists of three stacked G-quartets, however, it exhibits a "slipped-loop" with two asymmetric diagonal loops with two additional dG residues not participating in the G-quartet formation. This folding topology was only observed with Na^+ and $NH_4{}^+$ and not K^+. The structural polymorphism observed by these telomeric sequences shows the dynamic nature of bimolecular scaffolds and also offers insight into the possible folding topologies of genomic sequences in which consecutive guanine tracts tend to be irregular.

1.6 Telomere End-Binding Proteins (TEBPs)

In vitro studies of telomeric G-quadruplex structures have demonstrated that they exist in a slow equilibrium between the highly stable four-stranded structure and the linear single-stranded form (Lane et al., 2008). It is therefore postulated that the promotion, formation and resolution of G-quadruplexes *in vivo* must be carefully regulated by specific proteins. Telomere end-binding proteins (TEBPs) bind to the G-rich single-strand overhang of telomeres and these proteins have been identified in almost every telomere species. The heterodimeric telomere binding protein of *Oxytricha nova* consists of the α- (56 kD) and β-subunits (41 kD) that bind specifically to the single-stranded d($T_4G_4T_4G_4$) overhang of each macronuclear DNA terminus. This forms a cap at the end of the chromosome protecting it from nuclease degradation or inappropriate repair mechanisms (Gottschling and Cech, 1984, Gottschling and Zakian, 1986), which

as a result inhibits telomere extension by telomerase (Froelich-Ammon et al., 1998). Under physiological conditions, *in vitro* studies have demonstrated that the β-subunit of the *Oxytricha nova* telomere-binding protein acts as a molecular chaperone that recognises and binds to the single-stranded telomeric overhang to promote the formation of tetramers and is therefore able to promote and accelerate the association of the sequences $(TTTTGGGG)_2$ and $(TTGGGG)_2$ into G-quadruplex structures (Cashman et al., 2008, Borman, 2009). In a similar manner, the budding yeast *Saccharomyces cerevisiae* protein RAP1 binds sequence-specifically to the guanine-thymine strand of the telomeric yeast sequences. This evidence shows that the protein also binds to and promotes the formation of G-quadruplexes (Qin and Hurley, 2008, Berberich and Postel, 1995).

Conversely, proteins that disrupt the formation of G-quadruplexes in favour of promoting telomere extension have also been identified; one of the most extensively studied is the human chromosome protection of telomeres 1 (hPOT1). Although the extreme 3' terminus of telomeres ranges between 100-200 nucleotides (approximately 16-33 repeats), most studies conducted on the human telomerase activity *in vitro* typically use synthetic DNA oligonucleotides, containing no more than three repeats. Zaug and co-workers (Zaug et al., 2005) demonstrated that, as soon as the DNA sequence includes a fourth block of guanines, as in $d[GGG(TTAGGG)_3]$, telomerase extension stalls because of the formation of an intramolecular G-quadruplex. They also showed that addition of hPOT1 generates a stable single-stranded DNA-hPOT1 complex and the G-quadruplex structure is disrupted, rescuing extension by telomerase. Aside from the hPOT1, other proteins that act as G-quadruplex resolvases have also been revealed, including the human RecQ helicases, Bloom's (Sun et al., 1998) and Werner's syndrome (Fry and Loeb, 1999) DNA helicases. These have also been shown to unwind intra and intermolecular G-quadruplexes of G-rich sequences in a process that is dependent on both ATP and Mg^{2+}.

1.7 G-quadruplexes and Gene Regulation

The human genome contains many G-rich regions with potential G-quadruplex-forming motifs. Many of these motifs are over-represented in gene promoter regions, yet unlike

the 3' telomeric sequences they are not neat tandem repeats of single-stranded DNA, but instead are irregular in nature with repetitive G-tracts varying in number and length, which normally exist as Watson-Crick base pairs (Huppert and Balasubramanian, 2005, Todd et al., 2005, Huppert and Balasubramanian, 2007, Todd and Neidle, 2011, Balasubramanian et al., 2011). Genomic G-quadruplexes have been implicated in a number of key biological functions such as transcription and replication (Huppert and Balasubramanian, 2005, Todd et al., 2005) hence the identification of human genes with relatively high or low potential to form these structures may provide insight into structural assembly, function and identify mechanisms that could account for genomic instability in human malignancies (Eddy and Maizels, 2006). The conformation of G-quadruplexes provides selective recognition sites that might allow for development of small molecules that can target and specifically bind to these DNA structures *in vivo* (Balasubramanian and Neidle, 2009, Qin and Hurley, 2008, Luedtke, 2009), although it is important to note that evidence to support their formation *in vivo* is still largely circumstantial. Many G-quadruplex-forming motifs have been found to be concentrated in regions immediately upstream of transcription start sites (TSS) (Huppert and Balasubramanian, 2005, Todd et al., 2005) and conservation of the sequences that form stable intramolecular G-quadruplexes has been observed in a number of human oncogenes, as summarised in Table 1.2:

Table 1.2 Some potential G-quadruplex-forming motifs identified in oncogenic promoter sequences.

Gene	**Sequence (5′-3′)**							**Reference**
VEGF	**GGG**	T	**GGG**	GA	**GGG**	T	**GGG**	(Sun et al., 2005, Agrawal et al., 2013)
PDGFB	**GGG**	T	**GGG**	T	**GGG**	CTCT	**GGG**	(Qin et al., 2007, Chen et al., 2012)
RET	**GGG**	C	**GGG**	GCG	**GGG**	C	**GGG**	(Sun et al., 2003, Tong et al., 2011)
Hif-1α	**GGG**	A	**GGG**	GAGAGG	**GGG**	C	**GGG**	(De Armond et al., 2005)
k-ras	**GGG**	AGA	**GGG**	AGA	**GGG**	G	**GGG**	(Cogoi et al., 2008)

This section discusses some of the most commonly researched G-quadruplexes in oncogenic promoter sequences.

1.7.1 Chicken-β-globin Gene

The chicken-β-globin gene provided the earliest evidence for a relationship between the formation of secondary DNA structures and transcriptional changes in chromatin structure and gene activity (Dexheimer et al., 2006a). However, the nature of this was not fully understood until the mid-90s when Woodford *et al*. (Woodford et al., 1995, Woodford et al., 1994) proposed the possible formation of a G-quadruplex structure within the G-rich region of its promoter. Using a primer extension assay, they described a novel potassium-dependent DNA synthesis arrest site in the chicken-β-globin gene G-rich DNA motif d[G_{16}CG(GGT)$_3$]. However, the results presented were not enough to suggest that the arrest of DNA synthesis was due to the formation of G-quadruplexes. With the aid of chemical footprinting, further analysis of the structure by Howell *et al*. (Howell et al., 1996) on a derivative sequence d[G_{16}CG(GGT)$_2$GG] revealed a "cinched" intramolecular quadruplex structure containing five stacked quartets. However, the model differed from the canonical G-quadruplex assembly and only contained one complete G-quartet, whilst the others were incomplete containing non-guanine bases. The authors then proposed that hydrogen-bonding interactions between the loop bases aid the stability of the structure along with the base pairing of the 5'- and 3'- flanking sequences which form a molecular "cinch" (Figure 1.13).

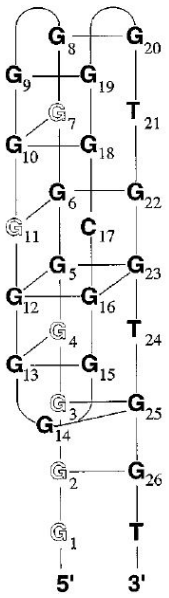


Figure 1.13 Intramolecular "cinched" G-quadruplex structure as proposed by Howell *et al.* (Howell et al., 1996) Bases adjacent to the four-stranded quadruplex structure interact to stabilize the structure.

1.7.2 c-myc

The human *c-myc* proto-oncogene is a transcription factor that is integral to cell growth, proliferation and apoptosis. Expression of the *c-myc* gene is tightly regulated in normal cells, and over expression of this gene is associated with a significant number of human cancers (Dang, 1999, Marcu et al., 1992). The nuclease-hypersensitivity element III_1 (NHE III_1) is an important region within the promoter and is responsible for ~ 85% of *c-myc*'s transcriptional activity. The NHE III_1 is a 27 nucleotide sequence located 115 base pairs upstream from the P1 promoter (Berberich and Postel, 1995). The coding stranding is pyrimidine-rich whilst the non-coding strand is purine-rich. The 27 nucleotide G-rich strand (*Pu27*) consists of six separate and unequal guanine tracts; three runs consisting of four guanines each, two runs consisting of three guanines and one run consisting of two guanines (Table 1.3). *In vitro* studies of this region have revealed that the NHE III_1 sequence experiences sensitivity to digestion by the single-strand specific S1 nuclease because it is in a slow equilibrium between the classic duplex helical structure and non-B-form regions of DNA, (Boles and Hogan, 1987).

A G-rich oligonucleotide, complementary to the coding strand of the *c-myc* gene targeted to the NHE III_1 was first found to reduce *c-myc* transcription in HeLa cell extracts (Postel et al., 1991). The same oligonucleotide was later shown to inhibit *c-myc* transcription in live HeLa cells (Olivas and Maher, 1995).

1.7.2.1 Folding topology of c-myc

Over the years there have been conflicting reports on the precise folding topology of the 27 nucleotide G-quadruplex sequence of the *c-myc* NHE III_1, due to the fact that it consists of six unequal G-tracts. The irregular nature of this sequence means that it has the potential to exhibit numerous folding topologies involving any of the six G-tracts, making it difficult to characterise a precise structure. In studies conducted by Simonsson *et al*. (Simonsson et al., 1998) they deduced that only the first, second, fourth and fifth G-tracts contributed to the G-quadruplex formation. Whilst, Siddiqui-Jain and co-workers (Siddiqui-Jain et al., 2002) used DMS footprinting to identify two major *c-myc* quadruplex structures that coexist. The first was a "basket-type" conformation consisting of two edgewise loops and a central diagonal loop, whilst the second structure adopted the "chair" configuration comprising of all edgewise loops.

NMR analysis of modified and truncated sequences derived from the G-rich strand of the NHE III_1 revealed an alternate double-chain reversal structure different to the basket and chair topologies previously proposed (Phan et al., 2004a). Both myc-2345 and myc-1245 (Table 1.3) contain four guanine tracts and form a three quartet stacked intramolecular parallel-stranded G-quadruplex in the presence of K^+. Each guanine is oriented in the same direction; with all in the *anti* glycosidic bond conformation and all three loops are double-chain-reversal. The loop compositions of these two structures are however different; the myc- 2345 has a di-nucleotide central loop (GA), while the two flanking loops contain only a single nucleotide (T). In the structure of myc-1245, each of the two flanking loops contains a single nucleotide (first loop containing adenine and the second thymine), while the central loop contains six residues (T_5A) (Phan et al., 2004a) (Table 1.3).

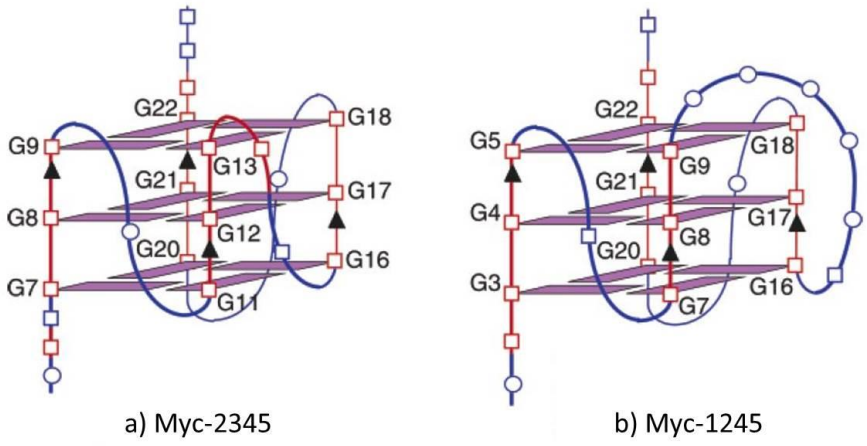


Figure 1.14 Schematic structures of human c-myc promoter: a) Myc-2345 and b) Myc-1245**.** Modified from (Phan et al., 2004a). Red lines indicate G-quadruplex strands, whilst blue lines indicate G-quadruplex loops. Guanine bases are represented by red squares, adenine by blue squares and thymine by blue circles.

Studies conducted on a 24 nucleotide NHE III_1 sequence consisting of a five-guanine-tract G-quadruplex motif (Phan et al., 2005) reveal a distinct parallel-stranded topology in K^+ solution that is composed of a three G-quartet stacked structure with three double-chain-reversal loops. However, this motif differs from the previously reported topologies as the G24 guanine residue from the 3'-end is "plugged" back into the G-tetrad core. By participating in the G-tetrad formation it displaces another guanine (G10) from a continuous guanine tract forcing it into a loop. This configuration is stabilised by a diagonal loop which contains a G:A:G triad stacking on the bottom of the structure which caps the G-tetrad core.

Yoon and co-workers (Yoon et al., 2010) were able to address this structural ambiguity by using single-molecule fluorescence resonance energy transfer (FRET) techniques. By preparing various partial DNA duplexes with 18 base pairs in the duplex region and 23 bases of the *c-myc* $NHEIII_1$ sequence containing five tracts of guanines in the single-stranded region they concluded that all the guanine regions were involved in the formation of different G-quadruplexes and that it is possible that physiologically relevant G-quadruplex structures can be dynamically selected depending on specific conditions such as ionic concentration or temperature of the cell.

Table 1.3 DNA sequences of the NHE III_1 G-rich strand from the c-myc promoter and its truncated and modified derivatives: **a)** c-myc (*Pu27*): the 27 nucleotide wild-type sequence; **b)** myc-1245: a "basket-type" topology containing G-tracks number 1, 2, 4, and 5; the G-track number 3 is substituted by T_4, **c)** myc-2345: a "chair-type" topology containing G-tracks number 2, 3, 4, and 5. T1-T6 correspond to the G-tracts and **d)** Pu24: Five G-tract topology containing tracts 2, 3, 4, 5, and 6. Structural modifications highlighted in bold. G-residues participating in G-quartets are underlined. [a] This sequence was studied in reference (Phan et al., 2004a), [b] This sequence was studied in reference (Phan et al., 2005)

Name	Sequence
	T1 T2 T3 T4 T5 T6
a) c-myc (Pu27)	5'-TGGGGAGGGTGGGGAGGGTGGGGAAGG -3'
	1 5 10 15 20 25
b) Myc-1245[a]	TGGGGAGGG**TTTTTT**AGGGTGGGGA
	1 5 10 15 20
c) Myc-2345[a]	**T**GAGGGTGGGGAGGGTGGGGAA
	5 10 15 20 25
d) Pu24[b]	**T**GAGGGTGGGGAGGGTGGGGAAGG
	5 10 15 20 25

1.7.3 c-kit

The *c-kit* proto-oncogene encodes a 145-160 kDa membrane-bound glycoprotein belonging to a family of tyrosine kinase receptors, which regulates key signal changes that control cell growth and proliferation (Tuveson et al., 2001). Due to the critical role that is played by the *c-kit* protein, any mutations or overexpression could be detrimental to its function. Mutations resulting from the over expression of the *c-kit* oncogene have been related to human gastrointestinal stromal tumours (GIST), making the *c-kit* proto-oncogene an extremely important therapeutic target in treating GIST. Two G-rich sequences separated by 31 nucleotides; *c-kit1* and *c-kit2* have been identified in the promoter region of the human *c-kit* gene, and biophysical data such as NMR and CD, have confirmed that these sequences can readily adopt G-quadruplex structures (Rankin et al., 2005, Fernando et al., 2006). The 22-mer wild-type *c-kit1* quadruplex-forming sequence d(AGGGAGGGCGCTGGGAGGAGGG), which is 87 nucleotides upstream from the transcription start site of the human *c-kit* gene, forms a single G-quadruplex species in K^+ solution (Rankin et al., 2005). The sequence contains four tracts of three consecutive guanines separated by linkers of one (A) or four (CGCT and AGGA) residues and although the sequence consists of four consecutive tracts of guanines it does not form the archetypal G-quadruplex structure with G-tracts building the G-tetrad core and the linkers building the loops. Biophysical analysis of modifications to the *c-kit1* quadruplex linker regions (Rankin et al., 2005) resulted in a significant reduction in

the formation of the G-quadruplex structure, demonstrating that a sequence with high guanine content is not always a sufficient determinant of G-quadruplex formation, but that loop sequence and length also play a crucial role by contributing cooperatively (to some extent) to G-quadruplex formation and stability (Rankin et al., 2005, Phan et al., 2007). Initial studies of a modified sequence containing four equally sized G-tracts, suggest that the assembly would involve a typical "parallel-type" formation, but NMR analysis (Phan et al., 2007) revealed a unique four-looped structure which is conserved in the crystal structure (Wei et al., 2012). This structure (Figure 1.15) consists of a three G-tetrad core, with one of the loop guanines flipping back to participate in one of the quartets. It also contains two single-nucleotide double-chain-reversal loops spanning three G-quartet layers, one lateral loop and a novel five-nucleotide (AGGAG) loop that terminates at the guanine insertion (Figure 1.15). This five-nucleotide loop forms a cleft in the structure that has possible drug binding implications (Rankin et al., 2005, Phan et al., 2007).

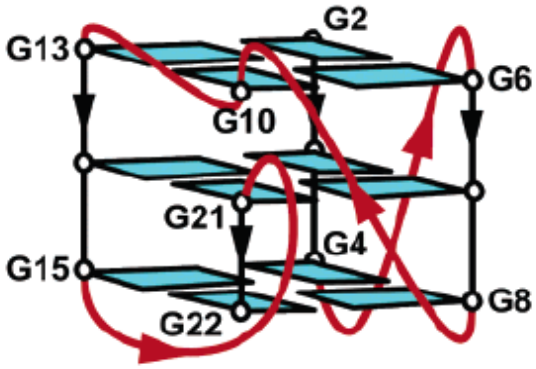


Figure 1.15 Topology of *c-kit1*scaffold**.** The guanine bases are all in the *anti* conformation indicated in blue and the loops are indicated in red (Phan et al., 2007).

Conversely, the 21-mer wild-type c-kit2 G-quadruplex-forming sequence d(CGGGCGGGCGCGAGGGAGGGG) is more complex. This sequence is unable to form a single G-quadruplex structure in solution without mutations to remove sequence ambiguity (Fernando et al., 2006). Each structure appears to adopt an all parallel strand topology with three double-chain-reversal loops, where the first, second and third loops consist of d(C), d(CGCGA) and d(A) respectively (Hsu et al., 2009, Kuryavyi et al., 2010) as depicted in Figure 1.16.

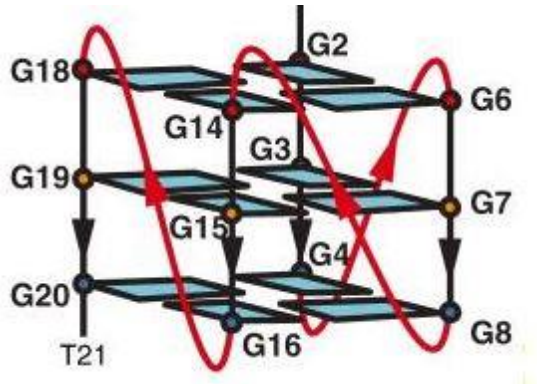


Figure 1.16 Topology of *c-kit2* G-quadruplex fold. The structure is parallel folded, with all double-chain-reversal loops. The guanine bases are all in the *anti* conformation indicated in blue and the loops are indicated in red (Kuryavyi et al., 2010).

1.7.4 Bcl-2

The B-cell lymphoma gene 2 (*Bcl-2*) is a member of a family of proteins that are central regulators of apoptosis because they incorporate a variety of survival and death signals that are generated outside and inside the cell (Borner, 2003). One of the unique features of the *Bcl-2* protein family is heterodimerisation between anti-apoptotic and pro-apoptotic proteins, which is considered to inhibit the biological activity of their partners (Oltvai et al., 1993, Yang et al., 1995, Tsujimoto and Shimizu, 2000). It encodes a 25 kDa membrane protein that functions to prevent apoptosis (Vaux et al., 1988). Deregulation of the *Bcl-2* proto-oncogene can lead to overexpression in cancer cells, as well as to autoimmune and degenerative diseases (Adams and Cory, 2007, Cory and Adams, 2002). The *Bcl-2* gene contains several promoter regions, with P1 and P2 promoters being the main regulators of transcription of the *Bcl-2* gene. In numerous cell types a considerable amount of *Bcl-2* transcripts are derived from the P1 promoter, whereas the P2 which is down-regulated by the p53 protein shows little or no activity. However, usage of the P2 promoter is activated in lymphoma cells (Wu et al., 2001). At the 5'-end, located upstream of this regulatory region (-58 to -19 base pairs) is a region of DNA that is highly GC-rich (Table 1.4) (Young and Korsmeyer, 1993). This 39 nucleotide sequence (*bcl2Pu39*) contains six guanine tracts, with one run of five guanines, two runs of four guanines each, and three runs of three guanines that in the presence of potassium forms a mixture of stable intramolecular quadruplexes (Table 1.4) (Dai et al., 2006b). The G-quadruplex formed from the middle four consecutive

guanine runs (*bcl2MidG4Pu23* – tracts II to VI) is the most stable (Dai et al., 2006a, Dexheimer et al., 2006b, Dai et al., 2006b) and is suggested to be the major G-quadruplex structure formed in the *Bcl-2* promoter region. The *bcl2MidG4Pu23* (Table 1.4) contains a run of five guanines, that can form different loop isomers. Each of these can be isolated by introducing specific G-to-T mutations which have been shown to adopt a mixed parallel/antiparallel-stranded G-quadruplex, creating a structure with the same (3+1) core topology as that of the four-repeat *Tetrahymena* telomeric sequence. NMR, CD and DMS footprinting analyses have shown that the G15T/G16T dual mutant was the most stable (Table 1.4) (Dai et al., 2006b).

Table 1.4 Promoter sequences of the *bcl-2* gene and its modifications. The top sequence is the wild-type *bcl-2* 39-mer sequence with the six G-runs (red). *Bcl2MidG4Pu23* represents the 23mer sequence containing the middle four consecutive runs of guanines, which forms the most stable G-quadruplex structure. *Bcl2MidG4Pu23-G15T/G16T* represents the mutant with G15 and G16 to-T15 and T16 mutations (Dai et al., 2006b)

	Sequence
Bcl-2 (bcl2Pu39)	I II III IV V VI 5’-AGGGGCGGGCGCGGGAGGAAGGGGGCGGGAGCGGGGCTG
bcl2MidG4Pu23	5’ –GGGCGCGGGAGGAAGGGGGCGGG
bcl2MidG4Pu23-G15T/G16T	5’ –GGGCGCGGGAGGAATTGGGCGGG

1.8 RNA Quadruplexes

While the focal point of this thesis is the formation of DNA G-quadruplexes under negatively supercoiled conditions, it is also important to mention that G-rich RNA sequences can also assemble into these tetramolecular structures. Although RNA quadruplexes have not been studied as extensively as DNA quadruplexes, biophysical studies on various RNA quadruplex-forming sequences have shown that RNA structures are relatively more stable than their DNA counterparts and exhibit faster association and slower dissociation rates (Mergny et al., 2005). To date there are only a small number of high resolution structures available for RNA quadruplexes, however in most, if not all studies it is found that these structures adopt parallel conformations (Joachimi et al., 2009, Martadinata and Phan, 2009, Mergny et al., 2005) that are independent of

surrounding conditions (Zhang and Zhi, 2010, Zhang et al., 2010). This is in contrast to the polymorphic nature exhibited by DNA quadruplex structures due to the fact that RNA quadruplexes only permit the *anti* conformation of ribose and guanine bases (Halder and Hartig, 2011). RNA quadruplexes have been speculated to form more readily *in vivo* unlike genomic DNA structures as these are derived from transcribed genes and are essentially single-stranded therefore eliminating competition between the duplex formation of the complementary C-rich strand (Halder and Hartig, 2011, Joachimi et al., 2009, Neidle, 2012).

1.8.1 Stability of RNA G-quadruplexes RNA Quadruplexes in Telomeres

Loop size and composition play a key role in stabilising RNA quadruplex structures. Just as single thymine loops form the most stable DNA G-quadruplexes, similar results were found with RNA structures with single uracil loops forming the most stable complexes (Halder and Hartig, 2011, Halder et al., 2009, Martadinata and Phan, 2009, Mergny et al., 2005, Xu et al., 2008). Similar to DNA quadruplexes, short loop sequences have higher stability and as loop length increases the stability of the structure decreases (Joachimi et al., 2009). It was also noticed that the shorter looped RNA structures formed more stable structures compared to their equivalent DNA counterparts, but in contrast longer looped DNA sequences tend to exhibit greater stability than long-looped RNA structures. The loss of stability by the longer looped structures may be a consequence of RNA quadruplexes propensity to fold into parallel topologies (Joachimi et al., 2009, Pandey et al., 2013). Additionally, structures containing three stacked G-quartets are considerably more stable than those with two (Halder and Hartig, 2011, Joachimi et al., 2009, Wieland and Hartig, 2007, Pandey et al., 2013). In the same way that DNA G-quadruplexes require a centrally located cation to stabilise the structure this is also necessary for RNA quadruplex formation. In general RNA quadruplexes are selectively stabilised by potassium over sodium with a 10-30 ºC difference in melting temperature (depending on loop length and sequence) compared to DNA structures under similar conditions (Halder and Hartig, 2011, Joachimi et al., 2009).

1.8.2 RNA Quadruplexes in Telomeres

For many years it was considered that telomeres were transcriptionally inactive, however recent analysis has demonstrated that telomeres are regularly transcribed by RNA polymerase II, (Luke and Lingner, 2009, Schoeftner and Blasco, 2009) giving rise to short single-stranded telomeric repeat-containing RNAs (TERRA). TERRAs are large non-coding RNAs which form an integral component of telomeric heterochromatin although their function and role still remains unclear (Azzalin et al., 2007, Schoeftner and Blasco, 2008). In a recent study, a human TERRA RNA G-quadruplex structure was found to localise to chromosome ends in cell nuclei, suggesting a possible association between TERRA RNA and telomeric DNA (Xu et al., 2010). In another study, a long TERRA RNA fragment of approximately 640 nucleotides exhibited a beads-on-a-string model for the long telomeric sequence, where individual $r(UUAGGG)_4$ units formed parallel quadruplex beads, comprising of UUA loops with each structure connected by UUA linkers (Randall and Griffith, 2009). Recently, the bimolecular telomeric RNA quadruplex of sequence r(UAGGGUUAGGGU) was described by NMR (Martadinata and Phan, 2009) and crystallography (Collie et al., 2010) and both analysis show that in the presence of potassium ions this sequence forms a parallel strand topology, which is in agreement with studies using CD spectroscopy (Martadinata and Phan, 2009, Collie et al., 2011) and unlike the DNA equivalent this topology is conserved in the presence of sodium (Xu et al., 2008) and also in the four repeat sequence (Martadinata and Phan, 2009).

1.8.3 mRNA Quadruplexes

Bioinformatics searches of the human genome have also revealed G-rich sequences in untranslated regions (UTRs) of messenger RNA (mRNA) which is speculated to promote transcriptional termination, cleavage and polyadenylation (Huppert et al., 2008, Wieland and Hartig, 2007). The majority of studies are conducted on the 5'-UTR sequences, where like genomic DNA quadruplexes such as the *Bcl-2* (Shahid et al., 2010), *VEGF* (Morris et al., 2010) and TRF2 (Gomez et al., 2010) many have uneven G-tracts. However, unlike DNA quadruplexes all appear to form parallel structures. Small molecules that target G-quadruplexes in the 5'-UTR of RNA transcripts have shown that by binding to the G-quadruplex structures translational activity is significantly reduced (Beaudoin and Perreault, 2010, Bugaut and Balasubramanian,

2012). *In vitro* transcription and translation assays of the TRF2 quadruplex showed that whilst transcription was unaffected a decrease in the translation was observed due to the presence of a G-quadruplex structure (Gomez et al., 2010). Similarly the NRAS (Kumari et al., 2008, Kumari et al., 2007) and the oestrogen receptor α (Derecka et al., 2010) quadruplex-forming sequences have also been shown to repress translation when located closer to the 5'-end of the UTR of the genes.

1.9 Bioinformatics

The availability of relatively complete genomic datasets coupled with high performance bioinformatics tools allows the search for other sequences that have the potential to form G-quadruplex structures in human and other genomes *in vivo* (Eddy and Maizels, 2006, Huppert, 2008, Huppert and Balasubramanian, 2005, Huppert et al., 2008, Todd, 2007, Todd et al., 2005, Todd and Neidle, 2011, Hershman et al., 2008, Du et al., 2007, Beaudoin et al., 2013). This is particularly useful when designing potential therapeutics that can specifically target these structures. G-rich sequences that have the potential to form these tetramolecular structures are often referred to as potential/putative quadruplex-forming sequences (PQS). Yet, predicting three-dimensional folding of G-quadruplexes from linear genomic sequences using bioinformatics analysis can be extremely problematic. This is because the polymorphic nature of these structures makes it difficult to distinguish which bases will or will not participate in the quadruplex structure (Todd, 2007). This section describes some of the studies that have been used to identify the majority of sequences exploited in G-quadruplex research today.

1.9.1 Prevalence of G-quadruplexes in the Human Genome

For the identification and investigation of novel G-quadruplex-forming sequences, it is first essential to develop an algorithm for recognising these putative quadruplex-forming sequences (Huppert, 2006, Huppert, 2008). An early study conducted by Todd *et al*. (Todd et al., 2005) defined a PQS as a sequence with four runs of guanines each between three and five bases long, separated by segments of DNA of between one to seven bases, characterising the three loop regions (L1,L2 and L3).

$$G_{3\text{-}5}N_{L1}G_{3\text{-}5}N_{L2}G_{3\text{-}5}N_{L3}G_{3\text{-}5}$$

Where $N_{L1\text{-}3}$ are loops of $1 < N_{L1\text{-}3} < 7$ nucleotides.

The algorithm employed by Todd and co-workers (Todd et al., 2005) had similarities to the "Folding Rule" implemented by Huppert and Balasubramanian (Huppert and Balasubramanian, 2005), however, the fundamental difference from Todd *et al*. (Todd et al., 2005) was that no upper limit was set for the continuous runs of guanine, which could exceed five bases, leading to a sequence search of:

$$G_{3+}N_{1\text{-}7}G_{3+}N_{1\text{-}7}G_{3+}N_{1\text{-}7}G_{3+}$$

Where N refers to any base including guanine

The initial search of the human genome conducted by Todd *et al.* (Todd et al., 2005) identified 5,713,900 such sequences, though once the overlapping sequences were rejected this reduced to 375,157 potential sequences, similar to the 376, 446 PQS calculated by Huppert and Balasubramanian (Huppert and Balasubramanian, 2005) although the value by Todd *et al*. was reduced to 226,157 for the number of unique quadruplexes. It should be emphasised that this number merely reflects the total number of PQS and does not predict how many might exist at one time, as there is likelihood that many coexist in dynamic equilibrium between quadruplex and duplex formation.

1.9.2 Loop Length and Sequence Distribution Patterns

The loops linking the runs of guanines play an important role in determining the quadruplex stability and topology (section 1.2.3). Early bioinformatics studies conducted by Huppert and Balasubramanian (Huppert and Balasubramanian, 2005, Rachwal et al., 2007b, Hazel et al., 2004) revealed that the first and third loops

displayed the same properties, whilst the central loop showed a greater tendency to be three or four bases longer than the others. They found that loop lengths of the PQS were non-random and that the most common loop length was <1,1,1> which accounted for 8% of all PQS. Table 1.5 lists the 20 most common length combinations, which between them account for 32% of all PQS. Table 1.5 also lists the 20 most uncommon loop length combinations which account for 1.7% of all PQS and generally consists of longer loops (Huppert and Balasubramanian, 2005, Huppert, 2006, Huppert, 2008). By comparison, early work conducted by Todd *et al.* (Todd et al., 2005) focused primarily on the sequences of the loops by identifying the frequency of various loop sequences and the extent to which an individual sequence is found in one loop compared to the others. Shorter loops were found more frequently than longer loops (Table 1.5); while analysis of the sequences showed that CCTGT and CCTAT occurred most frequently and especially in the first loop with CCTGTCA being the most frequent, followed by CCTGTT (Table 1.6).

However, not all known G-quadruplex structures follow the rules of four runs of guanine that are separated by three loop regions which results in either false positives or false negatives. A prime example of this is the *c-kit* quadruplex-forming sequence, as described in section 1.7.3. NMR analysis has shown that this sequence does not follow the rules presented by these algorithms as it forms a snapback parallel-stranded structure with an isolated non G-tract guanine involved in the G-quartet despite the presence of four three-guanine tracts (Phan et al., 2007), whilst the NHE III_1 of the *c-myc* promoter sequence exists in dynamic equilibrium with several different folding topologies (Phan et al., 2005). Likewise, these searches typically look for loops that are 1-7 bases in length for two main reasons; shorter loops are more stable and also because an upper limit is set for practical reasons. If longer loop sizes were considered in these searches then the number of PQS would dramatically increase (Todd, 2007). For example, the CEB25 minisatellite has a central 9-nt loop which was not identified by these algorithms but has been shown to form a G-quadruplex (Amrane et al., 2012) whilst another study demonstrated that intramolecular G-quadruplexes with two short first and third loops of 1-2 nt could tolerate a central loop more than 20 nt in length. In a study by Pandey *et al*. (Pandey et al., 2013) they showed that RNA structures with loops of 15 nt were still able to form G-quadruplex structures. Recently, *in vitro* studies conducted by

Mukundan and Phan (Mukundan and Phan, 2013) reported formation of bulges on G-quadruplex-forming sequences that contained discontinuous runs of G-tracts. They found that even structures with three isolated guanines were still able to form stable G-quadruplexes. In contrast, some sequences based on the 5'-UTR mRNA quadruplex that fit the prediction patterns were in fact unable to form G-quadruplex structures due to the presence of C-tracts flanking the quadruplex sequence which most likely increased the ability to form stable stem structures resulting from GC Watson–Crick base pair formation (Beaudoin and Perreault, 2010).

Table 1.5 PQS loop lengths: The twenty most common and twenty least common sets of observed PQS loop lengths as reported by Huppert and Balasubramanian (Huppert and Balasubramanian, 2005).

Most common loop length					**Least common loop length**			
Loop 1	***Loop 2***	***Loop 3***	***Number***		***Loop 1***	***Loop 2***	***Loop 3***	***Number***
1	1	1	47,475		6	5	7	441
1	4	1	11,328		7	6	5	441
1	2	1	10,656		7	6	3	447
1	1	2	10,415		5	6	7	447
2	1	1	10,040		6	6	7	449
2	2	2	9411		6	7	7	450
1	3	1	9127		7	5	6	452
1	5	1	7799		5	7	6	484
5	1	1	7379		5	5	7	501
1	1	5	7337		5	6	3	505
3	3	3	6827		7	7	6	505
3	1	1	6458		6	6	5	506
1	1	3	6403		5	6	6	511
1	1	4	6196		3	6	7	521
4	1	1	6189		7	6	6	523
2	2	1	5123		6	7	3	525
1	2	2	5046		7	7	4	528
2	1	2	4780		3	7	6	533
1	6	1	4556		5	7	7	536
6	1	1	4462		6	7	5	538

***Note**: Loops are numbered from 5' to 3' of the G-rich strand*

Table 1.6 Loop sequence frequency found in PQS. The most frequent loop sequences found in PQS, separated by which loop the sequences was found in, as reported by Todd (Todd et al., 2005)

Sequence	Population	1st loop	2nd loop	3rd loop
A	193,756	51,361	63,872	78,523
T	121,406	53,234	37,657	30,515
C	44,020	14,983	14,907	14,130
AA	400,026	12,778	13,717	13,531
CT	32,472	11,637	10,554	40,281
CA	32,070	10,781	10,846	10,443
G	29,623	7183	8375	14,065
AT	19,957	6789	7242	5926
AGA	19,144	5377	6919	6848
TT	17,089	7437	5530	4122
TA	12,641	4744	4329	3568
CC	10955	3646	3726	3583
AGT	9896	2767	4447	2682
AGGA	9463	1932	3559	3972
AGGT	9434	1516	6448	1470
TGA	9237	3006	2849	3382
AAA	7839	2393	2970	2476
CCT	7151	2540	2298	2313
TGT	6619	2530	2307	1782
CCA	6269	2105	2048	2116

1.9.3 G-quadruplex Motif Distribution Patterns

A second enhanced search arising from the same analysis showed that 42.7% of the gene promoters contained at least one PQS and that within the promoters the probability of finding a PQS is directly related to its proximity to the transcription start site (TSS) (Huppert and Balasubramanian, 2007). It was found that the highest concentration of PQS occurs in close proximity to the TSS, with the PQS density in the first 100 upstream bases being over 12 times higher than that of the genome average. It was discovered that there was a progressive reduction in PQS density away from the TSS

until it decreases to below the genome average at extreme distances from the TSS (>20,000 bases). In these sites G-quadruplexes are highly represented at nuclease hypersensitive sites and G-quadruplexes at both nuclease hypersensitive regions and promoters are enriched by a factor of 230, compared to the rest of the genome. It was also observed that there is a progressive reduction in PQS density away from the TSS (Huppert and Balasubramanian, 2007) supporting the hypothesis that quadruplexes play an important role in gene regulation throughout the genome as shown in Figure 1.17 (Huppert and Balasubramanian, 2007).

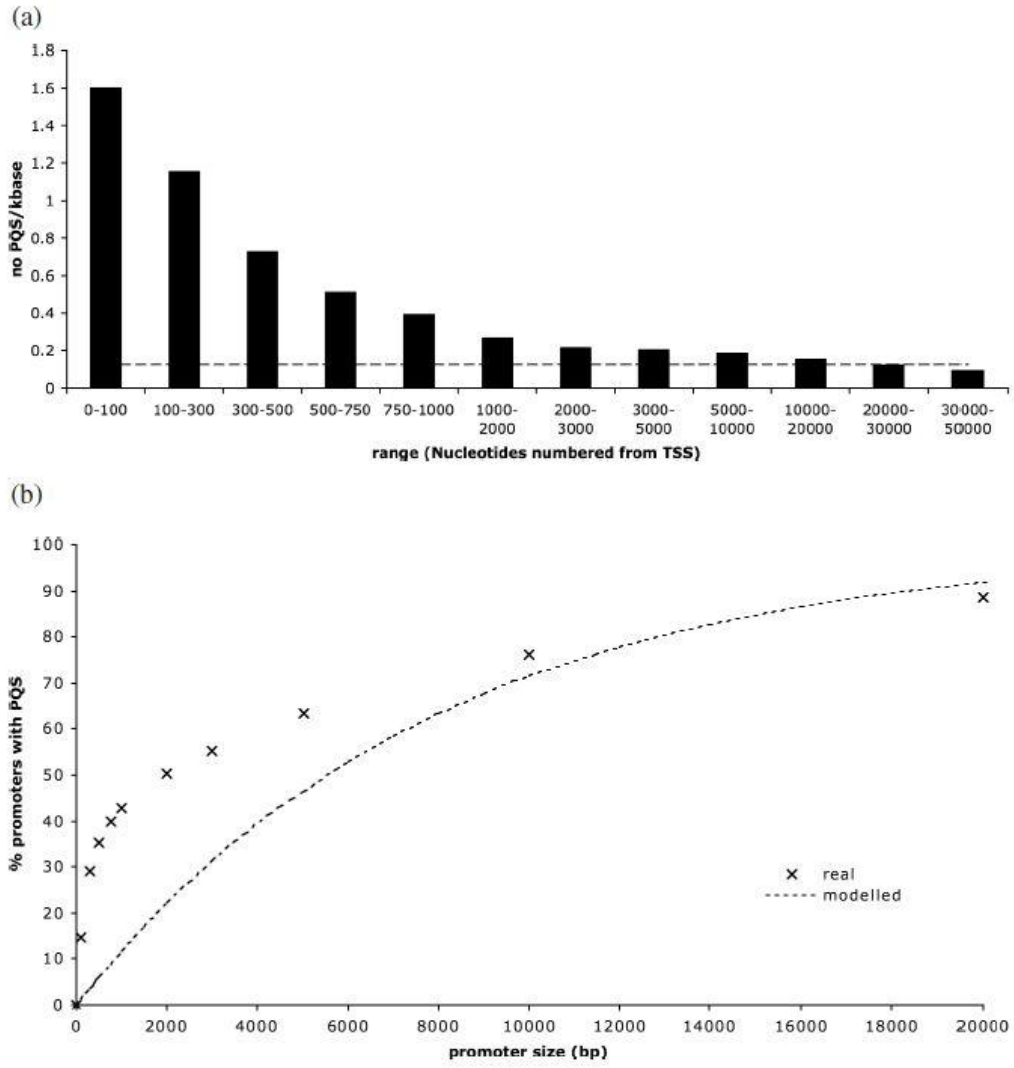


Figure 1.17 A graph to show that the probability of finding PQS is directly related to its proximity with the TSS. a) Density of PQS with distance upstream from the TSS. The genome as a whole has a density of 0.13 PQS/kb, shown by the dashed line. b) The percentage of promoter regions containing at least one PQS increases as the size of the promoter increases. This increase is extremely fast over the first 1000 bases. The dashed line shows what percentage would be predicted if the density of PQS were equal to that across the genome as a whole. Data collected by (Huppert and Balasubramanian, 2007).

In a study correlating the potential to form G-quadruplex structures (G4P) with specific gene function Eddy and Maizels (Eddy and Maizels, 2006) found that tumour suppressor genes are characterised by lower G4P than their flanking sequences and proto-oncogenes by higher G4P than their flanking sequences. This therefore suggests that the potential for G-quadruplex formation correlates with gene function, rather than local genomic environment. It appears that the low G4P of tumour suppressor genes minimises the potential instability of the genes that function to maintain genomic stability, while the high G4P in proto-oncogenes may contribute to their instability (Eddy and Maizels, 2006).

G-quadruplex bioinformatics studies have been invaluable in identifying potential motifs in the human genome for therapeutic targets. However, it is worth remembering that although the current search patterns have been determined by known quadruplex folding topologies, the estimated number of PQS may be over or underestimated.

1.10 Quadruplex Stabilising Ligands as Therapeutic Targets

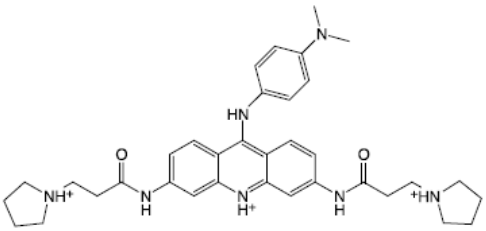
Telomere synthesis is performed by the enzyme telomerase. This is essentially a ribonucleoprotein reverse transcriptase. Telomerase is responsible for maintaining telomere length and has been shown to be active in ~ 85% of cancer cells. It was first recognised in the early 90s that the formation of potassium stabilised G-quadruplex structures inhibited telomerase, thereby preventing telomere elongation (Zahler et al., 1991). The therapeutic possibilities of targeting G-quadruplexes in telomeric DNA in order to inhibit telomerase were first reported by Sun *et al*. in 1997 (Sun et al., 1997). The regulatory potential of G-quadruplexes at the end of telomeres and in G-rich promoter regions has generated vast interest in researching and developing drugs that target and stabilise these structures as an entirely novel approach to anticancer therapeutics (Monchaud and Teulade-Fichou, 2008, Tan et al., 2008, Yang and Okamoto, 2010). The majority of cytotoxic cancer chemotherapeutic agents target DNA in a somewhat indiscriminate manner, merely inhibiting replication and cell division. There is therefore considerable interest in designing drugs that target DNA in a sequence-specific manner. Biophysical, NMR and crystallographic structural analyses of G-quadruplex-ligand complexes (Gowan et al., 2002, Burger et al., 2005, Monchaud

and Teulade-Fichou, 2008, Sun et al., 1997, Yang and Okamoto, 2010) have universally identified two main binding sites for G-quadruplex ligands; the first and primary binding site is by co-facial π-π end-stacking or 'hemi intercalation' of (typically) a flat aromatic ligand onto one or both the terminal G-quartets. The G-quadruplex ligand should have a large aromatic surface that is much broader than duplex DNA binders in order to enhance the aromatic-aromatic overlap and provide selectivity towards the quadruplex structure (Monchaud and Teulade-Fichou, 2008, Luedtke, 2009). The second involves the surface features of the grooves and/or loop regions (Luedtke, 2009) via electrostatic interactions between positively charged ligands and the quadruplex structure. Subtle changes in the topology of the G-quadruplex structure such as groove width and loop sequence can lead to selective binding interactions of these compounds (Arora and Maiti, 2008). However due to the polymorphic nature of the G-quadruplex backbone and the limited data available the electrostatic potential of the grooves is not as fully understood as with duplex DNA (Monchaud and Teulade-Fichou, 2008). Developing small molecules that predominantly bind to G-quadruplexes is primarily based on large planar polycyclic aromatic and heteroaromatic compounds. This is because the aromatic compounds are inclined to π stack effectively on the surface of the planar quartets thereby offering further stabilisation (Monchaud and Teulade-Fichou, 2008, Neidle, 2012, Gowan et al., 2002, Incles et al., 2004). Many of these ligands also contain side-chain substituents with at least one substituent terminating with a cationic group so that the compound exhibits both hydrophobic and hydrophilic properties. Also the cationic charges of these side-chains have the propensity to interact with the grooves of the G-quadruplex structure (Monchaud and Teulade-Fichou, 2008, Neidle, 2012, Tan et al., 2008, Yang and Okamoto, 2010). However, the hydrophobic faces of the chromophores leads to a tendency to bind indiscriminately to a wide range of cellular macromolecules (Sun et al., 1997, Guo et al., 1992, Chen et al., 1996). Neidle (Neidle, 2012) summarised five main categories for designing effective quadruplex-binding ligands:

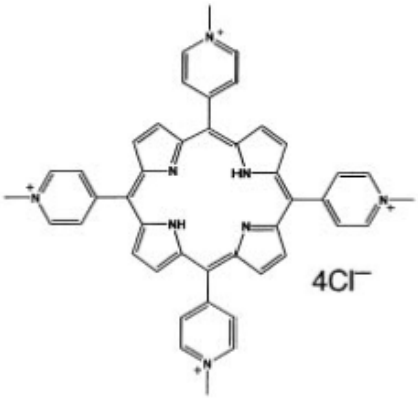
1. Improving drug-like features: by designing a drug with small molecular weight
2. Enhancing quadruplex affinity
3. Enhancing quadruplex selectivity over duplex DNA

4. Enhancing selectivity for a particular quadruplex target compound to other quadruplexes
5. Establishing *in vitro*, cell-based and *in vivo,* assay cascades for hit-to-lead selection

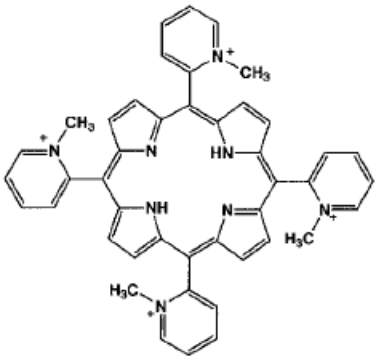
The majority of telomeric G-quadruplex research is performed on short sequences which produce just a singular structure, yet the human telomeric sequence could potentially form a number of structures simultaneously. Recently longer, extended sequences are being used for research, as it has been suggested that multiple structures may exist at one time (Yu et al., 2012). Designing drugs that can sandwich between quadruplexes at the quadruplex-quadruplex interface is highly attractive as these compounds can offer further stabilisation to the structure(s) (Gowan et al., 2002, Incles et al., 2004). The field of quadruplex ligand design is extremely vast and too extensive to cover in this thesis. The compounds discussed in this section have all been reported to bind the G-quadruplex structure and some even bind selectively over other structural forms of DNA. Different G-quadruplex-interactive compounds also appear to selectively bind different topological folds and this selectivity may relate to their biological activity. For example, telomestatin has been shown to be a highly potent inhibitor of telomerase by facilitating and stabilising the formation of the basket-type telomeric G-quadruplex.



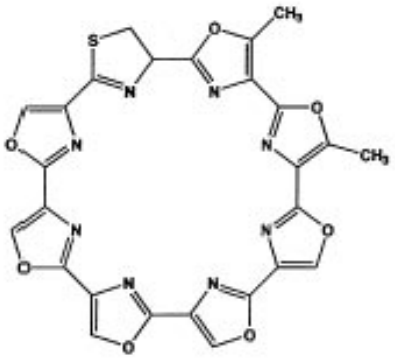
a) BRACO19



b) TMPyP4



c) TMPyP2



d) Telomestatin

Figure 1.18 Molecular structure of a) BRACO19, b) TMPyP4 c) TMPyP2 and d) Telomestatin

1.10.1 BRACO-19

The 3,6,9-trisubstituted tricyclic acridine compound BRACO-19 ((9-[4-(N,N-dimethylamino)phenylamino]-3,6-bis(3-pyrrolidino- propionamido)) (Figure 1.18a) is one of a series of potent acridine derivative inhibitors that were designed by computational modelling in order to target telomeres directly (Read et al., 2001). The positive charge on the acridine complexes (central ring nitrogen) was designed to

complement the channel of negative electrostatic potential of the G-quadruplexes. BRACO-19 was derived from a disubstituted acridine compound on the assumption that the three substituents would each occupy a groove in the G-quadruplex structures (Gowan et al., 2002, Incles et al., 2004) and in comparison to its starting model it is exceptionally low in cytotoxicity with a higher G-quadruplex binding affinity and telomerase-inhibitory activity. X-ray crystallographic analysis of the binding of BRACO-19 to a human bimolecular telomeric sequence d(TAGGGTTAGGGT) (Figure 1.19) reveals that the ligand is asymmetrically stacked forming a π-π overlap with just two guanine bases at the dimeric interface between two 5' to 3' stacked bimolecular folded G-quadruplex structures containing parallel d(TTA) loops forming a "biological unit". The tricyclic aromatic ring is sandwiched between the two structures bounded on one side by the 3'-end G-quartet and on the other by a 5'-end TATA quartet face formed from the adenine and thymine bases flanking the sequence forming a "biological unit". A 3'-end thymine base is flipped into the binding site so that it is able to interact with the BRACO-19 molecule which is mediated by two water molecules, whist the three side-chains interact with the grooves of the G-quadruplex structure. The 3- and 6-position substituents with the cationic termini each extend into a wide groove located on opposite sides of the G-quartet face, whilst the 9-position anilino substituent fits into a narrow hydrophobic pocket (Campbell et al., 2008) (Figure 1.19).

BRACO-19 has been shown to induce short-term growth arrest and replicative senescence in telomeres by inhibiting telomerase activity of prostate cancer cells (Incles et al., 2004). After two weeks of exposure to 2 μM BRACO-19 (well below that causing acute cytotoxicity), it was shown to be an inhibitor of cell growth and an inducer of senescence in 21NT human breast cancer cell line (Gowan et al., 2002). Evidence for telomere shortening in Xenograft studies at noncytotoxic concentrations show that the hTERT expression was significantly reduced after 24 hours and that complete termination of growth was seen after 15 days (Burger et al., 2005). Although BRACO-19 has good solubility, it is highly lipophilic causing it to have poor membrane permeability. Therefore the drug may accumulate in the cell membranes, which could affect the integrity of the cell wall (Taetz et al., 2006).

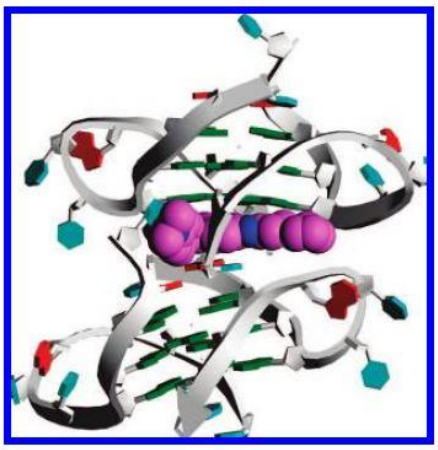


Figure 1.19 The biological unit in the crystal (PDB id 3CE5) with BRACO-19. A BRACO-19 molecule (mauve) is shown at the interface of the two quadruplexes in the unit, stacked between a G-quartet (top) and a TATA tetrad (bottom) (Campbell et al., 2008).

1.10.2 TMPyP4

Porphyrins are well known duplex-binding compounds and are well established as G-quadruplex binding agents. The cationic porphyrin derivative TMPyP4 (Figure 1.18b) (5,10,15,20-Tetrakis(N-methyl-4-pyridinio)porphyrin) is the most commonly studied porphyrin as a G-quadruplex binding ligand as it dramatically inhibits telomerase activity once bound to telomeric G-quadruplexes (Grand et al., 2002, Kim et al., 2003). The fused planar aromatic ring system, positive charge and the size of porphyrin analogues make these compounds ideal to bind to the G-quadruplex scaffold by stacking on top of the terminal G-tetrads. However, they are non-selective ligands and bind to both G-quadruplex and duplex DNA, with little selectivity. Visible absorption, circular dichroism, and fluorescent energy transfer studies indicate TMPyP4 is able to bind and stabilise both parallel and antiparallel G-quadruplex structures with only a two-fold affinity for quadruplex over duplex (Anantha et al., 1998), whereas a TMPyP4 derivative TMPyP2 [(5,10,15,20-tetra-(*N*-methyl-2-pyridyl)porphine] (Figure 1.18c) that has the *N*-methyl groups in sterically hindered 2 position showed extremely weak activity (Han et al., 1999a, Ou et al., 2008). X-ray crystal structure of TMPyP4 bound to the bimolecular human telomeric quadruplex sequence d(TAGGGTTAGGG) showed that the quadruplex topology is parallel-stranded with external double-chain-reversal loops. Surprisingly, the porphyrin molecules bind by stacking onto the TTA nucleotides, either as part of the external loop structure or at the 5’ region of the stacked

quadruplex. This involves stacking on hydrogen-bonded base pairs, formed from those nucleotides not involved in the formation of G-tetrads, thus resulting in no direct ligand interactions with G-tetrads (Parkinson and Neidle, 2007). However despite their poor selectivity to quadruplex-structure it is still a very popular ligand in G-quadruplex research.

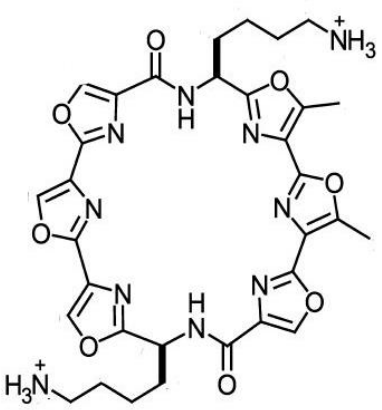
1.10.3 Telomestatin

Telomestatin (Figure 1.18d) is a macrocyclic natural product isolated from *Streptomyces anulatus* (Shin-ya et al., 2001). It consists of a pentaoxazole ring and one thiazoline ring. It is an extremely effective quadruplex-binding ligand that is a highly potent telomerase inhibitor with a $^{tel}IC_{50}$ of 5 nM (Shin-ya et al., 2001, Kim et al., 2002) although this has since been re-evaluated (De Cian et al., 2007). Structurally its dimensions are similar to that of a G-quartet and its potential to behave as a G-quadruplex ligand was confirmed by molecular modelling and experimentation (Kim et al., 2002). Telomestatin appears to interact with intramolecular G-quadruplexes rather than intermolecular structures, with a 70-fold selectivity for G-quadruplex over duplex DNA (Kim et al., 2003) and is a more potent and specific telomerase inhibitor than other G-quadruplex binding ligands such as BRACO-19. Different multiple myeloma cells lines were treated with minimal effective concentrations of telomestatin for a period of 3-5 weeks. Telomestatin treatment led to inhibition of telomerase activity, reduction in telomere length, and apoptotic cell death in all treated cell lines (Shammas et al., 2004). The effects of telomestatin on the biological function of telomeres and telomerase make it a highly desirable field of research. However, despite its high activity and selectivity further modifications and optimisation of the drug is restricted as it does not possess convenient functional groups (Chung et al., 2013), it has low water solubility that can inhibit its bioavailability and furthermore total synthesis of telomestatin (Doi et al., 2006, Doi et al., 2011, Linder et al., 2011) is fairly complex, meaning that it is not compatible with large industrial scale synthesis (Monchaud and Teulade-Fichou, 2008, Granzhan et al., 2010).

Recently, a series of telomestatin analogues have been synthesised exhibiting enhanced properties of the ligand (Rzuczek et al., 2008, Granzhan et al., 2010, Monchaud and

Teulade-Fichou, 2008). Despite these reports, the binding affinities towards G-quadruplexes are still unknown. A recent study (Chung et al., 2013) based on a highly effective telomestatin derivative called L2H2-6OTD that exhibited high selectivity towards G-quadruplexes with a high telomerase inhibitory activity synthesised by Rzuczek *et al.* (Rzuczek et al., 2008) offers detailed information on how these oxazole compounds interact with G-quadruplex structures. This compound called L2H2-6M(2)OTD (abbreviated to L2H) contains six oxazole rings and two alkyl amine side chains (Figure 1.20).

LH2 interacts with an intramolecularly (3+1) folded human telomeric sequence d(TTGGGTTAGGGTTAGGGTTAGGGA) by π-stacking and electrostatic interactions. The two side chains of the ligand are sufficiently close to the quadruplex structure for significant electrostatic interactions. The ligand was observed to interact with the top G-quartet of the quadruplex structure through π-stacking. The two cationic side chains of L2H are directed toward the negatively charged phosphate backbone of the structure, while the methyl groups on the oxazoles are positioned above the wide groove away from the bases at the 5'-end of the structure possibly to minimise steric clashes between bases in the edgewise loop. In addition to the π-stacking interactions, the presence of potassium cations were shown to enhance the stabilising ability of telomestatin and its analogues. The authors also suggest that small molecules designed to target G-quadruplexes may not necessarily have to be planar molecules for π- stacking interactions, because if the small molecule has sufficient flexibility it could mould into a specific conformation to maximise interactions with its target. Although L2H is highly aromatic, quantum calculations have shown it to be non-planar and to adopt a roof-like bent conformation in its free state. However, upon binding with the G-quadruplex structure, the compound becomes more planar to maximise the π- π stacking interaction with the top G-tetrad.



L2H2-6M(2)OTD

Figure 1.20 Chemical structure of the macrocyclic hexaoxazole L2H2-6M(2)OTD with proton naming indicated.

Over the past 20 years hundreds of small molecules varying in chemical structure and properties have been synthesised and used to investigate their interaction with G-quadruplexes (Monchaud and Teulade-Fichou, 2008, Georgiades et al., 2010) and many have exhibited *in vitro* and *in vivo* activities. Recently a database, G4LDB, was created to assist researchers in designing effective G-quadruplex ligands. The database consists of two major components; the first component contains over 800 reported G-quadruplex ligands that are associated with 4000 related activity study records and the second part contains an online prediction module, where the binding strength of a newly designed ligand can be predicted (Li et al., 2013).

1.11 Negative Supercoiling of DNA and G-quadruplexes

Even before the elucidation of the three-dimensional structure of the B-DNA (Watson and Crick, 1953), it was already known that DNA was a highly polymorphic molecule that can exhibit a great deal of sequence dependence conformations under specific experimental or *in vivo* conditions (Rich, 1993, Lilley, 1984). The conversion of B-form DNA to an alternative secondary structure *in vivo* requires the local unwinding and/or melting of the duplex. *In vivo* sequence-dependent structural perturbations can occur via localised under- or over-winding or melting of the DNA molecule generated by DNA supercoiling (Vinograd et al., 1965, Giaever et al., 1988). Natural genomic covalently-closed-circular DNA (cccDNA) in most prokaryotic cells exists in a negatively supercoiled state, that is the DNA is underwound (Bauer, 1978). DNA supercoiling is

the result of underwinding of the DNA helix (a reduction of the linking number, *Lk*, corresponding to the number of times that one strand wraps around the other, discussed in Chapter 5). This causes twisting and writhing of the DNA double helix into a higher order helical structure. The torsional stress generated by negative supercoiling results in a build-up of excess free energy within the plasmid (Mirkin, 2001, Bowater, 2005, Lilley, 1984, Peck and Wang, 1983). Since duplex B-DNA is a configuration of minimum energy (Bowater, 2005) structural transitions within this molecule are unlikely to occur, however, sequence-specific structural perturbations are favoured in cccDNA as the free energy generated by negative supercoiling may be enough to facilitate the localised unwinding of the DNA and the formation of alternative secondary structures. Thus, structural transitions induced by negative supercoiling can be measured directly and described by the number of topological parameters (Chapter 5) (Peck and Wang, 1983). Early studies have shown that the torsional stress generated from negative supercoiling can facilitate the formation of sequence-specific alternative secondary structures such as Z-DNA (Azorin et al., 1983, Nordheim et al., 1983), H-DNA (Vojtiskova et al., 1988, Mirkin and Frankkamenetskii, 1994, Lyamichev et al., 1986, Mirkin et al., 1987, Potaman et al., 2004) and cruciforms (Panayotatos and Wells, 1981, Murchie and Lilley, 1992). Generally these investigations rely on a combination of cleavage of strand-specific nuclease (Lilley, 1980), chemical probing (Lilley and Hallam, 1984, Onyshchenko et al., 2009, Onyshchenko et al., 2011, Sun and Hurley, 2009), or two-dimensional gel electrophoresis (Peck and Wang, 1983, Vologodskaia and Vologodskii, 1999, Jude et al., 2013, Potaman et al., 2004) which are discussed in further detail in Chapter 4 and Chapter 5.

Following on from these studies, it is therefore reasonable to suggest that negative supercoiling will facilitate the formation of quadruplex DNA within G-rich segments in supercoiled DNA. This would be consistent with suggestions that quadruplexes have a significant biological role and might explain their role in gene regulation and their presence at TSS (Kendrick and Hurley, 2010, Lipps and Rhodes, 2009).

Despite the postulated relationship between G-quadruplex formation and gene regulation, to date there have been very few reports providing evidence that negative

supercoiling can disrupt double-stranded DNA (dsDNA) to promote the formation of G-quadruplex and/or *i-motif* structures. With the aid of chemical footprinting a study based on the wild-type sequence of the polypurine/polypyrimidine tract of NHE III_1 of the *c-myc* promoter (Sun and Hurley, 2009) revealed that negative superhelicity, induced the formation of a G-quadruplex structure from the four 5'-end guanine tracts at near physiological pH and in the presence of 100 mM KCl. Interestingly it was also demonstrated that the *i-motif*, which typically forms at low pH, was also formed under neutral conditions under superhelical stress. *In vitro* footprinting of the cloned VEGF G-quadruplex-forming sequence also demonstrated the formation of quadruplex structures in supercoiled DNA with and without the presence of stabilising ligands (Sun, 2010, Sun et al., 2008, Sun et al., 2005). Conversely, a plasmid clone containing the sequence of the *BCl-2* gene (Onyshchenko et al., 2009) was unable to form a G-quadruplex structure under negative supercoiling without the addition of short peptide nucleic acids (PNAs) that bound to the complementary C-rich strand, displacing the G-strand, thereby making it available for quadruplex formation. In a further study they reveal that the invasion of the PNA on the complementary C-strand was dependent on the formation of the G-quadruplex structure (Onyshchenko et al., 2011).

1.11.1 DNA G-quadruplexes *in vivo*

Linear G-rich DNAs spontaneously form G-quadruplex structures under physiological conditions. Besides the single-stranded region of the extreme 3' termini of telomeres, genomic quadruplex-forming motifs must compete with Watson-Crick base pairing which must be separated to allow for quadruplex formation. Transient separation of the double strand occurs during DNA cellular processes such as replication, transcription and recombination, which results in the formation of DNA supercoils. Lui and Wang (Liu and Wang, 1987) demonstrated that transcription of right-handed double-helical DNA produces negative supercoiling behind a moving RNA polymerase and as the double-stranded DNA becomes temporarily denatured the torsional stress generated by the movement of the polymerase is energetically more favourable and is sufficient to stabilise the formation of Z-DNA. In a recent study conducted by Zhang *et al*. (Zhang et al., 2013) chemical probing and *in vitro* transcription were used to show that negative supercoiling induced by downstream transcriptional events facilitated the formation of G-quadruplex structure(s). Their data demonstrated that G-quadruplex formation in

dsDNA was triggered thousands of bases pairs away from the transcription event, which may prove essential when regulation at a specific locus is required. Interestingly, the report also demonstrated that transcription-generated positive supercoiling failed to induce G-quadruplex formation, thus providing further evidence to support the presence and abundance of genomic G-quadruplex-forming motifs upstream of TSS rather than downstream. The report also suggests that G-quadruplex-forming sequences can serve as a sensor or receiver to sense remote DNA tracking activity in response to propagation or mechanical torsion in DNA double-helix.

In an alternative study, Biffi *et al*. (Biffi et al., 2013) described the used of an antibody (BG4) with high selectivity and low nanomolar affinity for G-quadruplex DNA to quantitatively visualise G-quadruplex formation in the DNA of human cells. Initial studies conducted on the antibody demonstrated that it did not discriminate between inter or intramolecular structures. The human cells were incubated with the antibody tagged with a fluorescent dye and the formation of G-quadruplex structures was followed during the cell-cycle progression. They showed that the number of G-quadruplexes increased during S phase, indicating that their structural formation was dependent on DNA replication, a point at which the duplex DNA was being separated at the replication fork allowing the single strands to fold in to secondary structures (Figure 1.21). A similar study by Henderson and co-workers (Henderson et al., 2013) describes the development and characterisation of novel monoclonal antibodies specific for distinct structural variants of telomeric G-quadruplexes. *In vitro* analysis of one of the antibodies, designated 1H6, which demonstrated high specificity for G-quadruplex structures was also able to exhibit the same specificity and recognise structures in chromosomes and throughout the genome *in vivo* which is consistent with the studies by Biffi *et al*. (Biffi et al., 2013) using a phage antibody and supports the hypothesis that these structures do not only exist in telomeres.

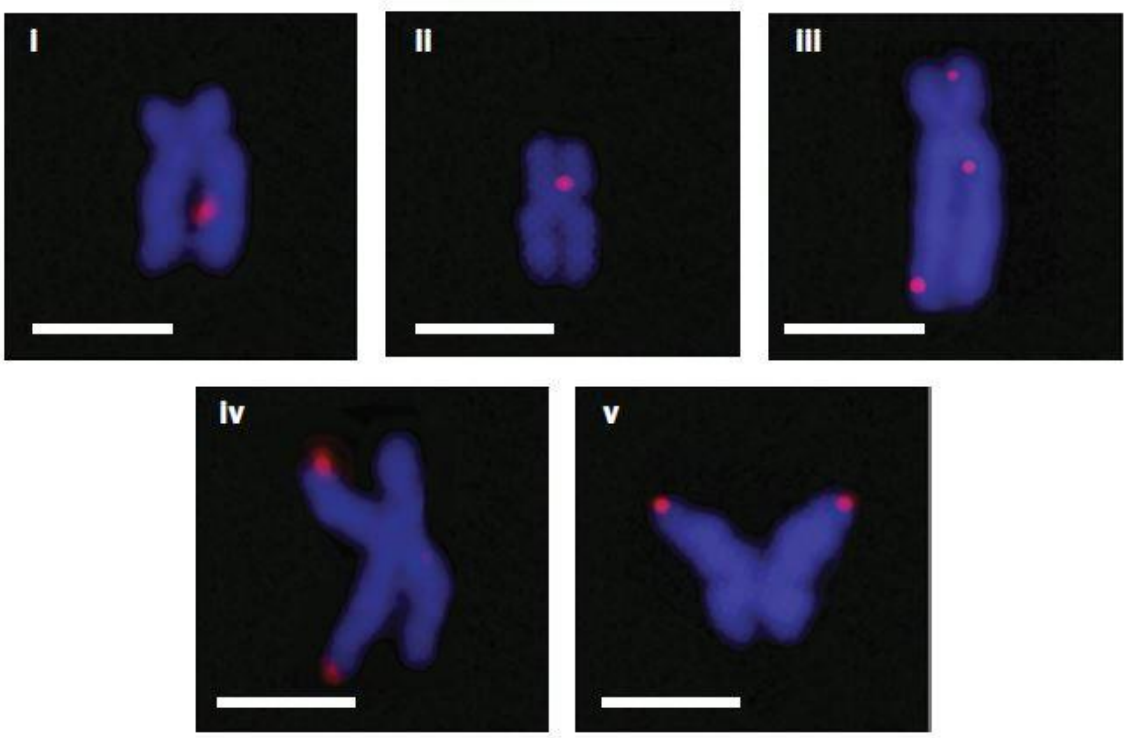


Figure 1.21 Immunofluorescence visualisation of G-quadruplex structures for the BG4 antibody on chromosomes isolated from HeLa cells. Discrete BG4 foci (red) were observed both within the non-telomeric regions and at the telomeres. i) to iii) corresponds to regions of known genes, whilst iv) and v) corresponds to telomeric DNA. Scale bars corresponds to 2.5 μm (Biffi et al., 2013).

1.12 Aims

G-rich DNA is prevalent throughout the human genome and is frequently found in the proximal promoter region of many growth-related genes, particularly many oncogenes. Although, the formation of G-quadruplex structures from G-rich DNA can occur spontaneously in single-stranded synthetic linear DNA fragments, double-stranded DNA such as genomic DNA must first dissociate from its complementary strand in order to form. In vivo, strand separation is known to be facilitated by negative supercoiling.

The aim of this research is to investigate the effects of DNA supercoiling in the formation of intramolecular G-quadruplexes. This will be examined by preparing several G-rich sequences that have the potential to form quadruplexes and cloning them into the plasmid pUC19. The chemical and enzymatic probes; dimethyl sulphate, potassium permanganate and S1 nuclease will be used to examine G-quadruplex formation on supercoiled and linear DNA fragments. Two-dimensional gel

electrophoresis of their DNA topoisomers will also be used to assess whether the excess energy of superhelical stress is sufficient to drive quadruplex formation. The aim of this research is to investigate the effects of DNA supercoiling in the formation of intramolecular G-quadruplexes. This has been examined by preparing several G-rich sequences that have the potential to form quadruplexes and cloning them into the plasmid pUC19. Chemical and enzymatic probes were used to examine G-quadruplex formation on supercoiled and linear DNA fragments. Two-dimensional gel electrophoresis of DNA topoisomers was also used to assess whether the build-up of superhelical stress is sufficient to drive quadruplex formation.

Chapter 2:
Materials and Methods

2.1 Instrumentation

Apparatus were obtained as follows: Two-dimensional gel electrophoresis tanks (Gel Electrophoresis Apparatus-GNA-200 Pharmacia or Jencons Scientific LTD: model number. H026), agarose gel electrophoresis tank (LKB Bromma-2197 power supply), CD spectropolarimeter (JASCO, J-720), CD power supply (JASCO, Spectropolarimeter PS 450), CD software, (J-700 for windows Standard Analysis, Version 1.00.01, © 1991-1992 JASCO Corporation), CD water supply (HAAKE UWK 45), centrifuge (Biofuge pico, Heraeus), gel drier pump (KNF, Labo Port), gel dryer (Model 583 gel dryer, Bio Rad), Geiger counter (Morgan, Mini Instruments Ltd), imager (Storm 860 phosphorimager), LightCycler (Roche), transilluminator (Syngene, G-Box), transilluminator software (GeneSnap, SynGene), X-ograph (SRX-101A, Konica Minolta).

2.2 Materials

2.2.1 Chemicals and Reagents

All chemicals used were purchased at the highest quality.

[α-^{32}P] dATP (25 mCi/ml, PerkinElmer), AccuGel (19:1 sequencing grade, ultra pure, National Diagnostics), acetic acid (Fisher Scientific, 99%, 1.04g/ml), agar (blood agar, BD), calcium chloride (Sigma-Aldrich), AMV reverse transcriptase (Sigma-Aldrich), ammonium persulphate, (Sigma-Aldrich), carbenicillin (Melford Laboratories LTD), DNA Ladder (Amersham Pharmacia, Biotech Inc), IPTG (Isopropyl β-D-1-thiogalactopyranoside, Sigma-Aldrich), Q1Aprep® spin mini prep kit 250 (Qiagen), pUC19 (BioLabs Ltd), SequaGel (sequencing grade, ultra pure, National Diagnostics), restriction enzymes: all enzymes were purchased from Promega unless stated otherwise, sodium chloride (Fischer Scientific), SURE® 2 Supercompetent cells (Stratagene), T7 sequencing kit (Amersham Pharmacia), sodium acetate (Sigma-Aldrich), sodium

hydroxide (Sigma-Aldrich), TG2 *E. coli* cells **(**Glycerol stocks), TEMED (N,N,N,N – tetramethyl- ethylethylenediamine, Sigma-Aldrich), Tryptone (Milford), Urea (99.0-100.5%, Sigma-Aldrich), wheat germ topoisomerase I (Inspiralis), Yeast extract (Bacto™ yeast extract, Technical, BD), X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside, Sigma-Aldrich), X-ray film (Fuji RX Film, 18 x 43 cm).

2.2.2 Buffers and Solutions

All solutions were prepared using high quality deionised water (Millipore Milli-Q Water System).

5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (Xgal) – 2% solution	X-gal (0.2 g) dissolved in dimethylformamide (10 ml)
5x TBE Running Buffer	Tris (108 g), boric acid (55 g), EDTA (9.4 g) in 2 L water
Annealing buffer: Tris-NaCl	10 mM Tris-HCl pH 7.4 containing 10 mM NaCl)
Carbenicillin	Carbenicillin (100 mg/ml) in water, filter sterilised
Diluent	Urea (50%, w/v)
DNase I stop	1 mM NaOH, 10 mM EDTA, 80% formamide, 0.2% (w/v) bromophenol blue
DNA Elution Buffer (TE1)	10 mM Tris-HCl pH 7.4 containing 1 mM EDTA
Incubation buffers	Tris-NaCl: 10 mM Tris-HCl pH 7.4 containing 10 mM NaCl or 100 mM NaCl or 10 mM KCl or 100 MM KCl
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	IPTG (100 mM), dissolved in water and filter sterilized
Lithium phosphate buffer	LiOH (10 mM), pH adjusted to 7.4 with H_3PO_4
Loading dye (20%, w/v)	Ficoll® (20% w/v)), 10 mM EDTA, 0.1% (w/v) bromophenol blue
Re-suspension buffer (TE2)	10 mM Tris-HCl, pH7.4 containing 0.1 mM EDTA
Sterile Media	2YT - (16 g tryptone, 10 g yeast extract, 5 g sodium chloride (dissolved in 1 L water)
Transformation Buffer	10 mM Tris-HCl, pH 7.4 containing 50 mM $CaCl_2$

Any additional buffers used will be described where appropriate.

2.2.3 Oligonucleotide Sequences

Oligonucleotide sequences used for cloning (Table 2.1) and quadruplex melting experiments were synthesised by ATDBio (Department of Chemistry, University of Southampton) on an Applied Biosystems 394 DNA/RNA synthesiser. The oligonucleotides used in the cloning experiments were provided with a 5'-phosphate group, whilst oligonucleotides used for fluorescence melting (discussed in Chapter 5) were labelled with 6-aminohexylfluorescein (FAM) at the 5'-end and dabcyl at the 3'-end and all oligonucleotides were dissolved in water and stored at -20 °C.

Please note: Oligonucleotide formula (G_n,L,X)

Where:

G_n is the number of guanine tracts,

L is the number of thymines in the loop

and X is the number of G-quadruplex forming sequences

Table 2.1 **List** of G-rich oligonucleotides sequences with complimentary C-rich strand.The oligonucleotides were synthesised by ATDBio (Department of Chemistry, University of Southampton) used for the cloning experiments. Where: G_n is the number of guanine tracts, L is the number of thymines in the loop and X is the number of G-quadruplex forming sequences. All sequences contained 5'-GATC overhangs for cloning into the *Bam*HI site

Oligonucleotide formula (G_n,L,X)	Sequence
G_4,1,1	5′-GATCGGGTGGGTGGGTGGG CCCACCCACCCACCCTAG-5′
G_5,1,1	5′-GATCGGGTGGGTGGGTGGGTGGG CCCACCCACCCACCCACCCCTAG-5′
G_4,1,2	5′-GATCGGGTGGGTGGGTGGGGATCGGGTGGGTGGGTGGG CCCACCCACCCACCCCTAGCCCACCCACCCACCCTAG -5′
G_5,1,2	5′-GATCGGGTGGGTGGGTGGGTGGGGATCGGGTGGGTGGGTGGGTGGG CCCACCCACCCACCCCACCCCTAGCCCACCCACCCACCCACCCCTAG -5′
G_4,4,1	5′-GATCGGGTTTTGGGTTTTGGGTTTTGGG CCCAAAACCCAAAACCCAAAACCCCTAG-5′
G_5,4,1	5′-GATCGGGTTTTGGGTTTTGGGTTTTGGGTTTTGGG CCCAAAACCCAAAACCCAAAACCCAAAACCCCTAG-5′
G_4,4,2	5′-GATCGGGTTTTGGGTTTTGGGTTTTGGGTTTTGATCGGGTTTTGGGTTTTGGGTTTTGGG CCCAAAACCCAAAACCCAAAACCCAAAACTAGCCCAAAACCCAAAACCCAAAACCCCTAG -5′
G_5,4,2	5′-GATCGGGTTTTGGGTTTTGGGTTTTGGGTTTTGGGGATCGGGTTTTGGGTTTTGGGTTTTGGGTTTTGGG CCCAAAACCCAAAACCCAAAACCCAAAACCCCTAGCCCAAAACCCAAAACCCAAAACCCAAAACCCCTAG -5′
G_4,8,1	5′-GATCGGGTTTTTTTTGGGTTTTTTTTGGGTTTTTTTTGGG CCCAAAAAAAACCCAAAAAAAACCCAAAAAAAACCCCTAG-5′
G_5,8,2	5′-GATCGGGT$_8$GGGT$_8$GGGT$_8$GGGGATCGGGT$_8$GGGT$_8$GGGT$_8$GGG CCCA$_8$CCCA$_8$CCCA$_8$CCCCTAGCCCA$_8$CCCA$_8$CCCA$_8$CCCCTAG -5′

2.3 Methods

2.3.1 General Techniques

The sections below describe general experimental techniques that were frequently used throughout this research.

2.3.1.1 Preparation E. coli TG2 competent cells

A colony of TG2 cells was inoculated into 2YT sterile media (5 ml) and grown overnight at 37 °C under constant agitation. From the culture containing the TG2 cells 1 ml was taken and added to sterile media (100 ml) and then grown at 37 °C for approximately 2 hr until an optical density between 0.5 – 0.8 at 600 nm was reached. The TG2 cells were pelleted by centrifuging at 3,000 rpm for 10 min at 4 °C. The supernatant was removed and the pelleted cells re-suspended in transformation buffer (20 ml). The cell suspension was then placed on ice for 30 min, and centrifuged for a further 10 min at 3,000 rpm at 4 °C. The supernatant was removed and the pellet re-suspended in transformation buffer (5 ml).

2.3.1.2 Transformation

Competent cells (50 µl) were added to 1 µl of plasmid (~ 300 ng/µl). The solution was then placed on ice for 30 min. The samples were then heat shocked at 45 °C for 1 min and placed on ice for 10 min. The cells were plated onto agar plates containing carbenicillin (100 µg/ml) and grown overnight at 37 °C overnight. The plates were then stored at 4 °C for a maximum of 4 weeks.

2.3.1.3 Plasmid extraction and preparation

After transformation a single colony was picked and inoculated into 2YT sterile media (5 ml) and grown overnight at 37 °C with constant agitation. The culture was transferred into microcentrifuge tubes (1.5 ml) and the cells were harvested by centrifuging at 4000 rpm for 5 min. The supernatant was removed and the plasmid was extracted and purified using a QIAprep® spin mini prep kit 250 (Qiagen) according to the

manufacturer's instructions as described in the QIAp. DNA was eluted into 50 µl of water (or EB buffer) and stored at -20 °C.

2.3.1.4 3'-labelling of DNA with [α-32P] dATP

The plasmid was incubated with the appropriate restriction enzymes to remove the insert containing the G-rich region of interest, by adding 10x enzyme buffer (2 µl) and both enzymes (1 µl, 10 units) to 18 µl of plasmid. This was incubated at 37 ºC for at least 1 hr. The fragments were then labelled with [α-^{32}P]dATP (1.5 µl) using AMV (avian myeloblastosis virus 1) reverse transcriptase and placed in a heat block for 2-3 hrs. Loading dye (15 µl) was added and the samples were loaded to an 8% non-denaturing polyacrylamide gel. The DNA fragments were then subjected to electrophoresis at 800 V until the bromophenol blue had migrated to the bottom of the plates. The plates were then separated and the gel was exposed to X-ray film for between 5 and 30 min, depending on the amount of radiolabel that had been incorporated. After exposure, the film was developed using an X-ograph film developer. Using the film as a template, the radiolabelled DNA fragments were located, cut from the gel, covered in elution buffer TE1 (approximately 400-500 µl) so that the gel slice was fully submerged in the buffer and then gently agitated overnight. The elution buffer containing the DNA was separated from the gel slice by gentle centrifugation. Ethanol (100%, 1 ml) was added to the buffer, mixed thoroughly and placed on dry ice for 30 min. This was then centrifuged for 15 min at 13000 rpm. The supernatant was removed (checking to confirm that the supernatant was not radioactive and that the majority of the radioactivity was in the pellet). The DNA pellet was washed with ethanol (70%, 200 µl) and further centrifuged for 2 min at 13,000 rpm. The supernatant was removed and the DNA pellet was dried in a speed-vacuum for 5 min to remove any residual ethanol. The purified DNA was then dissolved in TE2 buffer at a concentration of approximately 10 counts per second per µl, (as determined on a hand held Geiger counter).

2.3.1.5 Ethanol precipitation of DNA

Ethanol precipitation was used to recover DNA from either reaction mixtures or elution from polyacrylamide gels. The volume of ethanol (100%) added was 4x the original sample volume. The samples were mixed thoroughly and placed on dry ice for 30 min.

The samples were then centrifuged for 15 min at 13,000 rpm. The supernatant was removed and the pellet was dried under speed-vacuum for 5 min to remove any residual supernatant. The pellet was washed with ethanol (70%) and re-centrifuged for 2 min. The supernatant was removed and the pellet dried under speed-vacuum to remove any residual ethanol.

2.3.1.6 Preparation of GA tract

DNase I stop solution (4 µl) and water (20 µl) was added to 1.5 µl of radiolabelled DNA in a microcentrifuge tube. The sample was heated at 100 °C for 30 min, with the cap open, and then placed immediately on ice. The sample was loaded to a denaturing polyacrylamide gel containing 8 M urea. See section 2.51 for details.

2.3.2 Cloning

The G-rich sequences (Table 2.1) were inserted into pUC19 cloning vector (Figure 2.1) at the *Bam*HI restriction site. The samples were then transformed and grown overnight on x-gal blue-white selection plates and the clones were identified by the white colonies.

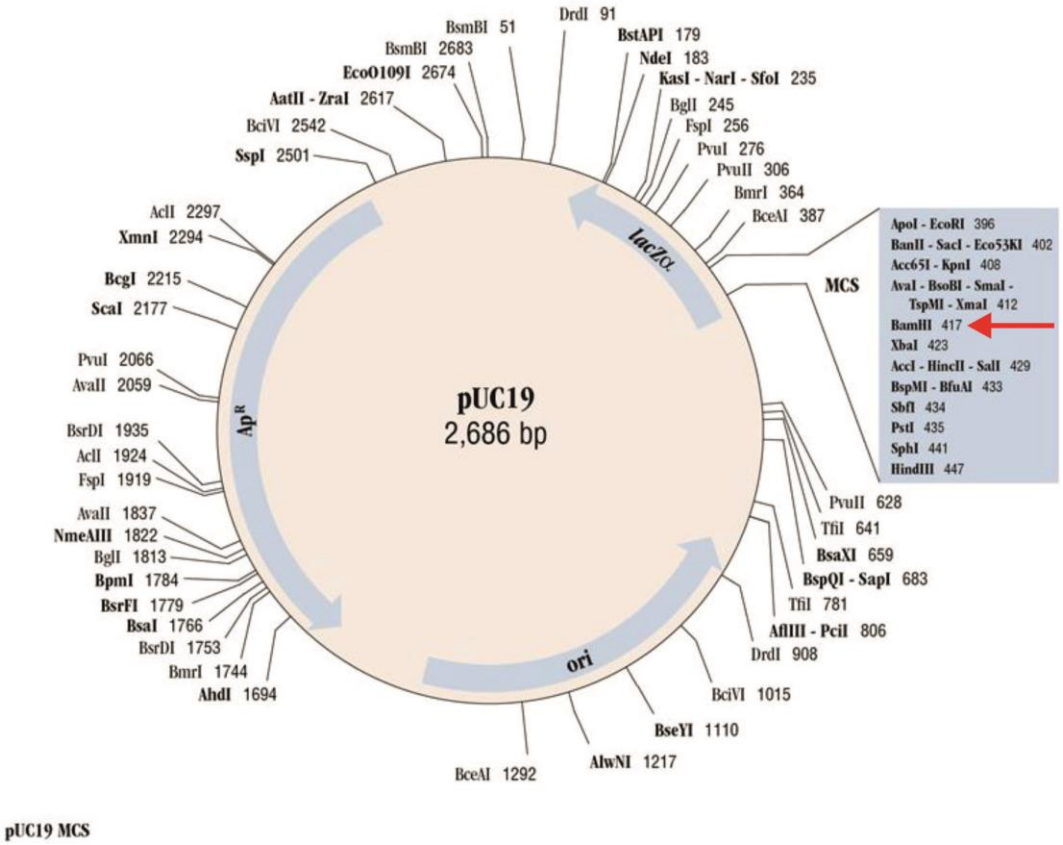


Figure 2.1 pUC19 cloning vector. (Yanisch-Perron et al., 1985). Image taken from NEW ENGLAND BioLabs Inc, with the *Bam*HI cloning site highlighted with a red arrow.

2.3.2.1 Annealing of oligonucleotides (1:1 mixture)

The oligonucleotides were annealed by mixing equimolar amounts of each of the complementary strands: G-rich strand (1 µl, ~60 µM) was added to the complementary C-rich strand (1 µl, ~60 µM) and 18 µl Tris-NaCl annealing buffer. This was incubated for 10 min at 95 °C and then left to slowly cool to room temperature.

2.3.2.2 Preparation of cloning vector

1 µl of plasmid pUC19 (300 ng/µl) was added to 2 µl of 10x *Bam*HI buffer, water (16 µl) and *Bam*HI (1 µl, 10 units). This was then incubated at 37 °C for > 1 hr. The plasmid was ethanol precipitated as described in section 2.3.1.5.

2.3.2.3 Ligation of oligonucleotide to cloning vector

The plasmid pellet was re-dissolved in water (18 µl). 1 µl of annealed oligonucleotide, 10x ligase buffer (2 µl) and 1 µl of T4 DNA ligase (10 units) was then added and left to ligate overnight at room temperature. 200 µl of TG2 competent cells were added to the ligated mixture and the plasmid was transformed as described above. 20 µl of the transformation mixture was plated onto agar plates containing X-gal (0.02%), IPTG (0.1 mM) and carbenicillin (100 µg/ml). The remainder of the transformation mixture was plated onto a separate plate, thereby giving low and high density plates, and grown overnight at 37 °C.

2.3.2.4 Blue-White selection

From the agar plates grown overnight, white colonies were picked and grown in 2YT sterile media (5 ml) containing carbenicillin (100 µg/µl) overnight at 37 °C with constant agitation. The plasmid was extracted and purified as described in section 2.3.1.3.

2.3.3 Sequencing

2.3.3.1 T7 sequencing kit (Amersham Pharmacia)

Early in this project the sequences of plasmid inserts were determined using a T7 sequencing kit (Amersham Pharmacia) according to the manufacturer's instructions, as described below. Subsequently all sequences were confirmed by commercial DNA sequencing.

Denaturing the plasmid DNA – Sodium hydroxide (2 M, 10 µl) was added to plasmid DNA (40 µl, 300 ng/µl) prepared as described in section 2.3.2 and left at room temperature for 10 min. 15 µl of sodium acetate (3 M, pH 4.8), water (35 µl) and ethanol (100 %, 300 µl) were then added, mixed and placed on dry ice for 30 min. The DNA was ethanol precipitated as described above, and the DNA was washed with 70% ethanol.

Annealing of primer - The plasmid was re-dissolved in water (10 µl) and annealing buffer (2 µl) and universal primer (2 µl) were then added. The solution was incubated at 37 °C for 20 min, and then left to stand at room temperature for at least 10 min.

Sequencing reaction - For each sequence, 4 tubes containing 2.5 µl of the dideoxy mixes G-short, C-short, T-short, A-short were set up.

Preparation of polymerase-enzyme mix – polymerase enzyme mix contained: 12 µl of label mix A, 7.5 µl of water and 1.5 µl [α-^{32}P] dATP. This was kept on ice whilst the enzyme mix was prepared. Enzyme mix contained: 6.5 µl of enzyme buffer and 1.5 µl T7 polymerase, this was added to the polymerase mix and placed on ice.

Dideoxy reaction - To each pot of annealed DNA 6 µl of polymerase-enzyme mix was added and left at room temperature for 5 min. The tubes containing the dideoxy mixes were incubated at 37 °C and 4.5 µl of the reaction mixture was added to each of the dideoxy mixes and left to react for 5 min. The reaction was terminated by adding 5 µl of stop solution to each. This was boiled for 3 min at 100 °C then immediately placed on ice. The samples were then loaded onto an 8% polyacrylamide denaturing gel containing 8 M urea and run for approximately 2 hr at 1500 V. The gel was fixed with acetic acid (10%, v/v), and transferred to Whatman 3MM paper and dried under vacuum for 1.5 hr at 80 °C. The dried gels were then subjected to phosphorimaging and visualised using ImageQuant 5.0.

2.3.3.2 Commercial sequencing by Eurofins MWG Operon

In addition to the manual dideoxy sequencing experiments, the plasmids containing the cloned sequences were also confirmed by commercial sequencing by Eurofins MWG Operon.

2.4 Probing of G-quadruplex Formation

Two experimental probing techniques were used to test for the presence of G-quadruplex structures and to determine whether negatively supercoiled DNA can provide enough free energy to facilitate the formation of these secondary structures. The first technique used chemical probing assays. Chemical probing was used to assess the accessibility of different bases: dimethylsulphate to identify the guanines participating in the quartets, potassium permanganate to identify exposed thymidines within the loops and diethylpyrocarbonate to probe for the accessibility of adenosine. The second technique used S1 nuclease to map the single-stranded regions that would be formed in the loops and the displaced C-rich strand.

2.4.1 Chemical Probing: Chemical Footprinting of G-rich Strand

Chemical probing with dimethyl sulphate (DMS) and potassium permanganate ($KMnO_4$) was employed to identify the structural changes from B-form DNA to the four-stranded G-quadruplexes within supercoiled DNA. Dimethyl sulphate methylates guanine residues at the N7 position in the major groove; guanines participating in G-quadruplex formation will not be accessible for methylation so should be protected from reaction with DMS. Guanines that do not participate in the G-quadruplex will be methylated, creating an adduct for which the DNA backbone can be cleaved using piperidine, as described by Maxam-Gilbert (Maxam and Gilbert, 1977). Alternatively, $KMnO_4$ reacts with exposed thymines, by an out-of plane attack across the C5-C6 bond. This is not accessible in duplex DNA and so permanganate can be used to identify any exposed thymines that are not participating in Watson-Crick base pairing. As in the case of dimethyl sulphate, modified thymine bases can also be cleaved by piperidine.

2.4.1.1 Methylation protection assay: reactions with dimethyl sulphate

Supercoiled plasmid - 4 µl of plasmid DNA (~300-400 ng/µl) was incubated overnight in 97 µl of 10 mM Tris-HCl pH 7.5 containing 100 mM KCl. 50 µl was removed and added to β-mercaptoethanol (1 µl). 0.5 µl of DMS (1%, v/v) was added to the remainder of the plasmid solution (50 µl) and after 1 min β-mercaptoethanol (1 µl) was added to halt the reaction. Ethanol (200 µl, 100 %) was then added to each sample and the plasmid was precipitated with ethanol as previously described. Each pellet was re-

dissolved in water (18 µl). 2 µl of the appropriate Promega 10x buffer was added and 1 µl (10 units) of restriction enzymes (either: *Hind*III or *Eco*RI plus *Sac*I or *Pst*I, depending on which strand was being studied) were then added to remove the insert and the mixture was digested for at least 1 hr at 37 °C. The modified plasmid was labelled with ^{32}P at the 3'-end as described in section 2.3.1.4.

Linearised plasmid - 4 µl of plasmid (~300-400 ng/µl) was added to water (13 µl), 2 µl of the appropriate Promega buffer and 1 µl (10 units) restriction enzyme (either *Sac*I or *Pst*I, depending on which strand was being studied) and left to digest for 1 hr at 37 °C. The plasmid was then precipitated with ethanol (100 µl, 100 %) and purified as previously described. The plasmid was re-dissolved in water (10 µl). 90 µl of 10 mM Tris-HCl pH 7.5 containing 100 mM KCl was then added and left to incubate overnight. 50 µl was removed and added to β-mercaptoethanol (1 µl). 0.5 µl of DMS (1%, v/v) was added to the remainder of the plasmid solution (50 µl) and after 1 min β-mercaptoethanol (1 µl) was added to stop the reaction with DMS. Ethanol (200 µl, 100 %) was then added to both samples and the DNA was precipitated as previously described. Each pellet was re-dissolved in water (18 µl). 2 µl of the appropriate Promega 10x buffer was added, followed by 1 µl (10 units) of restriction enzyme (*Hind*III or *Eco*RI, depending on the strand being examined) and the mixture was digested for at least one hr at 37 °C. The modified plasmid was then 3'-labelled as described in section 2.3.1.4.

2.4.1.2 Oxidation assay: reaction with potassium permanganate

The permanganate assays were conducted similarly to the DMS reactions.

Supercoiled plasmid - 4 µl of recombinant plasmid (~300-400 ng/µl) was incubated overnight in 97 µl of 10 mM Tris-HCl pH 7.5 containing 100 mM KCl. 50 µl was removed and added to β-mercaptoethanol (1 µl). 0.5 µl of 100 mM $KMnO_4$ was added to the remainder (50 µl) of the plasmid (giving a final concentration of 1 mM). After 20 min 1 µl of β-mercaptoethanol was added to halt the reaction. Ethanol (200 µl, 100 %) was then added to each sample and the plasmid was precipitated as previously

described. Each pellet was re-dissolved in water (18 µl). 2 µl of the appropriate Promega 10x buffer (2 µl), and 1 µl (10 units) of restriction enzymes (either: *Hind*III or *Eco*RI plus *Sac*I or *Pst*I, depending on which strand was being studied) were then added to remove the insert and the plasmid left to digest for at least 1 hr at 37 °C. The modified plasmid was labelled with ^{32}P as described in section 2.3.1.4.

Linearised plasmid - 4 µl of recombinant plasmid (~300-400ng/µl) was added to water (13 µl), 10x the appropriate Promega buffer (2 µl), restriction enzyme whose cleavage of DNA does not allow 3'-end radio-labelling (1 µl, 10 units) and left to digest for 1 hr at 37 °C. The plasmid was ethanol precipitated (100 µl, 100 %) as previously described and re-dissolved in water (10 µl). 90 µl of 10 mM Tris-HCl pH 7.5 containing 100 mM KCl was then added and left to incubate overnight. 50 µl was removed and added to β-mercaptoethanol (1 µl). 0.5 µl of 100 mM $KMnO_4$ was added to the remainder of the plasmid solution (to give a final permanganate concentration of 1 mM) and after 20 min, β-mercaptoethanol (1 µl) was added to stop the reaction. Ethanol (200 µl, 100 %) was then added to both samples and the plasmid was precipitated as previously described in section 2.3.1.5. Each pellet was re-dissolved in water (18 µl). 1x the appropriate Promega 10x buffer was added, followed by 1 µl (10 units) of restriction enzyme (*Hind*III or *Eco*RI, depending on the strand being examined) and the mixture was digested for at least 1 hr at 37 °C. The modified plasmid was then 3'-labelled as described in section 2.3.1.4.

2.4.2 Piperidine Cleavage of Modified Bases

Piperidine (10 µl) was added to each sample of modified DNA (to give a final concentration of 10%, v/v) and the samples were heated for 30 min at 100 °C to cleave the DNA at the modified bases. After boiling, the samples were centrifuged for a few seconds (200 rpm) and immediately placed on dry-ice. The piperidine was removed by lyophilisation. In the initial experiments the piperidine was removed by speed vacuum, but freeze drying proved the most effective. The plasmid was washed with water (50 µl) which was also removed by lyophilisation. This washing step was repeated several times until all traces of piperidine had been removed.

2.4.2.1 Denaturing Polyacrylamide Gel Electrophoresis

Each DNA pellet was then dissolved in 4 µl of TE2 buffer. 4 µl of DNase I stop solution was then added and the samples heated for 3 min at 100 °C. The samples were placed on ice to cool and then loaded on to an 8% denaturing polyacrylamide gel containing 8 M urea and electrophoresed for approximately 2 hr at 1500 V. The gels were fixed in acetic acid (10%, v/v), transferred to Whatman 3MM paper and dried under vacuum for 1.5 hr at 80 °C. The dried gels were then subjected to phosphorimaging and visualised using ImageQuant 5.0.

2.4.3 S1 Mapping

S1 nuclease digests single-stranded DNA and not duplex DNA, so by utilising this enzyme it is possible to identify either the potential single-stranded loop regions of the G-quadruplex structure (together with the single stranded complementary C-rich strand) or the G-quadruplex-DNA duplex interface as shown in Figure 2.2a. By digesting the supercoiled DNA with S1 nuclease, the plasmid would be linearised if it contains any single-stranded regions. The location of these cleavage sites can then be mapped by digesting with a second enzyme (*Sca*I) and examining the sizes of the DNA fragments that are produced. If there are no specific S1 sites then cleavage with the second enzyme will merely linearise the circular plasmid DNA. If the plasmid is linearised with *Sca*I before digestion with S1 nuclease then this will eliminate any supercoil-dependent single stranded regions (such as those that might be produced on G-quadruplex formation) and only one band should be visible. *Eco*RI was used to map the region of quadruplex formation as the *Eco*RI restriction site is situated in close proximity to the *Bam*HI cloning site and so will be adjacent to the putative quadruplex-forming region. Figure 2.2b shows a schematic representation of the expected results.

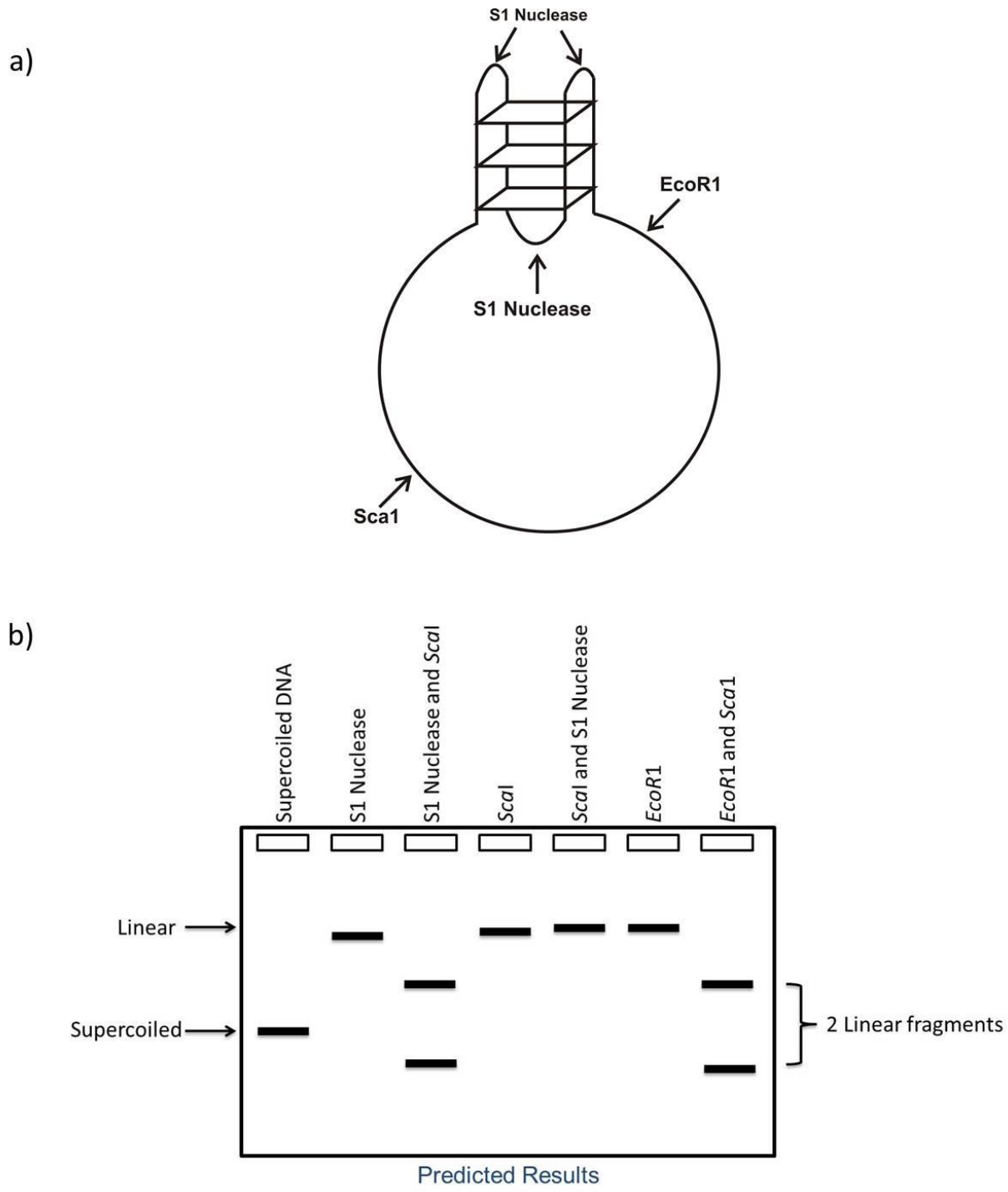


Figure 2.2 Diagram depicting the potential region of S1 cleavage of the G-rich strand of the DNA plasmid. a) The quadruplex-duplex junction is also a potential site for cleavage by S1 nuclease. The *EcoR*I and *Sca*I cleavage sites are used to map the location of the S1 nuclease cleavage sites of the quadruplex-forming sequences that have been cloned into the polylinker site of plasmid pUC19. The complementary C-rich strand is not shown, but this too will be cleaved with S1 nuclease, as it is unlikely to form the four stranded i-motif. b) A Schematic representation of the mapping of the G-quadruplex region within the pUC19 cloning vector. The fragment sizes indicated are not intended to be completely accurate.

2.4.3.1 Digestion of supercoiled plasmid

A sample of supercoiled plasmid was either incubated in the absence of KCl, or in the presence of 10 mM Tris-HCl pH 7.4 containing 100 mM KCl, S1 nuclease (8 units) and 1x S1 nuclease buffer for 30 min. DNA was ethanol precipitated as described in section 2.3.1.5 and re-suspended in water (20 µl). The DNA was then digested with 1 µl (8-12 units) of *Sca*I restriction enzyme in 1x *Sca*I buffer for 1 hr at 37 °C.

2.4.3.2 Digestion of linear plasmid

The supercoiled plasmid was first linearised by digesting with 1 µl (8-12 units) of *Sca*I. The plasmid was ethanol precipitated as previously described in section 2.3.1.5 and then re-suspended in water (20 µl) or 20 µl of 10 mM Tris-HCl containing 100 mM KCl. The plasmid was then digested with S1 nuclease (8 units) and 1x S1nuclease buffer for 30 min.

Plasmids were also digested with *Sca*I and *Eco*RI. The samples were loaded onto 1% agarose gels containing GelRed^{TM} (0.1 µl/ml) and visualised using a transilluminator.

2.5 Two-Dimensional Gel Electrophoresis

Two-dimensional gel electrophoresis was employed to reveal whether any topology-dependent structural transitions occurred within closed circular plasmids containing guanine-rich inserts. Experiments were performed as described by Bowater (Bowater et al., 1992). The initial step required for these experiments was to prepare plasmid samples that contained a good distribution of topoisomers.

2.5.1 Preparation of Topoisomer Distributions

Supercoiled plasmid samples (~ 300 ng/µl) with varying concentrations of ethidium bromide (range between 0 and 2.0 µg/ml) were prepared according to Table 2.2. The samples were incubated with 1 µl of wheat germ topoisomerase I for 1 hr at 37ºC. The DNA was then ethanol precipitated as previously described and the pellets re-suspended in 20 µl of TE2 buffer. These samples were then stored at -20 ºC.

Table 2.2 Volumes (µl) required for the preparation of topoisomer distributions

	Final EtBr Concentration (µg/ml)								
	0	**0.125**	**0.25**	**0.375**	**0.5**	**0.75**	**1.0**	**1.5**	**2.0**
DNA[a] (µl)	9	9	9	9	9	9	9	9	9
EtBr (5 µg/ml)	-	0.5	1	1.5	2	3	4	6	8
10 x Buffer[b]	2	2	2	2	2	2	2	2	2
Water	8	7.5	7	6.5	6	5	4	2	-
Topoisomerase 1 (µl)	1	1	1	1	1	1	1	1	1

[a] Stock solution of supercoiled DNA (~ 300 ng/µl) was incubated with wheat germ topoisomerase 1 in the presence of a final concentration of ethidium bromide (EtBr) of between 0 and 2 µg/ml.[b] The 10x topoisomerase 1 buffer used for the incubation contained 50 mM Tris-HCl, pH 7.9, 50 mM NaCl, 1 mM EDTA, 1 mM DTT and glycerol (20 %, v/v)

Each of the sample incubations generates a Gaussian distribution of topoisomers, in which the ethidium bromide concentration determines the position of the mean. Each set of distributions was examined by one-dimensional (1D) gel electrophoresis using a 1% agarose gel run in 1x TBE. This was then stained with GelRed™ (1 µl/100 ml). Samples containing an even distribution of topoisomers were prepared by combining these samples in appropriate ratios as judged from the intensity of the various bands in the 1D gel. Each mixture was prepared so as to cover the complete range of *ΔLK* (change in linking number, see Chapter 5). This sample was then used for two-dimensional gel electrophoresis analysis.

2.5.2 Experimental Procedure for Two-Dimensional Gel Electrophoresis

A 1% agarose gel (20 x 20 x 1 cm) was prepared with a single circular well (2 mm in diameter) at the top right hand corner. A sample of the complete topoisomer distribution containing ~ 2.5 µg of DNA in a volume of 10 µl was prepared. Ficoll (5 µl) was added to the DNA sample, which was then loaded into the well. Electrophoresis in 1x TBE buffer was carried out for approximately 16-18 hr at 3.0 V/cm. The gel was then

disassembled and soaked for 8-10 hr in TBE buffer containing the required concentration of chloroquine diphosphate (2-10 µg/ml) in the dark. After soaking, the gel was then reassembled in the tank at a rotation of 90º and run in the second dimension in 1x TBE buffer containing the same concentration of chloroquine. Electrophoresis was conducted for approximately 16-18 hr at 3.0 V/cm. After electrophoresis the gel was removed from the tank and stained in GelRed^{TM} (1µl/100ml) in water overnight. The gel was then visualised and photographed using a Syngene, G-Box transilluminator.

2.6 Circular Dichroism

Circular dichroism (CD) is a spectroscopic technique which is used to study chiral molecules and can measure the two-state conformational structural changes in DNA oligonucleotides in solution (Blackburn G. Michael., 2006, Eriksson and Norden, 2001, Randazzo et al., 2013). CD measures the difference in absorption of right-handed circularly polarised light and left-handed polarised light. Although CD does not give clear information about absolute conformations it is a powerful tool for assessing difference and changes in conformation. Prior to cloning, the G-rich inserts were examined using CD spectroscopy to determine their folding topologies. The CD spectra of parallel and antiparallel G-quadruplexes are different. Parallel structures typically have a positive band at 260 nm and a negative band at 240 nm while antiparallel structures have a positive band at 295 nm and a negative band at 265 nm.

2.6.1 CD Measurements

All CD measurements were carried out using a Jasco J-720 spectropolarimeter.

2.6.1.1 Single G-rich Strand

The G-rich strands of the oligonucleotides used for cloning (Table 2.1) were diluted to 5 µM in 10 mM Tris-HCl, pH 7.5, supplemented with varying concentrations of potassium chloride. All samples were first heated to 95 °C for 10 min and left to anneal slowly overnight on the bench to room temperature to allow for folding of the G-quadruplex structures. Spectra were recorded between 220 and 320 nm in a 1 mm path length quartz cuvette. Spectra were averaged over 16 scans. Each scan was recorded at a

scan rate of 100 nm/min with a response time of 2 s and 1 nm bandwidth. A spectrum of the Tris buffer was taken to gain a buffer baseline which was subsequently subtracted from each spectrum and the spectra was then normalized to have zero ellipticity at 320 nm.

2.6.1.2 Effects of Watson-Crick base pairing and G-quadruplex formation

Equimolar mixtures of the G-rich and C-rich strands (5 µM) of the oligonucleotides used for cloning (Table 2.1) were prepared in 10 mM Tris-HCl pH 7.5 containing 100 mM KCl. The solutions were heated to 95 ºC for 10 min and left to anneal slowly overnight on the bench to room temperature. Spectra were recorded between 220 and 320 nm in a 1 mm path length quartz cuvette. Spectra were averaged over 16 scans. Each scan was recorded at 100 nm/min with a response time of 2 s and 1 nm bandwidth. A spectrum of the Tris buffer was taken to gain a buffer baseline which was subsequently subtracted from each spectrum and the spectra was then normalized to have zero ellipticity at 320 nm.

Chapter 3:
The Cloning and Characterisation of pUC19 G-quadruplex-Forming Plasmids

3.1 Introduction

G-quadruplex forming motifs are of particular importance due to their high abundance and function throughout the genomes of almost all prokaryotic and eukaryotic species. Bioinformatic studies have suggested that these occur most frequently in the proximal promoter regions of genes which influence transcription, and at the end of telomeres (Huppert and Balasubramanian, 2005, Todd et al., 2005, McCarthy and Heywood, 1987). However, while the last 100-200 nucleotides of telomeric G-quadruplexes are readily available in a single-stranded DNA, gene-promoter quadruplexes are constrained to the duplex formation (Balasubramanian et al., 2011). Therefore, conformational transitions from right-handed B-form DNA to a G-quadruplex structure would require the separation of Watson-Crick hydrogen-bond base pairing and the association of Hoogsteen-bond base-pairing. In order for this to occur *in vivo,* the localised unwinding and/or melting of the duplex DNA strands would be required. This is typically known to be facilitated by the free-energy change associated with the negative supercoiling of DNA, termed the 'free energy of supercoiling' (Wang et al., 1983). Studies of negatively supercoiled DNA have shown that DNaseI or S1 nuclease hypersensitive sites are found within regions of DNA that contain G-quadruplex forming motifs (McCarthy and Heywood, 1987, Sun et al., 2005, Sun and Hurley, 2009) and that these are also structurally dynamic and have the potential to convert from the canonical Watson-Crick double-stranded helix conformation to alternative secondary structures (McCarthy and Heywood, 1987, Sun, 2010, Sun and Hurley, 2009, Sun et al., 2005). *In vitro* studies of negatively supercoiled DNA have demonstrated that the resultant free energy can give rise to structures such as cruciforms (Panayotatos and Fontaine, 1987, Panayotatos and Wells, 1981, Mizuuchi et al., 1982), left-handed DNA (Nordheim et

al., 1983, Wang et al., 1983) and as mentioned above, G-quadruplexes (Sun, 2010, Sun and Hurley, 2009, Sun et al., 2005) which has also been mentioned in Chapter 1. Free energy can be naturally generated by the twisting and bending of DNA and as the B-form of DNA is a configuration of minimum energy, any stress generated from bending and twisting will increase its free energy which in turn may then be used to drive the conformational change from duplex to G-quadruplex (Bowater, 2005). Although it can be speculated that G-quadruplex structures may occur *in vivo* and play a significant role in biology, evidence to support this theory is somewhat sparse.

This chapter concentrates on the experimental techniques employed to clone and characterise specifically designed G-rich oligonucleotides into the DNA cloning vector pUC19 in order to mimic supercoiled *in vivo* conditions, to thereby address the question: can negative superhelicity provide the torsional energy required to induce G-quadruplex formation in unwound DNA? This question will be investigated by chemical and enzymatic probing (Chapter 4) and 2-dimensional gel electrophoresis (Chapter 5).

3.1.1 Circular Dichroism

The first part of this chapter examines the folding abilities of the oligonucleotide sequences selected for cloning using circular dichroism (CD). CD is a spectroscopic technique used to study chiral molecules particularly nucleic acids and nucleic acid-ligand interactions (Eriksson and Norden, 2001, Blackburn G. Michael., 2006, Randazzo et al., 2013). A CD signal results from the difference in left and right circularly polarised light that is conducted over a range of wavelengths. Different conformations of nucleic acids exhibit different spectral patterns, so that structural topology may be determined by referencing known solution structural motifs.

CD spectroscopy is an indirect method for determining the characterisation of folding topologies of DNA in solution. It is a not a precise technique and cannot be used unambiguously to determine structural characteristics in the manner of strand connectivity; however, it can offer insight into the effects of sequence, cations and ligand binding of these structures. Therefore additional characterisation techniques are

generally required to compliment and corroborate the results. In G-quadruplex research CD is routinely used to identify strand orientation of the tetrahelical motifs that can be categorised as either parallel or antiparallel. The differences in their CD spectra originates from the different stacking interactions of the guanine residues, in which the glycosidic bonds of the guanine can be oriented in either the *syn* or *anti* position. Thus, the position of the CD spectral peaks can offer an insight into loop arrangements in the G-quadruplex structure. The strands are characterised by distinct spectral properties where parallel quadruplexes exhibit a positive ellipticity maxima at 264 nm and a negative minima at 240 nm, whilst antiparallel complexes usually demonstrate a positive maxima at 295 nm and a negative minima at 265 nm (Figure 3.1) (Gray et al., 2008, Balagurumoorthy et al., 1992, Paramasivan et al., 2007, Vorlickova et al., 2005, Lu et al., 1993).

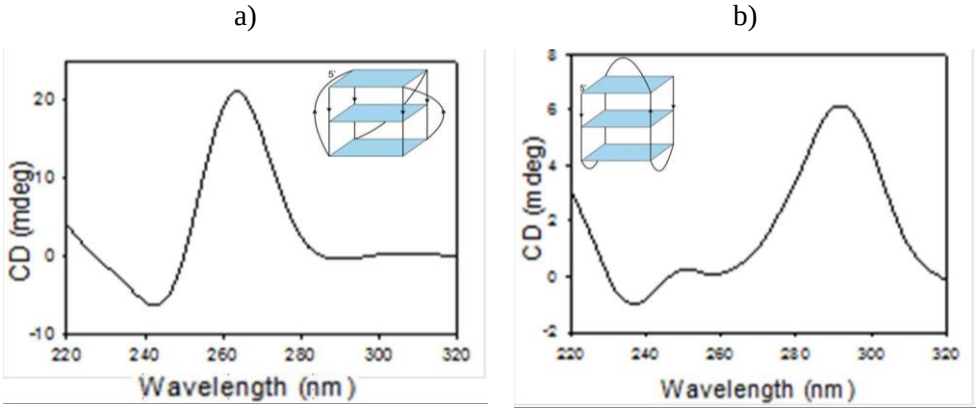


Figure 3.1 CD spectra of intramolecular G-quadruplexes. a) Parallel strand topology and b) Antiparallel Strand topology

3.1.2 Cloning: Blue-White Selection

The pUC19 vector (Yanisch-Perron et al., 1985) was designed with a set of unique restriction sites that can be used for the insertion of additional DNA fragments. The restriction sites are located within a copy of the *lacZ'* gene, where the gene provides a selectable marker that facilitates the identification of recombinant *E. coli* bacteria. The *lacZ'* gene codes for the first 146 amino acids of the *E. coli* β-galactosidase enzyme, which converts lactose into glucose and galactose. However, the segment coded by the *lacZ'* gene is not sufficient to catalyse the conversion, but is complemented by a second

peptide (made by the host bacterium) which contains the remainder of the β-galactosidase protein. The recombinant plasmid is then transformed into *E. coli* competent cells. After the transformation process, cells containing the recombinant plasmid are plated on agar plates that contain the lactose analogue X-gal (5-bromo-4-chloro-indolyl-β-D-galactopyranoside) and the plasmid was isolated by blue-white selection (Brown, 1994). Cells that contain a vector with an intact *lacZ* gene (i.e. without an inserted DNA fragment) are able to synthesise the β-galatosidase protein and convert X-gal to a deep blue coloured compound called bromochloroindole, therefore producing blue colonies. In contrast, cells that contain a disrupted *lacZ* gene (with a cloned insert in this region) cannot synthesise the β-galactosidase protein and subsequently produce white colonies.

3.1.3 Dideoxy Sequencing – The Sanger Method

In the 1970's Frederick Sanger (Sanger et al., 1977) developed a DNA sequencing technique, that is referred to by several terms; the dideoxy sequencing method, the Sanger method or the chain termination method. As the latter term suggests this method involves the use of chain-terminating dideoxynucleotides (ddNTPs) to cause base-specific termination of primed DNA sequences during replication (Hutchison, 2007). A ddNTP lacks the hydroxyl group attached to the 3' carbon sugar of a typical deoxynucleotide (dNTP) as shown in Figure 3.2.

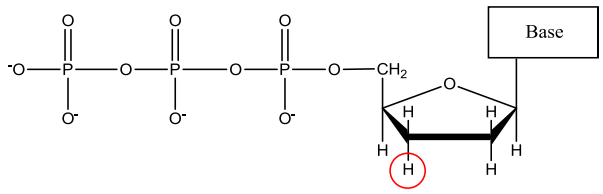


Figure 3.2 Diagram of a dideoxynucleotide. 2'3'-dideoxynucleotide 5'-triphosphate lacking the hydroxyl group attached to the 3' carbon. The H atom circled in red indicates the position where the -OH of a dNTP is replaced by -H

The principle behind this technique is that single stranded DNA molecules that differ in length by just a single nucleotide can be separated into distinct bands by polyacrylamide gel electrophoresis. This method requires DNA polymerase to copy specific regions of DNA under controlled conditions (Sanger et al., 1977). The initial step is to denature

the DNA plasmid followed by the annealing of a short oligonucleotide which acts as a primer. The chain elongation reaction is catalysed by DNA polymerase which requires the four deoxyribonucleotide triphosphates (dATP, dCTP, dGTP and dTTP) and one ddNTP. In a typical DNA polymerase elongation the DNA chain can be extended for thousands of nucleotides in length by introducing dNTPs to the chain, however in dideoxy sequencing the length of the chain is restricted by the incorporation of the ddNTP. This is due to the fact that the ddNTP which does not have a 3'-hydroxyl group, prevents the incorporation of another base as the hydroxyl group is required for the polymerase reaction and therefore the chain can no longer be elongated (Sanger et al., 1977, Brown, 1994). This polymerisation is carried out in four separate parallel reactions, with each reaction containing a separate ddNTP (i.e. ddATP, ddCTP, ddGTP and ddTTP). This results in four distinct groups of synthesised polynucleotide chains varying in lengths, with each strand truncated at different positions by one of the chain terminating ddNTPs. Each reaction is loaded onto a separate well on a polyacrylamide gel for electrophoresis; this allows the sequence to be read directly from the positions of each band on the gel.

3.2 Experimental Theory and Design

Although the majority of G-quadruplex research to date has focused on short, single-stranded linear fragments of DNA which readily fold into the quadruplex motifs, the relaxed nature of these linear sequences do not accurately reflect *in vivo* conditions. In fact, apart from the repetitive nature of telomeric DNA, genomic G-rich sequences are considerably diverse (Huppert and Balasubramanian, 2005, Todd et al., 2005, Todd and Neidle, 2011) and are flanked by large stretches of non-G-quadruplex forming motifs. Therefore the folding topology of a specific type of scaffold may not only be affected by the number of G-tracts (which can offer more than one form of folded structure) or loop sequence and size but also the flanking regions on either side (Risitano and Fox, 2004, Smirnov and Shafer, 2000, Hazel et al., 2004, Hazel et al., 2006). Therefore, to mimic *in vivo* style conditions, G-rich oligonucleotide sequences were cloned into pUC19 supercoiled DNA as described in Chapter 2.

3.2.1 Oligonucleotide Design Considerations

Each oligonucleotide sequence was designed with the expectation that at the very least they had the potential to form a standard three tetrad stacked G-quadruplex motif. Each DNA duplex was constructed so that it contained a GATC "sticky-end" at both the 5'-ends to allow for insertion into the *Bam*H1 restriction site (GGATCC) of the pUC19 vector. As each duplex oligonucleotide has a 5'-end GATC sequence complementary to that of the 3'-end, this allows the DNA molecules to be inserted in either orientation, but also allows the duplexes to anneal ensuring a library of plasmids with extended G-tracts can be generated.

3.2.1.1 Loops

The oligonucleotide sequences chosen for this study are based on either four or five d(GGG) repeats that are separated by thymine residues of $d(T)_n$: $d(G_3T_nG_3T_nG_3T_nG_3T_n)$ and $d(G_3T_nG_3T_nG_3T_nG_3T_nG_3)$ respectively: where n is either: one, four or eight. The thymine bases provide loops (or linkers) connecting the G-quartets that support the G-quadruplex structure and the difference in loop-length plays a key role in the stabilisation or destabilisation of the structure. As other nucleotides may prove detrimental to the G-quadruplex formation, thymine nucleotides were the only bases considered for the loop region for the following reasons:

1) Although single adenine base loops have the potential to form stable G-quadruplex structures, the selection of thymines over adenines is highly important in the stability of these structures. A previous study demonstrated that replacing the loop region in the human telomeric sequence d(TTA) with d(AAA) resulted in a complete destabilisation of the G-quadruplex structure (Risitano and Fox, 2003b).

2) Incorporating guanine residues into the loop region may lead to a sequence ambiguity as they can participate as part of the loop or the G-quartet.

3) Cytosine residues were also not considered, as they may generate Watson-Crick GC base pairs thereby disrupting the G-quadruplex structure.

Initially sequences with single T-loops were chosen for this study as several studies have shown that intramolecular quadruplexes with short thymine loop sequences tend to form the most stable quadruplexes, while longer loops form less stable structures (Risitano and Fox, 2003b, Hazel et al., 2004, Rachwal et al., 2007b). By substituting the single T-loops with longer T_4 and T_8 sequences, the importance of loop length on quadruplex folding and stability under supercoiled conditions can then be examined.

3.2.1.2 Sequence polymorphism and topology

G-quadruplex formation is not solely based on sequences containing the standard four repetitive tracts of at least two or more guanines. Due to the polymorphic nature of G-quadruplexes, in theory G-rich DNA with more than four tracts of guanine repeats can form multiple folding topologies. This type of structural ambiguity is often observed in genomic quadruplex forming sequences such as the NHE III_1 region of the *c-myc* promoter gene (Simonsson et al., 1998) or the G-rich sequence of the *bcl-2* gene (Dai et al., 2006a, Dai et al., 2006b). To avoid any possible structural ambiguity, the standard four G_3 repeat sequence was chosen to ensure the formation of the typically reported 3-stacked structure. However, by introducing an extra run of G_3 residues to create a fifth guanine tract and/or by increasing the number of G-quadruplex forming motifs within the plasmid, the folding abilities associated with these longer sequences can be investigated. For instance, by using a five guanine tract sequence, it is possible to explore the role in which either the two external G_3-tracts at the 5'- and 3'-end (first and fifth G_3-tract respectively) or the central G_3-tract (tract three) plays as tracts two and four must be included in the structure. Therefore it is possible that these DNA molecules will produce a sample that contains a mixed species of different loop isomers with structures consisting of a 5'-GATCGGGT$_n$ or a 3'-T$_n$GGGGATC overhang or that the central G_3-tract may be accommodated into an extended loop sequence with both the flanking 3'- and 5'- thymines. An example of this was demonstrated by Simonsson *et.al* (Simonsson et al., 1998) using the *c-myc* oncogene promoter sequence (Pu27) found in the NHE III_1 containing six runs of guanines d($TG_4AG_3TG_4AG_3TG_4A_2G_2$). They

discovered that four unpaired guanines G11–G14 were incorporated in a longer extended loop of the G-quadruplex structure.

By creating plasmids that contained sequences that are varied by number of consecutive G-tracts and/or loop length, it can be determined whether these differences play a vital role in the extrusion of G-quadruplex structures from supercoiled DNA. These findings will be discussed in Chapters 4 and 5.

3.2.2 Circular Dichroism

Prior to cloning these sequences for the subsequent analysis of G-quadruplex formation in supercoiled DNA, it was important to demonstrate that the selected sequences can form G-quadruplex structures and that the GATC flanking regions do not inhibit quadruplex formation, as flanking regions can influence the stabilisation/destabilisation of the structures (Zhang et al., 2005, Luu et al., 2006, Phan et al., 2006, Gaynutdinov et al., 2008, Yue et al., 2011). The oligonucleotide sequences used in the CD experiments are listed in Table 2.1. The CD spectra for each of the strands were determined as previously described in Chapter 2.

3.2.3 Cloning and Sequencing

Supercoiled G-quadruplex forming plasmids were constructed as previously described in Chapter 2, by inserting G-rich oligonucleotide DNA sequences (Table 2.1) into the *Bam*HI restriction site of the vector pUC19 (Fig. 2.1). Each recombinant plasmid was transformed into TG2 cells, isolated and characterised by the Sanger dideoxy sequencing method either via manual sequencing using a T7 DNA sequencing kit purchased from Amersham Pharmacia and/or commercial sequencing by Eurofins.

Bioinformatic studies have identified G-quadruplex forming motifs throughout the human genome (Huppert and Balasubramanian, 2005, Todd et al., 2005); therefore if these structures are formed *in vivo* it is probable that multiple G-quadruplexes will exist in dynamic equilibrium at any given time. Therefore the final part of this cloning chapter focuses on the generation of multiple inserts of the G-quadruplex forming sequences in order to achieve multimeric structures within one plasmid. By ensuring that each duplex oligonucleotide contains a GATC sticky-end at the 5'-end of both

strands this allows for specific ligation into the *Bam*HI site of pUC19. Therefore, multiple copies of this insert can also be assembled, thus creating extended chains of G-quadruplex forming DNA.

3.3 Results

3.3.1 Circular Dichroism

CD spectra for the intramolecular folding of the G-rich oligonucleotides used in the cloning experiments are displayed in Figure 3.3 to Figure 3.8.

Oligonucleotide formula (G_n,L,X)

Where:

n is the number of consecutive G_3 tracts,

L is the number of thymines in the loop

and X is the number of repeats of the G-quadruplex forming sequence

For example: G_5,4,2 = $[GATC(G_3T_4)_4G_3]_2$

3.3.1.1 The effect of potassium on the formation of G-quadruplex structures

The CD spectra of the monomeric single T-loop sequences containing both four (**G_4,1,1**) and five (**G_5,1,1**) G_3-tracts are shown in Figure 3.3a and b. Both samples displayed strong positive peaks at around 260 nm and a negative peak at around 240 nm even in the absence of potassium. These results suggest that these sequences can readily adopt a parallel-strand topology and are consistent with previous research that reported G-quadruplex structures with single thymine residues in the loop typically adopt very stable parallel stranded propeller-type structures (Hazel et al., 2004). The CD spectra

also indicated that the absence of potassium did not prevent the formation of the G-quadruplex structure. A similar trend was observed with the dimeric single T-loop G-quadruplex sequences of four and five G_3-tracts (**G_4,1,2** and **G_5,1,2** respectively) as shown in Figure 3.3c and d. A strong positive peak was observed at 260 nm and a strong negative peak at 240 nm. Although these peaks which are characteristics of a parallel folded topology were clearly defined, a small shoulder in the absence of potassium was observed around the 280-285 nm region. However, the addition of as little as 1 mM KCl was sufficient to induce a change to the CD spectrum and the shoulder at around 280-285 nm was no longer present. The presence of this shoulder may be due to the single stranded GATC region that separates the two tetrahelical structures. It is also possible that in the absence of a monovalent cation the dimeric structures may not be as stable as those formed in the presence of cations, suggesting there may be some single stranded, unfolded regions present as well.

The CD spectra for the monomeric four G_3-tract T_4-loop sequence **G_4,4,1** displays potassium dependent conformational changes, shown in Figure 3.4a. In the absence of potassium two positive bands were observed at around 280 nm and 260 nm, and a negative band at around 240 nm which becomes more pronounced after the addition of 1 mM KCl. As these are not the standard G-quadruplex peaks this implies that the sample may be predominantly single stranded. Increasing the potassium concentration to 5 mM, induces a large change in the CD spectra resulting in a sharp transition to two distinctive peaks which becomes more pronounced as potassium concentration is increased. A positive peak at around 295 nm and a negative peak at around 260 nm consistent with an antiparallel topology were observed. In this instance the presence of a monovalent cation appears to be essential for the formation of the G-quadruplex and plays a key role in stabilising the structure with a longer loop. A similar trend was also observed for the five repeat sequence **G_5,4,1** as shown in Figure 3.4b.

A slightly different spectral pattern was observed for both dimeric sequences **G_4,4,2** and **G_5,4,2** (Figure 3.4 c and d respectively). In the absence of potassium, positive bands were observed at around 280 nm and 260 nm and a negative band at 240 nm. Again these bands are not typical of an antiparallel G-quadruplex topology, but of single-

stranded DNA. The addition of 5 mM potassium changes the CD spectra and the positive peak at 285 nm becomes more pronounced, the second positive peak at 260 nm is converted to a negative peak at the same wavelength, whilst the negative peak at 240 nm remains. Increasing the potassium concentration to 10 mM, introduces a sharp shift to the spectral peaks and a positive peak at 295 nm and a negative peak at 260 nm is observed. The peaks are more pronounced as potassium concentration is increased and the spectra are consistent with the formation of a potassium-dependent antiparallel structure.

Finally the CD spectra of **G_4,8,1** and **G_5,8,2** are shown in Figure 3.5a and b. Both sequences display positive bands in the region of 270-280 nm and as the ionic concentration increases, these bands shift towards the 270 nm region. Both samples also display negative bands at around 245 nm. Neither of these samples suggests that G-quadruplex formation has occurred as they do not exhibit peaks that are characteristic of either a parallel or antiparallel topology. They do however exhibit the typical peaks associated with classical B-form DNA which are characterised by a positive band at around 260-280 nm and a negative band at around 245 nm.

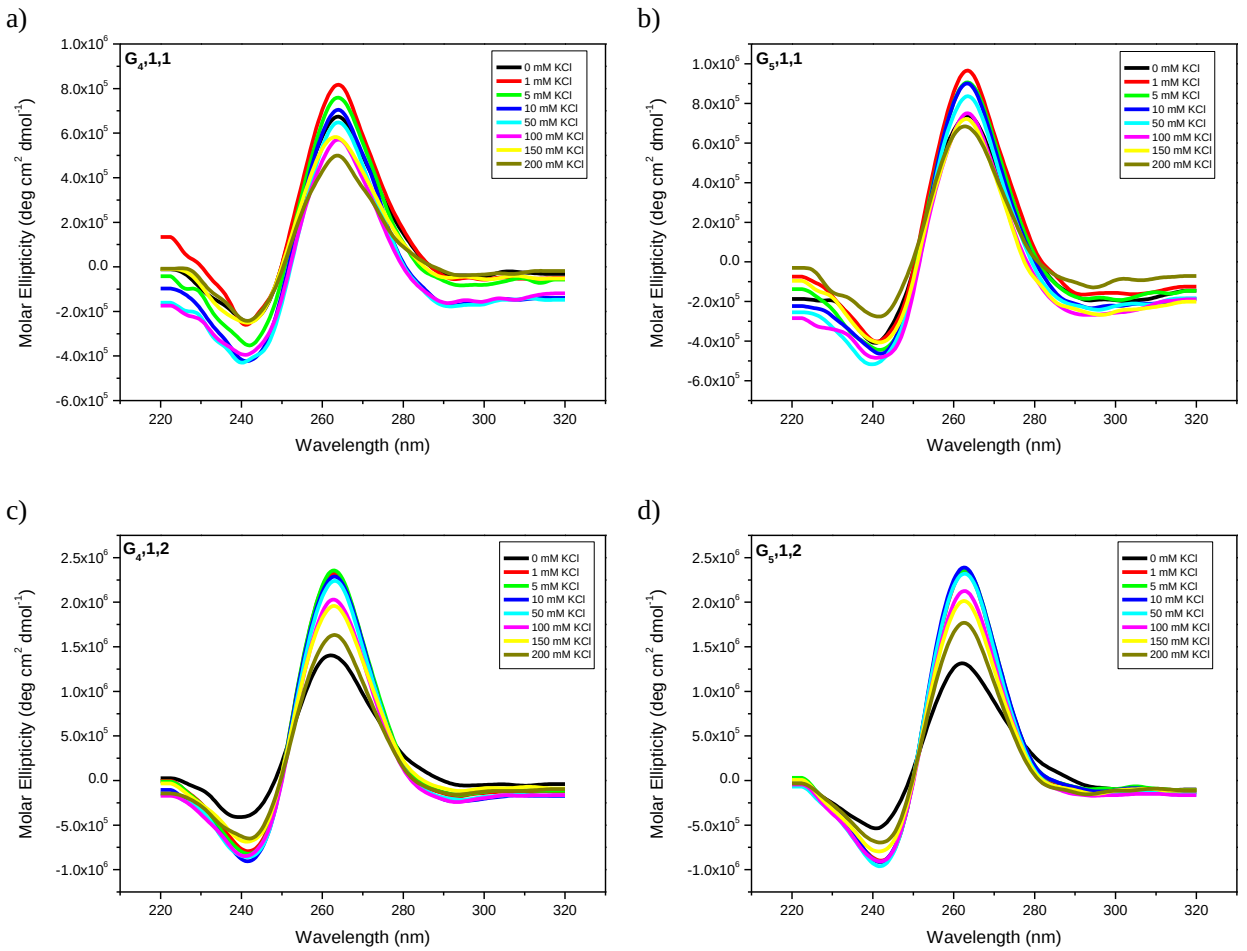


Figure 3.3 CD spectra of the T-loop quadruplex-forming oligonucleotides. i) G_4,1,1, ii) G_5,1,1, iii) G_4,1,2 and iv) G_5,1,2. CD spectra of quadruplex-forming oligonucleotides (5 µM) were conducted in 10 mM Tris-HCl buffer (pH7.4) in the presence of increasing concentrations of KCl (0 to 200 mM). The spectra were measured at room temperature, a buffer baseline was subtracted and then normalised to have zero ellipticity at 320 nm.

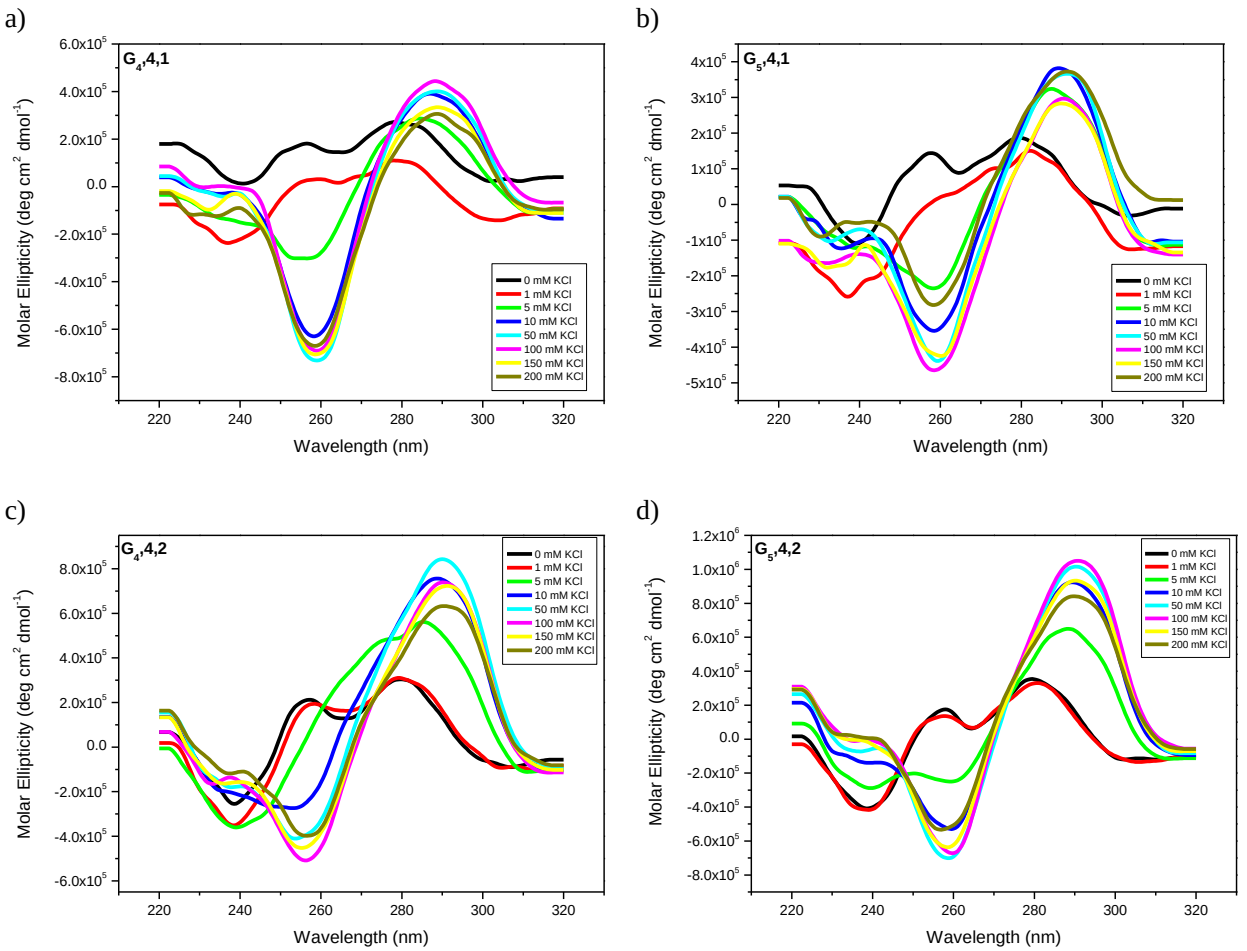
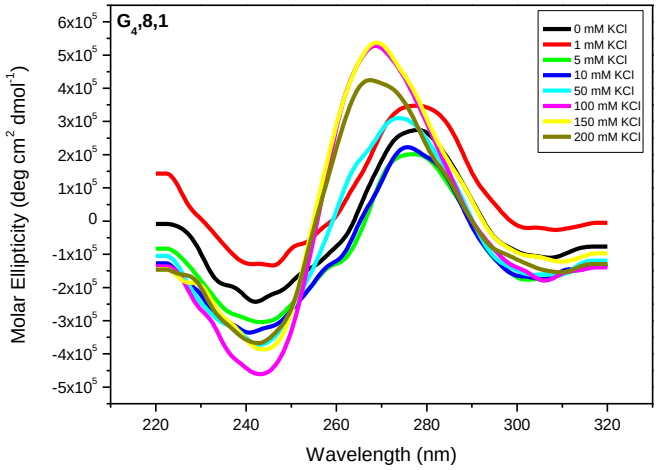


Figure 3.4 CD spectra of the T_4-loop quadruplex-forming oligonucleotides CD spectra of quadruplex-forming oligonucleotides (5 µM), i) G_4,4,1, ii) G_5,4,1, iii) G_4,4,2 and iv) G_5,4,2 were conducted in 10 mM Tris-HCl buffer (pH7.4) in the presence of increasing concentrations of KCl (0 to 200 mM). The spectra were measured at room temperature, a buffer baseline was subtracted and then normalised to have zero ellipticity at 320 nm.

a)



b)

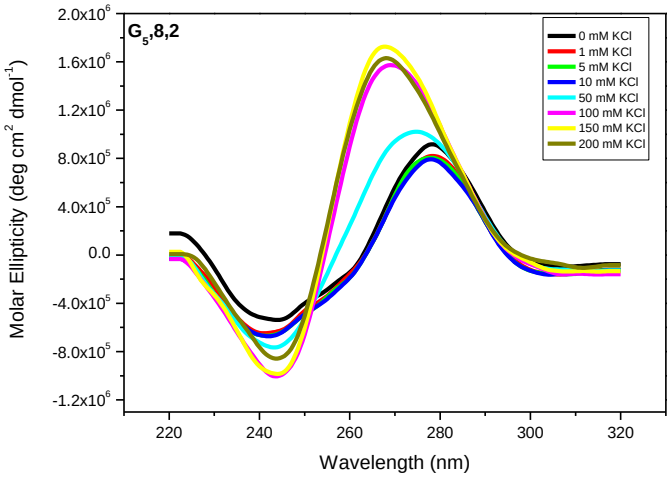


Figure 3.5 CD spectra of the T_8-loop quadruplex-forming oligonucleotides. CD spectra of the quadruplex-forming oligonucleotides (5 µM), i) G_4,8,1, ii) G_5,8,2 were conducted in 10 mM Tris-HCl buffer (pH7.4) in the presence of increasing concentrations of KCl (0 to 200 mM). The spectra were measured at room temperature, a buffer baseline was subtracted and then normalised to have zero ellipticity at 320 nm.

3.3.1.2 G-quadruplex-duplex competition

Aside from the single stranded segments of guanine DNA at the end of telomeres, most genomic G-rich sequences are present along with their C-rich complementary strand. This generates competition with the double helix that occurs via Watson-Crick complementary base pairing and Hoogsteen hydrogen-bonding of the G-quadruplex.

Moreover, successful cloning of these sequences relies on the formation of duplex structures. Therefore the final part of the CD experiments investigates the interconversion from G-quadruplex to duplex DNA on addition of the complementary strand.

The CD spectra of the G-rich strands (5 µM) were first recorded in Tris-HCl buffer (pH 7.5) supplemented with 150 mM KCl. An equimolar concentration of the C-rich strand was then added, the solution was heated to 95 ºC to destabilise any G-quadruplex formation of the G-rich strand and then left to anneal by slowly cooling to room temperature. The CD spectra were then recorded as previously described in Chapter 2.

Figure 3.6a shows the CD spectral results for the sequence **G_4,1,1** in the presence and absence of its complementary strand**.** As previously shown in Figure 3.3a, the CD spectrum of the solitary G-strand shows a strong positive peak at around 265 nm and a negative peak at 240 nm, which is indicative of a parallel arrangement. On addition of its complementary C-strand, the CD spectrum showed very little change. This suggests that under these conditions the G-quadruplex structure is preferred over the duplex, which is consistent with previous quadruplex fluorescence melting studies (Risitano and Fox, 2003b). The five G-tract monomer sequence (**G_5,1,1**) (shown in Figure 3.6b) also displayed strong positive peaks around 265 nm and a negative peak at 240 nm both with and without the complementary strand. However, a slight shift in the positive peak from 265 nm to 270 nm was observed on addition of the C-strand. This shift may reflect the fact that only four of the five G_3 tracts can be involved in the quadruplex. For example if the second, third and fourth G-tracts contribute to the G-quadruplex, then it is only possible for either the first or the fifth G_3 tract to also be included, thus leaving a G_3 overhang at either the 5'- or the 3'-end. Alternatively the central G_3 tract maybe excluded from quartet structures and instead participates in an extended loop formation. Therefore it may be possible that this region of single-stranded DNA interacts with a fragment of the C-rich strand, causing a portion of the spectrum to move toward that of duplex DNA.

The single T-loop dimeric sequences; **G_4,1,2** and **G_5,1,2** Figure 3.6c and d show similar results to that of the monomeric sequences. The CD spectra in the absence of the complementary C-strand showed strong peaks at both 265 nm and 240 nm. In the presence of the complementary C-strand a shift in the positive peak towards 270 nm was observed. Again it appears as though a significant proportion of the DNA is retained in a quadruplex structure, though the shift in peaks suggests some duplex formation, which may result from annealing with the central GATC or the flanking G_3 sequences.

Figure 3.7a-d illustrates the CD spectra for the T_4 loops (**G_4,4,1**; **G_4,4,2**; **G_5,4,1** and **G_5,4,2)**. In the absence of the complementary C-rich strand, each G-rich oligonucleotide exhibited positive peaks at around 295 nm and negative peaks at around 260 nm (as shown in Figure 3.4) indicative of an antiparallel topology, although the dimeric sequences displayed less pronounced peaks. On addition of the complementary strands the CD spectra changed dramatically showing positive bands between 260 and 280 nm and a negative peak at 245 nm. This suggests that the G-quadruplex structure had not folded and that duplex formation was dominant.

The CD spectral data for T_8-loop sequences; **G_4,8,2** and **G_5,8,2** are displayed in Figure 3.8 a and b. These sequences showed broad maxima peaks at 280 nm and broad minima peaks at around 245 nm (see Figure 3.5). While these peaks are not typical of G-quadruplex structures, addition of their complementary strand produces peaks in similar positions to those observed with the G_3T_4 sequences (Figure 3.7), suggesting that for these sequences the G-quadruplex formation was unfavourable and in the presence of the C-rich strand the duplex configuration was dominant under these conditions.

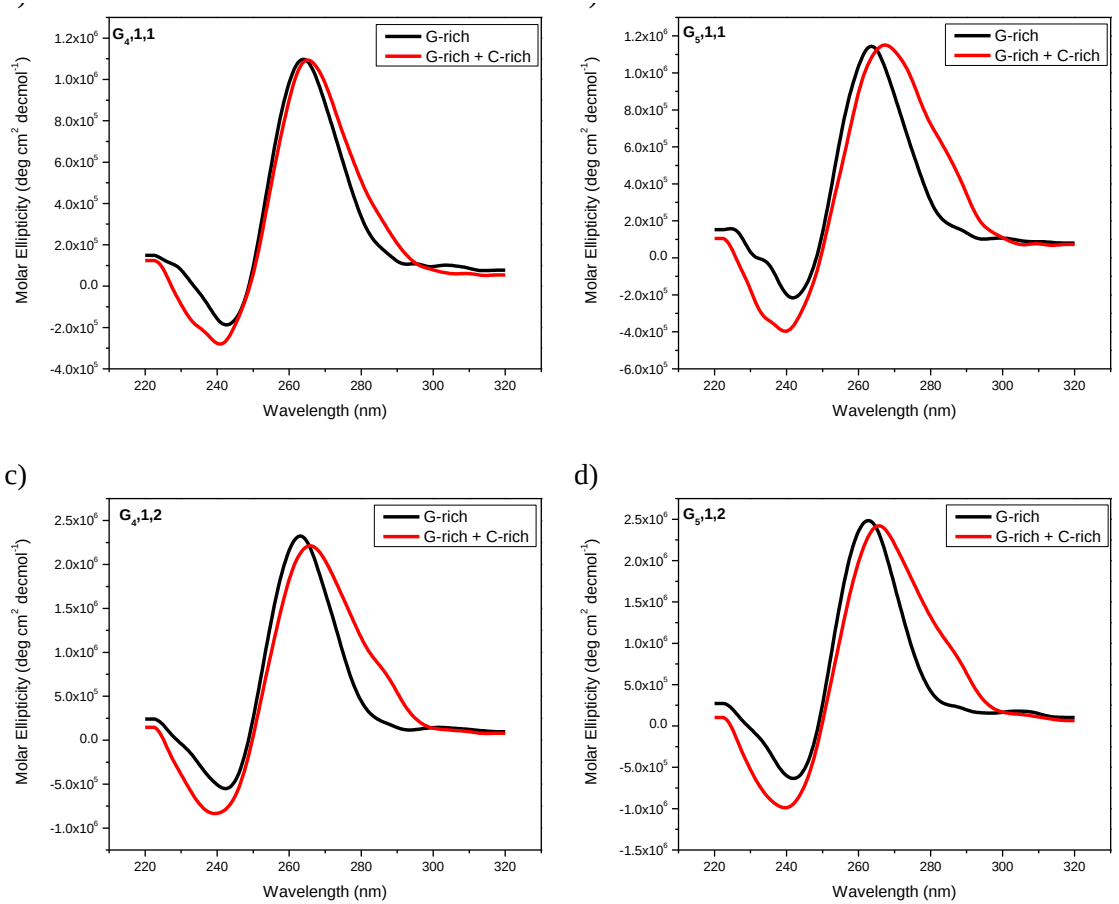


Figure 3.6 CD spectra of the T-loop quadruplex-forming oligonucleotides in the presence of its complementary C-rich strand. CD spectra of quadruplex-forming oligonucleotides (5 µM), i) G_4,1,1, ii) G_5,1,1, iii) G_4,1,2 and iv) G_5,1,2 in 10 mM Tris-HCl buffer (pH7.4) supplemented with KCl (150 mM). The black spectra are the G-rich DNA in the absence of its complementary strand. The red spectra are the DNA after annealing equimolar concentrations of the G-strand with the complementary C-strand. The spectra were measured at room temperature, a buffer baseline was subtracted and then normalised to have zero ellipticity at 320 nm.

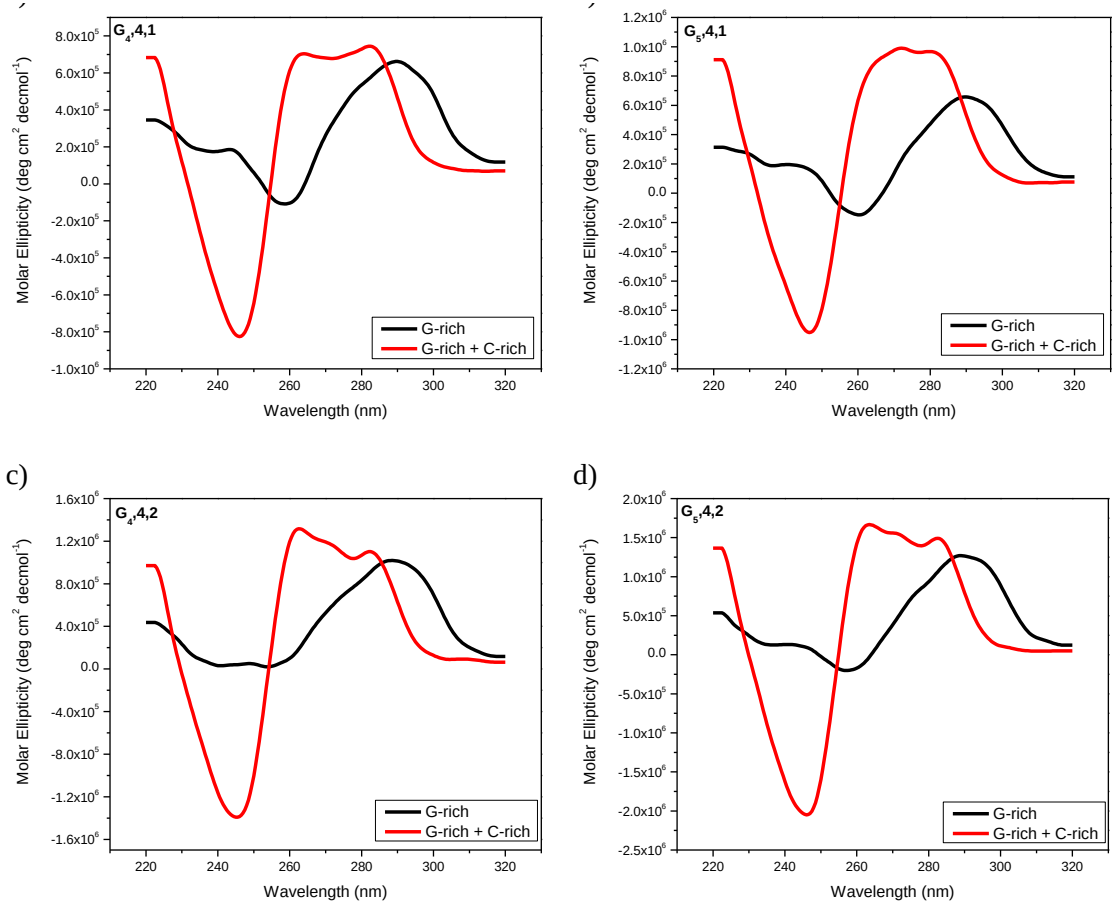


Figure 3.7 CD spectra of the T_4-loop quadruplex-forming oligonucleotides in the presence of its complementary C-rich strand. CD spectra of quadruplex-forming oligonucleotides (5 µM), i) G_4,4,1, ii) G_5,4,1, iii) G_4,4,2 and iv) G_5,4,2 in 10 mM Tris-HCl buffer (pH7.4) supplemented with KCl (150 mM). The black spectra are the G-rich DNA in the absence of its complementary strand. The red spectra are the DNA after annealing equimolar concentrations of the G-strand with the complementary C-strand. The spectra were measured at room temperature, a buffer baseline was subtracted and then normalised to have zero ellipticity at 320 nm.

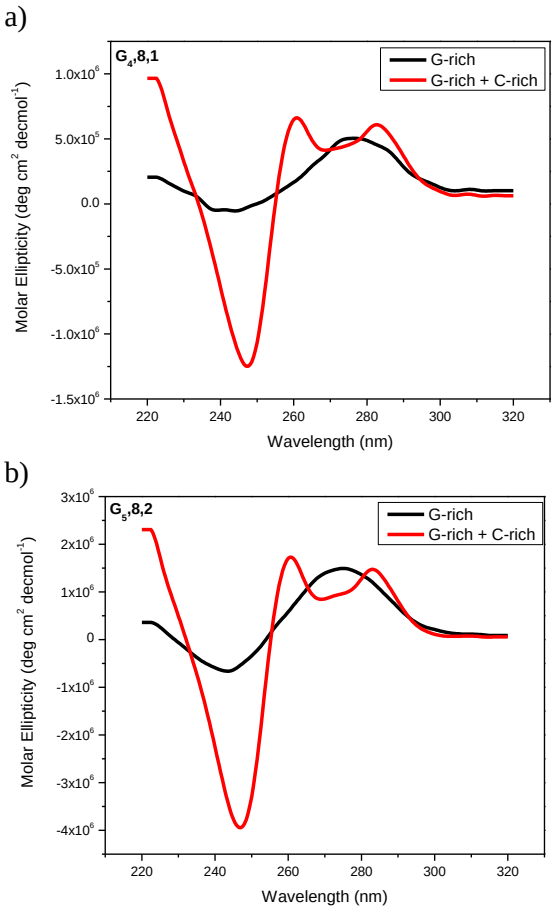


Figure 3.8 CD spectra of the T_8-loop quadruplex-forming oligonucleotides in the presence of its complementary C-rich strand. CD spectra of quadruplex-forming oligonucleotides (5 µM), i) G_4,8,1 and ii) G_5,8,2 in 10 mM Tris-HCl buffer (pH7.4) supplemented with KCl (150 mM). The black spectra are the G-rich DNA in the absence of its complementary strand. The red spectra are the DNA after annealing equimolar concentrations of the G-strand with the complementary C-strand. The spectra were measured at room temperature, a buffer baseline was subtracted and then normalised to have zero ellipticity at 320 nm.

3.3.2 Cloning of Potential G-Quadruplex Forming Oligonucleotide Sequences

3.3.2.1 Monomer inserts

In vitro recombination techniques were used to construct G-quadruplex forming plasmids. Initially the cloning experiments (described in Chapter 2) were conducted using oligonucleotides **G_4,1,1**, **G_5,1,1**, **G_4,4,1**, and **G_5,4,1** shown in Table 2.1. Equimolar concentrations of G-strand and C-strand were mixed together and annealed and inserted into the pUC19 cloning vector. The plasmids containing each of the inserts were characterised either by manual dideoxy sequencing or commercial sequencing by Eurofins.

Figure 3.9 and Figure 3.10 shows the commercial sequencing of the oligonucleotide fragments **G_4,1,1** and **G_4,4,1** cloned into pUC19 (**pG_4,1,1** and **pG_4,4,1**), whilst Figure 3.11 displays manual characterisation by dideoxy sequencing of the cloned DNA fragments **G_5,1,1**, and **G_5,4,1** (**pG_5,1,1** and **pG_5,4,1** respectively). In principle, due to the presence of the GATC sticky-ends at the 5'- ends of both strands, the annealed duplexes can be cloned in either orientation. Lane 1 in Figure 3.11 shows the sequencing results of the complementary C-strand whilst lane 2 shows the G-strand. All plasmids in Figure 3.11 show the C-strand. The clones: **pG_4,1,1**, **pG_4,4,1**, **pG_5,1,1** and **pG_5,4,1** were later used in the subsequent research.

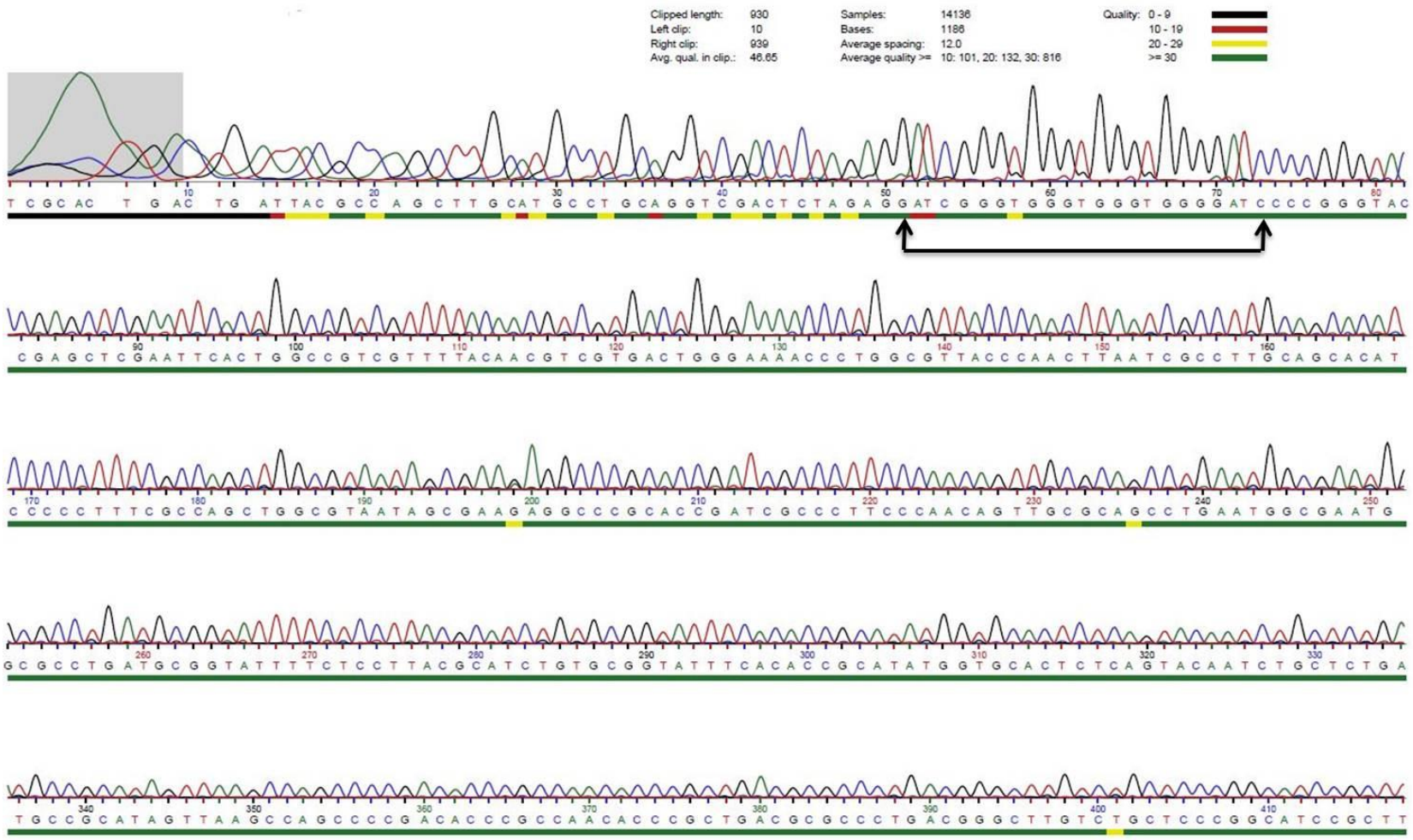


Figure 3.9 Automated chain-termination DNA sequencing of a p$(G_3T)_n$ clone. Commercial DNA sequencing of the plasmid containing oligonucleotide sequence **pG_4,1,1**. The cloned sequence indicated between the arrows. The plasmid was constructed by cloning the duplex insert into *BamH*I restriction site located in the MCS of pUC19.

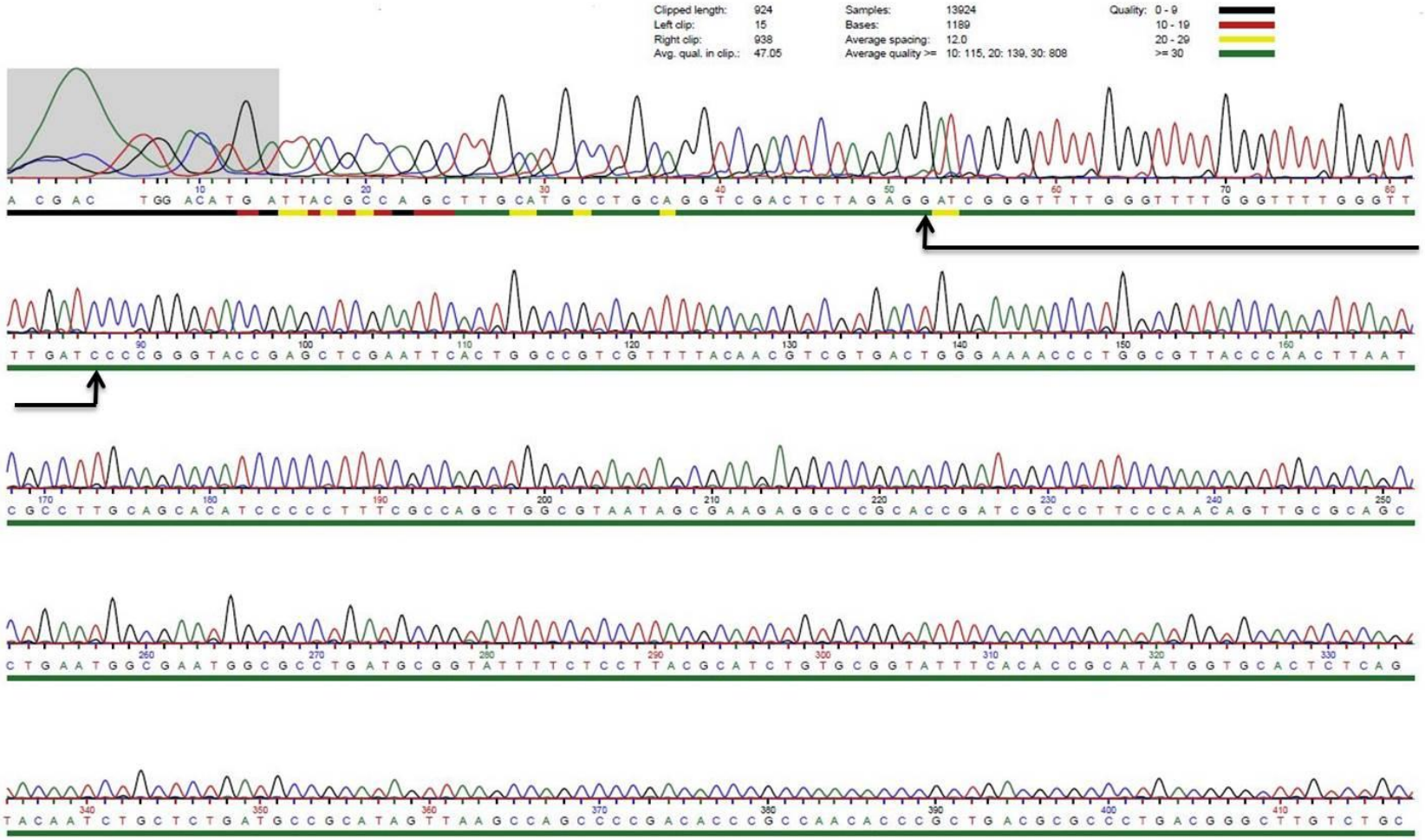


Figure 3.10 Automated chain-termination DNA sequencing of a $p(G_3T_4)_n$ clone. Commercial DNA sequencing of the plasmid containing oligonucleotide sequence **pG_4,4,1**. The cloned sequence indicated between the arrows. The plasmid was constructed by cloning the duplex insert into *BamH*I restriction site located in the MCS of pUC19.

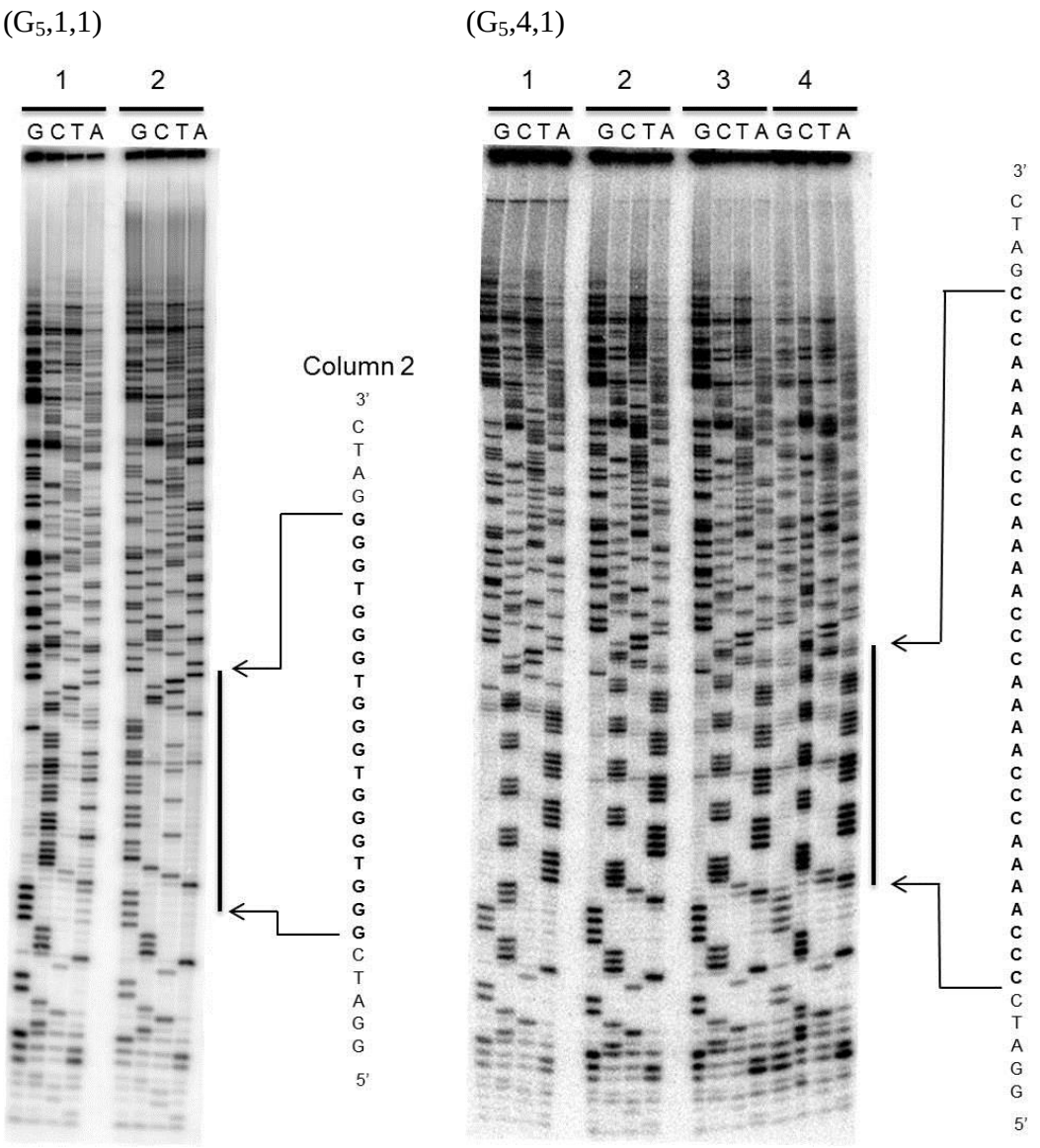


Figure 3.11 Manual dideoxy sequencing of G-rich plasmids. The experiments were conducted using a T7 sequencing kit and the samples were loaded to a polyacrylamide gel (8%). The gel reads from bottom to top (5' to 3'). **a)** shows (G_5,1,1) sequences 5'- d[GATC($G_3T)_3G_3$] column one indicates the C-strand and column 2 shows the G-strand and **b)** (G_5,1,1) sequence 5'- d[GATC($G_3T)_3G_3$].

3.3.3 Generating Plasmids with Extended G-Quadruplex Forming Sequences

The aim of this chapter was to generate a library of G-rich plasmids that have the potential to form either monomeric, dimeric or multimeric G-quadruplex structures within a single plasmid. Due to the propensity for G-rich DNA to assemble into quadruplexes, incorporating these linear fragments of DNA into the vector can be extremely problematic. This was particularly noticeable with the single T-loop G-quadruplex forming sequences, where relatively few clones were generated.

Furthermore each monomer G-rich duplex was designed to contain complementary 5'-GATC sticky-ends to allow for self-ligation, the sequencing analysis revealed that the four and five tract single T-loop sequences predominantly generated monomeric motifs. Although the four and five tract T_4-loop fragments did produce a modest number of dimer constructs again the monomeric motifs were dominant.

Therefore to increase the probability of inserting predominantly longer fragments (*i.e.* multiple inserts) into pUC19, extended repeat sequences varying in lengths were created by ligating the annealed duplexes overnight using T4 DNA ligase, prior to the insertion into the pUC19 plasmid. In theory, the ligated fragments should consist of multiple G-quadruplex forming units each separated by a GATC sequence, with 5'-GATC overhangs. The DNA was transformed into TG2 *E. coli* competent cells, the white colonies that indicated the presence of clones were selected and the DNA extracted and purified as described in Chapter 2. This protocol was able to produce several dimeric quadruplex-forming regions for all $(G_3T)_n$ and $(G_3T_4)_n$ complexes, however the monomeric plasmids were still the dominant species. No sequences longer than dimers were generated. Figure 3.12a and b shows the manual dideoxy sequencing results of the dimeric inserts of **G_5,1,1** and **G_5,4,1** (**pG_5,1,2** and **pG_5,4,2**). Lane 1 in Figure 3.12a shows the sequencing results of the C-strand of plasmid **pG_5,1,2**, whilst lane 2 shows that this plasmid did not contain any insert. Lane 4 in Figure 3.12b shows sequencing of the C-strand of the dimeric plasmid **pG_5,4,2** whilst lanes 1 and 2 show only monomer inserts (**p$G_5$4,1**), lane 3 contains no insert.

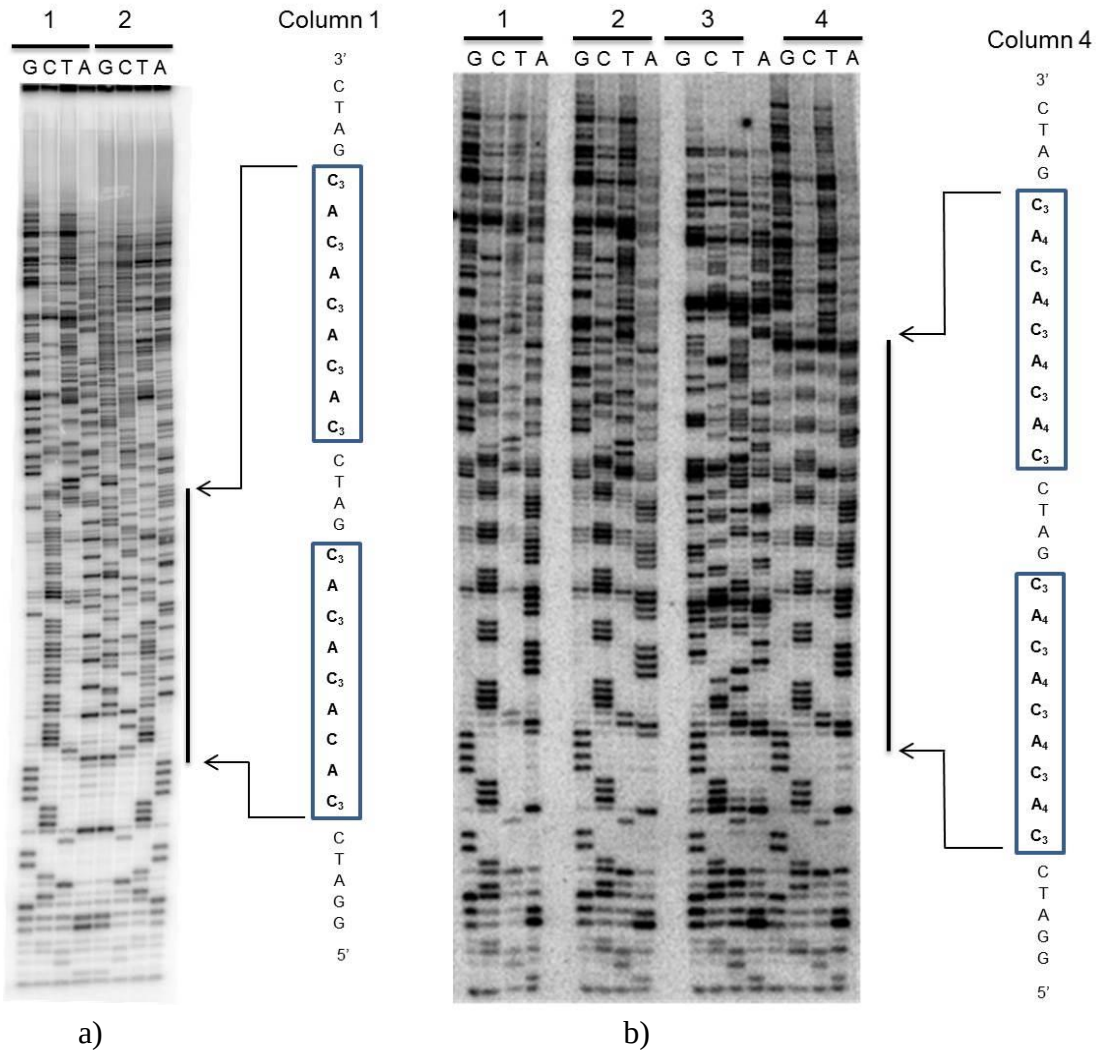


Figure 3.12 Dideoxy sequencing of plasmids containing G-rich oligonucleotides (extended inserts). a) Shows the sequencing results of the C-strand of G_4,1,1. Column 1 contains a dimeric insert whilst lane 2 does not contain an insert. b) Shows the sequencing results of the C-strand of G_5,1,1. Column 1 and 2 contain monomeric inserts of G_5,4,1, column 3 contains no insert whilst column 4 contains a dimer (G_5,4,2).

As the single T-loop sequences proved the most difficult to not only clone but to create plasmids with extended G-quadruplex forming sequences, the **G_5,4,1** (T_4-loop) oligonucleotide was selected to optimise the cloning experiments. Attempts under various conditions to insert the longer fragments were made. These included cooling fast rather than slowly when annealing the complementary strands, exchanging the annealing buffer from Tris-Na to Tris-EDTA to eliminate the sodium cation in an attempt to destabilise the G-quadruplex structures. Despite these changes the majority of clones generated were monomeric with only a small number of dimers. No clones longer than dimers were created. Interestingly, it was also observed that several of the white colonies selected by the blue-white screening (described in Chapter 2) contained DNA that was difficult to sequence. Due to the propensity for G-rich DNA to form non-conventional DNA structures, it was therefore considered that the interruptions observed during the sequencing process may have been introduced by the *E. coli* strain during DNA replication. This is because the formation of unusual DNA structures can be removed by the host cell's DNA repair system, by "repairing" the DNA through deletion or rearrangement during the transformation or replication process. Stratagene, is a company that has identified the bacterial pathways that are responsible for this process and by removing or altering the *E. coli* genes involved in the rearrangement and deletion of DNA have engineered super competent SURE® (Stop Unwanted Rearrangement Events) cells to help alleviate this problem. In order to determine whether the difficulties in the sequencing experiments were a result of the *E. coli* strain chosen for the transformation process, SURE® competent cells, TG2 and XL1-Blue cells (another commonly utilised strain that allows for blue-white screening) were employed in the cloning experiment of the **G_5,4,1** sequence, in an attempt to obtain multiple inserts.

After ligating the G-rich duplexes, the ligation mixture was added to linearised pUC19 and inserted into the *Bam*HI restriction site. This mixture was transformed into three separate types of competent cells; TG2, XL1-Blue and SURE® *E. coli* competent cells. Plasmid DNA was extracted from single white colonies from these plates and subjected to dideoxy sequencing. The sequencing results of C-rich strand are shown in Figure 3.13. Although the SURE® cells appear to provide a better plasmid quality, two major arrest sites (AS1 and AS2, Figure 3.13) by DNA polymerase were observed. Both arrest

sites appear to occur within the region of cloned DNA and after the second arrest site (AS2) the DNA polymerase is unable to continue with chain elongation and the process is halted.

A separate uncharacterised clone **$pG_n,4,X_n$** that was generated from the monomer oligonucleotide **$G_5,4,1$** was sent for commercial sequencing. By sequencing both the G-rich and C-rich strands using forward and reverse universal primers it was anticipated that results would reveal the plasmid sequence and locate any sites at which sequencing had previously paused. Figure 3.14 shows the sequencing results of the G-rich strand using the reverse primer. The sequencing results of the C-rich strand were inconclusive however, the G-rich strand showed a similar arrest pattern to those obtained by manual sequencing.

The sequencing results show that, there are at least four G_3-tracts at the start of the cloning site that are separated by T_4 loops, but it is not possible to read the starting GATC. These are followed by five G_3-tracts also separated by T_4-loops after the following GATC the sequencing then fails. Although the read beyond this point is poor (several weak G_3- and T_4-tracts), the results appear to suggest that there are further inserts.

As the arrest of DNA polymerase was observed by manual and commercial sequencing of both the C-rich and G-rich strands the results suggest that the arrest of DNA polymerase may be the result of either the DNA undergoing self-cleavage or the formation of an interstrand rather than intra secondary structure.

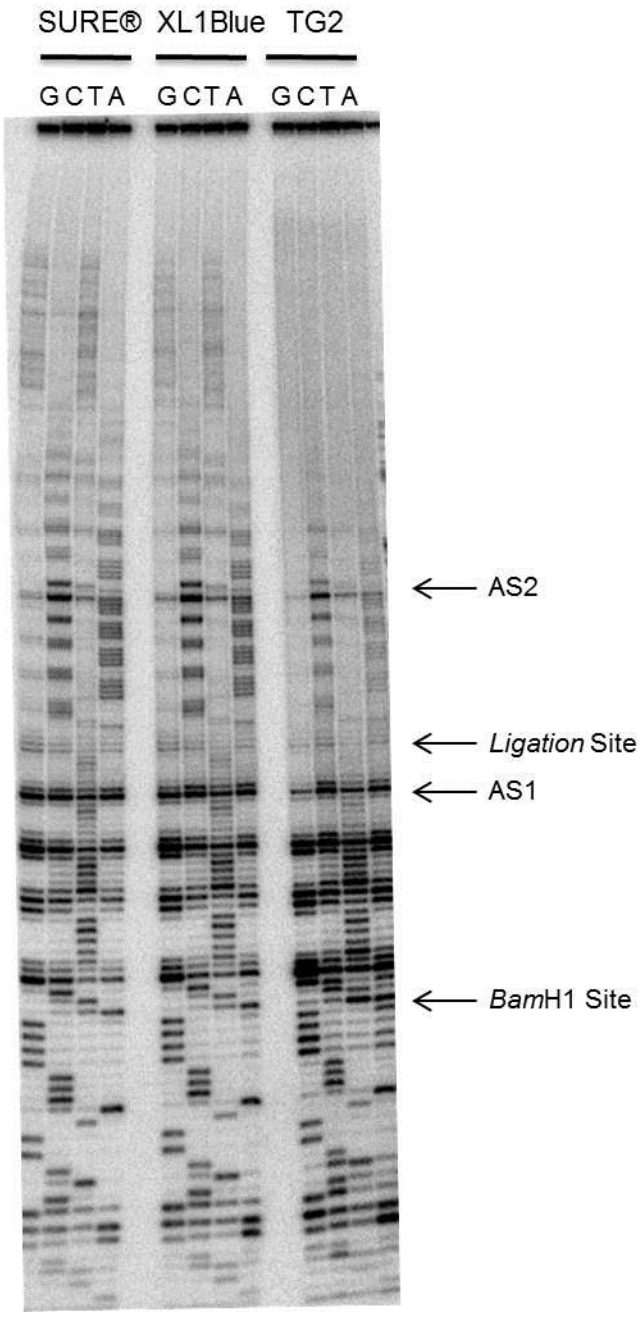


Figure 3.13 Dideoxy sequencing of plasmids containing the sequence **$pG_n,4,X_n$.** This plasmids was transformed and isolated from TG2, XL1-Blue and Sure competent cells as indicated. Two DNA polymerase arrest sites were observed at position AS1 and AS2. The *Bam*H1 and ligation sites are also indicated.

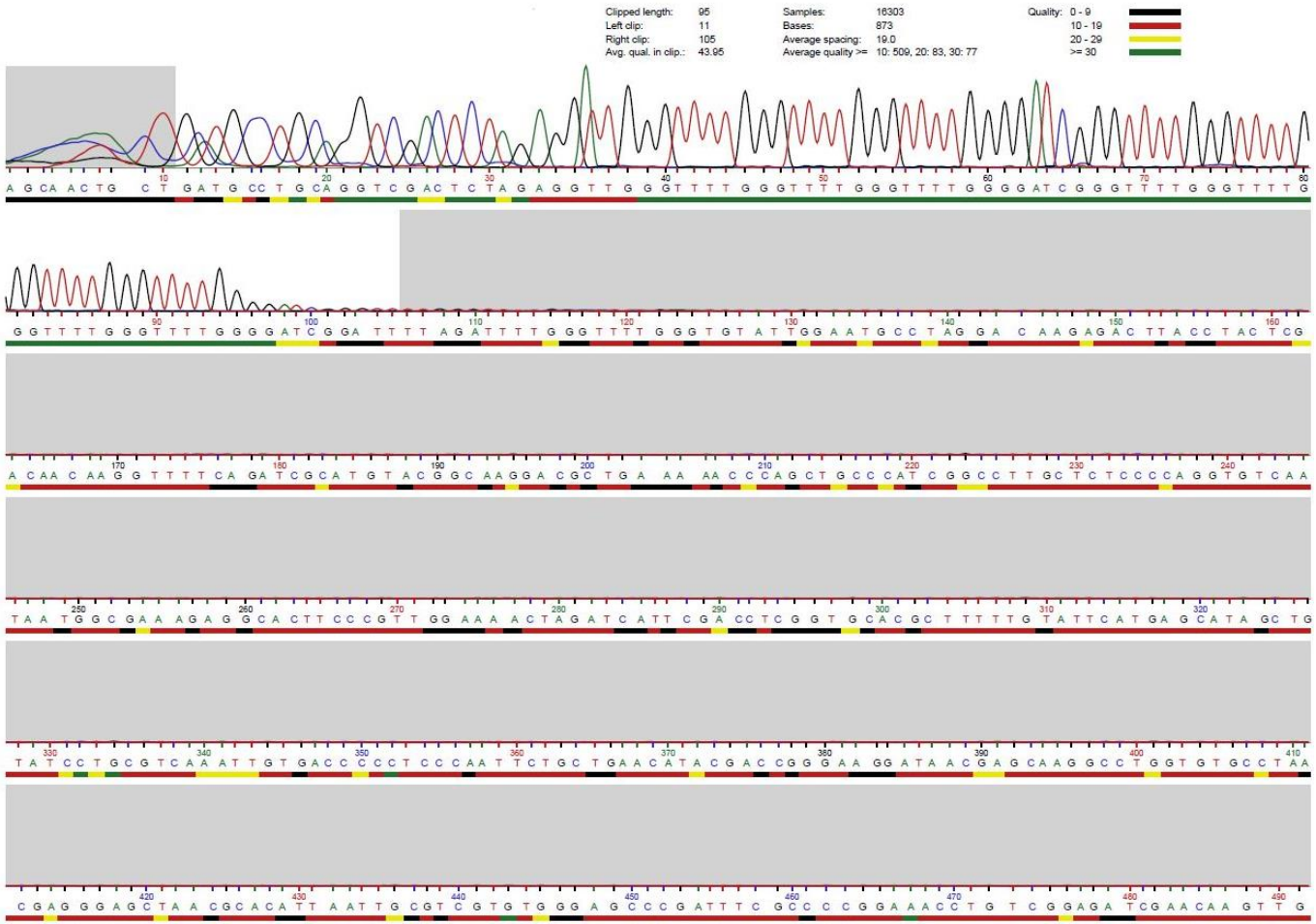


Figure 3.14 Automated DNA sequencing of an extended G-quadruplex forming sequence **p(G_n,4,X_n)** cloned into pUC19 using a universal reverse primer**.** The results show that sequencing failed after the second insert, indicated by a dramatic loss in peak intensity.

3.3.4 Investigations into Whether the Sequencing Reaction Induces G-Quadruplex Self-Cleavage

$MgCl_2$ is a common component in polymerase chain reaction (PCR) and PCR sequencing buffers and it has been found that in the presence of $MgCl_2$/ $MgSO_4$ and histidine certain structural assemblies of artificially designed G-quadruplexes can undergo self-cleavage (Liu et al., 2008, Li et al., 2008, Ng et al., 2009).

It was then speculated that the abrupt change in the sequencing signal, as exhibited in Figure 3.13 and Figure 3.14, may have been a result of a self-cleaving reaction at the G-quadruplexes induced by the sequencing buffers. If this was the case then DNA polymerisation would inevitably stop, as the polymerase would reach the end of a DNA template. To determine whether the difficulties in sequencing **$pG_n,4,X_n$** were due to self-cleavage, some simple tests were performed in which the plasmid DNA was subjected to various conditions. The samples were then examined by agarose gel electrophoresis.

A change in temperature can trigger a denaturing in the DNA in the same way alkaline denaturing does at the start of the sequencing process. Therefore, the plasmid clone **$pG_n,4,X_n$** was initially subjected to heat denaturation/renaturation to facilitate the formation of the G-quadruplex structure. The DNA sample (20 µl, 150 ng/µl) was boiled for 10 min at 100 °C in the presence and absence of potassium then cooled fast by placing immediately on to ice. A second sample was taken and heated in the same way, but this time left to slowly cool to room temperature. These steps were repeated using the native pUC19 plasmid as a control. Both the pUC19 vector and the clone displayed identical banding patterns shown in Figure 3.15. Lanes 3 and 4 display the results after fast cooling and lanes 5 and 6 for slow cooling. If a self-cleavage reaction had taken place within the cloned region of the plasmid, bands differing from those observed by the native pUC19 would have been visible.

In the following step each plasmid was alkaline denatured under the same conditions of the sequencing protocol. The plasmid samples were exposed to NaOH (2M, 10 µl) for 10 minutes followed by ethanol precipitation as described in chapter 2. The plasmid pellets were then re-dissolved in water (2 ml) and Tris-Na (18 µl), heated for 10 minutes at 100 °C and re-annealed either by fast cooling (lanes 7 and 8) in Figure 3.15 or by slow cooling (lanes 9 and 10). The banding pattern for the DNA in lanes 7 and 8 show that there is a small amount of open circular DNA at the top of the gel whilst the two bands lower down the gel that were also seen in lanes 9 and 10 correspond to fully annealed and partially annealed DNA. Again the cloned plasmid showed identical banding patterns to the pUC19 vector indicating that it was unlikely that a self-cleavage reaction had taken place.

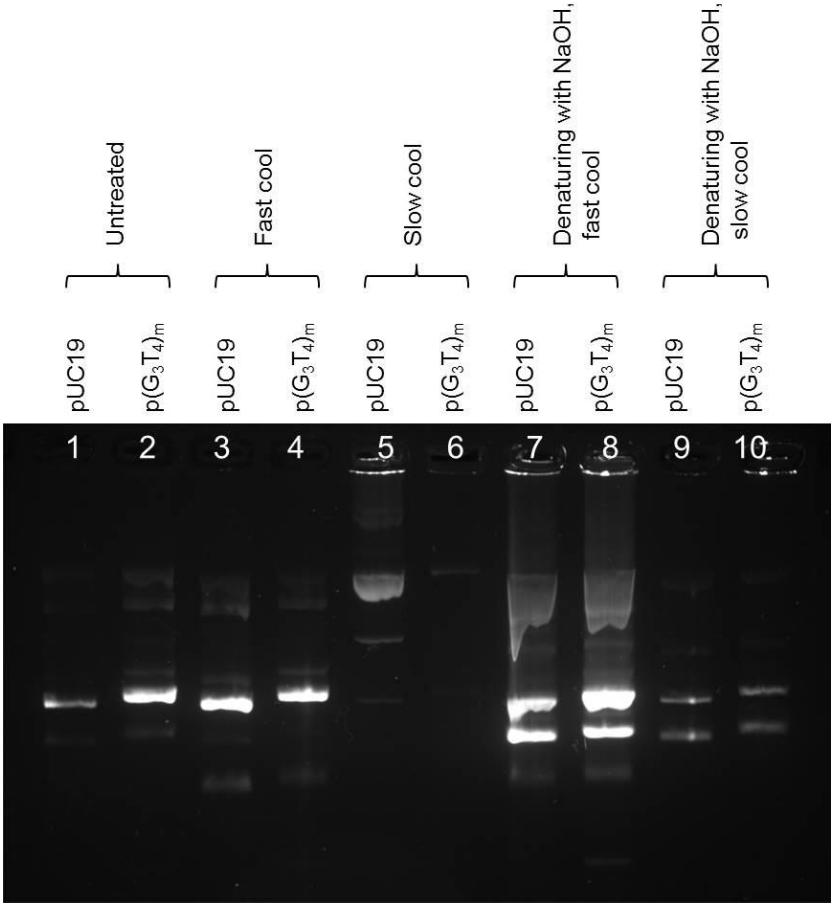


Figure 3.15 Determination of plasmid self-cleavage using agarose gel electrophoresis**.** Plasmids: pUC19 and $p(G_4,4,X_n)$ were used in the self-cleavage investigations. Lanes 1 to 2 : Untreated plasmids; Lanes 3 to 4: plasmids rapidly cooled on ice after boiling to 100 °C for 10 min; Lanes 5 to 6: plasmids slowly cooled on at room temperature after boiling to 100 °C for 10 minutes; Lanes 7 to 8: plasmids denatured with NaOH, boiled to 100 °C for 10 min and rapidly cooled on ice small amount of open circular DNA is seen at the top of the gel whilst the two bands lower down the gel correspond to fully annealed and partially annealed DNA; Lanes 9 to 10: plasmids denatured with NaOH, boiled to 100 °C for 10 min and slowly cooled at room temperature. The two bands lower down the gel correspond to fully annealed and partially annealed DNA.

Eventually, the clone **$pG_n,4,X_n$** was sent for commercial sequencing by 'special chemistry' a protocol that is the intellectual property of Eurofins. The results are shown in Figure 3.16. Although the sequencing results do not show a 'clean' trace as the starting GATC is not evident and the sequencing trace fades towards the 3'-end of the sequence. The sequencing of the DNA revealed a mutant with the following sequence:

$GATC(G_3T_4)_3G_3GATC(G_3T_4)_4G_3GATC(G_3T_4)_4G_3GATCC_3A_4C_3A_4C_3GATC$.

The plasmid contains three G-quadruplex forming sequences; the first sequence contains a four G_3-tract structure, whilst the second and third contain a five G_3-tract structure. The flanking sequence contains three C_3-tracts separated by $d(A_4)$ that has the potential to form a cruciform structure with the terminating G-quadruplex forming sequence.

In a final attempt to achieve plasmids with multiple G-quadruplex forming regions, dimeric oligonucleotides of the monomeric sequences were synthesised (Table 2.1, Chapter 2). The ligation protocol was repeated and the longer fragments were subsequently cloned. The results of the cloning experiment proved again that plasmids containing just one insert of the dimeric sequences were the dominant species. Any plasmids that appeared to have longer chains either had guanine tracts deleted or bases missing within the sequence. As a consequence of this inconsistency, only monomeric and dimeric sequences, along with the mutant plasmid **$pG_n,4,X_n$** were used for the remainder of this research.

DNA clones containing G_3T_8 inserts were also prepared, but as G-quadruplex formation was not detected by the CD analysis and due to the inability for longer loop sequences to form stable G-quadruplex structures, these were not considered in the remainder of this thesis.

Table 3.1 lists a summary of the clones obtained.

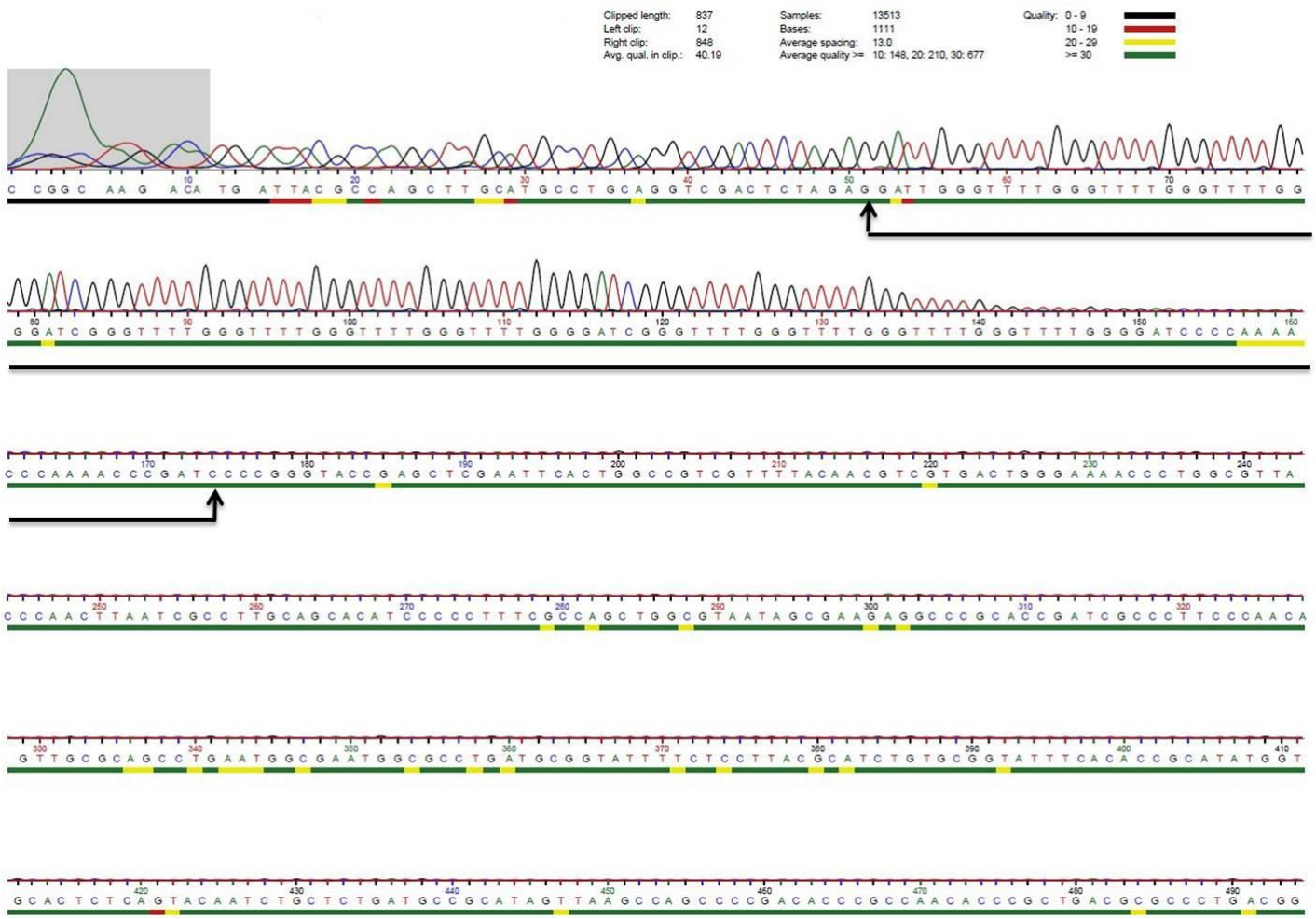


Figure 3.16 Automated DNA sequencing of $p(G_4,4,X_n)$. The black arrows indicate the sequence of the cloned region. $GATC(G_3T_4)_3G_3GATC(G_3T_4)_4G_3GATC(G_3T_4)_4G_3GATCC_3A_4C_3A_4C_3GATC$.

Table 3.1 Summary of Cloned G-quadruplex forming plasmids (see appendices for sequencing).

Name $p(G_n,L,X)$	**Oligonucleotide Sequence**
$pG_4,1,1$	5'-GATCGGGTGGGTGGGTGGG
$pG_5,1,1$	5'-GATCGGGTGGGTGGGTGGGTGGG
$pG_4,1,2$	5'-GATCGGGTGGGTGGGTGGGGATCGGGTGGGTGGGTGGG
$pG_5,1,2$	5'-GATCGGGTGGGTGGGTGGGTGGGGATCGGGTGGGTGGGTGGGTGGG
$pG_4,4,1$	5'-GATCGGGTTTTGGGTTTTGGGTTTTGGG
$pG_5,4,1$	5'-GATCGGGTTTTGGGTTTTGGGTTTTGGGTTTTGGG
$pG_4,4,2$	5'-GATCGGGTTTTGGGTTTTGGGTTTTGGGTTTTGATCGGGTTTTGGGTTTTGGGTTTTGGG
$pG_5,4,2$	5'-GATCGGGTTTTGGGTTTTGGGTTTTGGGTTTTGGGGATCGGGTTTTGGGTTTTGGGTTTTGGGTTTTGGG
$pG_4,8,1$*	5'-GATCGGGTTTTTTTTGGGTTTTTTTTGGGTTTTTTTTGGG
$pG_n,4,X_n$	5'-GATC$(G_3T_4)_3G_3$GATC$(G_3T_4)_4G_3$GATC$(G_3T_4)_4G_3$GATC$C_3A_4C_3A_4C_3$GATC

* This sample was not used in the remainder of the research.

3.4 Discussion

The results presented in this chapter describe the preparation and characterisation of potential G-quadruplex containing plasmids, which were obtained by cloning linear G-rich oligonucleotides into the *Bam*H1 site of pUC19. Apart from the T_8 sequences, the CD spectroscopy of the T- and T_4-loop guanine sequences chosen for this study demonstrated that they readily form G-quadruplexes structures. However the ease at which these structures form led to difficulties in cloning into the pUC19 vector and this was particularly apparent when trying to generate plasmids with multiple inserts.

3.4.1 Circular Dichroism

CD is often used as an empirical technique to determine the quadruplex folding topologies; parallel or antiparallel. Parallel folds display a positive molar ellipticity at approximately 260 nm and a negative at approximately 240 nm, whilst antiparallel folds have a positive signal at approximately 295 nm and a negative at 260 nm (Balagurumoorthy et al., 1992, Giraldo et al., 1994). As a rule sequences that exhibit neither of these spectral patterns are not considered to be G-quadruplex structures.

3.4.1.1 Folding topology of guanine sequences

Since flanking sequences and loop lengths (Smirnov and Shafer, 2000, Hazel et al., 2004, Rachwal et al., 2007b, Risitano and Fox, 2003a) are known to affect the formation of quadruplexes, CD spectroscopy was used to confirm the suitability of the oligonucleotides designed for the cloning experiments. Their ability to form G-quadruplexes and their strand orientation were assessed under ionic conditions. The single thymine loop sequences **G_4,1,1**, **G_5,1,1**, **G_4,1,2** and **G_5,1,2** all exhibited CD spectra that is characteristic of parallel-stranded G-quadruplexes. Even with the 5'-end GATC flanking sequence, each single T-loop complex exhibited a positive molar ellipticity at around 265 nm and a negative molar ellipticity at around 240 nm.

These results are in agreement with previous studies that have demonstrated that short loops containing single nucleotides are topologically restricted and can only form double-chain-reversal linkers because they are unable to span the surface of a G-tetrad

(Hazel et al., 2004). Although cation concentration increased the signal strength, all G_3T sequences revealed the same distinctive peaks even in the absence of potassium, suggesting that all complexes readily adopt the parallel folded topology.

Increasing the loop length from one to four thymines increases conformational flexibility. Rachwal *et al.* (Rachwal et al., 2007b) reported that intramolecular G-quadruplexes that contain two single T-loops, and a centrally or laterally located T_4-loop, produced CD spectra typical of a parallel topology, whilst when all loops contained four thymines the G-quadruplex adopted an antiparallel configuration. As expected in this study the G_3T_4 sequences displayed different CD spectra to the single T-loop sequences, which were highly dependent on the cation concentration. At low KCl concentrations (0 to 1 mM), the CD spectra for the monomeric four and five tract sequences did not exhibit any distinct spectral peaks that were indicative of G-quadruplex formation (Figure 3.4) and appeared to remain as unstructured single strands. On addition of 5 mM KCl two distinct bands were observed, at around 295 nm and 260 nm which, as expected, are typical of an antiparallel topology. The dimeric sequences **G_4,4,2** and **G_5,4,2** also displayed a similar pattern, in which the folding of the G-quadruplex did not occur until 5 mM KCl.

Generally, intramolecular G-quadruplexes are relatively stable under near physiological conditions such as pH and ionic concentration and have a melting temperature above 37 ºC. However, both G_3T_8 sequences did not appear to form any G-quadruplex structures (Figure 3.5) due to the fact that they did not exhibit spectral peaks that are indicative to either the parallel or antiparallel topology. This is because G-quadruplex stability can be affected by the loop length and an increase in loop length leads to a decrease in stability (Rachwal et al., 2007b, Hazel et al., 2004). Also CD analysis of long loop structures is not ideal as long loops (> 7 nt) largely behave as single strands resulting in spectral ambiguity (Hazel et al., 2004).

3.4.1.2 Quadruplex-duplex

The second part of the CD studies investigates whether G-quadruplex formation can compete with duplex formation. By annealing equimolar concentrations of the G-rich

strand with its complementary C-rich strand, CD spectral data was obtained to determine whether under physiological conditions of pH and ionic strength G-quadruplex formation can compete with the duplex. The CD spectra collected for all oligonucleotides that contained a single T-loop exhibited the characteristic parallel G-quadruplex spectral peaks which persisted even in the presence of the complementary C-rich strands (Figure 3.6). This suggests that the four stranded parallel structure is the predominant species, which is consistent with the melting studies data collected by Risitano and Fox (Risitano and Fox, 2003b). In contrast, all sequences containing G_3T_4 and G_3T_8 repeats (Figure 3.7 and Figure 3.8 respectively) showed that the duplex was the dominant species.

3.4.1.3 Cloning and sequencing: Sanger method

The dideoxy chain termination sequencing technique was employed to determine the precise sequence of the DNA fragments inserted into pUC19. The manual and commercial sequencing of the DNA fragments proved successful for all monomeric inserts (Table 3.1). However, it was difficult to generate longer extended sequences. Longer length multimeric inserts were still not observed after the synthetic duplexes were pre-ligated before inserting into the vector. The sequencing continually revealed arrest sites in which DNA polymerase was unable to continue chain elongation after the first four-to-six guanine tracts.

To investigate whether the difficulty in obtaining clones was due to removal of the insert by the bacterial repair processes the ligation mixtures were transformed into SURE® competent cells. SURE® cells were engineered for this type of cloning as they carry mutations that inactivate many of the *E. coli* repair genes that are involved in the rearrangement and deletion of "unwanted" DNA arrangements. However, this made no noticeable difference on the cloning process and again generated plasmids containing monomeric inserts, with only a relatively small number of dimers. On several occasions, manual dideoxy sequencing revealed that some plasmids appeared to contain longer inserts but these were difficult to sequence accurately due to the arresting of DNA polymerase at particular locations, suggesting possible quadruplex formation.

Having established that the choice of competent cells was not affecting the generation of clones other reasons for the difficulty in obtaining G-rich plasmid inserts were examined. The first hypothesis was that the quadruplex-forming sequences were undergoing a self-cleavage mechanism similar to the studies conducted by Liu *et.al* (Liu et al., 2008). In the presence of $MgCl_2$ and histidine they demonstrated that a particular assembly of a G-quadruplex structure was able to perform a self-cleaving action in a site specific manner. This self-cleaving action was later demonstrated on the repeat sequences of the human telomere d(TTAGGG) (Li et al., 2008) and *Oxytricha nova* d(GGGGTTTT) (Ng et al., 2009). By separately boiling both pUC19 and **$pG_n,4,X_n$** for 10 minutes at 100 °C and annealing either rapidly or slowly the results revealed that both plasmids had identical banding patterns, suggesting that self-cleavage had not occurred. When the plasmids were alkaline denatured with NaOH and re-annealed with Tris-Na, again identical results were observed. No additional bands were observed to indicate that a self-cleavage reaction had occurred (Figure 3.15) and therefore cannot account for the stalling that was detected during the sequencing reaction.

For that reason it seems likely that high GC-content of the newly constructed vector may have assembled into a higher order secondary structure, particularly as the CD data revealed that oligonucleotides selected for this research readily fold into the G-quadruplex structure(s) in little or no potassium. The CD results also reveal that the G-quadruplex structure of the single T-loop sequences persists even in the presence of its complementary C-rich strand, which is consistent with the data published by Risitano and Fox (Risitano and Fox, 2003b).

Due to the presence of potassium in the sequencing buffers it was possible that a G-quadruplex structure had been induced and stabilised by the cation. Reports suggest that templates containing such G-rich regions are often poor substrates for PCR reactions due to the formation of stable secondary structures and often result in little or no yield of the expected product (Woodford et al., 1995, Woodford et al., 1994). *In vitro* analysis, conducted by Woodford *et al.*, on the G-rich sequence of the chicken *β*-globin promoter was reported to act as a strong barrier to DNA synthesis by a variety of DNA polymerases (Woodford et al., 1994). They demonstrated that the arrest of DNA

synthesis was K^+-dependent and that this was probably due to the formation of a G-quadruplex structure. They also demonstrated that the use of a K^+-free buffer eliminated any premature chain termination in PCR and PCR sequencing by DNA polymerase (Woodford et al., 1995). Thus determination of the G-rich sequences by the dideoxy method may have proved challenging as this technique requires DNA polymerase to use linear fragments of DNA.

Dimeric oligonucleotides of all monomeric single T and T_4-loop sequences along with a dimer T_8-sequence were later synthesised to increase the probability of inserting longer chains. The pre-ligation experiments were repeated and the annealed fragments cloned into pUC19. The clones from this experiment also did not generate any longer plasmid inserts. Eventually the unusual mutant clone **$pG_n,4,X_n$** was sequenced commercially using special chemistry as shown in Figure 3.16 and this sample was also used in the subsequent experiments.

3.5 Conclusion

The cloning experiments in this chapter successfully generated nine G-rich plasmids that could potentially form G-quadruplex structures. These plasmids are summarised in Table 3.1. These plasmid constructs (unless indicated otherwise) were used in the studies reported in the remainder of this thesis, using a range of techniques to probe their ability to form G-quadruplex structures under negative supercoiling.

Chapter 4:
Chemical and Enzymatic Probing of G-Quadruplex Formation within Supercoiled DNA

4.1 Introduction

It is widely accepted that the majority of DNA present in cells is in the Watson-Crick right-handed helical B-form conformation, and in most biological systems, B-DNA exists in a negatively supercoiled state. Supercoiled DNA is in a higher energy state relative to the relaxed or linear form. Supercoiling results from unwinding of the double helix to an energetically less favourable state, therefore any process that lessens the degree of negative supercoiling is energetically favourable (Wang et al., 1983, Nordheim et al., 1983, Panayotatos and Wells, 1981). Reductions in supercoiling can be generated by DNA intercalators or structural changes triggered by local strand denaturation of the double helix and the formation of alternative DNA conformations within particular sequences such as the left-handed Z-DNA (Nordheim et al., 1983) or cruciforms (Panayotatos and Wells, 1981). Known poly(G_3T_n) G-quadruplex forming sequences were cloned into the pUC19 vector as reported in Chapters 2 and 3, and detection of possible structural transitions from B-DNA to G-quadruplex structures within these G-rich sequences under negative supercoiling was investigated using chemical and enzymatic probes and is reported in this chapter.

4.1.1 Investigating G-Quadruplex Structures in Supercoiled DNA using Chemical and Enzyme Probes

The use of chemicals and enzymes is extremely important in the study of DNA structures. Localised structural changes in DNA can be identified and characterised by either chemical modifications or enzyme cleavage (Lilley, 1980, Panayotatos and Wells, 1981, Lilley, 1992). The use of enzyme and chemical probing techniques have provided remarkable insight into alterations of DNA assemblies (Lilley, 1992), such as the

transformation from B-DNA to structures such as cruciforms (Lilley, 1983, Lilley and Palecek, 1984), left handed Z-DNA (McLean et al., 1988, McLean and Wells, 1988), and G-quadruplexes (Sen and Gilbert, 1988, Sundquist and Klug, 1989). Whilst enzyme nuclease probing can be useful in mapping structural perturbations they may cause problems because of the release of topological constraint with the first cleavage (Lilley, 1992). Chemical probes are more effective and have many advantages over enzyme probes; i) they can be used over a wide range of conditions such as; pH, temperature, and varying ionic strengths, ii) they are molecularly smaller and can diffuse into various parts of the DNA, iii) their structures and reaction mechanisms are known and are less likely to induce secondary structures with large protein molecules, iv) they do not induce scissions within the DNA chain, thus allowing the chemical to react simultaneously at several sites of the DNA molecule, v) their reaction sites can be identified under various techniques and vi) they have the potential to probe DNA *in vivo* (Palecek, 1991).

The use of chemical modification assays and enzymatic cleavage of specific DNA sequences has demonstrated that alternative DNA structures such as Z-DNA (Peck and Wang, 1983, Nordheim et al., 1983) and cruciforms (Panayotatos and Wells, 1981) are stabilised by negative supercoiling and that the formation of these structures involves the local unwinding of DNA. In this chapter, the use of both chemical (dimethyl sulphate and potassium permanganate) and enzyme (S1 nuclease) probing techniques were employed to investigate whether negative supercoiling can facilitate the formation of G-quadruplex structures in G-rich DNA supercoiled DNA.

4.2 Experimental Theory and Design

Guanine-rich supercoiled plasmids were subjected to investigations to determine whether G-quadruplex structures can be induced as a result of negative supercoiling. This chapter is divided into two sections. The first investigates the use of chemicals dimethyl sulphate (DMS) and potassium permanganate ($KMnO_4$) to identify any DNA bases that may be involved in an alternative secondary structure, whilst S1 nuclease was used to map the single stranded regions resulting from any structural transitions. These

two techniques combine to provide complementary evidence for the formation of G-quadruplexes.

4.2.1 Chemical Probing: The use of Dimethyl Sulphate and Potassium Permanganate as Chemical Probes to Investigate the Formation of G-Quadruplexes

The assembly of guanines to form a quartet structure is achieved by Hoogsteen-hydrogen bonds that involve the N7 of deoxyguanosine. These quartet units are then stacked and held together by individual strands that are linked by single-stranded loop regions. If a particular guanine base is involved in a G-quartet or if a nucleotide (in this case thymine) is participating in the loop region, its reactivity to various chemical agents will be altered relative to duplex- or single-stranded DNA. These changes in reactivity can pinpoint the exact site and nature of any structural transitions. The studies in this chapter describe chemical footprinting using DMS and $KMnO_4$ to identify G-quadruplex formation in supercoiled DNA.

4.2.1.1 Dimethyl sulphate: methylation protection assay

Dimethyl sulphate (DMS) methylates the N7 of guanosine or the N3 of cytidine both (Maxam and Gilbert, 1977, Mattes et al., 1986) of which are accessible in single-stranded DNA and the N7 of guanosine in Watson-Crick duplex DNA. In the G-quartet array, DMS is unable to attack the N7 of guanine when it is hydrogen-bonded with the exocylic amino group of a neighbouring guanine. However, in duplex DNA methylation of the N7 guanine can occur and causes the glycosidic bond to become unstable when heated. By heating the modified DNA, an abasic site is created at the modified guanine, which results in a strand break at the phosphodiester backbone when treated with a mild alkali (Figure 4.1). These strand breaks can then be identified by gel electrophoresis (Mattes et al., 1986). This methylation protection assay was used to identify those guanines that are participating in the Hoogsteen-hydrogen bonding of the G-quadruplex structure.

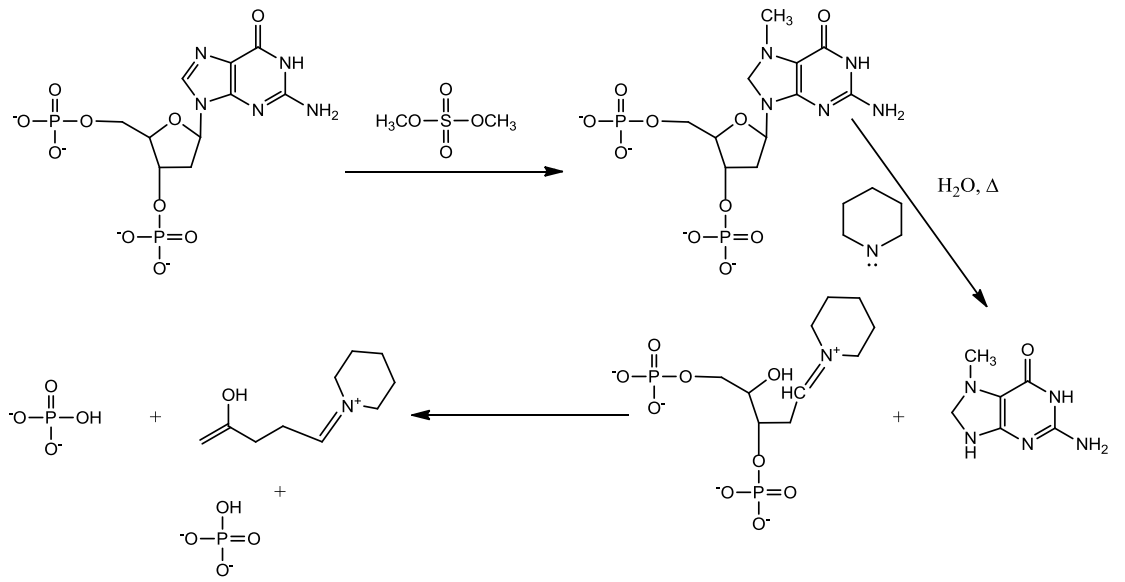


Figure 4.1 Schematic reaction of the proposed Maxam and Gilbert guanine base modification. Guanine base methylation using DMS followed by cleavage (piperidine) from the phosphodiester backbone of a guanine (Mattes et al., 1986).

4.2.1.2 Potassium permanganate: protection assay

Potassium permanganate ($KMnO_4$) is one of the most effective and versatile chemicals used in DNA probing (Bui et al., 2003, Spicuglia et al., 2004). It is a bulky reagent and has been previously used in genomic footprinting assays to map unusual gene structures within DNA (Sasse-Dwight and Gralla, 1989, Zhang and Gralla, 1989, Spicuglia et al., 2004). There are many advantages in using permanganate as a chemical probe, for example; i) it is relatively safe, ii) it is soluble in aqueous solvent, and iii) under mild conditions it is relatively inert to duplex DNA, but highly reactive towards denatured DNA. It oxidises pyrimidine bases, primarily thymines, with a preference for single stranded regions (Spicuglia et al., 2004). $KMnO_4$ reacts with thymine more readily than guanine, cytosine and adenine, by oxidising the double bond to a vicinal diol (Figure 4.2). These double bonds are more exposed in thymines in the loops of quadruplex DNA than in the duplex DNA. The reaction mechanism occurs via an out-of-plane attack on the C5-C6 double bond of the thymine ring. The ring remains intact, although there is a loss of aromaticity due to the insertion of the hydroxyl groups on the 5 and 6 carbons. Subsequent treatment of the vicinal diol with a strong base results in ring opening and cleavage of the phosphodiester backbone (Kahl and Paule, 2001). $KMnO_4$ was used to probe the proposed single-stranded loop regions of the G-quadruplex structure and the regions of melted DNA that flank the structure (Figure 4.2).

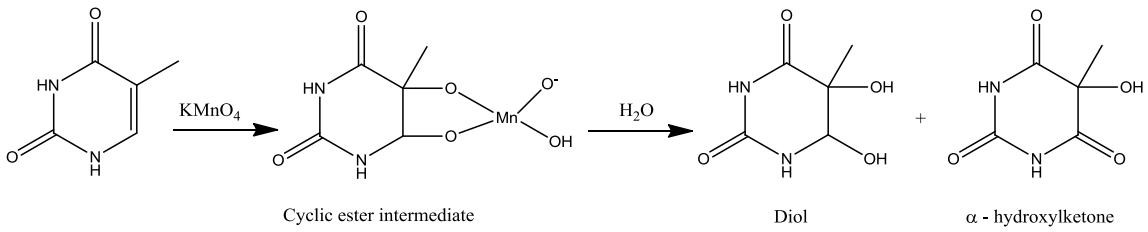


Figure 4.2 A Schematic reaction of the oxidation of thymine by potassium permanganate. (Bui et al., 2003)

4.2.1.3 DNA strand breakage by piperidine

The Maxam-Gilbert chemical method of DNA sequencing (Maxam and Gilbert, 1977) is commonly used to study sequence specific DNA damage. Aqueous piperidine provides a basic medium to promote strand cleavage at these sites (Mattes et al., 1986). For example the Maxam-Gilbert proposed mechanism for strand breakage on guanine bases after methylation with DMS has three crucial steps: (1) the alkaline conditions cause rupture of the bond between C8 and N9, resulting in the formamido-pyrimidine structure; (2) the formamido-pyrimidine structure is displaced from the sugar by piperidine and (3) piperidine then catalyses the β-elimination of phosphates from the sugar, which results in breaking the phosphodiester backbone. Figure 4.1 depicts the reaction mechanism for piperidine strand breakage at the N7-alkylguanine sites as proposed by Maxam and Gilbert (Mattes et al., 1986, Maxam and Gilbert, 1977).

4.2.2 S1 Nuclease

S1 nuclease is a single strand specific endonuclease isolated from *Aspergillus oryzae* which can be used to explore localised DNA structural changes (Lilley, 1980, Sunter et al., 1985). The main advantage in exploiting S1 nuclease to explore local DNA structures is that whilst linear and relaxed plasmids are not susceptible to the S1 digestion, supercoiled DNA may expose regions of single-stranded DNA which are ideal for S1 cleavage. Another advantage is that this enzyme is virtually sequence-neutral, whilst others such as restriction enzymes are sequence-dependent (Lilley, 1992). S1 nuclease was employed in this research to map the location of G-quadruplex

formation under negative supercoiling, by cleaving the single-stranded DNA that are formed in the quadruplex-loops and the melted regions of DNA at the duplex-quadruplex interface, depicted in Figure 4.3a and b.

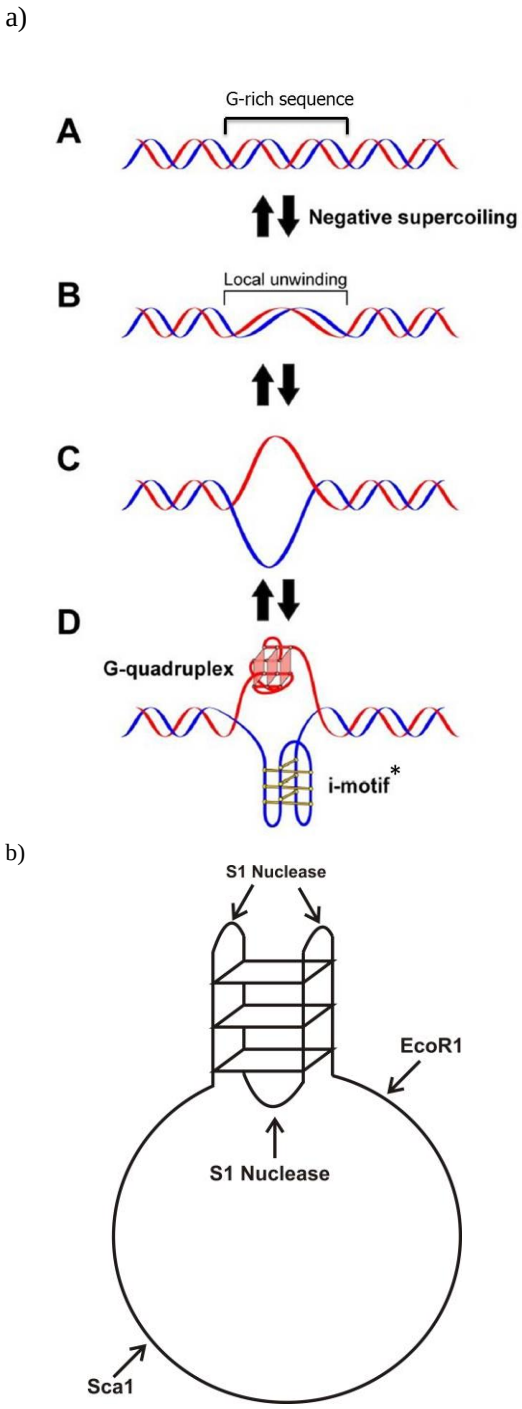


Figure 4.3 The proposed model for the formation of a G-quadruplex structure induced by negative supercoiling within a duplex plasmid. Modified from (Kendrick and Hurley, 2010) (A) Negative supercoiling within duplex, (B) induced local unwinding, (C) melting of the duplex strand to single-strand (D) Formation of the G-quadruplex and *i*-motif structures. b) Diagram depicting the potential areas of S1 sensitive cleavage sites of the G-rich strand of the supercoiled DNA.
* *i*-motif not expected to form under physiological conditions.

4.3 Results

In this study the formation of G-quadruplexes under negatively supercoiled conditions in cloned G-rich DNA was investigated by chemical probing using DMS and $KMnO_4$ modification assays and structural mapping using S1 nuclease cleavage. Each G-rich clone contained the sequence G_n,L,X, where: n is the number of consecutive G_3-tracts, L is the number of thymines in the loop and X is the number of repeats of the G-quadruplex forming sequence e.g. G_5,4,2 = $[GATC(G_3T_4)_4G_3]_2$.

In vitro footprinting analysis using DMS and $KMnO_4$ was employed as described in Chapter 2 on previously prepared G-quadruplex forming DNA (Chapters 2 and 3) to identify specific nucleotides that were participating in the formation of the G-quadruplex structure. In each of the chemical modification assays the sites of modification and the extent of modification were compared between contrasting reference conditions. The DNA concentration for each lane on a single gel was not identical. However, quantification methods where possible were used to make direct comparisons. All modification experiments were conducted at room temperature and at pH 7.4.

S1 nuclease digests were conducted via the protocol described in Chapter 2.

To determine whether supercoiling plays a vital role in the extrusion of G-quadruplex structures, a comparison between linear versus natively supercoiled G-rich DNA was made for each experiment. Where applicable some experiments were conducted in the presence and absence of 100 mM KCl.

4.3.1 *In Vitro* Chemical Footprinting using DMS

Methylation of the N7 guanine was conducted to elucidate any Hoogsteen-hydrogen bonding that may be associated with G-quadruplex formation. DNA samples treated in

the absence and presence of DMS are indicated with a (+) and (-) respectively in the figures that follow.

4.3.1.1 G-quadruplex formation with single T-loop Sequences: four d(GGG) tracts - monomeric and dimeric sequences

Figure 4.4a and b illustrate the results of DMS *in vitro* footprinting on the cloned DNA sample **pG$_4$,1,1** under supercoiled and linear conditions in the presence (Figure 4.4a) and absence of KCl (Figure 4.4b). Figure 4.4c shows the densitometric scans of the DMS modified lanes. In the presence of 100 mM KCl, the results reveal that all four G_3-tracts display considerably reduced reactivity by DMS, compared with other guanines in the rest of the plasmid. These observations indicate the possible presence of a G-quadruplex structure. The guanine bases located within the 5' and 3' flanking regions showed the highest degree of reactivity. Interestingly, G_3-tracts at the 5'- (upper) end of the insert appear to be more protected than those at the 3'-end. These variations in intensity can be more readily seen in the densitometric scans shown in Figure 4.4c. The G_3-tracts in position 1 and 2 were modified approximately 80% less than the guanines in the immediate flanking regions. Tract 3 showed an increase in modification from the first (5') to the third (3') guanine ranging from 20-50%, whilst tract 4 showed reactivity of approximately 50-60%. A similar modification pattern was observed in the linear sample and for the experiments conducted in absence of KCl (Figure 4.4b). This data provides suitable evidence to suggest the presence of a G-quadruplex in this G_3T insert, but surprisingly this did not depend on either negative supercoiling or ionic strength.

Figure 4.5a shows the DMS methylation protection assay on the dimeric sequence **pG$_4$,1,2**. In the presence of 100 mM KCl, a similar pattern to that of its monomeric structural counterpart was observed. Within each of the four G_3-tracts the extent of DMS modification is reduced in both the supercoiled and linearised plasmids. Again the two G_3-tracts at the 5'- (upper) end (tracts 1 and 2 of the first structure and tracts 5 and 6 of the second) showed the greatest degree of protection whilst the tracts at the 3'-end (tracts 3 and 4 of the first structure and tracts 7 and 8 of the second) displayed a slightly increased reactivity to the probe. Figure 4.5c shows the densitometric scans of each of the modified lanes. The guanines within the 3'- and 5'-GATC flanking sequences of the proposed G-quadruplex forming sequence showed the most reactivity towards DMS.

Figure 4.5b shows that in the absence of the potassium cation, similar banding patterns were observed albeit not as pronounced as those in the presence of 100 mM KCl. The results obtained for the dimeric sequence appear to indicate the presence of two G-quadruplex structures.

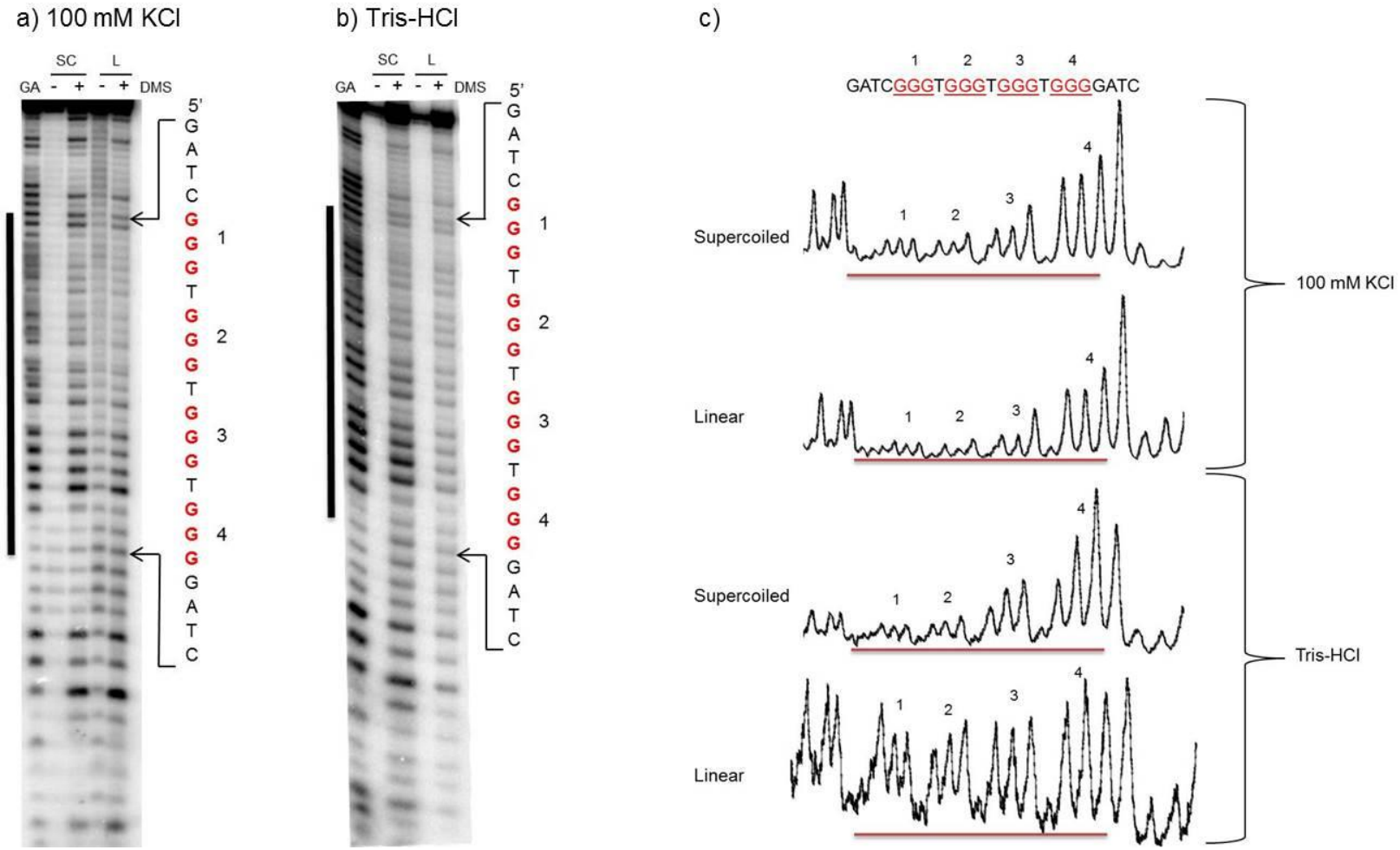


Figure 4.4 *In vitro* footprinting of plasmid **pG$_4$,1,1** with DMS. Both Supercoiled (SC) and linearised (L) plasmids were modified with DMS. a) In the presence of 100 mM KCl. b) In the absence of KCl. Treatment without and with DMS are indicated by (-) and (+) respectively. The plasmids were incubated in buffer overnight before adding DMS and the reactions were conducted at room temperature. c) Densitometric scans of the DMS modified lanes, the guanine tracts numbered 1-4 are indicated by the red horizontal bars on the scan. The G-quadruplex-forming guanines were normalised to a hyperreactive guanine outside the quadruplex-forming region.

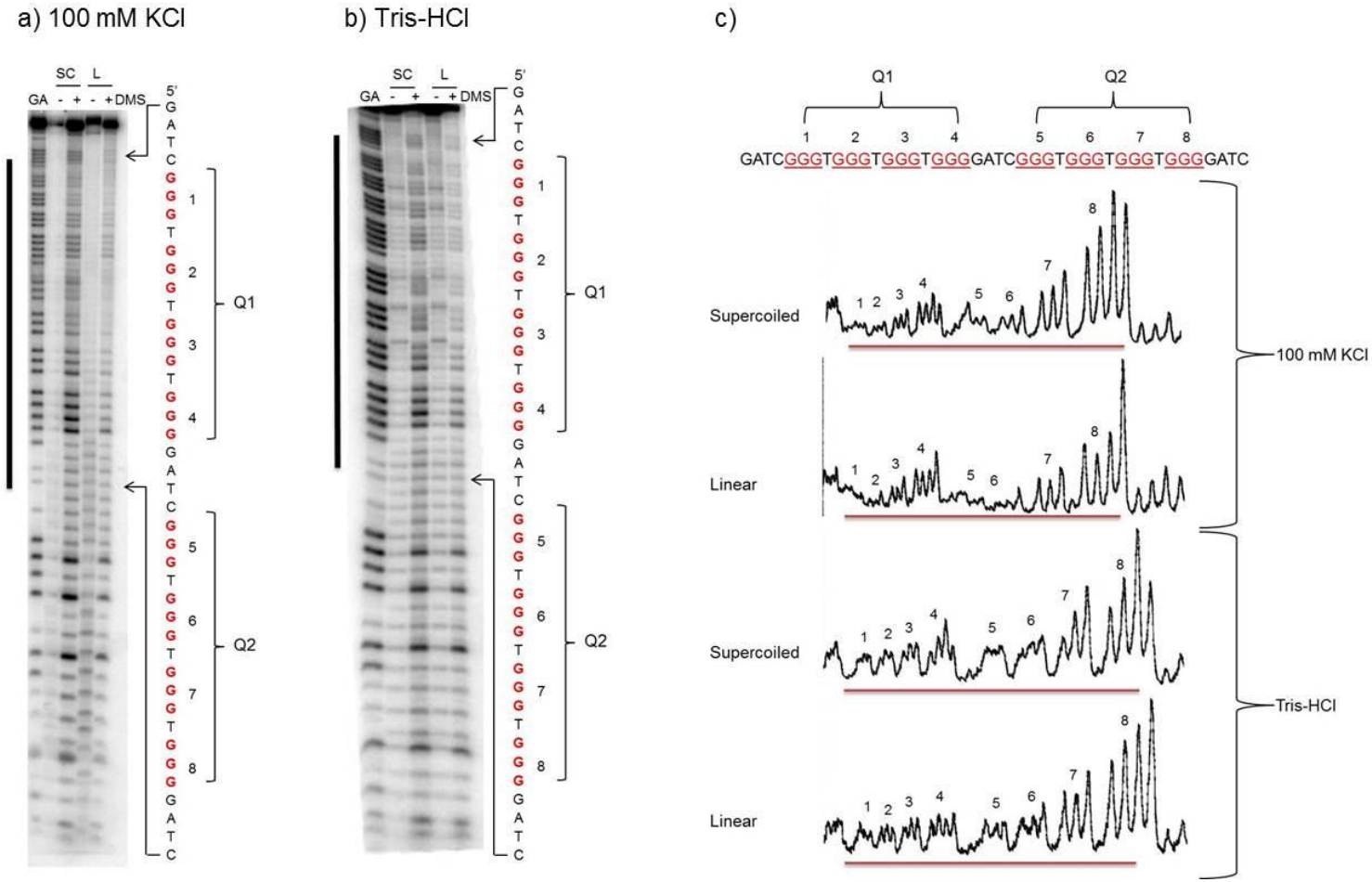


Figure 4.5 *In vitro* footprinting of plasmid **pG$_4$1,2** with DMS. Both Supercoiled (SC) and linearised (L) plasmids were modified with DMS. Potential quadruplex-forming regions are indicated by Q1 and Q2. a) In the presence of 100 mM KCl. b) In the absence of KCl. Treatment without and with DMS are indicated by (-) and (+) respectively. The plasmids were incubated in buffer overnight before adding DMS and the reactions were conducted at room temperature. c) Densitometric scans of the DMS modified lanes, the guanine tracts numbered 1-8 are indicated by the red horizontal bars on the scan. The G-quadruplex-forming guanines were normalised to a hyperreactive guanine outside the quadruplex-forming region.

4.3.1.2 G-quadruplex formation with single T-loop Sequences: five d(GGG) tracts - monomeric and dimeric sequences

DMS modification patterns in the absence and presence of 100 mM KCl for the plasmid containing five consecutive G_3-tracts **pG$_5$,1,1** in both the supercoiled and linear states are shown in Figure 4.6a and b respectively. Figure 4.6c shows the densitometric scans of these DMS reactions. In the presence of 100 mM KCl the methylation patterns for the supercoiled plasmid show that guanines in the quadruplex forming region are protected from modification relative to the guanines in the immediate flanking sequence. On closer inspection of the methylation patterns it can be seen that tracts 2 to 4 show a higher degree of protection than tracts 1 and 5, with only 20-25% of the central three guanine tracts being modified compared to 35-45 % for the external tracts. A similar pattern is also evident in the linearised sample. A sequence containing five tandem G_3-tracts can adopt several isomeric loop topologies, although it is most likely that two distinct quadruplex conformations, with either a 5'-GATCGGGT overhang or a 3'-TGGGATC overhang (*i.e.* excluding the first or fifth G_3-tract from the quadruplex) may exist in equilibrium. The methylation pattern suggests that the DNA sample contains a mixture of both structures. When the plasmid was linearised a similar modification pattern was observed. The five G_3-tracts were significantly protected from DMS modification, compared to guanines in the remainder of the fragment, with the external 5' and 3' G_3-tracts showing distinctly more reactivity than the central three tracts. In the absence of KCl a similar protection pattern to the experiments conducted in the presence of KCl was observed (Figure 4.6.) However, the exclusion of KCl did not appear to significantly enhance DMS reactivity towards the five G_3-tracts and only a 10% increase in methylation was observed, suggesting that the five tract sequence is only slightly more ionic-dependent than the four tract sequence.

The DMS modification patterns and densitometric scans for the dimeric G-quadruplex forming sequence **pG$_5$,1,2** are shown in Figure 4.7. As expected the cleavage pattern is similar to the monomeric sample (**pG$_5$,1,1**) and the protection of guanine modification was the same. In both the supercoiled and linear samples the central three G_3-tracts (tracts 2 to 4 and tracts 7 to 9), showed a higher degree of protection than the two external G_3-tracts (quadruplex 1: tracts 1 and 5, quadruplex 2: tracts 6 and 10) which

show weaker modification. The same structural polymorphism exhibited by the monomeric sequence is also exhibited by the dimeric DNA. In the absence of KCl no distinct changes were observed in the DMS modification pattern and the G_3-tracts do not appear to be protected compared to other guanines in the sequence.

In all of the above experiments none of the reactions in the absence of DMS revealed any unusual modification patterns.

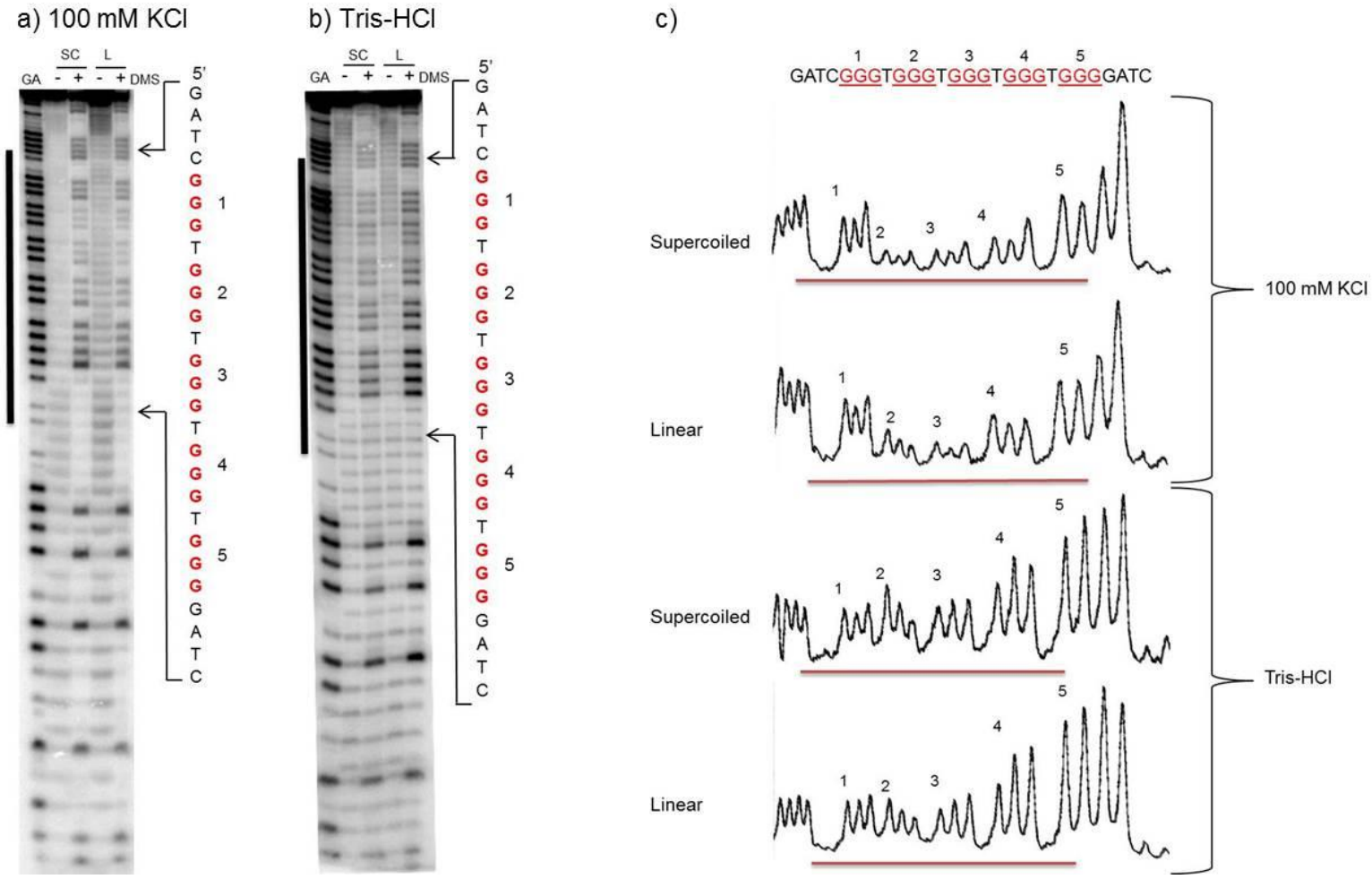


Figure 4.6 *In vitro* footprinting of plasmid **pG$_5$,1,1** with DMS. Both Supercoiled (SC) and linearised (L) plasmids were modified with DMS. a) In the presence of 100 mM KCl. b) In the absence of KCl. Treatment without and with DMS are indicated by (-) and (+) respectively. The plasmids were incubated in buffer overnight before the reactions were conducted at room temperature c) Densitometric scans of the DMS modified lanes, the guanine tracts numbered 1-5 are indicated by the red horizontal bars on the scan. The G-quadruplex-forming guanines were normalised to a hyperreactive guanine outside the quadruplex-forming region.

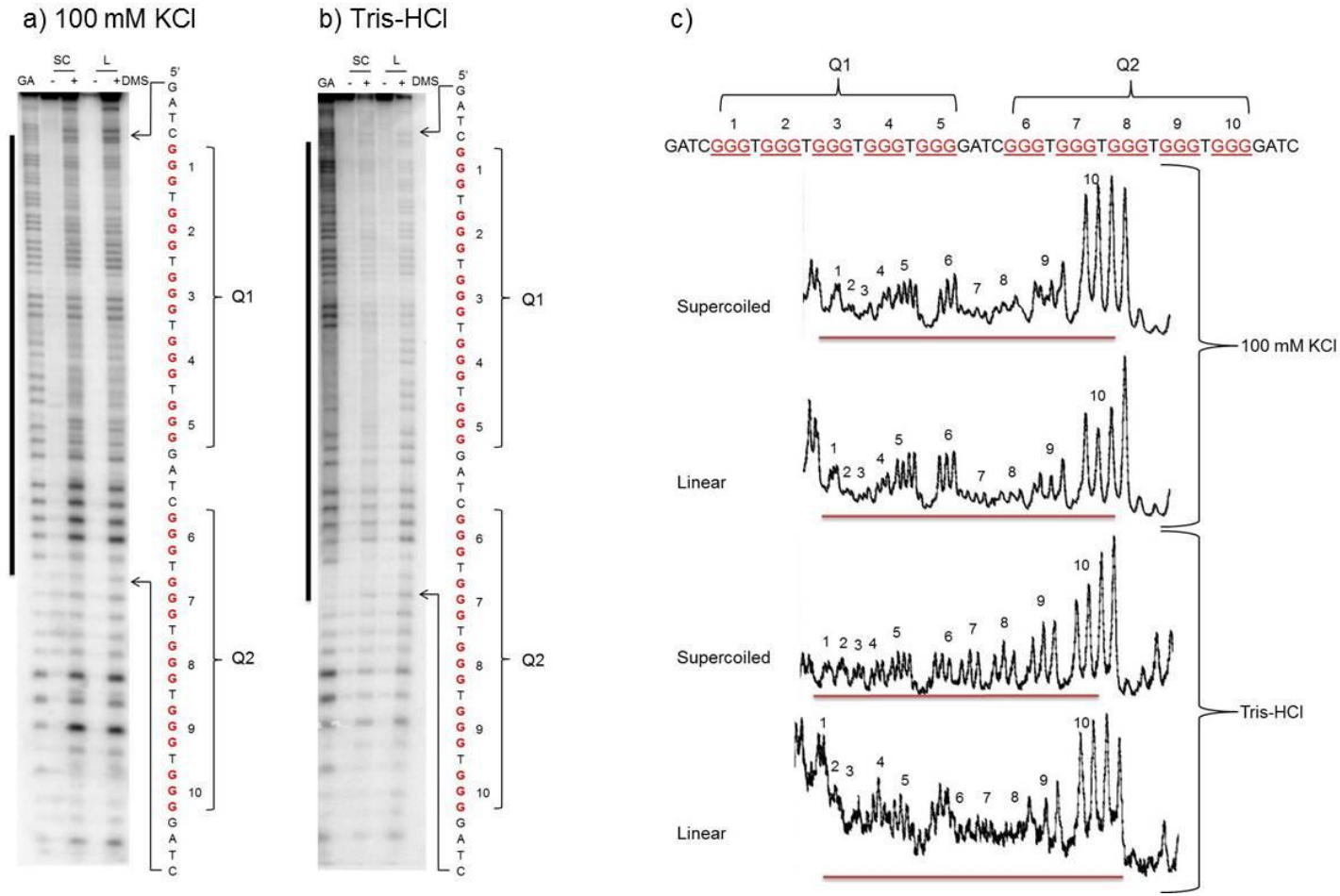


Figure 4.7 *In vitro* footprinting of plasmid **pG$_5$,1,2** with DMS. Both Supercoiled (SC) and linearised (L) plasmids were modified with DMS. Quadruplex structure 1 and 2 are indicated by Q1 and Q2 respectively. a) In the presence of 100 mM KCl. b) In the absence of KCl. Treatment without and with DMS are indicated by (-) and (+) respectively. The plasmids were incubated in buffer overnight before the reactions were conducted at room temperature. c) Densitometric scans of the DMS modified lanes, the guanine tracts numbered 1-10 are indicated by the red horizontal bars on the scan. The G-quadruplex-forming guanines were normalised to a hyperreactive guanine outside the quadruplex-forming region.

4.3.1.3 G-quadruplex formation with T_4-loop sequences: four d(GGG) tracts - monomeric and dimeric sequences

Figure 4.8a and b show the results of DMS modification of plasmid **pG$_4$,4,1,** which contains four G_3-tracts separated by T_4-loops in the presence and absence KCl, and Figure 4.8c shows densitometric scans of these data. In both the supercoiled and linear reactions every guanine band in the G-quadruplex forming sequence has the same intensity as those in the surrounding sequences. The same result is seen in the absence of KCl. These methylation patterns imply that N7 of each guanine residue within this sequence is equally accessible to DMS methylation and similar to that in duplex DNA, therefore suggesting that this sequence does not form a G-quadruplex structure under negative supercoiling.

The dimeric sample of this sequence **pG$_5$,4,2** also showed a similar cleavage pattern in both the supercoiled and linear states (Figure 4.9). Once again every guanine residue produces a band of equal intensity. In both the monomeric and dimeric sequences, the N7 position of the guanines shows the same reactivity as those in duplex DNA, suggesting that none of these guanine bases are participating in G-quadruplex structures.

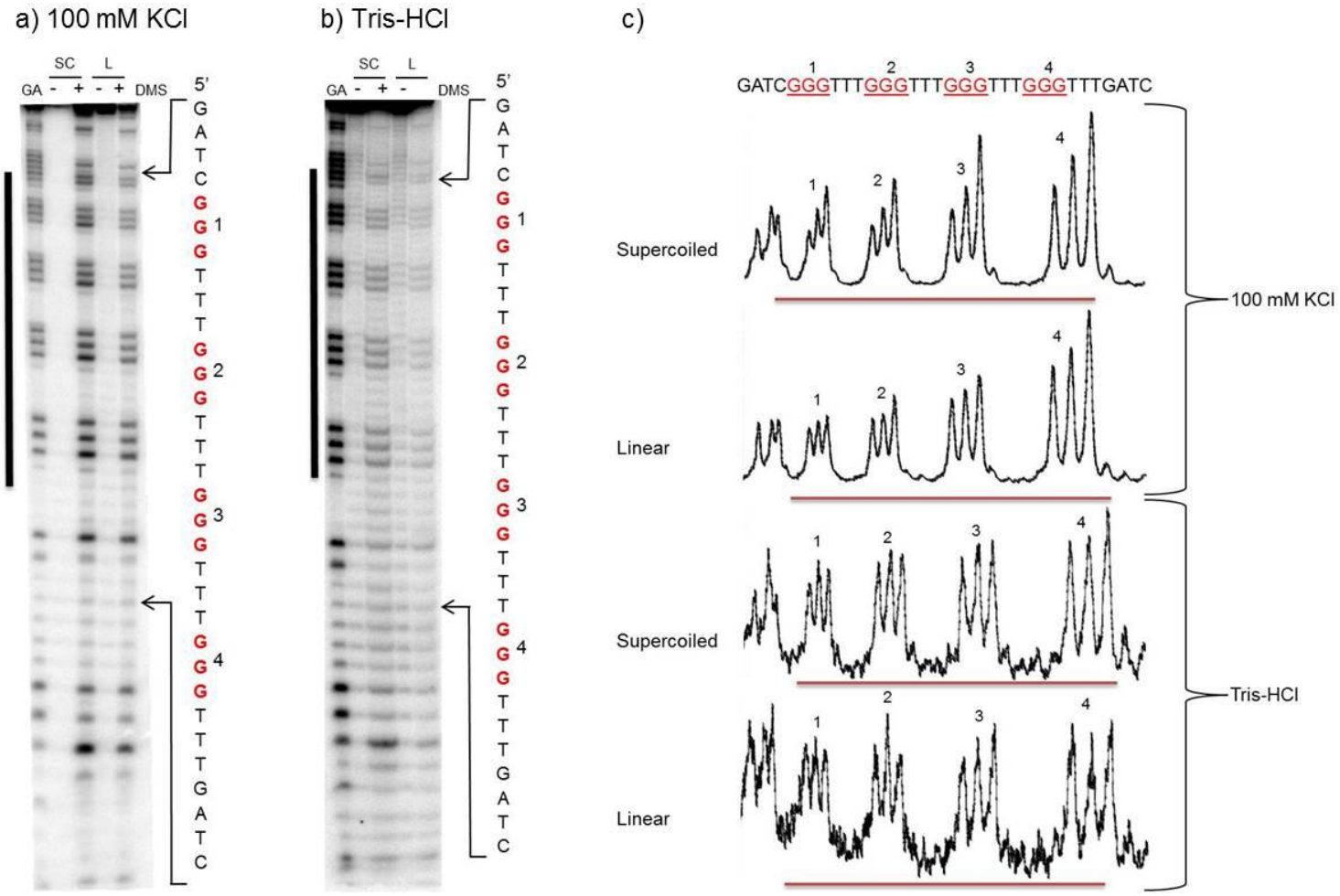


Figure 4.8 *In vitro* footprinting of the plasmid **pG$_4$,4,1** with DMS. Both Supercoiled (SC) and linearised (L) plasmids were modified with DMS. a) In the presence of 100 mM KCl. b) In the absence of KCl. Treatment without and with DMS are indicated by (-) and (+) respectively. The plasmids were incubated in buffer overnight before the reactions with DMS, which were conducted at room temperature c) Densitometric scans of the DMS modified lanes, the guanine tracts numbered 1-4 are indicated by the red horizontal bars on the scan. The G-quadruplex-forming guanines were normalised to a hyperreactive guanine outside the quadruplex-forming region.

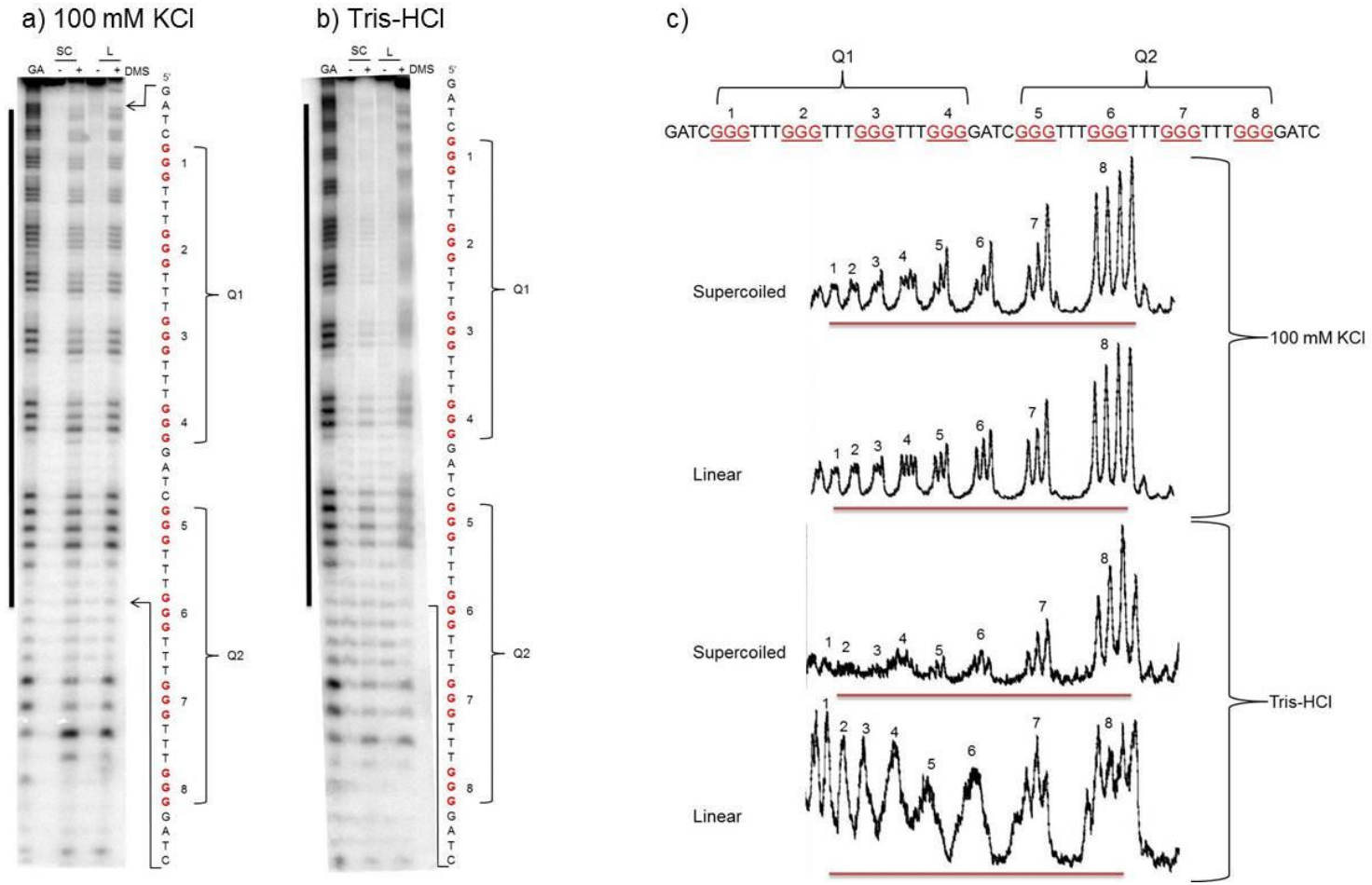


Figure 4.9 *In vitro* footprinting of the plasmid **pG$_4$,4,2** with DMS. Both Supercoiled (SC) and linearised (L) plasmids were modified with DMS. Quadruplex structures 1 and 2 are indicated by Q1 and Q2 respectively. a) In the presence of 100 mM KCl. b) In the absence of KCl. Treatment without and with DMS are indicated by (-) and (+) respectively. The plasmids were incubated overnight in buffer before reaction with DMS, which was conducted at room temperature. c) Densitometric scans of the DMS modified lanes, the guanine tracts numbered 1-8 are indicated by the red horizontal bars on the scan. The G-quadruplex-forming guanines were normalised to a hyperreactive guanine outside the quadruplex-forming region.

4.3.1.4 G-quadruplex formation with T_4-loop Sequences: five d(GGG) tracts - monomeric and dimeric sequences

The results of DMS modification of plasmid **pG$_4$,4,1,** which contains five d(GGG) tracts separated by T_4-loops in the presence and absence KCl, are shown in Figure 4.10a and b. Figure 4.10c shows the densitometric scans of the data for both the supercoiled and linear plasmids. In the presence of KCl every guanine base within the G-quadruplex forming sequence showed the same reactivity as guanines in the remainder of the quadruplex-forming sequence. This was not only observed in the linear sample but also in the absence of KCl. The reactivity of the guanines within the potential quadruplex-forming regions, suggests that it is extremely unlikely that any Hoogsteen bonding had occurred. Therefore these results do not present any evidence for extrusion of the G-quadruplex structure.

As expected the dimeric sequence **pG$_5$,4,2** in both the supercoiled and linear states showed a similar cleavage pattern (Figure 4.11) to that of the monomeric sequence with every guanine undergoing methylation by DMS. Again, this implies that in both the monomeric and dimeric sequences, the N7 position of the guanines in the G_3-tracts is as accessible as guanines in the remainder of the fragment, suggesting that none of the guanine residues are participating in the formation of G-quadruplex structures under negative supercoiling.

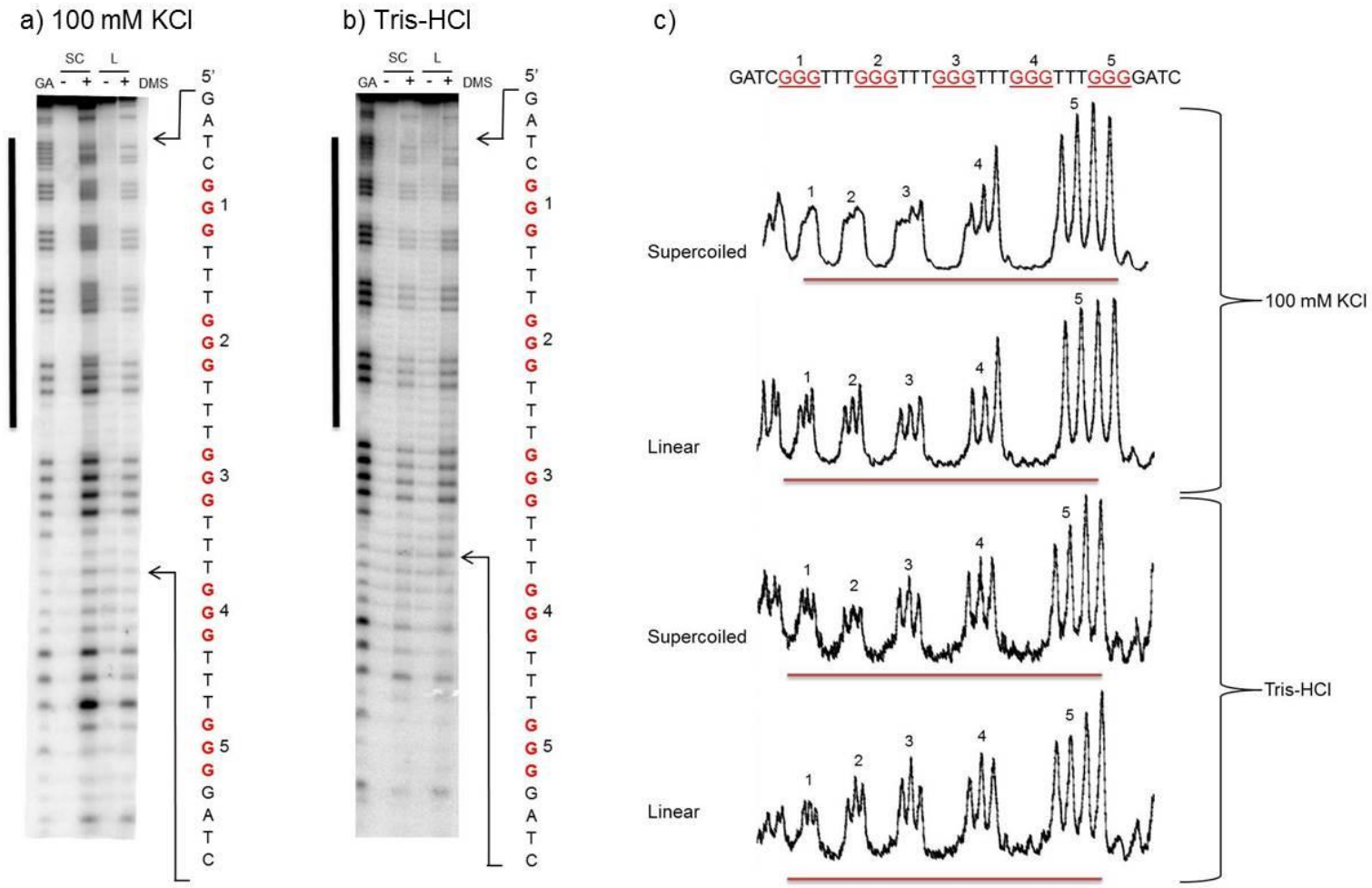


Figure 4.10 *In vitro* footprinting of the plasmid **pG$_5$,4,1** with DMS. Both Supercoiled (SC) and linearised (L) plasmids were modified with DMS. a) In the presence of 100 mM KCl. b) In the absence of KCl. Treatment without and with DMS are indicated by (-) and (+) respectively. The plasmids were incubated overnight in buffer, before reacting with DMS at room temperature. c) Densitometric scans of the DMS modified lanes, the guanine tracts numbered 1-5 are indicated by the red horizontal bars on the scan. The G-quadruplex-forming guanines were normalised to a hyperreactive guanine outside the quadruplex-forming region.

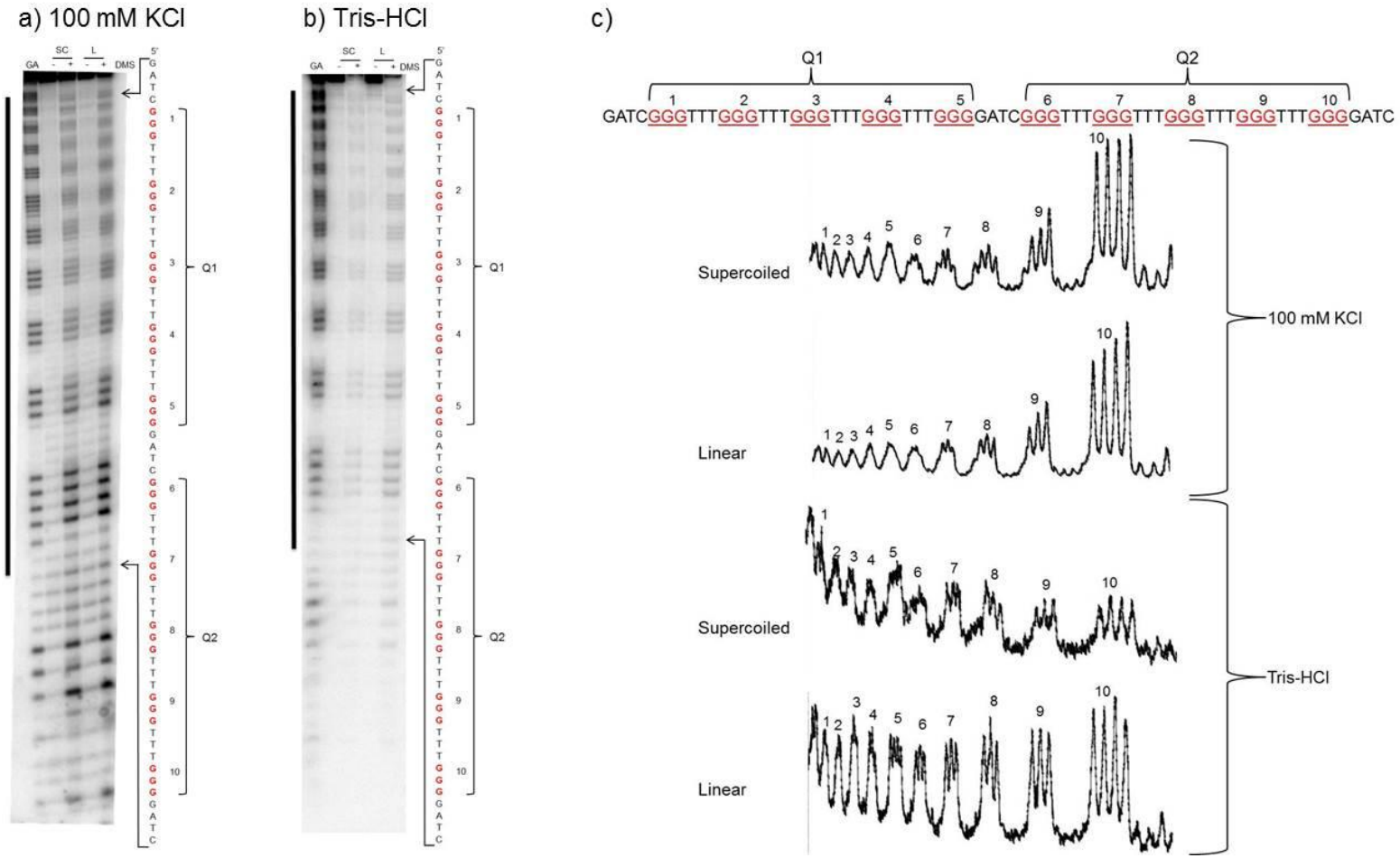


Figure 4.11 *In vitro* footprinting of the plasmid **pG$_5$,4,2** with DMS. Both Supercoiled (SC) and linearised (L) plasmids were modified with DMS. Quadruplex structure 1 and 2 are indicated by Q1 and Q2 respectively. a) In the presence of 100 mM KCl. b) In the absence of KCl. Treatment without and with are indicated by (-) and (+) respectively. The plasmids were incubated overnight and the reactions were conducted at room temperature. c) Densitometric scans of the DMS modified lanes, the guanine tracts numbered 1-10 are indicated by the red horizontal bars on the scan. The G-quadruplex-forming guanines were normalised to a hyperreactive guanine outside the quadruplex-forming region.

4.3.2 *In Vitro* Chemical Footprinting using $KMnO_4$

$KMnO_4$ reactivity assays were utilised as an alternate chemical probe for investigating the formation of G-quadruplex structures within supercoiled DNA. $KMnO_4$, which preferentially oxidises the 5,6 double bond of thymines not participating in the traditional Watson-Crick hydrogen base pairing (Hayatsu and Ukita, 1967, Bui et al., 2003), was employed to assess the presence of any single-stranded regions that might be associated with the loops within the quadruplex scaffold as well as any melted DNA in the region surrounding the structure.

4.3.2.1 G-quadruplex formation with single T-loop Sequences

Figure 4.12-1a illustrates the cleavage patterns of the $KMnO_4$ reactions on the supercoiled and linear sample of the four G_3-tract plasmid **G_4,1,1** after incubation with 100 mM KCl. Clear hyperreactivity can be seen in the thymines that immediately flank the guanine repeat sequence in the natively supercoiled sample, whilst no reactions were observed on any of the thymine residues associated with the proposed loop region (indicated in pink in Figure 4.12). A similar cleavage pattern was observed with the linearised sample of the same plasmid. The reaction observed on the thymine residues in the sequences 5'-(C**T**AGAGGA**T**) and 3'-(GA**T**C) that immediately flank the G-quadruplex forming sequences (indicated in bold in Figure 4.12) demonstrates that these thymines are more accessible than those in the remainder of the fragment and suggests that these regions may contain a segment of locally unwound DNA. However, it is not clear why $KMnO_4$ does not react with any of the thymines in the proposed loop regions, and at first sight this contrasts with the results obtained with DMS probe. It is possible that these single T-loops are sterically inaccessible with the thymidines stacked against the quadruplex in such a way that the C5-C6 bond is not accessible to reaction with the probe. In the absence of potassium a similar, though much weaker, cleavage pattern is evident.

Figure 4.12-2a and b show the results of the $KMnO_4$ probing experiments conducted with the dimeric version of the single T-looped, four G_3-tract plasmid **G_4,1,2**. In the

presence of 100 mM KCl, the supercoiled sample displayed similar band enhancements at thymine residues in the DNA flanking the 5'-and 3'-ends of the G-quadruplex structures (Figure 4.12-2a). Once again, none of the thymines associated with the G-quadruplex loops (indicated in pink in Figure 4.12) displayed any reactivity towards $KMnO_4$. In the presence of the potassium cation, the linearised sample displayed a similar pattern to that of the supercoiled sample, with similar band intensity. This may suggest that the sample contains a mixture of plasmids in dynamic equilibrium in which either one or both of the neighbouring structures are formed. Figure 4.12-2b shows that in the absence of KCl the supercoiled samples show a similar pattern of reactivity to thymines as those in the presence of 100 mM KCl (indicated in bold). There is no modification of thymines within the proposed loops of the G-quadruplex structure in the absence of the KCl. However, in the linear sample a significant reduction in intensity was seen (Figure 4.12-2b). This may indicate that the structural transitions within the G-rich region of this plasmid may not solely rely on negative supercoiling, but also the presence of the potassium ions.

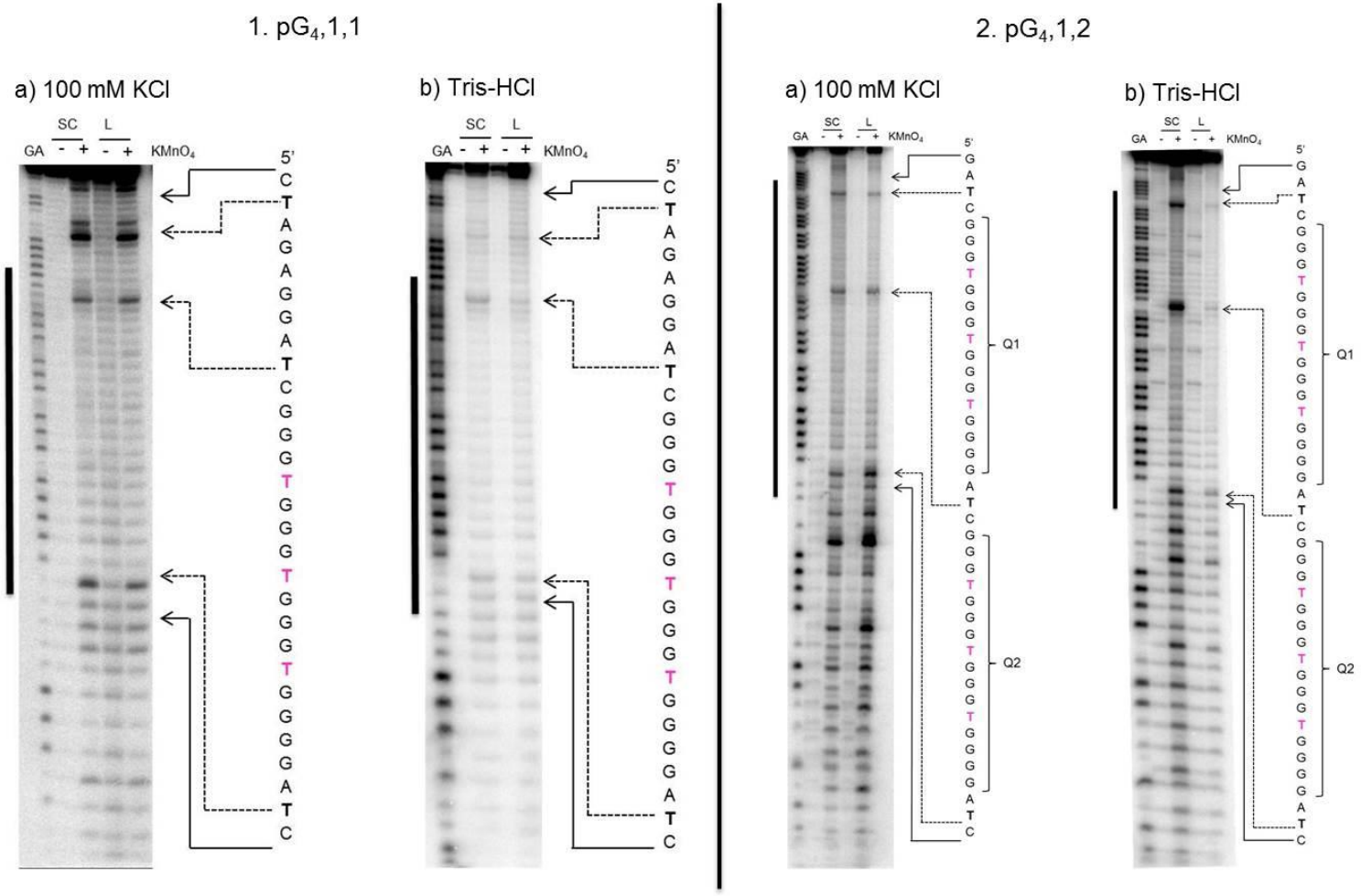


Figure 4.12 $KMnO_4$ *in vitro* footprinting on the plasmids: **pG_4,1,1** and 2. **pG_4,1,2**. Both supercoiled (SC) and linearised (L) plasmids were modified with $KMnO_4$. Quadruplex structure 1 and 2 are indicated by Q1 and Q2 respectively. a) In the presence of 100 mM KCl. b) In the absence of KCl. Treatment without and with $KMnO_4$ are indicated by (-) and (+) respectively. The plasmids were incubated in buffer overnight at room temperature before the reaction with $KMnO_4$. G-quadruplex thymines are indicated in pink.

Figure 4.13-1a shows the results of the $KMnO_4$ assays on supercoiled and linear plasmid **pG$_5$,1,1** that contains five G_3-tracts. The supercoiled sample reveals clear hyperreactivity at the thymines that immediately flank the G-quadruplex forming sequence as seen in Figure 4.13-1a (indicated in bold). In a similar trend to the four G_3-tract samples, no modification patterns was observed at any of the thymine residues within the proposed loop region (indicated in pink in Figure 4.13-1a). A similar cleavage pattern was observed for the linearised form of the same plasmid as previously observed in the four G_3-tract sample. This therefore suggests that there are regions of locally unwound DNA on either side of the G-quadruplex-forming sequence. The results obtained in this experiment contrast with those obtained in the DMS assays, which suggested the presence of an alternative secondary structure in the second, third and fourth G_3-tracts. The failure for $KMnO_4$ to modify any of the thymines in the proposed loop regions could be attributed to the C5-C6 double-bonds of the single T-loops which are sterically inaccessible to the $KMnO_4$. In the absence of potassium ions a similar banding pattern was observed for the supercoiled plasmid (Figure 4.13-1b). However, the linearised sample displayed extremely weak band intensities.

For the dimeric form of the five G-tract plasmid **pG$_5$,1,2**; similar band enhancements were observed in the presence of 100 mM KCl, in both the supercoiled and linear plasmids at thymines flanking the 5'- and 3'-ends of the G-rich sequences as shown in Figure 4.13-2a. Again, none of the thymines associated with the G-quadruplex loops (indicated in pink) displayed any reactivity to $KMnO_4$.

In the absence of KCl the thymines in the flanking region of the five G_3-tract dimeric plasmid showed reduced reactivity compared to that in the presence of 100 mM KCl, in both the supercoiled and linear plasmids (Figure 4.13-2b) The results suggest evidence of a partially single-stranded region in both the supercoiled and linear samples. Again, this suggests that the transitional structural changes within the G-rich region of this plasmid may not solely rely on negative supercoiling, but also the presence of the monovalent cation potassium.

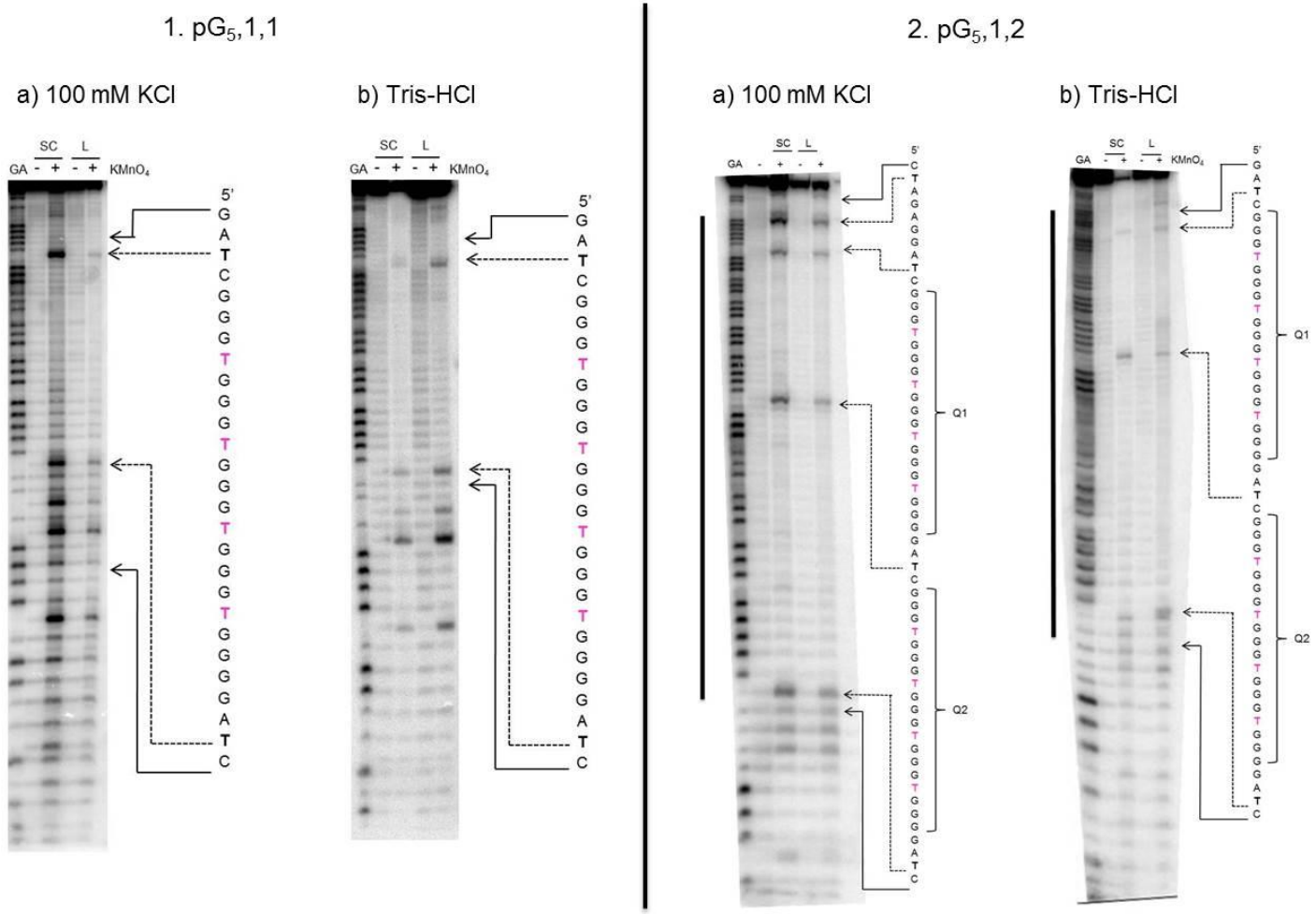


Figure 4.13 $KMnO_4$ *in vitro* footprinting on the plasmids: **pG$_5$,1,1** and 2. **pG$_5$,1,2**. Both supercoiled (SC) and linearised (L) plasmids were modified with $KMnO_4$. Quadruplex structures 1 and 2 are indicated by Q1 and Q2 respectively. a) In the presence of 100 mM KCl. b) In the absence of KCl. Treatment without and with $KMnO_4$ are indicated by (-) and (+) respectively. The plasmids were incubated in buffer overnight at room temperature before the reactions with $KMnO_4$. G-quadruplex thymines are indicated in pink.

4.3.2.2 G-quadruplex formation with T_4-loop Sequences

Figure 4.14 illustrates the cleavage patterns of $KMnO_4$ chemical probing experiments on supercoiled and linear G-rich DNA containing T_4-loop sequences in the presence and absence of 100 mM KCl. Supercoiled and linear samples of the plasmid **pG$_4$,4,1** which contains four G_3-tracts was incubated overnight in 100mM KCl. Figure 4.14-1a and b show distinct hyperreactivity of thymines that immediately flank both sides of the guanine repeat sequence. Both the supercoiled and linear DNA samples showed hyperreactivity towards thymines in the sequences at the 5'-(C**T**AGAGGA**T**) and 3'-(GA**T**C) regions (indicated in bold). No modification patterns were observed on any of the thymine loop residues in either the supercoiled or linear DNA (indicated in pink in Figure 4.14-1a and b), which is consistent with the results of DMS chemical probing of the same sample in which none of the guanines were protected from methylation. However, the fact that strong enhancements were still apparent in the flanking regions does suggest that these regions may be locally unwound. No reaction was observed in the absence of potassium (Figure 4.14-1b). In the experiments conducted without KCl, a similar pattern is evident in the flanking regions, but with decreased intensities, and no reaction is observed within the loops. These results suggest that even in the absence of potassium the region around the G-quadruplex forming sequence may be underwound.

Examination of **pG$_4$,4,2** the dimeric four G_3-tract sequence shows that in the presence of 100 mM KCl, the supercoiled and the linear samples each contain identical band enhancements (Figure 4.14-2a). In the presence of 100 mM KCl none of the thymines associated with the proposed G-quadruplex loops (indicated in pink) exhibited any reactivity to $KMnO_4$. However, some cleavage is evident in the regions that flank the G-quadruplex forming sequences, including the thymine within the junction between the two structures which is indicated in bold in Figure 4.14-2a.

In the absence of the KCl, no cleavages were seen in either the supercoiled or linear samples shown in Figure 4.14-2b indicating that the structural changes within the G-rich region of this DNA sample require the presence of potassium ions.

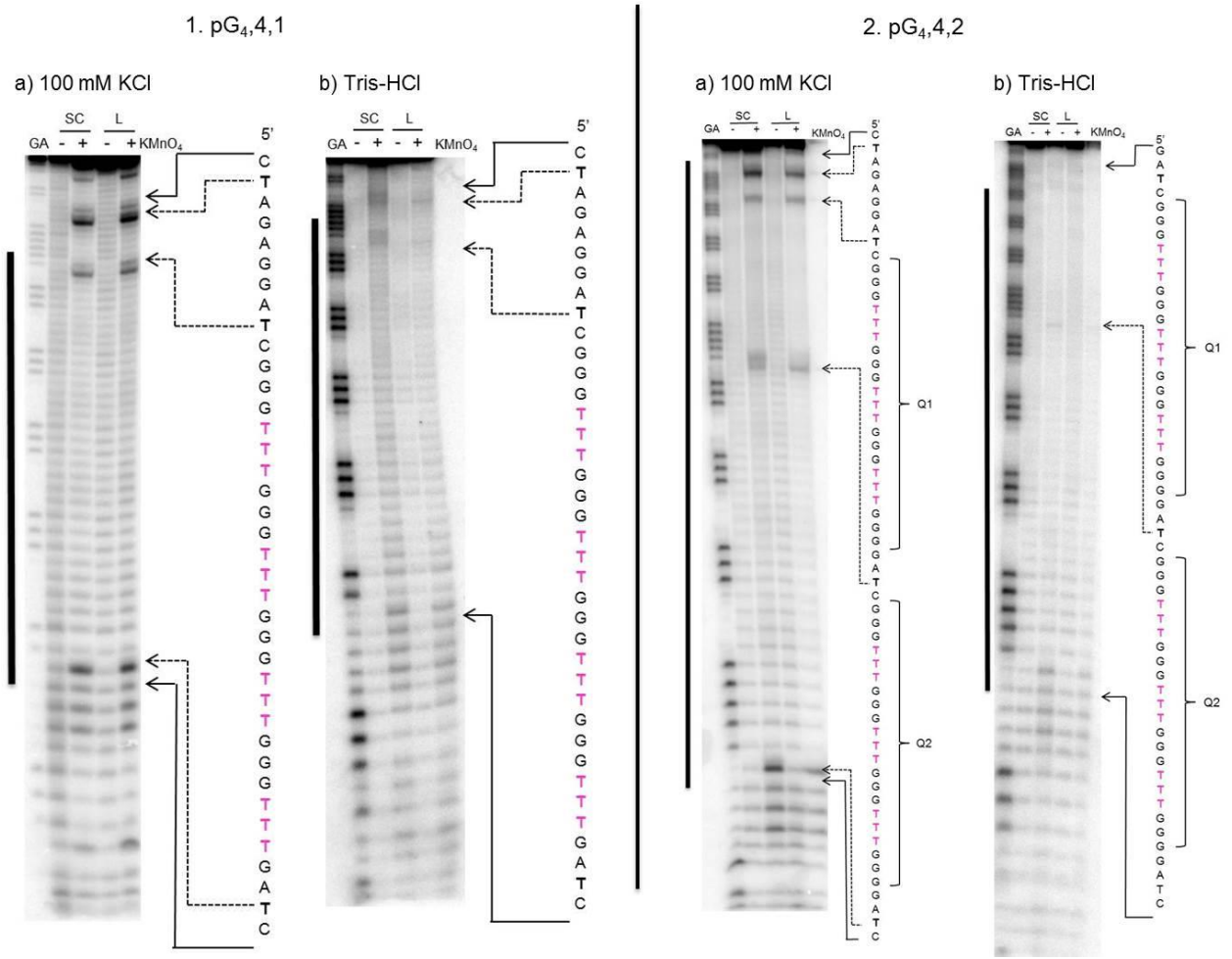


Figure 4.14 $KMnO_4$ *in vitro* footprinting of the plasmids: **pG_4,4,1** and **2. pG_4,4,2**. Both Supercoiled (SC) and linearised (L) plasmids were modified with $KMnO_4$. Quadruplex structures 1 and 2 are indicated by Q1 and Q2 respectively. a) In the presence of 100 mM KCl. b) In the absence of KCl. Treatment without and with $KMnO_4$ are indicated by (-) and (+) respectively. The plasmids were incubated in buffer at room temperature overnight before the reactions with $KMnO_4$. G-quadruplex thymines are indicated in pink.

Similar results were obtained for the plasmid containing five G_3-tracts with loops containing four thymines as shown in Figure 4.15. The results in Figure 4.15 illustrate the banding patterns observed on **pG$_5$,4,1** and **G$_5$,4,2** after chemical modification with $KMnO_4$.

Both the supercoiled and linear samples of the monomeric plasmid **pG$_5$,4,1** exhibit similar enhancement patterns in both the presence and absence of KCl (Figure 4.15-1a and b). No hyperreactive bands were observed within the loop regions indicated in pink in Figure 4.15-1a and b, but strong cleavage products were seen in the 5'- and 3'-flanking regions of the G-rich sequence (indicated in bold), though the samples treated without potassium, showed markedly less reactivity. The absence of any $KMnO_4$ modification of the thymine loop residues in either the supercoiled or linear DNA is consistent with the DMS footprinting results, which showed no evidence of G-quadruplex induced protection. However, the presence of strong enhancements in the flanking regions does imply that these regions may contain locally unwound DNA.

For the dimeric insert **pG$_5$,4,2** containing a five G-tract sequence, both the supercoiled and the linear samples exhibit similar bands in the presence of 100 mM KCl, at the thymine residues in the regions that flank the 5'- and 3'- termini of the G-quadruplex forming sequences (Figure 4.15-2a). As previously demonstrated with the four G_3-tract dimer, none of the thymines associated with the G-quadruplex loops (indicated in pink in Figure 4.15-2a) exhibited any reactivity to $KMnO_4$. In the absence of KCl (Figure 4.15-2b) no $KMnO_4$ modifications were detected on the supercoiled and linear sequences with this dimeric insert, in either the loops or the flanking regions, indicating that the unwound structure is dependent on the presence of potassium ions.

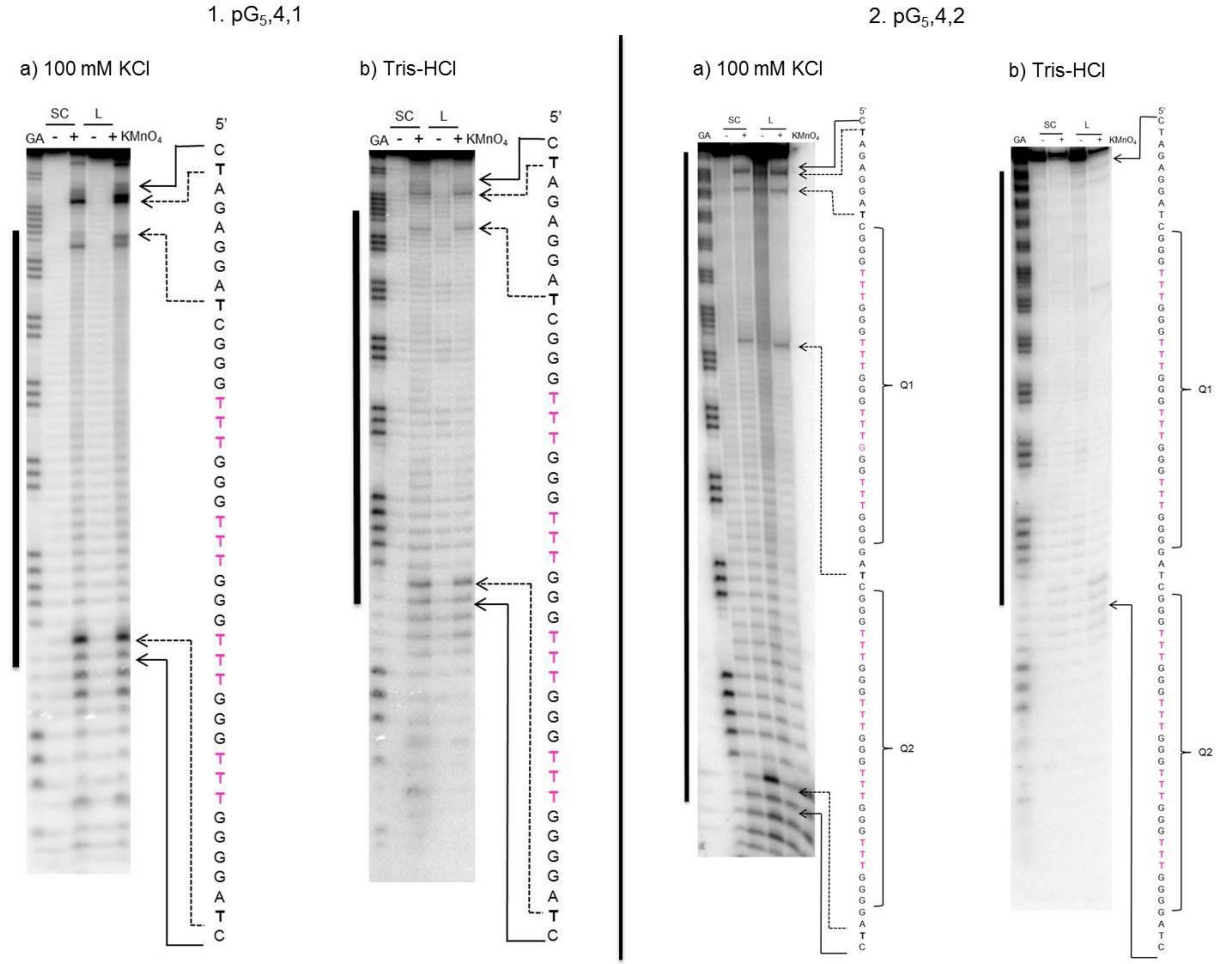


Figure 4.15 $KMnO_4$ *in vitro* footprinting on the plasmids: **pG$_5$,4,1** and **2. pG$_5$,4,2.** Both Supercoiled (SC) and linearised (L) plasmids were modified with $KMnO_4$. Quadruplex structures 1 and 2 are indicated by Q1 and Q2 respectively. a) In the presence of 100 mM KCl. b) In the absence of KCl. Treatment without and with $KMnO_4$ are indicated by (-) and (+) respectively. The plasmids were incubated in buffer overnight at room temperature before the reactions with $KMnO_4$. G-quadruplex thymines are indicated in pink.

4.3.3 The use of S1 Nuclease to Map the Sites of G-Quadruplex Formation Within Supercoiled DNA

S1 nuclease was used in order to supplement the results from the chemical probing experiments. As previously described it probes for single-stranded regions of DNA, and was used to identify the loop of the G-quadruplex and any single-strand regions at the quadruplex-duplex junction. Prior to the S1 nuclease cleavage each DNA sample was incubated overnight in 100 mM KCl as described in Chapter 2. Each supercoiled plasmid was digested sequentially with S1 nuclease (to identify the single-stranded regions) and then *Sca*I. Prior to digestion with S1 nuclease, an identical sample was digested with *Sca*I (to linearise the plasmid and eliminate any negative supercoiling). *Sca*I was chosen because it cleaves at a unique site on the opposite side of the plasmid outside of the multiple cloning site (MCS). As the *Eco*RI restriction site is in close proximity to the *Bam*HI cloning site, a further sample was digested with both *Eco*RI and *Sca*I and this restriction digest pattern was used as an aid to map the S1 cleavage sensitive sites. The results in this section show the cleavage of each plasmid in both supercoiled and linear states.

4.3.3.1 Mapping the G-quadruplex cleavage sites of single T-loops sequences with S1 nuclease

Figure 4.16a shows S1 cleavage of the G-quadruplex forming plasmid **pG$_4$,1,1** which consists of four G_3-tracts separated by a single T-loop. After the supercoiled DNA (lane 1) had undergone cleavage with S1 nuclease the plasmid was converted to a linear species (lane 2). This suggests that the enzyme has cleaved the plasmid in a specific location, as random cleavage would likely result in a smear of any overlapping fragment. After S1 treatment, the DNA was digested with *Sca*I, two distinct fragments can be seen at approximately 1900 bp and 760 bp (lane 3). This again provides evidence that S1 has cut in a specific location, as random double stranded breaks would lead to a smear of fragments of different lengths on subsequent cutting by *ScaI*. When the plasmid was linearised with *Sca*I (lane 4) prior to digestion with S1 nuclease a single linear fragment was observed, which remained after the subsequent treatment with S1 nuclease (lane 5). This suggests that the conformational change that occurred in the supercoiled DNA is no longer retained in the linear plasmid. Figure 4.16a also shows the restriction digests of *Eco*RI (lane 6) which cleaves at position 395 of the vector and

the double digest with *Sca*I which cleaves at position 2177 (lane 7) yielding two distinct DNA restriction fragments with lengths of 1780 bp and 900 bp (lane 7). Although these results show that negative supercoiling has induced a conformational transition from duplex-to-single-stranded DNA at a specific location, the fragment lengths observed (1900 and 760) do not correspond to the theoretically calculated results for G-quadruplex formation of approximately 1760 bp and 925 bp. This implies that there may be another region within the plasmid that experiences conformational change under superhelical stress, thus revealing an S1 sensitive site.

The dimeric plasmid **pG$_4$,1,2** which contains two G-quadruplex forming sequences separated by the short DNA sequence d(GATC) displayed similar results (Figure 4.16b). After digestion with S1 nuclease, the supercoiled DNA was converted to a linear fragment (lane 2), again, indicating the presence of S1 sensitive site(s). Following digestion with *Sca*I, two distinct fragments at approximately 1900 bp and 760 bp were seen (lane 3). When the plasmid was initially digested with *Sca*I to produce linear DNA (lane 4) and then cleaved with S1 nuclease (lane 5) an identical fragment to that of the *Sca*I digest was observed, showing that the S1 nuclease sensitive site is only present in the supercoiled species.

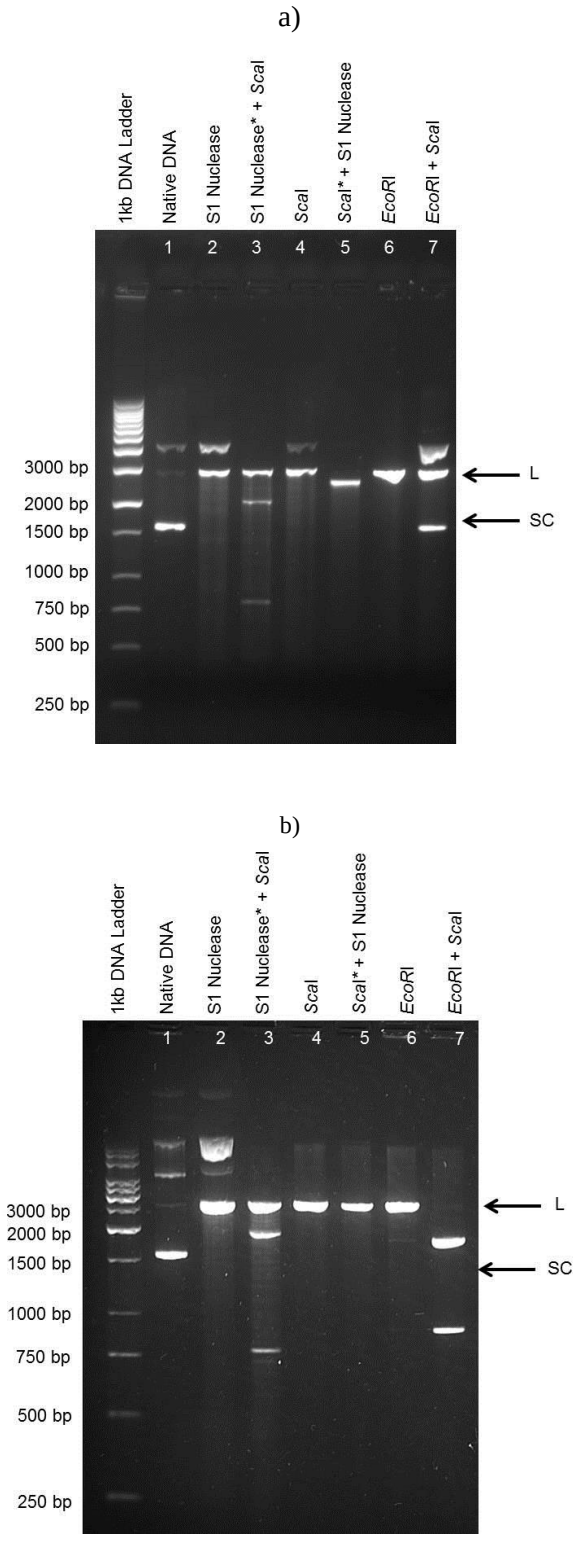


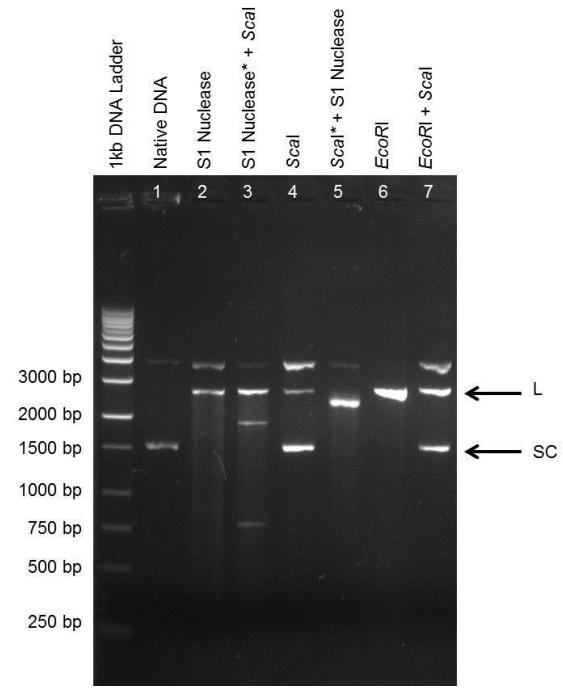
Figure 4.16 Mapping of S1 nuclease-sensitive sites in plasmids a) **pG$_4$,1,1** and b) **pG$_4$,1,2**: 1. Natively supercoiled DNA, 2. S1 nuclease cleavage, 3. Plasmids digested sequentially with S1 nuclease and then *Sca*1, 4. *Sca*I restriction digest, 5. Plasmids digested sequentially with *Sca*1 and then S1 nuclease, 6. *Eco*RI restriction digestion, 7. *Eco*RI and *Sca*I restriction digestion. Where SC indicates supercoiled plasmids and L indicates linearised plasmids.

4.3.3.2 Mapping the G-quadruplex cleavage sites of T_4-loops sequences with S1 nuclease

Figure 4.17a) illustrates S1 cleavage of the G-quadruplex forming plasmid **pG$_4$4,1** which consists of four G_3-tracts separated by T_4-loops. The results obtained for this experiment were similar to those presented above for the single T-loop sequences. After treatment with S1 nuclease, the supercoiled DNA was converted to a linear fragment (lane 2) indicating a region of locally unwound DNA. Following a restriction digest with *Sca*I, two DNA fragments at approximately 1900 bp and 760 bp were observed (lane 3). When digested with *Sca*I the plasmid was linearised, (lane 4) which remained as a single fragment after digestion with S1 nuclease (lane 5). This suggests that the conformational changes occurring within this plasmid are supercoil-dependent. Again, the banding pattern observed for this plasmid did not correspond to the expected results for G-quadruplex formation of approximately 1760 bp and 925 bp, indicating that a possible secondary region within the plasmid experiences conformational change under superhelical stress.

The dimeric plasmid **pG$_4$4,2** also displayed similar results (Figure 4.17b), to those of the monomeric sample (Figure 4.17a) and those observed in the single T-loop plasmids (Figure 4.16). After digestion with S1 nuclease, the supercoiled DNA was converted to a linear fragment (lane 2), identifying the presence of an S1 sensitive site. Following digestion with *Sca*I, two fragments of sizes approximately 1900 bp and 760 bp were seen (lane 3). When the plasmid was initially digested with *Sca*I to linearise the DNA (lane 4) and then cleaved with S1 nuclease (lane 5) only one fragment was visible, implying that the single-stranded region of DNA that is S1 nuclease sensitive is only present in the supercoiled species. Again, as none of the theoretical calculations were compatible with those for G-quadruplex formation, it is unlikely that the S1 nuclease experiments were able to identify and map the structures.

a)



b)

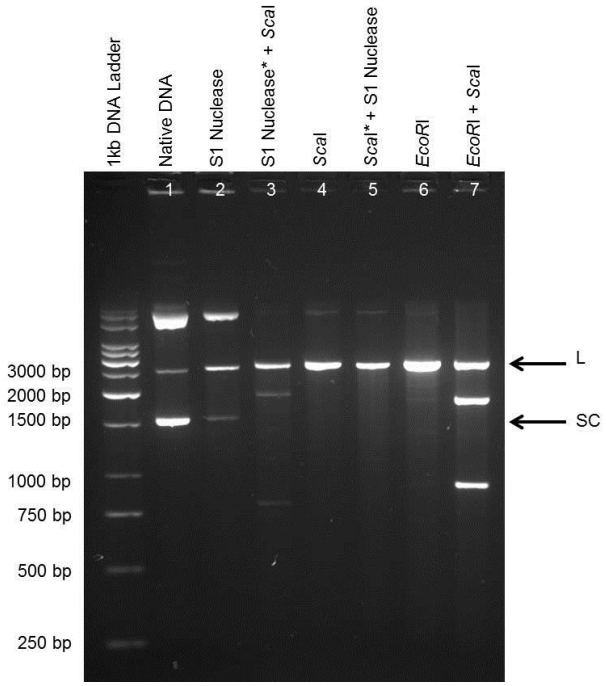


Figure 4.17 Mapping of S1 nuclease-sensitive sites in plasmids a) **pG$_4$,4,1** and b) **pG$_4$,4,2**: 1. Natively supercoiled DNA, 2. S1 nuclease cleavage, 3. Plasmids digested sequentially with S1 nuclease and then *Sca*1, 4. *Sca*I restriction digest, 5. Plasmids digested sequentially with *Sca*1 and then S1 nuclease, 6. *Eco*RI restriction digestion, 7. *Eco*RI and *Sca*I restriction digestion. Where SC indicates supercoiled plasmids and L indicates linearised plasmids.

4.3.3.3 Cleavage of pUC19 with S1 nuclease

To determine whether the cleavage patterns obtained for the cloned G-rich plasmids at specific loci correspond to the formation of intramolecular G-quadruplexes, the S1 nuclease experiments were also performed on the native pUC19 cloning vector. Figure 4.18 illustrates the results of this experiment. Natively supercoiled pUC19 (lane 1) was digested by S1 nuclease (lane 2), which was then cleaved by *Sca*I, which only cuts once within the plasmid molecule (lane 3). A separate sample of the plasmid was also initially cut with *Sca*I (lane 4) to linearise the plasmid to remove torsional stress prior to digestion with S1 nuclease (lane 5). The results also depict the digestion with *Eco*RI and *Sca*I as a mapping tool (lane 7). Although the original starting material contained a considerable amount of nicked DNA, after digestion with S1, the majority of the supercoiled plasmid was converted to a mixture of nicked and linear DNA fragments, which was noted by the slower migration. After digestion with *Sca*I a second cleavage product was observed (lane 3). When the supercoiled plasmid was first linearised with *Sca*I the expected single DNA fragment at approximately 3000 bp was produced (lane 4); no further bands were produced when this linearised DNA was digested with S1 nuclease. These results suggest the pUC19 plasmid itself also contains a supercoil-dependent S1 nuclease sensitive site that is eliminated when the plasmid is linearised.

pUC19 is a cloning vector derived from the bacterial plasmid pBR322 (Yanisch-Perron et al., 1985) and it therefore contains the same palindromic 11 bp inverted repeat sequence that is separated by three nucleotides (Bolivar et al., 1977). This sequence has previously been shown to form a cruciform structure (see Figure 4.19) under superhelical stress (Panayotatos and Wells, 1981, Lilley, 1980, Panayotatos and Fontaine, 1987). Therefore, it is possible that the cleavage pattern observed for the pUC19 control experiments corresponds to the cruciform structure in the pUC19 plasmid. The theoretical calculations for the banding pattern of the cruciform structure digested with S1 nuclease and *Sca*I should produce DNA fragment sizes of 1890 and 790 bp as *Sca*I cleaves the DNA at 2177 bp and the cruciform structure is at 1385 bp. This estimation reflects the banding pattern observed in the S1 nuclease experiments and as the cruciform structure is supercoil-dependent, it therefore offers an explanation as to why this same banding pattern was not observed after the plasmid was linearised.

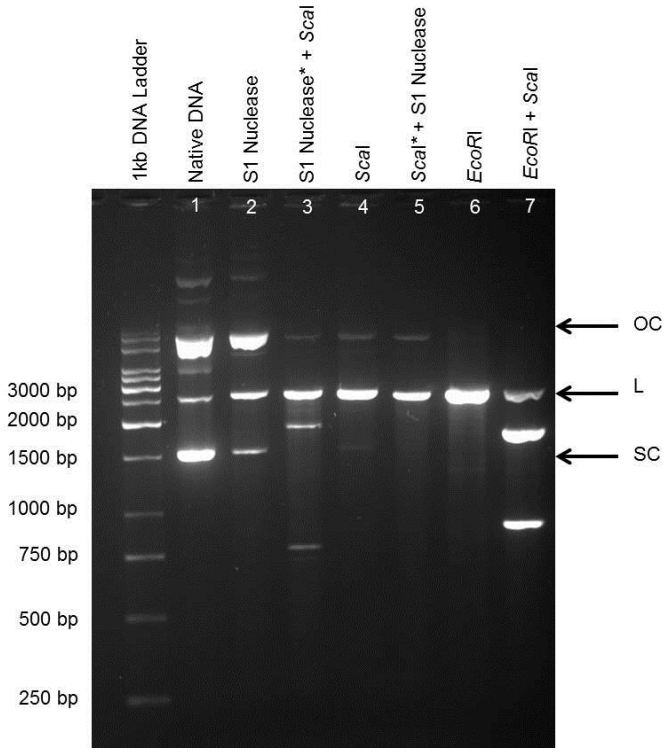


Figure 4.18 Mapping of S1 nuclease-sensitive sites in pUC19. 1. Natively supercoiled DNA, 2. SI nuclease cleavage, 3. Plasmids digested sequentially with S1 nuclease and then *Sca*1, 4. *Sca*I restriction digest, 5. Plasmids digested sequentially with *Sca*1 and then S1 nuclease, 6. *Eco*RI restriction digestion, 7. *Eco*RI and *Sca*I restriction digestion. Where OC and SC indicate open circular and supercoiled plasmids respectively and L indicates linearised plasmids.

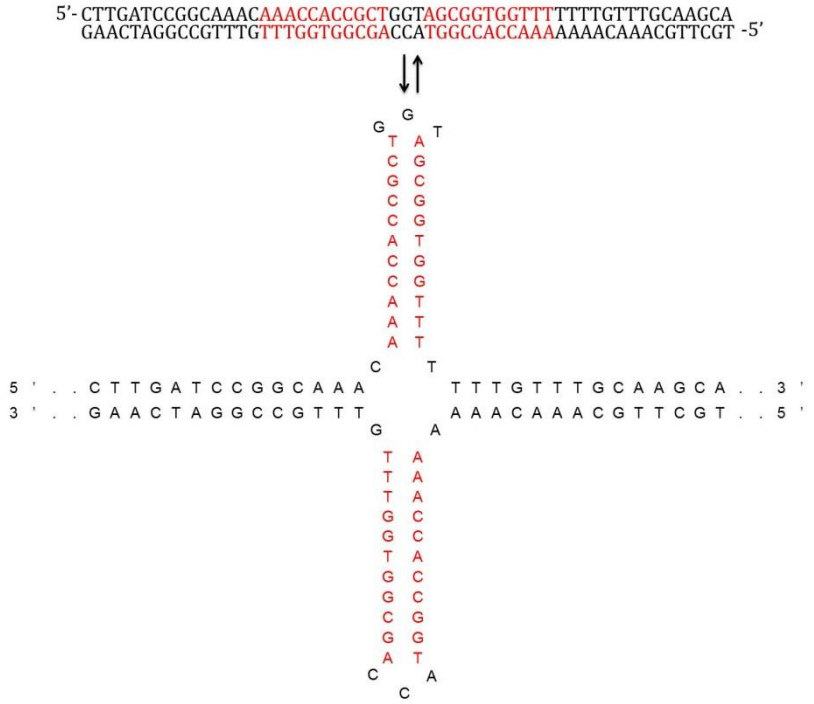


Figure 4.19 Formation of a pUC19 cruciform. Formation of the cruciform occurs at the 11 bp inverted repeat sequence (highlighted in red).

4.3.4 Probing an Extended G-rich Sequence Within Supercoiled DNA: $pG_n,4,X_n$

During the cloning of the various G-rich plasmids described in this work (chapter 3) an unusual mutant was isolated that contained an imperfect combination of $(G_3T_4)_n$ sequences. This mixed orientation plasmid **$pG_n,4,X_n$** is potentially a highly polymorphic sequence that can adopt several isomeric quadruplex structures. It was therefore subjected to the same chemical modifications and S1 nuclease analyses as used on the previous plasmids in this chapter. This DNA molecule is a trimeric G-quadruplex-forming plasmid, in which each structure has the potential to form a three quartet, three T_4-looped scaffold. The first part of the sequence contains four G_3-tracts, whilst the second and third probably contain five G_3-tracts. The 5'-end contains the sequence $GATCC_3A_4C_3A_4C_3GATC$ which may form a hairpin with one of the other G-rich tracts. Each probing experiment was conducted in the presence of 100 mM KCl.

4.3.4.1 DMS

Figure 4.20 shows the results of the DMS footprinting on the G-rich fragment from **$pG_n,4,X_n$**. Densitometric scans of the DMS modified lanes are shown in Figure 4.20b. As in previous T_4-loop experiments no DMS protection was observed in either the supercoiled or the linear experiments and all the guanine bands are still present. Although the DMS footprinting in this gel appears to indicate that guanines at the 3'-(lower) end of the sequence are more intense than those at the 5'- (upper) end, this is merely a consequence of the samples being over digested by the DMS. Also as there was no increase or decrease in band intensity on any guanines within the sequence it can be concluded that the DMS assay did not detect any Hoogsteen-hydrogen bonding.

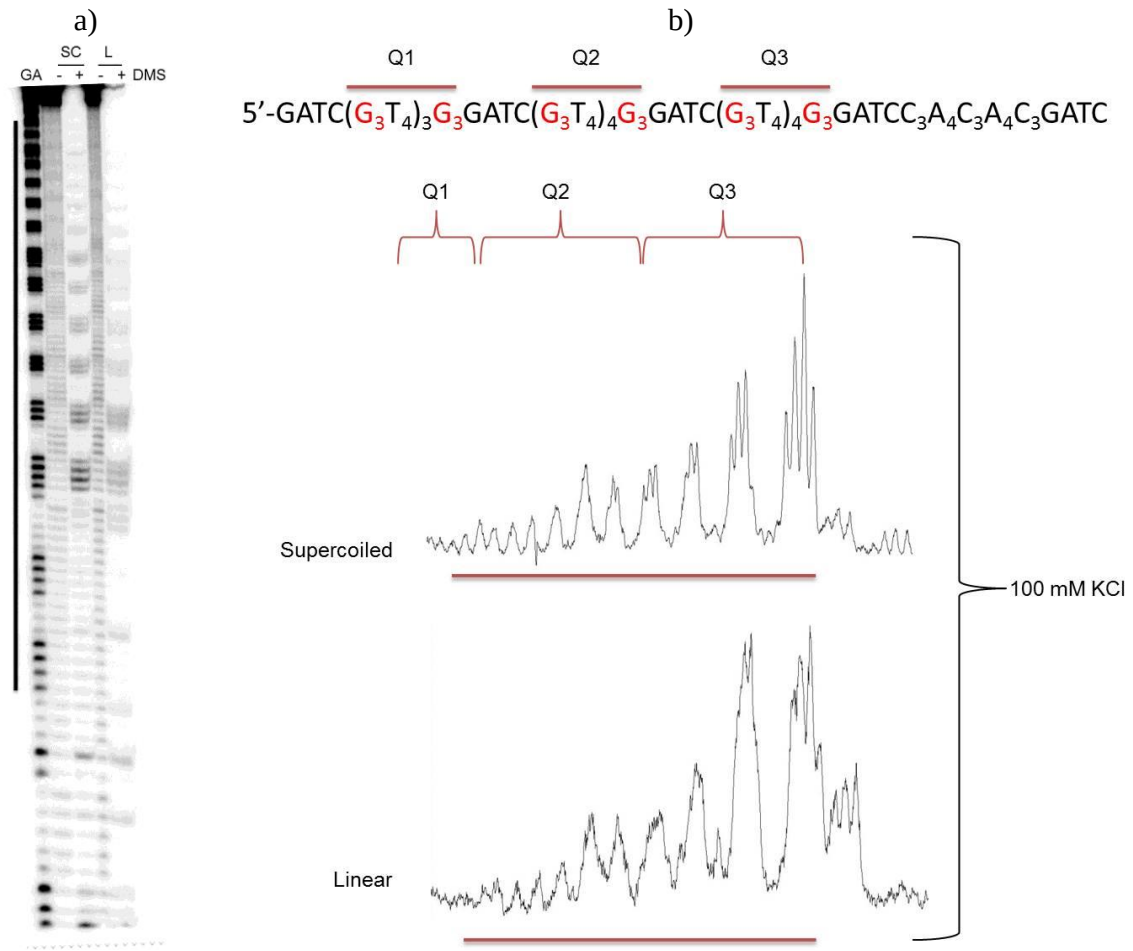


Figure 4.20 *In vitro* footprinting on the plasmid **pG_n4X_n** with DMS. Both supercoiled (SC) and linearised (L) plasmid were modified with DMS. Quadruplex structures 1, 2 and 3 are indicated by Q1, Q2 and Q3 respectively. Treatment without and with DMS are indicated by (-) and (+) respectively. The plasmids were incubated overnight in 100 mM KCl before the reaction with DMS, which was conducted at room temperature. c) Densitometric scans of the DMS modified lanes, the guanine tracts numbered are indicated by the red horizontal bars on the scan.

4.3.4.2 $KMnO_4$

Figure 4.21 shows the results of the reaction of $KMnO_4$ with the G-rich DNA sample **$pG_n,4,X_n$**, in which it can be seen that the supercoiled and linear samples produced different modification patterns to that seen in the other inserts in this chapter. Each thymine in the (GATC) regions between each series of G_3-tracts exhibits hyperreactivity towards $KMnO_4$ (numbered 1, 2, 3 in Figure 4.21). There is no reaction with thymines in the potential T_4-loop regions. This contrasts with the DMS experiments where no conformational changes were observed. There is also a clear difference between the reaction with supercoiled and linear DNA, in which bands 4, 5 and 6 are present only in the supercoiled sample. While band 4 corresponds to a hyperreactive thymine, the weaker bands 5 and 6 are at cytosine residues. These results suggest that there is a supercoil-dependent conformational change in this region.

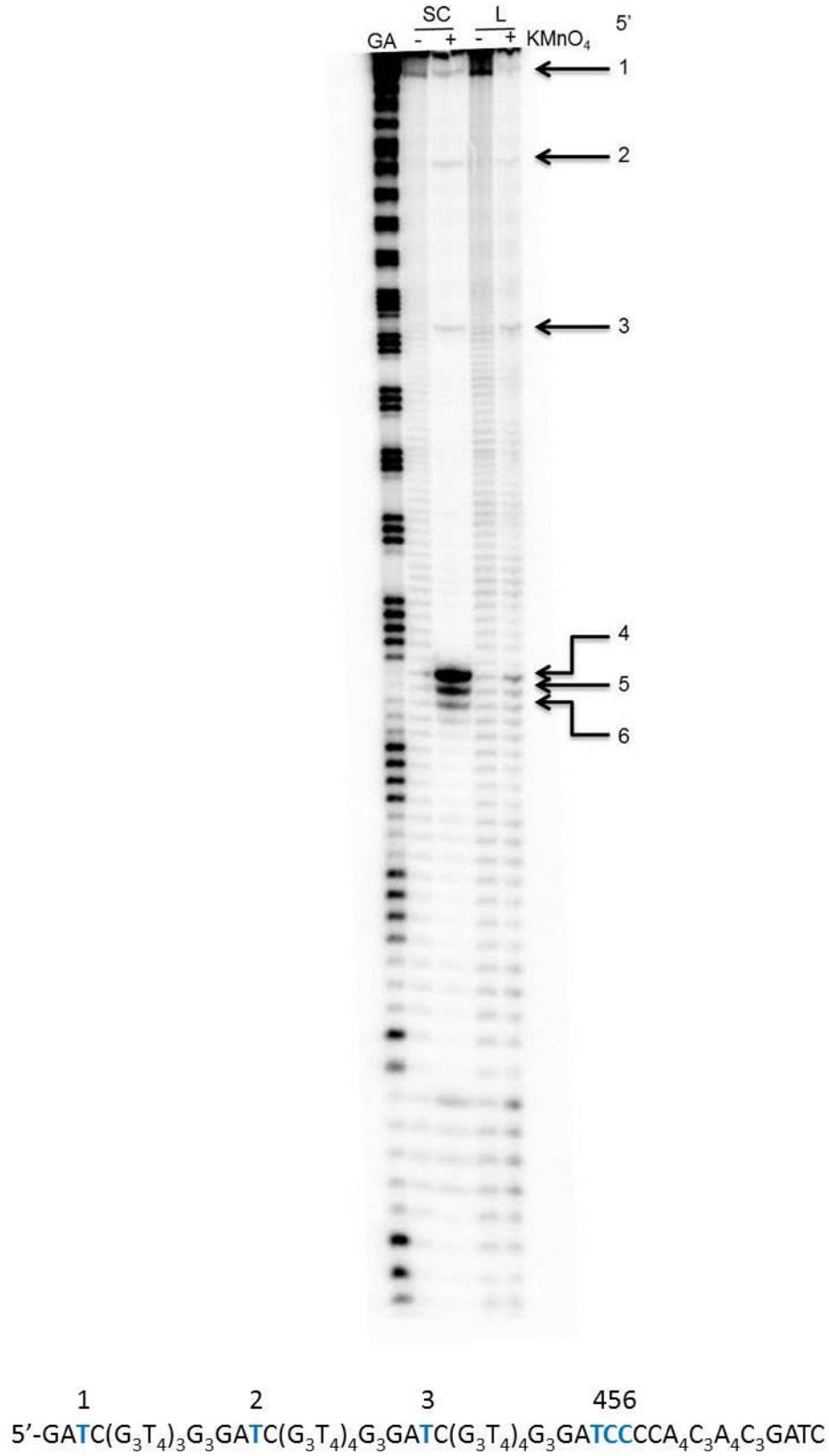


Figure 4.21 $KMnO_4$ *in vitro* footprinting of plasmid **$pG_n,4,X_n$.** Both supercoiled (SC) and linearised (L) plasmid were modified with $KMnO_4$. Individual Quadruplex structures are indicated at positions 1, 2 and 3, whilst 4,5,6 are thymine and two cytosine bases respectively. Treatment without and with $KMnO_4$ are indicated by (-) and (+) respectively. The plasmids were incubated overnight in 100 mM KCl, before the reactions with $KMnO_4$, which were conducted at room temperature.

4.3.4.3 S1 Nuclease

Figure 4.22 shows the results of S1 cleavage of the G-quadruplex forming plasmid **$pG_n,4,X_n$**. When the supercoiled DNA (lane 1) was digested with S1 nuclease the plasmid was converted to a species containing a mixture of open circular and linear DNA (lane 2), indicating the presence of a region of locally unwound DNA that is sensitive to S1 nuclease digestion. When the DNA was subsequently digested with *Sca*I a series of shorter fragments were generated (lane 3). The digestion pattern consisted of DNA fragments approximately 2300, 2000, 1800, 900 and 760 bp in size. Although some of the DNA fragments coincide with the expected banding pattern for S1 cleavage within the region of the three G-quadruplex structures (2000, 1900 and 900 bp) others do not and correspond to the patterns observed for the monomeric and dimeric single T- and T_4-looped samples (Figure 4.20 and Figure 4.21 respectively).

When the plasmid was first linearised with *Sca*I (lane 4), followed by the S1 nuclease digestion a single linear fragment was observed (lane 5). Indicating that any localised DNA unwinding that results in a conformational change was no longer retained in the linear plasmid. The DNA cleavage pattern of this sample suggests that there may be two regions within this plasmid that are sensitive to S1 cleavage in supercoiled DNA, one of which is in the vicinity of the G-rich insert.

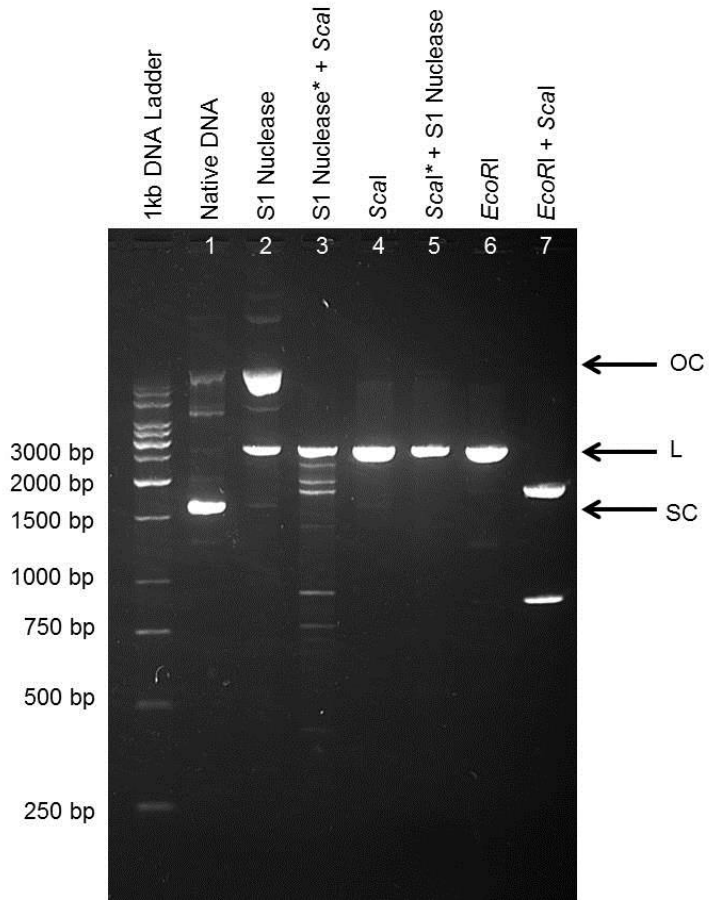


Figure 4.22 Mapping of S1 nuclease-sensitive sites in plasmid **$pG_4,4,X_n$.** 1. Natively supercoiled DNA, 2. S1 nuclease cleavage, 3. Plasmids digested sequentially with S1 nuclease and then *Sca*1, 4. *Sca*I restriction digest, 5. Plasmids digested sequentially with *Sca*1 and then S1 nuclease, 6. *Eco*RI restriction digestion, 7. *Eco*RI and *Sca*I restriction digestion. Where OC and SC indicate open circular and supercoiled plasmids respectively and L indicates linearised plasmids.

4.4 Discussion

Under physiological conditions such as pH, temperature and ionic concentration most G-rich sequences form duplex structures in preference to quadruplexes, in the presence of its complementary strand (Li et al., 2002, Phan and Mergny, 2002, Risitano and Fox, 2003b). In closed-circular DNA the free energy of negative supercoiling can provide the energy required for local unwinding and melting of the double helix. Conformational sequence-specific conversions within these sites can facilitate the transition from the canonical B-DNA to an alternative secondary structure such as cruciforms (Lilley, 1980) and Z-DNA (Nordheim et al., 1983). The free energy of the negative supercoiling in closed-circular plasmids mimics the kind of torsional stress that may occur *in vivo*.

Previous studies have demonstrated that the formation of G-quadruplexes can also be induced by negative supercoiling, which were identified using both chemical and enzymatic probes (Sun and Hurley, 2009, Onyshchenko et al., 2009, Onyshchenko et al., 2011). The probing of nucleic acids with either chemicals or enzymes is a popular and versatile approach for determining and characterising structural perturbations in both DNA and RNA (Lilley, 1992). This is achieved by examining the localised reactivity of chemical reagents for base-specific modification (Mattes et al., 1986, Howgate et al., 1968, Kahl and Paule, 2001) or enzymatic cleavage of nuclease sensitive sites (Viville and Mantovani, 1994, Sun, 2010).

Sun and Hurley (Sun and Hurley, 2009) reported that under physiological conditions negative supercoiling could facilitate the formation of a G-quadruplex structure in the nuclease hypersensitive element III_1 (NHE III_1) of the *c-myc* promoter. By comparing the wild-type sequence to a mutated sequence that disruptes the quadruplex structure, they were able to demonstrate, using DMS, $KMnO_4$ and S1 nuclease as probes, that in the presence of the potassium cations, negative supercoiling can induce the formation of a G-quadruplex. Interestingly they also demonstrated that on the complementary C-strand, the hemiprotonanted *i*-motif structure that typically forms at low pH was also able to form at neutral pH if the negative superhelicity was maintained. In another study

Onyshchenko *et al* (Onyshchenko et al., 2009) investigated the formation of a G-quadruplex structure in the human *Bcl-2* promoter region. They were not able to detect any evidence for local DNA melting or G-quadruplex formation using DMS, DEPC (diethyl pyrocarbonate) and OsO_4 (osmium tetroxide) as structural probes. However, by using C-strand-invading peptide nucleic acids (PNAs), they were able to induce the formation of the structure and offer a new mode of sequence-specific gene targeting. The authors later followed up this study by reporting that G-quadruplex formation in the human *Bcl-2* was a prerequisite for a stable invasion of the C-strand binding PNAs and vice versa (Onyshchenko et al., 2011).

With the aid of the chemical probes dimethyl sulphate and potassium permanganate and the enzyme S1 nuclease, the results presented in this chapter have investigated G-quadruplex formation and examined whether the torsional energy generated from negative supercoiling can assist the formation and stabilisation of G-quadruplex structures.

4.4.1 Investigating G-quadruplex Formation Using DMS and KMnO4 Chemical Probes and S1 Nuclease as an Enzymatic Probe

4.4.1.1 Single T-loops

The DMS assay on the plasmid **pG_4,1,1** revealed that in the presence of 100 mM KCl DMS showed significantly reduced reactivity towards all the guanines in the potential quadruplex-forming sequence in comparison to the guanine bases outside the cloned fragment. The results revealed that the relative reactivity towards the guanines within the quadruplex forming sequence increased in a 5'- to 3'-end direction. The G_3-tracts at positions 1 and 2, nearest to the 5'-end showed the greater protection from DMS methylation whilst the G_3-tract in position 4 showed the least. Unexpectedly, this effect was seen with both supercoiled and linear DNA. These results appear to suggest the existence of Hoogsteen or Hoogsteen-like hydrogen bonding. In general, intramolecular G-quadruplexes containing all single T-loop topologies assemble extremely fast and form very stable intramolecular G-quadruplexes (Rachwal et al., 2007b). Other studies

have shown that, in the presence of the C-rich complement, G_3T quadruplexes are formed in preference to the Watson-Crick duplex (Risitano and Fox, 2003b), which is consistent with the CD data presented in Chapter 3 and substantiates the DMS methylation patterns observed. A previous study has demonstrated that G-quadruplex structures within supercoiled DNA can exist in a dynamic equilibrium between duplex, unwound duplex, single-stranded DNA and G-quadruplex formation (Sun and Hurley, 2009). Therefore it is possible that the variation in protection of the different G_3-tracts may be an effect of the extremely fast association and dissociation of the G-quadruplex scaffold and that the G-quadruplex structure does not fully unfold at the 5'-end before it reassembles again. Nevertheless, it is clear that some degree of structural change had occurred within this particular sequence of DNA. Moreover, without further investigations, it is not possible to fully confirm G-quadruplex formation. When the same experiment was repeated in the absence of potassium, the methylation patterns for both the supercoiled and linear samples were identical to the experiments conducted in 100 mM KCl. However, in this instance each of the four G_3-tracts exhibited a 20% increase in susceptibility towards DMS methylation. Although the structural perturbation within this G-rich sequence appeared not to be dependent on negative supercoiling, the reduction in reactivity to DMS observed in the presence of potassium ions appears to play a moderate role in stabilisation of these structures. These results are consistent with those obtained by CD experiments in Chapter 3, which indicated that, even in the absence of KCl, the four G_3-tract, single T-looped sequence was still able to associate and form a stable intramolecular parallel stranded structure, although the results were prominent in the presence of K^+.

The insertion of a second G-quadruplex forming motif (**pG_4,1,2**) generated a similar methylation in which the guanine tracts were significantly protected from DMS methylation, with the two 5' G_3-tracts showing a relatively reduced reactivity to DMS compared to two 3'-end tracts. These results suggest that both these regions contain segments of melted DNA that may correspond to G-quadruplex formation, although it is not possible to conclude whether both structures form simultaneously or at independent times or whether the presence of two structures has either a stabilising or destabilising effect.

By extending the G_3-tracts to include a fifth tract of d(GGG), the G-quadruplex forming sequence **pG$_5$,1,1** can adopt several loop isomers. The most likely structure(s) consisting of a 5'-d(GATCGGGT) overhang or a 3'-d(TGGGGATC) overhang. It is also possible, but less likely that a central d(TGGGT) (G-tract 3) may be accommodated into an extended loop sequence with both the flanking 3' and 5' thymines similar to the *c-myc* oncogene promoter sequence found in the NHE III_1 which contains six consecutive G-rich tracts (Simonsson et al., 1998). The results with the **pG$_5$,1,1** clone show reduced reactivity at all the guanines, with the three central G_3-tracts (2,3 and 4) in Figure 4.6 having reduced reactivity towards DMS, compared to the 5'- and 3'-end flanking G_3-tracts (1 and 5). It is worth noting that DMS can only measure the average protection of any specific guanine; in the simplest structures the central three G_3-tracts will always be included in the quadruplex scaffold, while the G_3-tracts at positions 1 and 5 will feature in only one of the folded topologies. The methylation patterns are consistent with this reasoning, suggesting that the DNA solution contains an isomeric mixture of intramolecular quadruplexes that differ by flanking sequence. The same methylation pattern was observed when the plasmid was linearised, suggesting that formation of this G-quadruplex structure is not dependent on negative supercoiling. In the absence of KCl, the supercoiled and linear samples displayed similar methylation patterns to the experiments conducted in 100 mM KCl and as seen with four tract sample **pG$_4$,1,1**, the reactivity towards DMS was increased by approximately 20%, under ionic conditions suggesting that the stabilisation of the structure is potassium dependent.

Methylation patterns consistent with the monomeric **pG$_5$,1,1** sample was observed with the dimeric five G_3-tract sequence, **pG$_5$,1,2** in both the supercoiled and linear samples. For each structure the central three G_3-tracts again were the most protected with an increase in reactivity towards the external 5'- and 3'- guanine tracts. Both methylation patterns suggest the presence of Hoogsteen-hydrogen bonding and the possible formation of G-quadruplex structures. These results were replicated in the absence of KCl.

To substantiate and supplement the results gained from the DMS experiments, $KMnO_4$ footprinting assays were employed to identify the loop and flanking thymine bases that should be more accessible to this probe under negative superhelicity. Whereas the DMS protection assays require longer lived changes in a much greater proportion of the DNA molecules, the $KMnO_4$ enhancements might be detected if only a small proportion of the sequences are in an altered conformation for a small amount of time. It was therefore expected that thymines within the single-stranded G-quadruplex loops and in the flanking region between the duplex DNA and the quadruplex structure would show the greatest susceptibility to $KMnO_4$ oxidation. Yet, the results revealed the same modification patterns for both linear and supercoiled plasmids, and surprisingly showed that the potassium cation did not a play significant role. For the four single T-loop samples (Figure 4.12 and Figure 4.13), no reaction with $KMnO_4$ was observed for any of the potential loop regions and all the thymidine residues within the loop regions were completely protected from oxidation. This implies that this region probably adopts a duplex conformation. These results appear to contradict those obtained from the DMS assays, which showed some evidence of quadruplex formation. However, it is still possible that the single-loop thymines may be arranged in a way that still protects the C5-C6 bond from reaction with this probe. Surprisingly, the thymines that immediately flank the G-quadruplex forming sequence showed the most reactivity, with pronounced cleavage at both the 5'- and 3'-ends and in the (GATC) region between dimeric inserts.

Supercoiled DNA molecules have been known to undergo cleavage by single strand specific nucleases (Beard et al., 1973, Lilley, 1980). In a complementary study, enzymatic probing using the single-strand specific S1 nuclease was used to map any single-stranded regions associated with G-quadruplex formation within supercoiled G-rich plasmids. The data presented in this section suggests that under negative supercoiling, all the single T-looped G-rich plasmids, irrespective of sequence length, revealed identical banding patterns. In the supercoiled state S1 nuclease cleaved the DNA in a specific location, evidenced by the fact that further digestion by *Sca*1 produced two distinct DNA fragments. The lengths of the DNA fragments were identical in every sample. After the torsional stress had been relieved from the plasmid, the plasmids were no longer cleaved by S1 nuclease. This indicates that the single-stranded region of DNA was only present under negative supercoiling. However, a

control experiment with the pUC19 vector showed a similar digestion pattern. pUC19 is derived from the bacterial plasmid pBR322 (Yanisch-Perron et al., 1985) and is known to contain an 11 bp palindromic repeat sequence separated by 3 bp. This sequence has been shown to form a cruciform structure (Figure 4.19) as result of the intra-strand base pairing in supercoiled DNA (Lilley, 1980). It is therefore probable that the S1 nuclease sensitive sites detected in this experiment correspond to the cruciform rather than the G-quadruplex structure and that, under negative supercoiling, the formation of the cruciform structure is favoured over G-quadruplex formation. Still, it should be noted that once the S1 nuclease had cleaved the cruciform structure, the excess free energy generated by negative supercoiling will not be available, as the DNA will no longer be supercoiled, which may in turn inhibit the formation of any G-quadruplex structures.

While the DMS methylation assays of the single T-loop plasmids seem to suggest the formation of G-quadruplex structures that are not supercoil-dependent, the $KMnO_4$ reactions did not, although the results appear to show evidence of some structural perturbations within the flanking regions of the cloned fragments that were also independent of negative supercoiling. However, the S1 nuclease analysis appears to conflict with both the chemical probing assays and instead appears to suggest that G-quadruplex and/or any other structural transitions do not occur with the G-rich sequences. Therefore, it is questionable whether G-quadruplex formation had occurred within these DNA molecules.

4.4.1.2 T_4-loops

In sharp contrast to the single T-looped sequences, none of the complementary T_4-looped sequences showed any significant protection from DMS in either the supercoiled or linear states or in the presence or absence of potassium ions. Every band persists indicating that every guanine base within the G-quadruplex forming sequences is accessible for methylation. Although the CD results with this G-strand in Chapter 3 provided evidence of the ability for these sequences to form stable intramolecular antiparallel structures, it appears that negative supercoiling does not induce G-

quadruplex formation in this sequence and that duplex DNA base-pairing is retained. This is not surprising as G-quadruplex structures that contain longer loop sequences assemble slower and form less stable quadruplexes than their shorter looped counterparts (Rachwal and Fox, 2007, Risitano and Fox, 2004).

In contrast, the $KMnO_4$ footprinting assays performed on the T_4-looped G-quadruplex DNA plasmids (Figure 4.14 and Figure 4.15) revealed virtually identical patterns to those of the single T-loop sequences (Figure 4.12 and Figure 4.13). Again, thymines in the adjacent 5'- and 3'- flanking regions were most reactive to $KMnO_4$, whilst none of the T_4-loops showed any sensitivity to the probe.

The lack of modification of the proposed loops is consistent with the DMS results, which indicated no quadruplex formation, though the $KMnO_4$ reactions in the flanking sequences were unexpected, as they imply that there were some structural changes. As longer loop sequences are less compact than single-T loops and can form both edgewise and diagonal loops, in theory this should make them relatively more accessible to $KMnO_4$. But no modification patterns were detected in these regions and therefore appears to corroborate the DMS assays, which also did not indicate any conformational changes within the longer loop sequences. Nevertheless the enhanced reactivity observed in the proximal and distal flanking regions suggests that the nature of the 5'- and 3'- flanking regions is altered so that the thymines are more accessible to $KMnO_4$. Therefore it is possible to speculate that either the inserted G-rich DNA between the two flanking sites cannot efficiently dissociate or that natural supercoiling of the plasmid is not enough to induce and stabilise the G-quadruplex. This is consistent with the findings of Onyshchenko *et al*. who reported that negative supercoiling alone was unable to induce the formation of the G-quadruplex structure without the aid of PNAs (Onyshchenko et al., 2009, Onyshchenko et al., 2011).

The data presented for the S1 nuclease digestions for each T_4-looped sequence, revealed identical banding patterns to those of the single T-loop sequences. In the supercoiled state S1 nuclease cleaved the DNA in a specific location. After the torsional stress had

been relieved from the plasmids, the plasmids were no longer cleaved by S1 nuclease, implying that the single-stranded region of DNA was only present under negative supercoiling. It is believed that the S1 nuclease sensitive sites detected in this experiment correspond to the cruciform structure of the pUC19 vector rather than the G-quadruplex structure and that under negative supercoiling the formation of the cruciform structure is favoured over G-quadruplex formation.

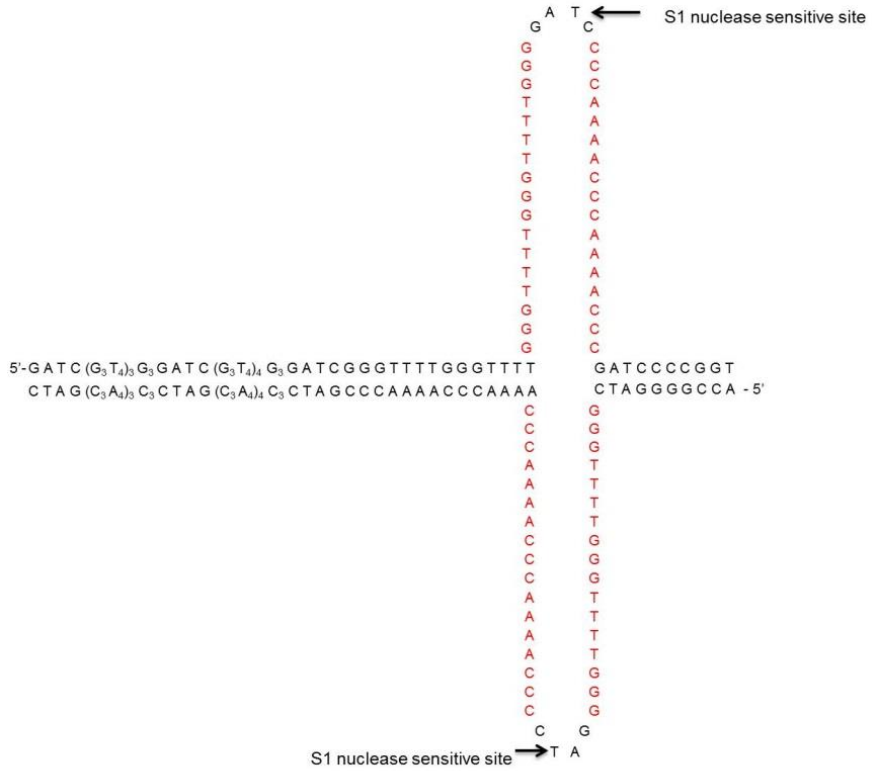
The chemical and enzymatic probing data collected for the T_4-looped sequences strongly suggest that G-quadruplex formation had not occurred within any of the cloned guanine sequences and that the duplex DNA was dominant.

4.4.1.3 $pG_n,4,X_n$

In a separate study a mutant, trimeric quadruplex-forming sequence **$pG_n,4,X_n$** was investigated using the same chemical and enzymatic probing assays. In the DMS assays none of the guanine residues within the quadruplex forming regions were protected from DMS methylation indicating that guanine bases were participating in duplex Watson-Crick base pairing rather than Hoogsteen hydrogen-bonding. Hence, it appeared unlikely that any of the possible three G-quadruplex structures had formed.

However, the results with $KMnO_4$ and S1 nuclease suggest that some degree of structural transition had occurred. Several thymine bases within the G-rich region were modified by $KMnO_4$ in both the supercoiled and the linear plasmids in the absence and presence of KCl. Still, these thymines were exclusively located within the GATC 5'- and 3'- flanking sequences for each of the three structures similar to the single T- and T_4-loop experiments. This signifies a region of locally unwound DNA at the beginning and end of the each potential quadruplex structure. The S1 nuclease mapping of this plasmid identified several regions of single-stranded DNA, which is consistent with the $KMnO_4$ footprinting assays. It was also noticed that these regions were no longer evident once the torsional energy from the negative supercoiling had been eliminated.

Analysis of the S1 nuclease digestion showed that among the fragments corresponding to cleavage at the pUC19 cruciform, the enzyme had cut the plasmid at specific locations close to or within the G-quadruplex-forming motifs. However, this may not necessarily reflect quadruplex formation as the third potential quadruplex structure ends with a 25 bp CA-rich sequence ($GATCC_3A_4C_3A_4C_3A_4$) that is separated by a 4 nucleotide GATC loop sequence which is an inverted repeat of the final G_3-tracts (Figure 4.23). This sequence therefore has the potential to form a G-quadruplex structure and/or a cruciform. Therefore it is plausible that the DNA fragments generated by the S1 nuclease in this experiment do not correspond to G-quadruplex formation, but instead represent cleavage of a G-rich cruciform which is consistent with the DMS analysis.



5'-GATC$(G_3T_4)_3G_3$GATC$(G_3T_4)_4G_3$GATC$(G_3T_4)_2$GGGTTTTGGGTTTTGGGGATCCCCAAAACCCAAAACCCCGATCCCCGGT

Figure 4.23 Formation of a possible cruciform structure from the inverted repeat sequence in the plasmid **p(G_4,4,X_n).** A cruciform structure is predicted to form in the inverted repeat sequence (highlighted in red) at the 3'-end of G-quadruplex forming sequence insert.

4.5 Conclusion

Localised unwinding of DNA is a prerequisite for the formation of G-quadruplex structures in supercoiled DNA. Although the single T-loop sequences appear to suggest the possibility of G-quadruplex formation, collectively the data provided in this chapter does not provide sufficient evidence to support the formation of G-quadruplexes under superhelical stress. However, this is not sufficient to refute this hypothesis as some unusual observations were made during these experiments, although it is possible that some structural perturbations occurred within each of the plasmids, the extent of this remains unclear. Whilst the DMS assays on the G_3T sequences appear to demonstrate some Hoogsteen-style hydrogen-bonding within the G-tracts the G_3T_4 sequences did not. The $KMnO_4$ experiments revealed localised unwinding (increased accessibility of the C5-C6 bond in thymine) in all of the plasmid samples, regardless of DNA sequence or length, though these were only observed at the duplex-quadruplex interface, suggesting that if a quadruplex is formed then their loops may be sterically inaccessible to the probe. The monomeric and dimeric structures appeared to yield similar results suggesting that the presence of a second structure did not have either a stabilising or destabilising effect. The level of protection and reactivity observed by the DMS and $KMnO_4$ chemical probing, leave open the possibility that the conformational transitions within these G-rich sequences generate structures that are not stable and which may be converting between duplex and quadruplex, with the duplex structure being the more dominant. In the S1 nuclease experiments the enzyme preferentially cut at the pUC19 cruciform structure, thus relieving the torsional stress required to induce the G-quadruplex structure. S1 is therefore not an ideal probe for studying the conformation changes in pUC plasmids using agarose gel electrophoresis.

Chapter 5:
Two-Dimensional Gel Electrophoresis of G-Quadruplex Forming Supercoiled DNA Topoisomers

5.1 Introduction

The double-stranded structure of the B-form DNA is usually depicted as a linear, rigid helix. In fact DNA is an incredibly flexible molecule that has the ability to bend, twist and coil within the cell and nucleus. DNA malleability is extremely necessary in packaging the very large molecule into a relative small space within the cell. During DNA replication, the two complementary strands of the double helix require separation in order to act as templates for synthesis of the daughter strands. Localised unwinding of the helix must take place ahead of the replication fork in order for the two strands to separate. However, if the rest of the DNA molecule is unable to unwind, tension (or stress) will occur causing the DNA to coil around itself as the strands are no longer able to separate further, this is known as "DNA supercoiling". Thus, the underwinding or overwinding of DNA results in either negative or positive supercoiling respectively and in all cells, cellular enzymes known as topoisomerases help to relieve the torsional stress generated by supercoiling. Most natural DNA exists in a negatively supercoiled state and a plasmid molecule purified from a bacterial cell will naturally occur as a covalently closed circular DNA molecule which can exist in a number of different topological configurations termed topoisomers (Sinden, 1994). The build-up of torsional energy from negative supercoiling destabilises the Watson-Crick bonds of the double helix, forming regions of locally unwound DNA. It is proposed that the within these regions the torsional energy can promote localised topological changes in the DNA structure. This chapter investigates the topological transitions of DNA topoisomers using two-dimensional gel electrophoresis experimentations to determine whether negative supercoiling can facilitate the formation of G-quadruplex structures.

5.1.1 Topology of Supercoiled DNA and Topoisomers

5.1.1.1 Early Studies

In the 1960s, sedimentation analysis conducted by Vinograd *et al*. (Vinograd et al., 1965) first revealed the supercoiled nature of the small circular DNA of tumour polyoma virus. When the virus was fractioned into two components; I and II. Component I had a higher sedimentation coefficient than component II indicating that component I was either more compact or larger in mass than component II. However, both components were shown to consist of double-stranded DNA with the same molecular weight. Closer examination by electron micrographs (Figure 5.1) revealed that component I consisted of a twisted circular form of DNA known as closed circular DNA (Figure 5.1a), whilst the circular double-stranded polyoma DNA in component II was converted to a less compact circular duplex by the introduction of single-strand breaks known as open circular DNA (Figure 5.1b).

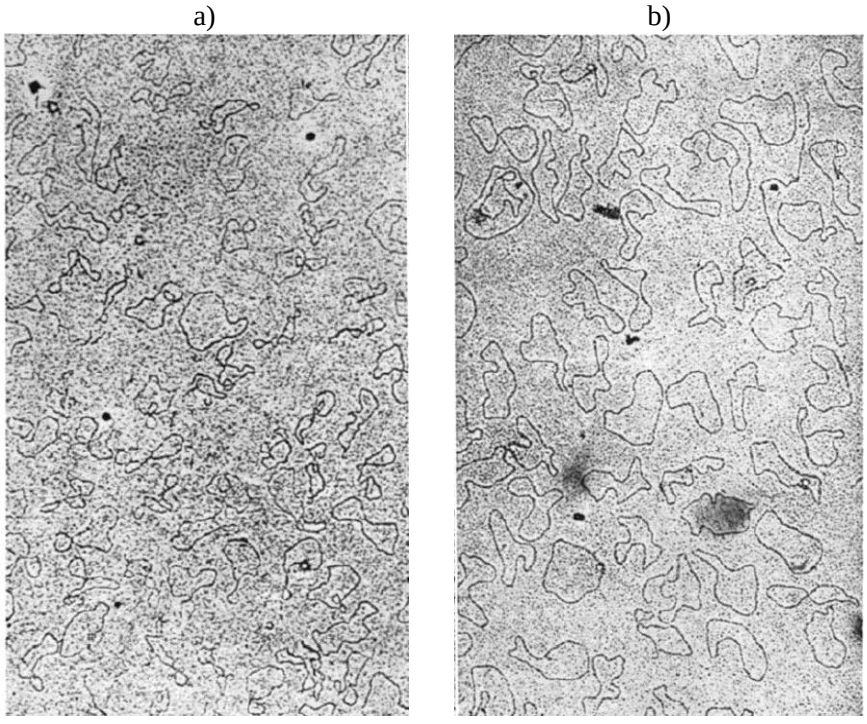


Figure 5.1 Electron micrographs of Polyoma virus DNA. a) Component I and b) Component II. Reported by Vinograd *et al*. (Vinograd et al., 1965).

5.1.1.2 DNA Supercoiling

The B-form of linear duplex DNA exists in a topologically relaxed state of minimum energy and any bending or twisting of the DNA molecule will increase its free energy (Bowater, 2005). The two strands of a linear DNA molecule are repeatedly interwound and have the freedom to rotate around each other (Bowater, 2005, Mirkin, 2001). The free rotation of the linear DNA ends can allow for the unravelling of the two strands. In short DNA sequences this can be achieved relatively easy where both strands can completely dissociate from one another, however, *in vivo* the complete dissociation of longer genomic molecules is unlikely to occur (Bowater, 2005, Mirkin, 2001). Covalently joining the 5'- and 3'-ends of double-stranded linear DNA results in a circular DNA molecule that is typically referred to as covalently closed-circular DNA (cccDNA). The helical nature of the phosphodiester backbone means that upon closure the two strands of the DNA molecule are now topologically linked by the number of helical turns in the original linear DNA molecule and that neither strand can any longer be separated without breaking one or both of the backbone strands (Bowater, 2005, Bates and Maxwell, 2005).

cccDNA may exist in a number of topological isomeric states which are commonly known as topoisomers. Although they have the same chemical formula/number of base pairs and bond activities they differ in the number of times that one strand is linked around the other (Murchie and Lilley, 1992). The topological fundamentals of cccDNA is the linking number (*Lk*). *Lk* is described as the number of times one DNA strand intersects the plane of the other. Therefore in B-DNA this is the number of times one strand of DNA coils (winds) around the other in a right-handed direction. In cccDNA *Lk* is invariant and can only be changed by breaking one or both of the DNA strands. Notably, *Lk* can only be an integer, because circularisation of DNA requires joining of the 5'- and 3'-ends of the linear DNA (Bowater, 2005, Mirkin, 2001, Murchie and Lilley, 1992).

It is important to note that DNA has an inherent number of double-helical turns, which is equivalent to the number of base pairs per DNA molecule (*N*) divided by the number

of base pairs per helical turn (*h*) (which for B-DNA is approximately 10.5 bp per helical turn). *Lk* of a DNA molecule should be measured from the relaxed and unconstrained (hypothetical) state defined as Lk^0 which corresponds to the circular, planar molecule derived from sealing the ends of a linear molecule Equation (1), where Lk^0 is not required to be an integer as it is not the true or actual linking number (Bowater, 2005, Murchie and Lilley, 1992):

$$Lk^0 = N/h \qquad \text{Equation (1)}$$

In cccDNA, the number of helical turns per DNA molecule is referred to as the twist (*Tw*). *Tw* describes how the individual strands of DNA coil around the axis of the DNA helix and is the total number of helical turns in circular DNA under given conditions (Bowater, 2005, Mirkin, 2001). As B-form DNA is a right-handed helical structure with approximately 10.5 bp per helical turn, *Tw* is therefore a large positive number for any natural DNA (Mirkin, 2001). In relaxed DNA the *Lk* is equal to the *Tw*, but once the linear molecule is circularised, the helical nature of DNA allows the double helix to rotate around itself in three-dimensional space known as the writhe (*Wr*). This then forms a "superhelix" (Figure 5.2) i.e. "supercoiled DNA". Therefore, as *Lk* is invariant for cccDNA any change in *Tw* must be compensated by an equal and opposite change in *Wr* and vice versa and *Lk* can further be defined as Equation (2) (Bowater, 2005):

$$Lk = Tw + Wr \qquad \text{Equation (2)}$$

Writhing of the DNA molecule helps to alleviate changes in the helical twist. So the relationship governing the changes in *Lk* number, *Tw* and *Wr* is shown in Equation (3) (Bowater, 2005, Mirkin, 2001). Note, that although *Lk* is an integer and is topologically invariant, neither Tw nor Wr need to be (Mirkin, 2001).

$$\Delta Lk = \Delta Tw + \Delta Wr \qquad \text{Equation (3)}$$

However, if ΔTw is zero then the change in linking number is taken up by an altered Wr Equation (4):

$$\Delta Lk = \Delta Wr \qquad \text{Equation (4)}$$

So, when torsional stress is applied to the linear DNA before ligation via the addition or subtraction of helical turns, the linking number of the resultant circular DNA is changed. This is defined as a linking difference ΔLk. This linking difference can be either positive (DNA being overwound) or negative (DNA being underwound) (Bowater, 2005). Thus, when ΔLk is equal to zero this refers to the DNA being in the relaxed state. However, when ΔLk is not equal to zero then the geometry of the molecule is distorted from the relaxed conformation, which has been defined by Equation (1) (Bowater, 2005).

An example of this is illustrated in Figure 5.2; if a piece of linear DNA has 36 helical turns and is ligated together by the 5'- and 3'-ends, then the linking number for the relaxed cccDNA (Lk^0) will also equal 36. If the DNA is unwound by four turns before circularisation, the length of the DNA is unchanged but Lk is reduced to 32 for both the relaxed and supercoiled state. If the twist is unaltered (ΔTw = 36) then this is accommodated by adopting a writhed supercoiled conformation (ΔWr = -4). However, the writhing can be removed by opening up a section of DNA that is equivalent to 4 helical turns (lower part of Figure 5.2) so that the twist in the remainder of the molecule is now 36. There will of course be an energy penalty as a result of unpairing this section of duplex DNA, but under the right conditions the single stranded "loop" will have the potential to fold into alternative DNA structures, repaying some of the energy penalty.

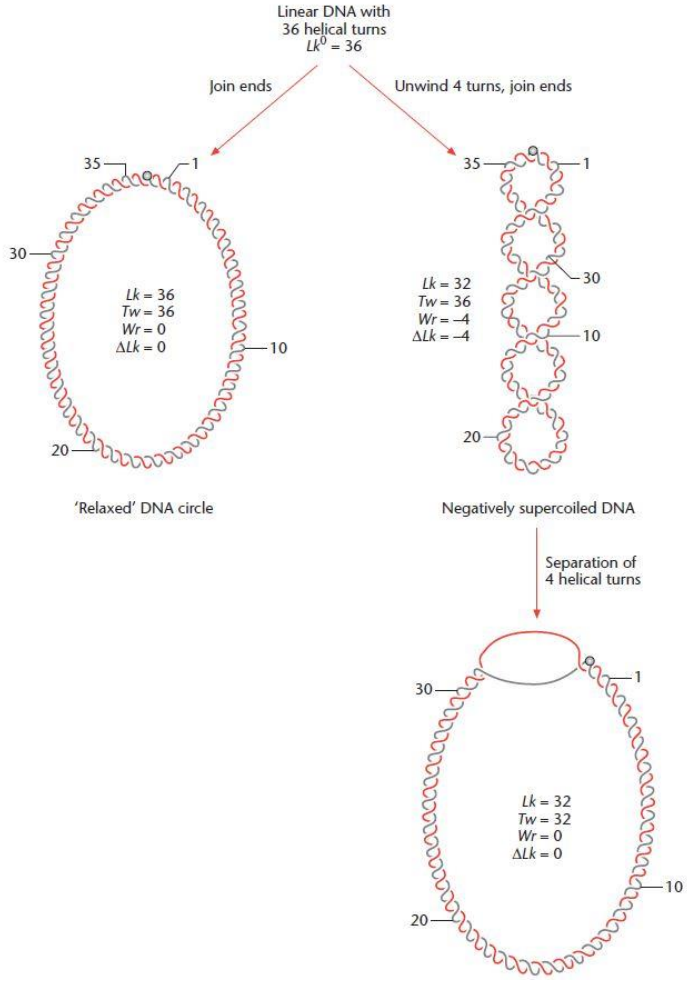


Figure 5.2 Relationship between linking number, twist and writhe for a linear DNA molecule. A linear molecule of DNA of 36 helical turns converted to cccDNA. Taken from Supercoiled DNA: Diagram by R. Bowater (Bowater, 2005)

5.1.2 Topoisomerases

Topoisomerases are a family of enzymes that are capable of catalysing changes in DNA topology, by altering the linking number of DNA, via a mechanism of transient strand breakages and re-ligation of the phosphodiester bonds (Bjornsti and Osheroff, 1999). These enzymes were first discovered by James Wang in the 1970's (Wang, 1971) and are used to regulate the level of DNA supercoiling in cells. Topoisomerases can alter the linking number of DNA by catalysing a three-step process: 1) the cleavage of one or both strands of DNA, during an initial transesterification reaction by forming a covalent linkage between their active site tyrosyl residues and one end of the cleaved DNA

strand, 2) passing a segment of the intact DNA through this break and 3) resealing the DNA breakage in a second transesterification reaction between the free hydroxyl terminus of the noncovalently bound DNA strand and the phosphotyrosine linkage (Bjornsti and Osheroff, 1999). Topoisomerase have been classified in two categories: type I and type II, and are distinguished by the mechanism in which they produce topological changes in DNA, and relieve supercoiling. Type I topoisomerases are ATP independent and catalyse the relaxation of supercoiled DNA by cleaving just one strand of DNA therefore altering the linking number of the DNA by one. Whereas type II are ATP dependent and utilise free energy from ATP hydrolysis to add negative supercoils to DNA by cleavage of both strands and altering the linking number of the DNA by two. Both type I and type II topoisomerases play important roles in DNA replication, transcription and recombination (Roca, 1995).

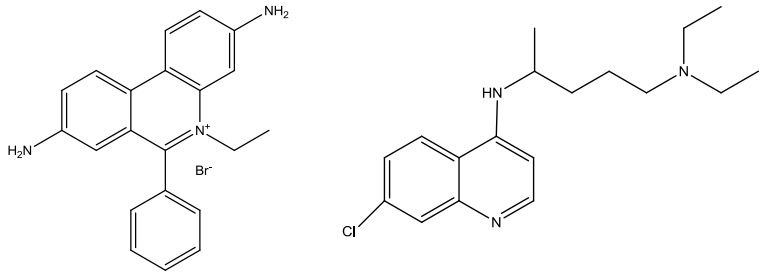
5.1.3 Two-Dimensional Gel Electrophoresis

Two-dimensional gel electrophoresis was first suggested by Wang *et al*. (Wang et al., 1983) as an ideal method for characterising local conformational changes in supercoiled DNA. Unlike one-dimensional gel electrophoresis, where migration is viewed as a function of topology (Bowater et al., 1992) and in which positive and negatively supercoiled DNAs have very similar mobility's, two-dimensional gels can unmask topology-dependent structural transitions resulting from torsional stress in supercoiled plasmids (Mirkin, 2001). Transitions from B-form to Z-DNA (Nordheim et al., 1983, Peck et al., 1982), cruciforms (Panayotatos and Wells, 1981, Lilley, 1980, Gellert et al., 1983) and H-DNA (Htun and Dahlberg, 1989) were amongst the first to be studied allowing other parameters relating to these transitions to be measured from the gels (Bowater et al., 1992). In the first dimension supercoil-driven structural transitions occur under the specific ionic conditions, whilst these are removed in the second dimension, helping to separate and resolve a wider range of topoisomer spots not seen in the first dimension.

5.1.3.1 Principles of two-dimensional gel electrophoresis

A distribution of topoisomers ranging from very negatively supercoiled to slightly positively supercoiled are prepared by incubating aliquots of DNA in varying concentrations of ethidium bromide (Figure 5.3a) and relaxing these with

topoisomerase. Each incubation generates a Gaussian distribution of topoisomers which is determined by the ethidium bromide concentration. A mixture of the topoisomers is prepared and loaded in a single well in the corner of a square agarose gel (typically 20 x 20 cm) and slowly electrophoresed to separate the topoisomers. This is the first dimension. The gel is soaked in a sufficient amount of an intercalating agent, typically chloroquine (Figure 5.3b) to induce local unwinding of the DNA, resulting in a distribution of positively supercoiled topoisomers. The degree of positive supercoiling is dependent on the chloroquine concentration and, as previously discussed, any change in *Tw* must be compensated by a change in *Wr*. Therefore, the gel is rotated 90º and electrophoresed once again in buffer containing an equal concentration of chloroquine. This is the second dimension.



a) Ethidium bromide b) Chloroquine

Figure 5.3 DNA intercalators: Ethidium bromide and Chloroquine. The chemical structure of these compounds makes it possible to alter the geometry of double strand DNA.

Figure 5.4 illustrates a schematic representation of two-dimensional gel electrophoresis in the presence of high and low chloroquine concentrations. The first dimension (D1) illustrates the ladder of topoisomers generated by topoisomerase relaxation in the presence of varying ethidium bromide concentrations. Here, each topoisomer band differs from its neighbour by a linking number of ± 1. The most supercoiled species migrate the fastest, however, in this dimension it is difficult to separate the most highly supercoiled species and they migrate together to form a broad band at the bottom of the gel. In the second dimension, the plasmids that had migrated the fastest in the first dimension, become the least positively supercoiled in the second dimension and migrate the slowest. In the presence of the high chloroquine concentration the first two spots on

the top right hand side of the gel correspond to the topoisomer species that were slightly positively supercoiled in the first dimension, therefore gaining a higher degree of positive supercoiling in the second dimension thus migrating the fastest. The spot at position *r1* corresponds to the species with the lowest migration in the first dimension and is referred to as the point of relaxation where *Wr* is zero in the first dimension, and is described by Equation (5) where *h* is in the absence of chloroquine (Bowater et al., 1992).

$$Lk = N/h \qquad \text{Equation (5)}$$

A higher resolution of topoisomers can be achieved if the gel is run at a lower chloroquine concentration in the second dimension, illustrated in Figure 5.4. The topoisomers with the lowest linking difference run similar to those in the high chloroquine concentration experiments. The spot in position *r1* is unchanged as this is determined by the migration in the first dimension. In this instance as the chloroquine concentration is lower not all the topoisomers will become positively supercoiled and the topoisomers with the highest $-\Delta Lk$ will remain predominantly negatively supercoiled. At low chloroquine concentration a new point of relaxation is introduced (*r2*). The spot at position *r2* corresponds to the point of relaxation in the second dimension, which now has the slowest migration and is described by Equation (6), where *h'* is in the presence of chloroquine (Bowater et al., 1992, Bowater, 2005).

$$Lk = N/h' \qquad \text{Equation (6)}$$

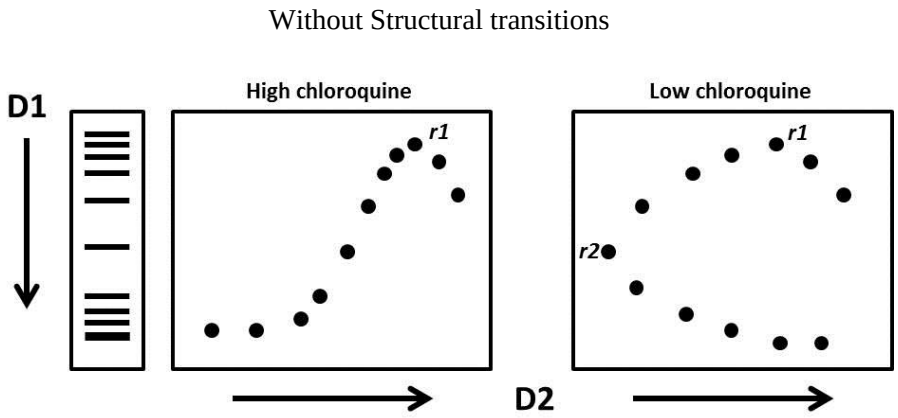


Figure 5.4 Schematic representation of two-dimensional gel electrophoresis of a circular DNA molecule that does not undergo supercoil-dependent structural transitions. Adapted from "Two-Dimensional Gel Electrophoresis of Circular DNA" by R. Bowater (Bowater et al., 1992). On the left-hand-side D1 indicates one-dimensional gel electrophoresis, whilst D2 indicates electrophoresis in the second dimension in the presence of high and low chloroquine concentrations.

As ΔLk and Lk are dependent upon length of a DNA molecule it is more convenient to refer to a normalised measure of supercoiling termed the superhelical density, which is also known as specific linking difference (σ). Thus the resultant banding pattern from these gels makes it possible to calculate σ, which corresponds to the average number of superhelical turns per helical turn of DNA, Equation (7) (Sinden, 1994), where τ is the number of helical turns. The degree of underwinding in naturally occurring cccDNA bacterial plasmid molecules generally falls into the range of σ = -0.05 to -0.06 (5 to 6%) (Bowater, 2005).

$$\sigma = 10.5\ \tau/N \qquad \text{Equation (7)}$$

5.1.3.2 Principles of two-dimensional gel electrophoresis in determining secondary structures

As previously mentioned the formation of alternative DNA structures in negatively supercoiled DNA requires localised unwinding of DNA. In order for a molecule of DNA to undergo localised structural transitions, a threshold value of -ΔLk supercoiling

is required. As the formation of alternative secondary structures involves a change in *Wr*, this can be demonstrated by two-dimensional gel electrophoresis, which is illustrated in Figure 5.5. Above the threshold level of -Δ*Lk* and under the correct experimental conditions alternative secondary structures are free to form. Topoisomers carrying these structures therefore have a reduced writhe and so migrate slower than they otherwise would have done, and are undetectable in the first dimension (D1), indicated by the blue dashed lines in Figure 5.5 (Bowater et al., 1992). It is important to note that it is necessary to run the first dimension under the conditions that favour the structural formation. When the gel is soaked in the chloroquine intercalator and electrophoresed in the second dimension and equilibrated in buffer containing the same concentration of chloroquine, the negative supercoiling is reduced, resulting in a destabilisation of the alternate secondary structure, causing these topoisomers to migrate with their original mobility in the second dimension. This can be seen as a discontinuous "jump" (*j*) which is identified by the blue spots in Figure 5.5 (Bates and Maxwell, 2005, Bowater et al., 1992).

With Structural transitions

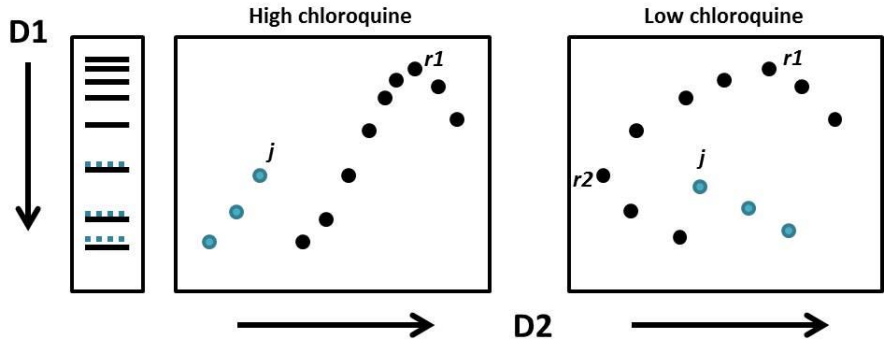


Figure 5.5 Schematic representation of two-dimensional gel electrophoresis of a circular DNA molecule that undergoes supercoil-dependent structural transitions.Adapted from "Two-Dimensional Gel Electrophoresis of Circular DNA" by R. Bowater (Bowater et al., 1992) On the left-hand-side D1 indicates one-dimensional gel electrophoresis, whilst D2 indicates electrophoresis in the second dimension in the presence of high and low chloroquine concentrations. The structural transitions leading to "jumps" (*j*) are indicated in blue.

5.1.4 The Energetics of Supercoiled DNA

Since, DNA will assume a configuration of minimum energy, any bending or twisting will increase its free energy (Bowater, 2005). Any process that decreases the degree of supercoiling is energetically favourable and can be used to drive other structural changes (Bowater, 2005) such as the extrusion of cruciforms (Gellert et al., 1983, Lilley, 1980, Panayotatos and Wells, 1981), Z-DNA (Peck et al., 1982, Wang et al., 1983) and more recently G-quadruplexes (Sun and Hurley, 2009, Onyshchenko et al., 2009, Onyshchenko et al., 2011). The free energy of formation (ΔG_f) of an alternative secondary structure in the presence of TBE buffer can be calculated by Equation (8), where R is the gas constant, T is temperature, L_j is the linking difference of topoisomer j and $L_{j\text{-}t}$ is the linking difference of the unshifted topoisomer of equal mobility.

$$\Delta G_f = 1050RT/N \times (L_j^2 - L_{j\text{-}t}^2) = 1050RT/N \times (\Delta Lk)^2 \qquad \text{Equation (8)}$$

An example of this is the transition of B-DNA to Z-DNA in negatively supercoiled DNA containing segments of alternating deoxyguanosine and deoxycytidine residues. The transition involves the "flipping" of the base pairs to form a left-handed helical structure, where the phosphodiester back-bone follows a zigzag path (Herbert and Rich, 1996). Z-DNA has a higher energy conformation than B-DNA and the torsional stress of negative supercoiling facilitates the formation and stabilisation of the alternative structure within the poly $d(CG)_n$ regions (Peck et al., 1982, Peck and Wang, 1983, Wang et al., 1983). Cruciform formation in palindromic sequences has also been shown to form under negatively supercoiled conditions in a similar manner to the B- to Z-DNA transitions (Lilley and Hallam, 1984, Lyamichev et al., 1983, Vologodskaia and Vologodskii, 1999, Vologodskii and Cozzarelli, 1994). Cruciform formation requires significant rearrangement of the DNA structure which includes the reorganisation of the base-pairing (Palecek, 1991) and the reduction of the interstrand twist number, although the reduction of the twist number is approximately half that of the B-to-Z DNA transitions.

5.2 Experimental Design

Previous chapters have discussed how DNA molecules that contain particular DNA sequences can undergo structural perturbations under negative supercoiling. However, these structural perturbations will only be possible if the plasmid contains sufficient superhelical density to drive the transition. Distributions of topoisomers of plasmids containing the G-rich inserts described in earlier chapters (with sizes ranging from approximately 2700 to 2800 bp (Table 5.1) were prepared by incubating the native plasmids in varying concentrations of ethidium bromide and relaxing with wheat germ topoisomerase I. These were subjected to two-dimensional gel electrophoresis as described above and in Chapter 2.

Table 5.1 List of G-quadruplex forming supercoiled DNA used in the two-dimensional gel electrophoresis experiments.

Name/p(G_n,L,X)	**Sequence**
pG_4,1,1	GATC$(G_3T)_3G_3$GATC
pG_4,1,2	[GATC$(G_3T)_3G_3]_2$GATC
pG_4,4,1	GATC$(G_3T_4)_3G_3$GATC
pG_4,4,2	[GATC$(G_3T_4)_3G_3]_2$GATC
pG_n,4,X_n	[GATC$(G_3T_4)_3G_{3]2}$[GATC$(G_3T_4)_4G_3]_5$GATC$(C_3A_4)_2C_3$GATC

5.3 Results

To assess that the extrusion of G-quadruplex structures in cccDNA is dependent on negative supercoiling, a distribution of topoisomers of pUC19 G-quadruplex forming clones were prepared by treating with wheat germ topoisomerase I in the presence of varying concentrations of ethidium bromide as described in Chapter 2. Intercalation by ethidium bromide introduces concentration-dependent positive supercoils, which after removal of the ligand revert back to being negatively supercoiled. Resolution of the DNA topoisomers of different linking numbers was conducted by two-dimensional gel electrophoresis also described in Chapter 2. The topoisomers with sufficient torsional energy to facilitate the extrusion of the G-quadruplex structures was expected to migrate at a reduced mobility in the first dimension, but migrate normally after the addition of chloroquine in the second dimensional. Each clone represented below contained the sequence G_n,L,X, where: G_n is the number of consecutive G_3-tracts, L is the number of thymines in the loop and X is the number of repeats of the G-quadruplex forming sequence e.g. G_4,4,2 = $[GATC(G_3T_4)_3G_3]_2$.

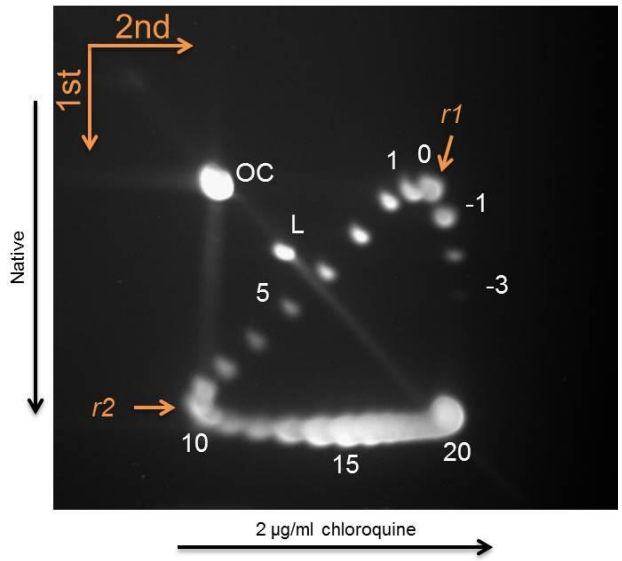
The majority of experiments were conducted under low KCl concentrations (1 mM), whilst others were repeated under a higher concentration (10 mM) where stated. Please note that ionic concentration did not affect elucidation of structural perturbations, therefore only the clear gels are presented in this thesis.

5.3.1 Two-Dimensional Gel Electrophoresis of Plasmid pUC19

Figure 5.6 illustrates the 2D gel electrophoresis of the pUC19 control supplemented with low (1 mM) and high (10 mM) KCl concentrations. Analysis of Figure 5.6a show that the topoisomers form a smooth curve with increasing superhelical density, where according to Equation (7) a change of 1 in linking number corresponds to a change of 0.004 (*i.e.* 10.5 x 1/2686) in the superhelical density for the plasmid size 2686. In the first dimension spots -3 to -1 (on the right-hand-side of the arc) are positively supercoiled, spot 0 corresponds to *r1* the position of relaxation in the first dimension in the absence of chloroquine, whilst spots 1 to 20 (left-hand-side of the arc) are negatively supercoiled. In the second dimension, spots -3 to -1 are more positively supercoiled,

spots 1 to 8 are positively supercoiled, spot 9 (*r2*) is the position of relaxation in the second dimension in the presence of chloroquine and spots 10 to 20 are negatively supercoiled. In this control experiment no structural transitions were observed.

a)



b)

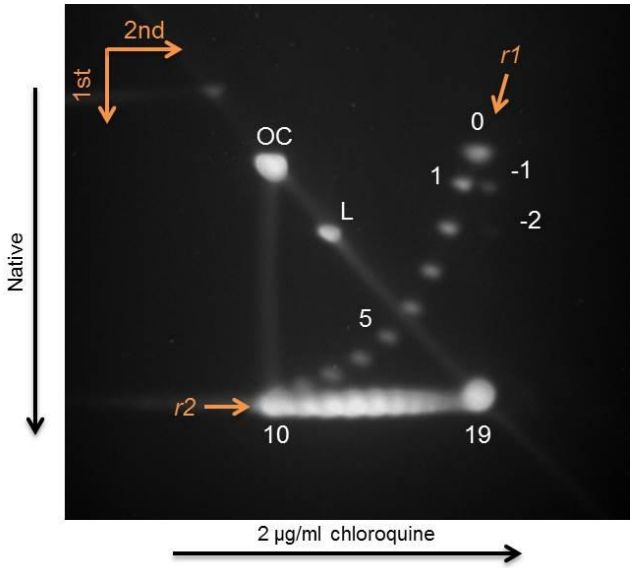
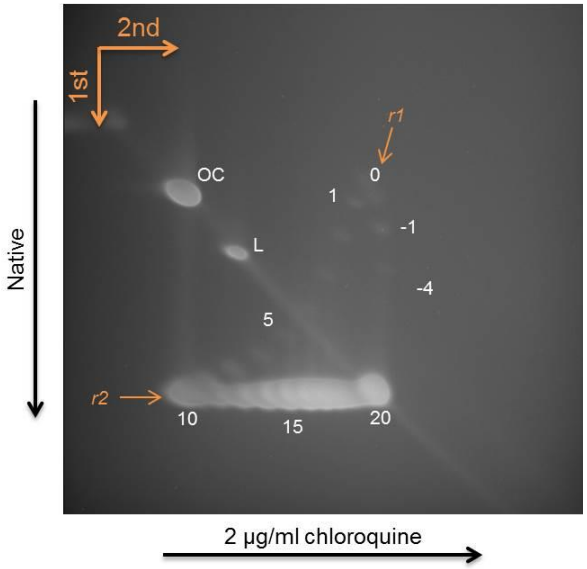


Figure 5.6 Separation of pUC19 DNA topoisomers by two-dimensional gel electrophoresis. The experiment was conducted on a 1% agarose gel supplemented with a) 1 mM KCl and b) 10 mM KCl. Topoisomers -1 to -3(or -2) have positive supercoiling, whilst 1-20(or -19) are negative supercoiled in the first dimension. In the top left hand corner a well was loaded with a distribution of topoisomers and electrophoresis was performed in the first dimension using a 1x TBE running buffer supplemented with 1 mM KCl. The gel was placed in a dark container and soaked in TBE buffer supplemented with chloroquine (2 µg/ml) to intercalate the double helix for approximately 7 hours. Then electrophoresis in the second dimension (from left to right) was conducted using the same 1x TBE running buffer supplemented with 2 µg/ml of chloroquine. The spot in the top left corner corresponds to the open circular DNA (OC), whilst the spot just below it corresponds to linear DNA (L).

5.3.2 Two-Dimensional Gel Electrophoresis of plasmid pG_4,1,1 and pG_4,1,2

Figure 5.7 illustrates the 2D gel electrophoresis of the plasmids **pG_4,1,1** (monomeric structural sequence) and **pG_4,1,2** (dimeric structural sequence) containing G_3T repeat sequences. Each topoisomer increases periodically with increasing superhelical density to form a smooth curve. In the first dimension spots -4 to -1 are positively supercoiled, spot 0 corresponds to *r1* the position of relaxation in the first dimension in the absence of chloroquine, spots 1 to 20 (left-hand-side of the arc) are negatively supercoiled. In the second dimension, spots -4 to -1 are more positively supercoiled, spots 1 to 8 are positively supercoiled, spot 9 (*r2*) is position of relaxation in the second dimension in the presence of chloroquine and spots 10 to 20 are negatively supercoiled. In this experiment it was expected that the G-rich segment of **pG_4,1,1** and **pG_4,1,2** could relax approximately 1.4 and 3.2 negative supercoils respectively however, no discontinuous jumps were seen indicating that no conformational transitions had taken place for either of these plasmids suggesting that supercoil-dependent G-quadruplex formation had not occurred.

a)



b)

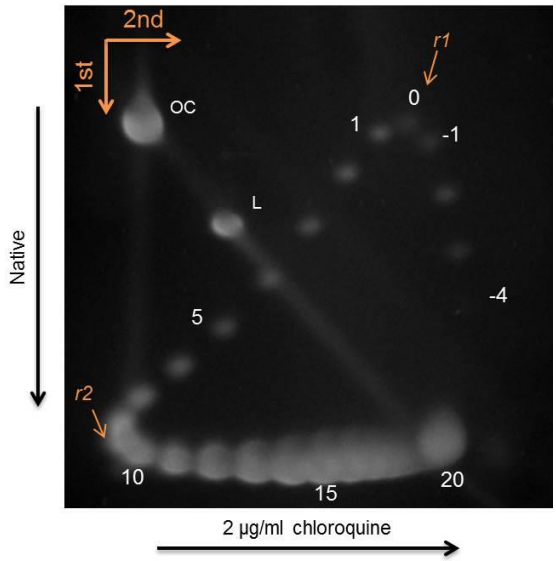
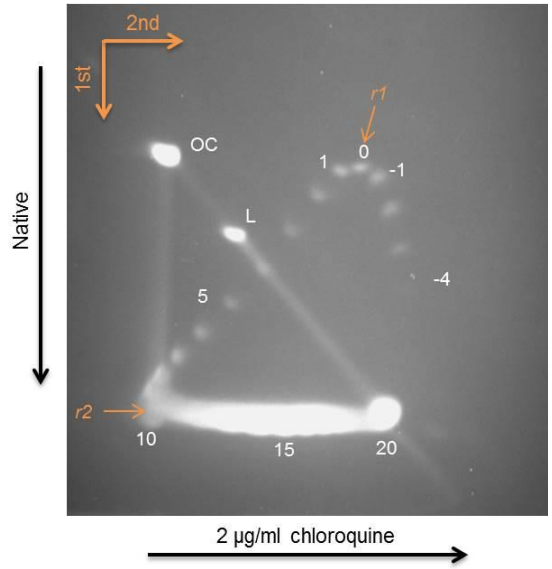


Figure 5.7 Separation of a) **pG$_4$,1,1** and b) **pG$_4$,1,2** DNA topoisomers by two-dimensional gel electrophoresis. Topoisomers -1 to -4 have positive supercoiling, whilst 1-20 are negative supercoiled in the first dimension. In the top left hand corner a well was loaded with a distribution of topoisomers and electrophoresis was performed in the first dimension using a 1x TBE running buffer supplemented with 1 mM KCl). The gel was placed in a dark container and soaked in TBE buffer supplemented with chloroquine (2 µg/ml) to intercalate the double helix for approximately 7 hours. Then electrophoresis in the second dimension (from left to right) was conducted using the same 1x TBE running buffer supplemented with 2 µg/ml of chloroquine. The spot in the top left corner corresponds to the open circular DNA (OC), whilst the spot just below it corresponds to linear DNA (L). The experiment was conducted on a 1% agarose gel supplemented with a) 10 and b) 1 mM KCl respectively.

5.3.3 Two-Dimensional Gel Electrophoresis of plasmid pG_4,4,1 and pG_4,4,2

Figure 5.8 illustrates the 2D gel electrophoresis of the G_3T_4 plasmids p**G_4,4,1** and p**G_4,4,2** for the monomeric and dimeric structural sequences respectively. As seen in the pUC19 control and the G_3T sequences, the mobility of each topoisomer increases steadily with increasing superhelical density to form a smooth curve. In the first dimension spots -4 to -1 are positively supercoiled, spot 0 corresponds to *r1* the position of relaxation in the first dimension in the absence of chloroquine, spots 1 to 20 (left-hand-side of the arc) are negatively supercoiled. In the second dimension, spots -4 to -1 are more positively supercoiled, spots 1 to 8 are positively supercoiled, spot 9 (*r2*) is position of relaxation in the second dimension in the presence of chloroquine and spots 10 to 20 are negatively supercoiled. Although it was expected that the G-rich segment of **pG_4,4,1** and **pG_4,4,2** could relax approximately 2.3 and 5.0 negative supercoils respectively, again no discontinuous jumps were seen in this experiment indicating that negative supercoiling had not induced any structural transitions in this DNA molecule.

a)



b)

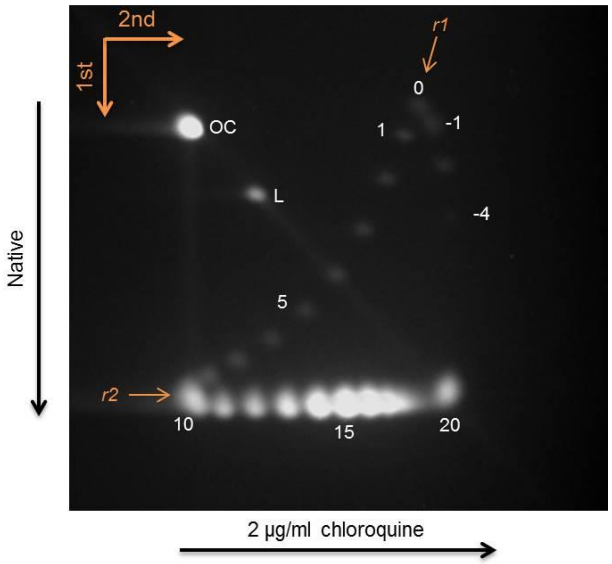


Figure 5.8 Separation of a) **pG$_4$,4,1** and b) **pG$_4$,4,2** DNA topoisomers by two-dimensional gel electrophoresis. Topoisomers -1 to -4 have positive supercoiling, whilst 1-20 are negatively supercoiled in the first dimension. In the top left hand corner a well was loaded with a distribution of topoisomers and electrophoresis was performed in the first dimension using a 1x TBE running buffer supplemented with 1 mM KCl (from top to bottom). The gel was placed in a dark container and soaked in TBE buffer supplemented with chloroquine (2 µg/ml) to intercalate the double helix for approximately 7 hours. Then electrophoresis in the second dimension (from left to right) was conducted using the same 1x TBE running buffer supplemented with 2 µg/ml of chloroquine. The spot in the top left corner corresponds to the open circular DNA (OC), whilst the spot just below it corresponds to linear DNA (L). The experiment was conducted on a 1% agarose gel supplemented with a) 1 mM and b) 10 mM KCl respectively.

5.3.4 Two-Dimensional Gel Electrophoresis of plasmid $pG_n,4,X_n$

The plasmid clone **$pG_n,4,X_n$** consists of 115 bp G_3T_4 sequence flanked by a GATC sequence at both the 5'- and 3'- ends (total 123 bp). As this clone consists of approximately three G-quadruplex forming motifs it was expected that the torsional energy generated by the native negative supercoiling of the plasmid would be sufficient to induce at least one or more of the quadruplex structures. The results of the two-dimensional gel electrophoresis for this plasmid are shown in Figure 5.9a, whilst Figure 5.9b illustrates a schematic representation of the results. The two-dimensional electrophoretic analysis of **$pG_n,4,X_n$** show an unusual pattern in bands 10-20, which appears as if the DNA sample has a topologically mixed populations; one with no discontinuity as for pUC19, and two others that possessed reduced mobilities in the first dimension, generating discontinuities around ΔLk = -10. A similar pattern was seen when the experiment was repeated using a higher concentration of chloroquine (10 µg/ml) in the second dimension (Figure 5.10). Since no structural transitions were observed in the pUC19 control, it can be concluded that the breaks in the pattern have occurred within the cloned G-rich region. This appears to indicate that the sample contains a mixed population of topoisomers which comprises of a plasmid that has not undergone any structural conformational transitions and is indicated by the black circles in Figure 5.9b, whilst the second and third populations which are represented by the blue and red circles in Figure 5.9b display sudden discontinuity "jumps". These jumps indicate a conformational transition induced by negative supercoiling within the G-rich insert in some of the more negatively supercoiled topoisomers. These structures reduce the negative superhelicity of the topoisomers resulting in retardation of electrophoretic mobility in the first dimension. After binding with chloroquine these topoisomers revert back to the B-form configuration and when electrophoresed in the second dimension, the electrophoretic mobilities are similar to that of the pUC19 control and therefore migrates as "normal". However, the two jumps imply the presence of two possible types of supercoil-dependent DNA structures. Analysis of the lower chloroquine concentrated gel (Figure 5.9) reveal that as the linking number progressively decreases, the topoisomers follow a smooth curve until topoisomer 10 which contains 10 negative supercoils (i.e. ΔLk = -10). In the first population (black circles in Figure 5.9b) the topoisomers continue to follow this smooth curve and displays a curve identical to that of the pUC19 vector, where in the first dimension spots -4 to -1 are positively

supercoiled, spot 0 corresponds to *r1* the position of relaxation in the first dimension in the absence of chloroquine, spots 1 to 20 (left-hand-side of the arc) are negatively supercoiled. In the second dimension, spots -4 to -1 are more positively supercoiled, spots 1 to 8 are positively supercoiled, spot 9 (*r2*) is position of relaxation in the second dimension in the presence of chloroquine and spots 10 to 20 are negatively supercoiled. In the second and third case the topoisomers follow a smooth curve as the linking number decreases until a sharp break in mobility is observed around topoisomers 11' (blue circle) and 12" (12", red circle) with 11 and 12 negative supercoils (ΔLk = -11 and ΔLk = -12 respectively) characterised by the discontinuity in the pattern. The presence of two jumps may correspond to the polymorphic potential of the G-rich insert which can potentially form several structural motifs. Whilst the first G-quadruplex forming region requires all four G_3-tracts to participate in the tetramolecular scaffold, the second and third regions have five G_3-tracts and therefore have the freedom to form a number of isomeric assemblies with any combination of four out of the five tracts. In the first dimension Topoisomer 11' migrates almost the same as a topoisomer with approximately 8 negative supercoils, which represents a loss of almost 3 helical turns and corresponds to approximately 31 bp being unwound. This is because the difference in gel electrophoretic mobilities of the two topoisomers reflects the difference in writhing number of their helical axis curve rather than difference in the way that the base pairs rotate along the axis (Wang et al., 1983, Peck and Wang, 1983), therefore $\Delta Tw = \Delta Lk$. So, for topoisomer 11':

$$\Delta Tw_{11} - \Delta Tw_{8} = \Delta Lk_{11} - \Delta Lk_{8} = 3 \text{ helical turns lost}$$

Similarly, topoisomer 12'' migrates as a topoisomer containing 7 negative supercoils, which results in a loss of 5 helical turns and the unwinding of approximately 52 bp.

As superhelical density is increased the DNA unpairing continues until topoisomer 20 is reached which has a superhelical density of ~ -0.07. At ΔLk = -20' approximately 126 base pairs are unpaired (*i.e.* (20 - 8 x 10.5), whilst at ΔLk = -20" approximately 136 base pairs are unpaired, which is consistent with the formation of all three G-quadruplex

structures including the duplex-quadruplex flanking regions. However, it is important to note that the terminal G-quadruplex structure is flanked by a C-rich sequence. This sequence has three C_3-tracts that are separated by A_4 loops, forming an inverted repeat of the final three G_3-tracts, therefore it is possible that at this location the 3'end of this insert can also form a cruciform structure containing a GATC loop. It is then probable that these calculations may include mixed G-quadruplex-cruciform formations.

Further analysis of these results suggest that extrusion of the secondary structure(s) occurs at superhelical density when $\sigma > -0.04$, which means that 4% of the helical turns present in the B-form DNA has been removed, which is in agreement with the theoretical calculated superhelical density of -0.04 for the negative supercoiling induced unwinding of the 123 bp G-rich insert (*i.e.* 123/2686 +123). Although in their study, Peck and Wang (Peck and Wang, 1983) report that the formation of Z-DNA in supercoiled plasmids had occurred in all the topoisomers with superhelical density above the threshold value, the analysis of $\mathbf{pG_n,4,X_n}$ suggests that extrusion of the secondary structures with 11 or 12 negative supercoils did not.

In the first dimension Lk^o lies at topoisomer 0 (*r1*). The extrusion of the unknown secondary structure(s) occurs between topoisomer 10 and 11'/12", so the transition from B-DNA occurs at ΔLk -11' and ΔLk -12" respectively. The length of the full G-rich insert plus the 5'- and 3'-end d(GATC) flanking regions is 123 bp. Therefore, the free energy of formation (ΔG_f) for untwisting the DNA in this region and the formation of the possible three quadruplex structures can be calculated according to Equation (8). Therefore, (ΔG_f) for 11' (blue circles) = 52.8 KJ mol^{-1} (12.6 kcal mol^{-1}) and for 12" (red circles) = 66.9 KJ mol^{-1} (15.9 kcal mol^{-1}).

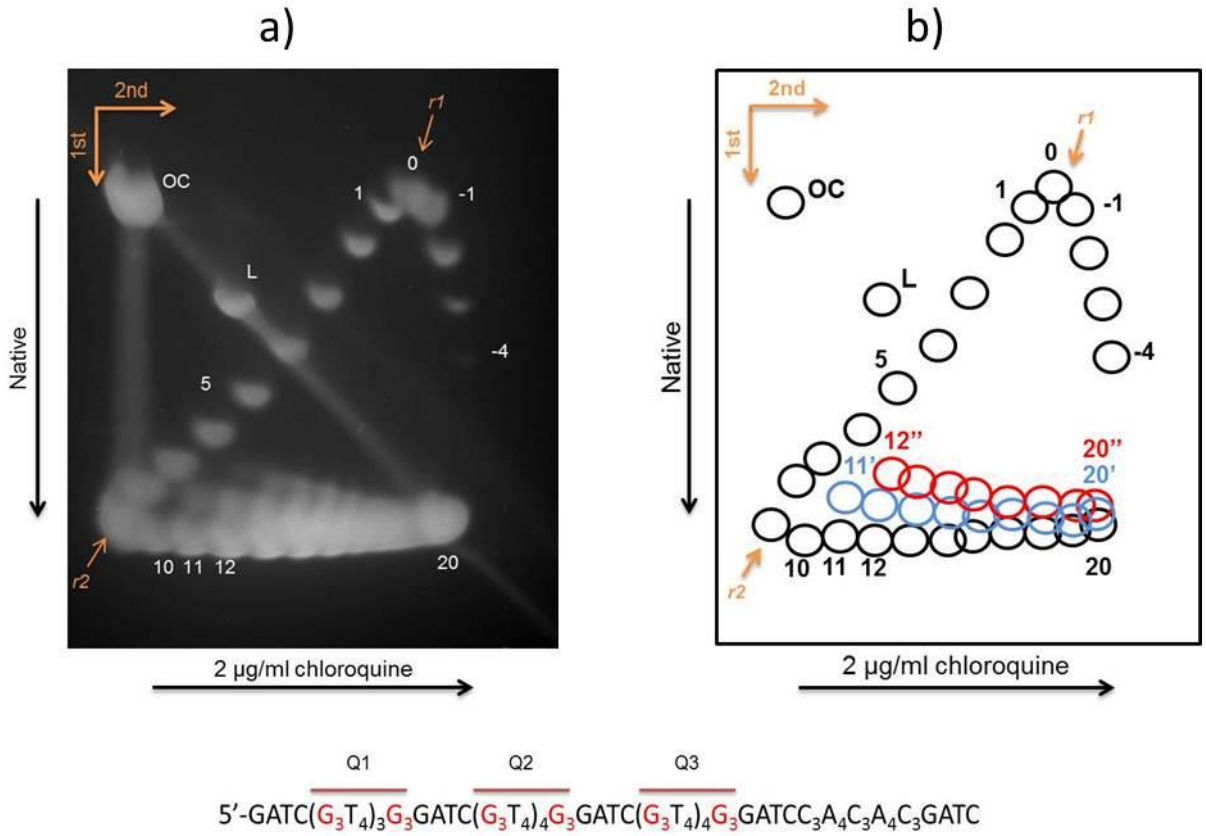


Figure 5.9 Separation of **$pG_n,4,X_n$** DNA topoisomers by two-dimensional gel electrophoresis. a) Positions 11' and 12" signify the discontinuity jumps. In the top left hand corner a well was loaded with a distribution of topoisomers and electrophoresis was performed in the first dimension using a 1x TBE running buffer supplemented with 1 mM KCl (from top to bottom). The gel was placed in a dark container and soaked in TBE buffer supplemented with chloroquine (2 µg/ml) to intercalate the double helix for approximately 7 hours. Then electrophoresis in the second dimension (from left to right) was conducted using the same 1x TBE running buffer supplemented with 2 µg/ml of chloroquine. The spot in the top left corner corresponds to the open circular DNA (OC), whilst the spot just below it corresponds to linear DNA (L). The experiment was conducted on a 1% agarose gel supplemented with 1 mM KCl. b) Schematic representation. The blue and red circles highlight the discontinuity jumps. Q1, Q2 and Q3 represent the first, second and third potential quadruplex forming regions respectively.

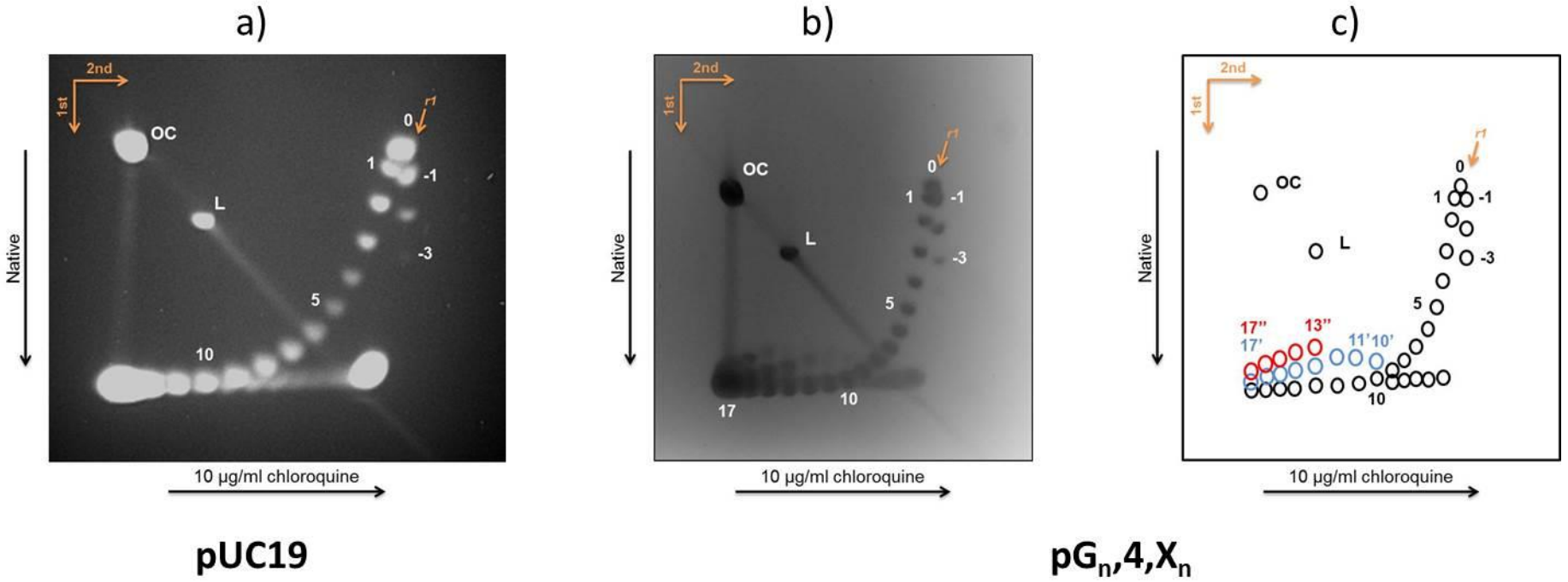


Figure 5.10 a) Separation of a) **pUC19** and b) **pG$_n$,4,X$_n$** DNA topoisomers by two-dimensional gel electrophoresis at high chloroquine concentration. . Positions 10' and 13" signify the discontinuity jumps. In the top left hand corner a well was loaded with a distribution of topoisomers and electrophoresis was performed in the first dimension using a 1x TBE running buffer supplemented with 1 mM KCl (from top to bottom). The gel was placed in a dark container and soaked in TBE buffer supplemented with chloroquine (10 µg/ml) to intercalate the double helix for approximately 7 hours. Then electrophoresis in the second dimension (from left to right) was conducted using the same 1x TBE running buffer supplemented with 2 µg/ml of chloroquine. The spot in the top left corner corresponds to the open circular DNA (OC), whilst the spot just below it corresponds to linear DNA (L). The experiment was conducted on a 1% agarose gel supplemented with 1 mM KCl. c) Schematic representation of pG$_n$,4,X$_n$ separation. The blue and red circles highlight the discontinuity jumps.

5.4 Discussion

Two-dimensional agarose gel electrophoresis is routinely used to detect localised structural transitions in supercoiled DNA. The effects of intercalators (typically chloroquine) on the mobility of topoisomers on an agarose gel that is run in two dimensions can be used to resolve a wider range of topoisomers and reveal any supercoil-dependent structural changes (Bates and Maxwell, 2005). Wang *et al*. (Wang et al., 1983) were one of the first groups to successfully use this technique to investigate the negative supercoiling induced B-Z transition of an alternating $d(CG)_{16}$ clone. The interconversion between the right-handed B-form DNA of the CG insert to the left-handed Z-form was visualised as a sharp break in the smooth curve that is typical of topoisomer separation.

In the previous chapter (Chapter 4) chemical and enzymatic probing was used to identify specific DNA bases involved in G-quadruplex formation of natively supercoiled G-rich plasmids. However, in this chapter, DNA migration as a function of topology to determine topological energetics required for this transition was investigated using two-dimensional gel electrophoresis.

The DNA topoisomers of varying superhelical densities were prepared by incubating the G-quadruplex forming monomeric and dimeric G_3T and G_3T_4 plasmids **G_4,1,1**, **G_4,1,2**, **G_4,4,1** and **G_4,4,2** with wheat germ topoisomerase I in the presence of varying concentrations of ethidium bromide before subjecting them to two-dimensional electrophoresis. By making comparisons to the parental pUC19 plasmid, it was therefore possible to identify any structural transitions that occurred within the G-rich insert. As the linking number of each topoisomer decreased for each T- and T_4-loop G-rich DNA sample a smooth curve progression in the spot patterns was observed. No discontinuity jumps in the pattern were seen and each experiment revealed an almost identical pattern to that of the pUC19 control. These results suggest that none of the G-quadruplex forming DNA had undergone any structural transitions and instead appeared to remain in the B-form configuration. To some extent, these results conflict with the data collected by DMS and $KMnO_4$ chemical probing in Chapter 4 assays. For every

G_3T or G_3T_4 monomeric or dimeric structural sequence investigated using $KMnO_4$ probing, hyperreactivity towards the thymines only occurred within the flanking regions of the sequences and not the loops. This suggests that, although the loops of the quadruplex structures were undetected the G-rich inserted segments may have undergone some structural transitions. Alternatively, in the DMS assays where protection from methylation on guanine bases indicates the presence of Hoogsteen-hydrogen bonds and G-quadruplexes, the data showed that although the guanine residues in the G_3T_4 sequences exhibited hyperreactivity towards the DMS, the guanines in single T-looped, G_3T sequences were protected indicating the presence of quadruplex formation, which was not evident by the two-dimensional electrophoretic analysis.

In contrast, two-dimensional gel analysis of the plasmids **G_4,1,1**, **G_4,1,2**, **G_4,4,1** and **G_4,4,2** appear to coincide with the S1 nuclease assays. Whilst, the enzyme was able to identify and cleave single-stranded regions in each of the four clones the cleavage fragments generated were identical to that of the pUC19 control. These results suggest that the enzyme was identifying a single-stranded region of a supercoil-dependent structure at a locale outside of the G-rich inserts and as pUC19 was derived from the plasmid pBR322 (Yanisch-Perron et al., 1985) it was therefore concluded that 11 bp inverted repeat sequence that forms a cruciform structure (an artefact of the pBR322 vector) was being preferentially cleaved, although this was neither detected in the pUC19 nor the G-rich clones in the two-dimensional gel experiments.

Therefore, some possible explanations as to why no structural transitions were detected by the two-dimensional gel analysis can be made. Firstly, it may be concluded that the chemical and enzymatic probes have a greater sensitivity to structural changes than two-dimensional gel electrophoresis analysis, particularly as probing with DMS, $KMnO_4$ and S1 nuclease demonstrated that each DNA sample had experienced some structural transitions that may or may not directly correspond to G-quadruplex formation.

A second reason is that G-quadruplex formation and stabilisation are highly dependent on the presence of a monomeric cation. A significant amount of G-quadruplex research

is carried out in the presence of high KCl concentration, typically 100 mM. The capabilities of this experiment meant that electrophoresis in the first dimension could only be conducted at a low salt concentration, therefore both the gel and running buffer were supplemented with just 1mM KCl (or 10 mM where stated). Nevertheless, this contrasts with the CD analysis (Chapter 3) and the chemical probing experiments (Chapter 4) of the G_3T sequences which were able to form G-quadruplex structures in the absence of salt and without the driving force of free energy from negative supercoiling.

It is also possible that the agarose gel used to resolve the negatively supercoiled topoisomers in the first dimension may be limited to a range of approximately 20 and topoisomers with more than -20 superhelical turns which correspond to a superhelical density of greater than 0.08 may therefore not be resolved. However, a more plausible explanation for this is that in the DMS and $KMnO_4$ probing experiments in particular with the intramolecular G_3T sequences, localised structural changes were identified not only in the supercoiled plasmids, but also the linear. It is possible that these structures may have been generated by the alkaline denaturing of the plasmids during the extraction and preparation process. Therefore once the G-rich strand had been dissociated from its complementary C-strand, it is then free to assemble into the G-quadruplex structures. Once formed, these structures are extremely thermodynamically stable, where quadruplexes with all single thymine loops have been shown to exhibit melting temperatures in excess of 90 ˚C under physiological-like ionic conditions (~100 mM KCl) (Risitano and Fox, 2003b). Subsequently, each $(G_3T)_n$ plasmid will contain the structurally altered G-rich region before experimental analysis. Therefore, as two-dimensional gel electrophoresis determines supercoiled dependent structural transitions as function of topology the topoisomers containing alternative DNA structures will migrate undetected as they are not supercoil-dependent and the spot pattern will appear as “normal”.

However, it is also important to mention that although the G_3T G-quadruplex structures may have been conserved after the plasmid extraction and preparation process, the S1 nuclease probing of the G_3T and G_3T_4 as well as the pUC19 control all exhibit

supercoil-dependent cleavage at the same position by the S1 enzyme, at the 11 bp pUC19 cruciform structure. Wang *et.al* (Wang et al., 1983) offer two explanations for this: they suggest that the agarose gel used to resolve the negatively supercoiled topoisomers in the first dimension is limited to a specific range due to the superhelical density of the plasmid. Therefore topoisomers with ΔLk = -11 or higher, which corresponds to a superhelical density of ~ -0.04 are not resolved and will not be detected on the gel. As most negatively supercoiled plasmids have a super helical density between -0.05 and -0.06 (Bowater, 2005) and the S1 nuclease experiments were conducted at around these values, it is unlikely that the cruciform structure will not be detected by two-dimension gel electrophoresis as a superhelical density threshold value of ~ 2.00 (10.5 x 2/11bp, where 2 is the minimum number of helical turns that needs to be unwound for the cruciform formation) is required. Another explanation is that the S1 nuclease may have formed an enzyme-substrate complex during experiment that may have inadvertently induced the formation of the cruciform structure.

The anomalous G-quadruplex forming plasmid **$pG_n,4,X_n$** was also subjected to two-dimensional gel electrophoresis. This trimeric quadruplex-forming plasmid consists of 115 bp G-rich insert where each potential quadruplex structure is separated and flanked by GATC sequence. Whilst the first structure consists of four G_3 repeats all of which must be included in the scaffold, the second and third structures are more ambiguous as they contain five G_3 repeats. Depending on their folding topologies the structures may have either a 5' (G_3T_4) or 3' (T_4G_3) overhang or the central G_3-tract may be included in an extended T_4GGGT_4 loop. As this clone consists of a longer G-rich region, it was expected that the torsional energy generated by the native negative supercoiling of the plasmid would be sufficient to induce at least one or more of the quadruplex structures.

The electrophoretic analysis revealed several different patterns indicating that the DNA sample contained a mixed population of non-interconverting topoisomers. The first population followed one continuous smooth curve in progression of topoisomers identical to the pUC19 control which is depicted by the black circles in the schematic representation in Figure 5.9b. This suggests that this population of topoisomers had not experienced any supercoiled-dependent structural transitions. However, two separate

discontinuity jumps were also observed indicating that some of the DNA topoisomers possessed stably unwound conformations that were retarded in mobility in the first dimension. The presence of more than one discontinuity jump at separate superhelical densities both of which were σ = -0.04 and is consistent with the unwinding of at least 123 bp of the G-rich duplex DNA and the assembly of the three G-quadruplex structures. These results appear to corroborate the data from the S1 nuclease probing experiments in Chapter 4 where multiple cleavage fragments were generated suggesting the formation of multiple structures.

Whilst the two-dimensional gel analysis of the monomeric and dimeric G_3T_4 plasmids; **G_4,4,1** and **G_4,4,2**, did not identify any B-DNA to G-quadruplexes transitions, it can be speculated that sharp jumps observed for trimeric plasmid **pG_n,4,X_n** were an effect of the palindromic d(pGpC) sequence at the 3'-end of the insert rather than the length of the sequence.

It is believed that the localised unwinding of the palindromic DNA in this region and the formation of the cruciform was enough to promote further melting of the G-rich DNA and induce the formation of the remaining two quadruplex structures. This is analogous to a study conducted on a cloned plasmid containing the G-rich segment of the *Bcl-2* gene (Onyshchenko et al., 2009). In this study, it was reported that under negative supercoiling and in the presence of either K^+, Na^+ and/or telomestatin the G-quadruplex formation was also inhibited. This is consistent with the results presented in this thesis for the chemical and enzymatic probing and two-dimensional analysis of the monomeric and dimeric G_3Tand G_3T_4 plasmids. However by using PNAs (peptide nucleic acids) that specifically bind to the C-rich strand, they were then able to release the G-rich strand and induce G-quadruplex formation. Which appears to be consistent with the results obtained for the p**G_n,4,X_n**, where in order for G-quadruplex-formation to take place, localised strand separation aided by factors other than negative supercoiling must first occur. They also found that quadruplex formation was essential for the effective binding of the PNAs and vice versa (Onyshchenko et al., 2009, Onyshchenko et al., 2011).

It is important to note that as the two-dimensional gel electrophoresis cannot provide structural analysis, it is impossible to identify what structures and how many were formed within the 123 bp insert. However, the free energy of formation (ΔG_f) required for the untwisting of the DNA in this region including the terminal GATC 5'- and 3' flanking sequences and the simultaneous formation of multiple structures (*i.e.* three quadruplexes or two quadruplexes plus a cruciform) was calculated at 52.8 KJ mol^{-1} (12.6 kcal mol^{-1}) 11' (blue circles) and 66.9 KJ mol^{-1} (15.9 kcal mol^{-1}) for 12" (red circles). To date, there do not appear to be any prior studies using this type of analysis to identify the extrusion of G-quadruplex structures in supercoiled DNA therefore a direct comparison cannot be made at this time. However, there have been several calculations made for the change in free energy of supercoiling required for the transition of B-DNA to alternative secondary structures and in an early study conducted by Wang *et al*. (Wang et al., 1983), the change in free energy for the B- to Z-DNA transition of a 32 bp d(CG) segment of a DNA was calculated to be 17.0 kcal mol^{-1} which is slightly higher than those calculated in this study, although it is important to note that these calculations do not include the B-Z junctions. These have been calculated separately and were estimated to be ~ 4 kcal mol^{-1} per B-Z junction, a significant energetic barrier for Z-DNA formation. In other studies, ~18.0 kcal mol^{-1} was calculated for cruciform formation (Lilley and Hallam, 1984, Lyamichev et al., 1983).

5.5 Conclusion

In conclusion, the evidence provided in this chapter did not support the theory that G-quadruplex formation can be induced by negative supercoiling.

Two-dimensional gel electrophoresis can only demonstrate supercoil-dependent changes in cccDNAs and none were detected in any of the monomeric and dimeric G_3T and G_3T_4 plasmids. The trimeric G-quadruplex forming plasmid **pG_n,4,X_n** which terminates with an inverted repeat sequence demonstrated supercoil-dependent structural changes that included the full G-rich sequence. It is believed that these structural changes were induced by the formation of the 3' cruciform that in turn allowed formation of the other quadruplex structures.

Chapter 6:
Final Discussion and Conclusions

6.1 Final Discussion

DNAs containing consecutive runs of guanines can readily self-associate through guanine-guanine Hoogsteen base pairing to form quadruple helices termed G-quadruplexes (Sen and Gilbert, 1988, Gellert et al., 1962). These structures are stabilised by centrally placed cations and are characterised by the stacking of G-quartets, strand orientation, and the glycosidic bond angles of the guanosine residues (Burge et al., 2006). Bioinformatic studies have found that G-rich sequences with the potential to adopt G-quadruplexes are widespread throughout the human genome (Huppert and Balasubramanian, 2005, Todd and Neidle, 2011, Huppert and Balasubramanian, 2007), and such structures have been implicated in a number of biological processes. Other than the single-stranded telomeric DNA, the majority of genomic G-rich sequences are present, along with their complementary C-rich strands, in the duplex conformation. However, DNA can exist in several different isomeric forms and sequence dependent interconversion between these is often facilitated by the torsional stress of negative supercoiling.

The studies presented in this thesis describe the construction and characterisation of recombinant G-rich plasmids and investigates their capabilities to form intramolecular G-quadruplex structures under negatively supercoiled conditions. Prior to cloning, each insert was subjected to CD analysis to confirm their folding potential. The fundamental aim of the cloning experiments was to generate a library of G-rich plasmids that differ in the number of G-quadruplexes and a series of G-quadruplex-forming oligonucleotides containing four and five repeats of the motifs $(G_3T)_n$, $(G_3T_4)_n$ and $(G_3T_8)_n$ (Table 2.1) were cloned into the *Bam*HI restriction site of pUC19 (Chapter 3).

After construction of the clones, conformational changes within the insert were investigated using chemical and enzymatic probing and two-dimensional gel electrophoresis. Chemical and enzymatic techniques are extremely important and widely used experimental methods for studying changes in DNA (Lilley, 1992). Chapter 4 describes experiments using chemical probing with DMS and $KMnO_4$ and S1 nuclease mapping to identify G-quadruplex formation in supercoiled DNA. Chapter 5 discusses the use of two-dimensional gel electrophoresis of the G-rich DNA topoisomers to resolve the supercoil-dependent topological transitions of negatively supercoiled cccDNA.

6.1.1 Analysis of the $p(G_3T)_n$ Sequences

CD analysis of all $(G_3T)_n$ oligonucleotide sequences demonstrated that they all displayed characteristic peaks associated with the formation of G-quadruplexes even in the absence of potassium. The monomeric four and five repeat G_3T sequences (**G_4,1,1** and **G_5,1,1**) each displayed strong peaks that are characteristic of parallel structures. Each sample exhibited a positive peak at around 260 nm and a negative peak at around 240 nm. These results were virtually mirrored by the dimeric G_3T samples (**G_4,1,2** and **G_5,1,2**), therefore suggesting that in this case both G-quadruplexes had formed.

As the majority of genomic G-rich sequences will be present in the Watson-Crick base-pairing configuration, G-quadruplex formation will inevitably be forced to compete with the double helix. Under near physiological conditions of pH, temperature and salt concentration, it has been reported that G-rich telomeric DNA predominantly remains in the duplex configuration, yet, at lower pH or higher temperature, the G-quadruplex and/or the i-motif are able to form (Li et al., 2002, Phan and Mergny, 2002). Therefore in a further study the duplex-quadruplex interconversion of the annealed duplexes were examined to determine whether these sequences could compete with duplex formation. The results presented demonstrated that the G-quadruplex formation was dominant, even in the presence of an excess of the complementary C-rich strand, which is in agreement with similar studies conducted by Risitano and Fox (Risitano and Fox, 2003b), which examined the melting profiles of fluorescently labelled G-rich oligonucleotides.

Following the CD analysis, the monomeric four and five repeat G_3T oligonucleotides were successfully cloned into pUC19. As each duplex oligonucleotide contained a 5'-GATC sticky-end it was hoped that these oligonucleotides could be ligated to produce longer multimeric inserts in the vector. However, only clones containing a maximum of up to two repeats were isolated. Inexplicably, cloning of the dimeric inserts also failed to produce plasmids with more than one inserted fragment and therefore only generated plasmids with just two G-quadruplex-forming structures. The difficulty in obtaining recombinant plasmids with longer fragments may be a consequence of the ease in which $(G_3T)_n$ DNA sequences readily form stable quadruplex structures. If G-quadruplex formation had occurred prior to insertion into pUC19, then ligation into the plasmid would be hindered. Therefore, due to the difficulty in cloning only monomeric and dimeric G_3T plasmids were studied in this research.

DMS protection analysis of the p(G_3T) plasmids **pG_4,1,1 and pG_4,1,2** demonstrated that G_3-tracts in the quadruplex-forming regions of the supercoiled plasmids had significantly reduced reactivity compared to the guanines flanking the sequence. These observations indicate the presence of Hoogsteen or Hoogsteen-like hydrogen bonding, consistent with the extrusion of a quadruplex structure. Noticeably, the first and second G_3-tracts at the 5'-end of the quadruplex revealed the greatest protection and as intramolecular G-quadruplex structures formed from linear G-rich DNA fragments assemble (relatively) quickly, particularly p(G_3T) sequences, it can be assumed that differences in reactivity between the 5' G_3-tracts and the 3' G_3-tracts indicated that the quadruplex structure does not fully unfold at the 5'-end before reassembly, and therefore exists in dynamic equilibrium between G-quadruplex and a partially formed duplex. Surprisingly, the same cleavage patterns were observed in the absence of negative supercoiling. Although the DMS footprinting of the **pG_4,1,1 and pG_4,1,2** plasmids indicated the presence of G-quadruplex structures, their formation did not require the torsional energy generated by negative supercoiling. Interestingly, almost identical results were seen in the absence of KCl, signifying that formation and stabilisation of these structures were not dependent on the presence of a centrally placed monovalent cation.

Probing of the thymine loops revealed that none of the proposed loops in either the supercoiled or the linear plasmids exhibited any reactivity towards $KMnO_4$ in either the presence or absence of KCl. However, the thymines that immediately flank the G-rich regions exhibited a significant sensitivity to the chemical reagent. Consequently, it is possible that, due to the compact nature of the single nucleotide loops, the bulky $KMnO_4$ molecule may have been sterically hindered from accessing the C5-C6 double bond of the thymines in these loop regions, although, the hyperreactivity observed by the thymines that immediately flank the G-rich sequence indicates that the DNA was experiencing some structural alteration around this location.

The five repeat p(G_3T) plasmids **pG_5,1,1** and **pG_5,1,2** revealed identical DMS footprinting patterns in both the presence and absence of KCl. Although each G_3-tract showed reduced reactivity to DMS, the central three G_3-tracts of each quadruplex forming sequence exhibited the greatest degree of protection from chemical modification. The overall cleavage pattern appeared to indicate the presence of more than one, possibly several G-quadruplex isomers, where the structures containing the central three tracts in the quartet formation being the most stable. This is in contrast to similar experiments conducted on the polymorphic wild type *c-myc* G-quadruplex forming sequence (Sun and Hurley, 2009). Whilst it was possible to produce several G-quadruplex isomers from the six G-tract motif it was found that only a single structure had formed using the four 3'-tracts. The $KMnO_4$ experiments exhibited identical cleavage patterns to those of the four G-tract sequences. The chemical also failed to identify the proposed thymine-loops of the quadruplexes in both monomeric and dimeric sequences. However, the thymines immediately flanking the inserted G-rich DNA were extremely hypersensitive to the probe, which may correspond to a region of locally melted DNA on either side of the G-quadruplex-forming sequence.

In contrast, the S1 nuclease mapping of the G-quadruplexes for all of the p$(G_3T)_n$ plasmids did not identify G-quadruplex formation. For each of the G-rich plasmids, the S1 nuclease appeared to cleave at the same site, which did not appear to be at the proposed location of the G-quadruplex, but preferentially at the putative cruciform of the pUC19 vector. Furthermore this S1 nuclease sensitive site was no longer apparent

after the native supercoiling of the DNA had been removed. This suggests that under the experimental conditions employed the cruciform structure was induced by the negative supercoiling of the plasmid. Therefore, if the cruciform structure had indeed been cut first by the enzyme, then the excess free energy generated by the superhelical stress of the plasmid would be consumed by cruciform formation and as a result the G-quadruplex region would revert back to the duplex formation. Although this theory is not consistent with the chemical probing assays, which indicated that the formation of the G-quadruplex structure(s) was independent of negative supercoiling, it is important to note that the chemical probing assays were conducted at room temperature, whilst the S1 nuclease digests were conducted at 37 ºC. Short linear single-thymine loop G-quadruplex forming oligonucleotides can readily form highly stable G-quadruplex structures in under physiological-like ionic conditions (~100 mM K^+) (Risitano and Fox, 2003b), it is therefore possible that, once these sequences are flanked by longer, non-G-quadruplex-forming sequences, their stability may be reduced.

Two-dimensional gel electrophoresis experiments with DNA topoisomers of the $p(G_3T)_n$ series did not identify any supercoil-dependent structural transitions in any of the G-quadruplex-forming DNA. This is consistent with the DMS and $KMnO_4$ chemical probing assays which demonstrated that the structural changes observed within the plasmids were independent of negative supercoiling. Consequently, as the topological changes in the DNA were pre-existent to the two-dimensional gel analysis, this experimental technique would not be able to detect their formation and consequently the DNA topoisomers would then migrate undetected and run "normally" as if it were not undergoing any structural transitional changes.

The structural changes observed with the $p(G_3T)_n$ plasmids imply that the G-quadruplexes were formed un-aided by the negative supercoiling. However, these sequences must exist in a duplex form at some stage in order to allow for their cloning and replication. It is plausible that the four-stranded structures may have been generated during plasmid extraction and purification, as this process involves alkaline denaturation. Under these conditions it is possible that, once the G-rich strand had been dissociated from its complementary C-strand, it was then free to assemble into the G-

quadruplex structure(s) upon renaturation to form a cccDNA. As single thymine-looped G-quadruplexes assemble relatively quickly and form extremely stable complexes (Risitano and Fox, 2003b), it is highly possible that the subsequent $p(G_3T)_n$ plasmid would then contain the structurally altered G-rich region before experimental analysis.

6.1.2 Analysis of the $p(G_3T_4)_n$ sequences

CD analysis of the monomeric four and five repeat $p(G_3T_4)$ sequences (**G_4,4,1** and **G_5,4,1**), revealed cation-dependent conformational changes. At low KCl concentration (< 5 mM), the CD results did not show peaks that were characteristic of G-quadruplex formation, suggesting that DNA remained in a single-stranded configuration. At >5 mM KCl concentration a change in the spectral pattern was observed that appeared to indicate the DNA sample was in dynamic equilibrium between single-stranded and quadruplex DNA. As the potassium concentration increased (>10 mM) a positive peak at 295 nm and a negative peak at 260 nm were observed. These peaks are characteristic of an antiparallel topology, signifying that higher KCl concentrations are needed to fully convert the single-stranded G-rich DNA into a stable intramolecular antiparallel structure. These findings were also observed for the dimeric G_3T_4 sequences (**G_4,4,2** and **G_5,4,2**).

In contrast to the $(G_3T)_n$ sequences the $(G_3T_4)_n$ sequences did not form G-quadruplex structures in the presence of the complementary C-strand and remained in the duplex formation. This is unsurprising, as longer loop sequences are less stable than short loop sequences (Rachwal et al., 2007b) and are less likely to form in the presence of their complements (Risitano and Fox, 2003b).

In contrast to the DMS chemical probing of the $p(G_3T)$ plasmids, which appeared to detect G-quadruplex formation, DMS probing of G_3-tracts of the monomeric and dimeric $p(G_3T_4)$ plasmids revealed that guanines within the G-rich regions exhibited the same reactivity towards the reagent as those guanines outside the cloned insert. These results suggest that G-quadruplex formation had not occurred and that the G-rich insert had remained in the duplex configuration. Surprisingly, the $KMnO_4$ results appeared to

reveal similar cleavage patterns to those of the $p(G_3T)_n$ sequences. Again, in both the presence and absence of negative supercoiling, only thymines immediately flanking the 5'- and 3'-ends of the G-rich insert exhibited reactivity towards the $KMnO_4$, whilst the T_4-loops remained unreactive to the probe. Although both the DMS and $KMnO_4$ experiments did not appear to suggest the formation of quadruplexes, the $KMnO_4$ experiments did reveal that the thymines flanking the inserted G-rich DNA were extremely hypersensitive to the probing experiment, possibly indicating a region of locally unwound DNA outside the G-quadruplex sequence, though the exact nature for this change remains unclear.

Digestion of the $p(G_3T_4)_n$ DNAs with S1 nuclease generated identical cleavage patterns to that of the single T-loop sequences. Again the enzyme did not generate any cleavage fragments corresponding to the formation of G-quadruplex structure(s), but products that appeared to correspond to the cleavage of the pUC19 cruciform. These results were consistent with the chemical probing assays as neither DMS nor $KMnO_4$ detected any G-quadruplex formation.

The two-dimensional gel electrophoresis of the $p(G_3T_4)_n$ DNA topoisomers also did not identify any supercoil-dependent structural transitions. This was consistent with the results obtained with $KMnO_4$ chemical footprinting analysis which only identified negatively supercoil-independent structural changes flanking the sequences of the G-rich insert. A similar explanation analogous to that of the $p(G_3T)$ plasmids can be made; as the structural transitions are independent of negative supercoiling, topological changes in the DNA will have no effect.

6.1.3 Analysis of the $p(G_3T_8)$ sequences

CD analysis of the $p(G_3T_8)$ oligonucleotides did not identify G-quadruplex formation. A previous study has reported that long quadruplex loops do not form very stable structures and that long loops (> 7 nt) tend to destabilise quadruplex structures (Hazel et al., 2004, Rachwal et al., 2007b). However, CD analysis of longer loop sequences is

complicated, as long loop lengths (> 7 nt) largely behave as single strands resulting in spectral ambiguity (Hazel et al., 2004).

As the longer T_8-looped G-rich sequences did not appear to form the G-quadruplex structure(s), it was not surprising that they also did not form in the presence of their complement, and like the p(G_3T_4) fragments they also remained in the duplex conformation. Although some pUC19 G_3T_8 clones were prepared, due to the difficulties in forming stable G-quadruplexes in the single stranded linear formation, these clones were subsequently not considered for the remainder of this research.

6.1.4 Analysis of the plasmid mutant $pG_n,4,X_n$

In a further study the mutant quadruplex-forming sequence **$pG_n,4,X_n$** was investigated using chemical and enzymatic probing assays and two-dimensional gel electrophoresis. The DMS experiments revealed that none of the guanine residues within the quadruplex-forming regions were protected from methylation, suggesting that it was unlikely that G-quadruplex formation in any of the three potential G-quadruplex forming sequences had taken place. In contrast, $KMnO_4$ revealed similar cleavage patterns to the p$(G_3T)_n$ and p$(G_3T_4)_n$ clones previously analysed. Cleavage differences were observed at thymines flanking each quadruplex-forming sequence, but none within the loops separating the G_3-tracts. Reactivity to the thymines in the 5'-GATC flanking region was observed in both the supercoiled and linear DNA and in the presence and absence of KCl. It was also observed that two 3' cytosine residues that preceded the thymine on the flanking region of the third (3') quadruplex-forming sequence also exhibited $KMnO_4$ sensitivity, that was inhibited once the plasmid had been linearised and the negative supercoiling had been eliminated. Sensitivity of the thymine and cytosines at this location signifies a region of locally unwound DNA, which exists within a region of palindromic DNA. Hence, it is possible that the cleavage pattern observed at 3'-end of this insert corresponded to the single-stranded loop region of a possible cruciform.

Unlike the $p(G_3T)_n$ and $p(G_3T_4)_n$ clones, the S1 nuclease mapping of the **$pG_n,4,X_n$** mutant plasmid identified several regions of single-stranded DNA, which was consistent with the $KMnO_4$ footprinting assays. These regions only existed under negative supercoiling. Analysis of the digestion fragments indicated that the S1 nuclease had cleaved at several different sites along the DNA. Whilst some of the fragments appeared to correspond to cleavage of the pUC19 cruciform, the other fragments were consistent with cleavage in the vicinity of the G-quadruplex sequence, though, it is important to note that these cleavage fragments may not necessarily correspond to formation of the four stranded structure but to the possible cruciform structure of the palindromic sequence at the 3'-end of the terminal G-quadruplex (Figure 4.23). Therefore it is plausible that the DNA fragments generated by the S1 nuclease in this experiment do not correspond to G-quadruplex formation, but possibly to cruciform extrusion, which is consistent with the DMS and $KMnO_4$ chemical probing experiments.

Two-dimensional gel electrophoresis of the mutant plasmid **$pG_n,4,X_n$** (Figure 5.9 and 5.10) revealed several different patterns indicating that the DNA sample contained a mixed population of (non-interconverting) topoisomers. The first population followed a smooth progression of topoisomers identical to the pUC19 control. This suggests that this population of topoisomers had not experienced any supercoiled-dependent structural transitions. Two separate discontinuity "jumps" were also observed indicating that the DNA topoisomer sample possessed a mixture of DNA containing different isomeric conformations that were retarded in mobility in the first dimension. The results also revealed that each "jump" occurred at separate superhelical densities both of which were around σ = -0.04. By the time topoisomer 20' and 20'' was reached, calculations revealed that approximately 125 bp were unwound which is consistent with the melting of the G-rich DNA within this region and/or the possible formation of either; three complete quadruplex structures or a combination of two quadruplexes, a cruciform and the melted region of DNA flanking the G-rich insert. These results appear to correlate with the data from the $KMnO_4$ and S1 nuclease probing experiments in Chapter 4 that identified multiple sites of single-stranded DNA, indicating regions of locally unwound DNA flanking the G-rich DNA. However, as two-dimensional gel electrophoresis cannot provide any structural data, it cannot be concluded that the jumps in the gel pattern were a result of G-quadruplex formation and it is also important to note that

these different structural forms must be non-interconverting, otherwise they would not generate discrete separate two-dimensional gel electrophoresis profiles. Nonetheless the ΔG_f required for untwisting the DNA in this region and the proposed formation of three quadruplex structures was also estimated to be 52.8 KJ mol^{-1} (12.6 Kcal mol^{-1}) and 66.9 KJ mol^{-1} (15.9 Kcal mol^{-1}) which is similar to the value of 17.0 Kcal mol^{-1} calculated for the B-Z transition of 32 bp d(CG) repeat (Wang et al., 1983) or ~18.0 Kcal mol^{-1} calculated for cruciform formation (Lilley and Hallam, 1984, Lyamichev et al., 1983).

Possible explanations as to why only the two-dimensional gel electrophoretic analysis of the mutant plasmid **$pG_n,4,X_n$** was able to identify structural perturbations within this DNA sample may not be due to presence of a higher number of G-tracts, particularly as the probing experiments demonstrated that the monomeric and dimeric samples exhibited identical results, but perchance the palindromic sequence at the 3'-end of the G-rich insert. It may be possible that the inverted sequence was able to induce the cruciform structure and as the formation would require the localised unwinding of the DNA it was able to facilitate further unwinding on the G-rich DNA and possible extrusion of a G-quadruplex structure(s). By locally unwinding and melting the DNA at the 3' terminus to form the cruciform structure, it may be possible to facilitate the formation of the two 5' G-quadruplexes as the G-rich strand will be temporarily denatured and free to form the intramolecular structures that are thermodynamically stable under the experimental conditions of two-dimensional gel analysis rather than the chemical probing experiments. This is analogous to the research conducted by Onyshchenko *et.al* (Onyshchenko et al., 2009, Onyshchenko et al., 2011), where peptide nucleic acids PNAs were used to sequester the complementary C-rich strand, to allow formation of a stable quadruplex structure in the G-rich strand.

6.2 The Effect of Negative Supercoiling and G-Quadruplex Formation

The results presented in this thesis were unable to provide direct evidence that the formation of G-quadruplex structures could be induced by the torsional stress generated by negative supercoiling. Whilst the experimental analysis of the T_4-looped molecules

did not identify any quadruplexes within the plasmids, DMS chemical probing of the single thymine-looped plasmids did identify G-quadruplex formation within the cloned G-rich inserts, albeit the formation of these structures were independent of negative supercoiling. However, neither $KMnO_4$ nor S1 nuclease were able to detect the formation of the four stranded structures in any of the $p(G_3T)_n$ plasmid clones. It must then be concluded that the experimental techniques employed by this research were not ideal for researching this hypothesis or that the chosen sequences for this research do not form G-quadruplexes under negative supercoiling.

Although this research was unable to detect negative supercoiling-induced formation of G-quadruplexes, there are a small number of publications which have demonstrated that negative supercoiling can facilitate the formation of quadruplex structures. Using chemical and enzymatic footprinting Sun and Hurley (Sun and Hurley, 2009) identified G-quadruplex formation under the influence of negative supercoiling in a DNA clone containing the G-quadruplex forming sequence of the NHE III_1 within the *c-myc* promoter. The authors used S1 nuclease and DNase I footprinting to finely map G-quadruplex formation in negatively supercoiled plasmids. They showed that under torsional stress and in the absence of KCl both enzymes reveal that NHE III_1 and its flanking regions are highly dynamic and exist in an equilibrium mixture of duplex, partially unwound and secondary DNA structures. On addition of 100 mM KCl, the enzymes' activities were inhibited as the dynamics of the flanking region were dampened by the formation of a potassium stabilised G-quadruplex structure. These results signify that, in the presence of a G-quadruplex stabilising cation, the four-stranded structure is capable of forming under negative supercoiling and that the formation of this structure acts as a buffer to absorb the torsional stress generated by negative supercoiling. In sharp contrast to the enzymatic probing, the chemical probing assays revealed that KCl had little effect on the reactivity of the chemical probes, revealing that the *c-myc* promoter region was able to form a G-quadruplex structure that was facilitated by negative supercoiling irrespective of the presence of a stabilising cation. Similar findings were reported for the DMS footprinting assays of the G-quadruplex-forming region of *VEGF* promoter clone (Sun et al., 2011). However, these results are in contrast to other previously reported studies. Whilst these results have shown that G-quadruplex formation within G-rich segments of the *c-myc* and *VEGF*

promoters under the influence of negative supercoiling is independent of ionic concentration, other studies have shown that this only occurs with the aid of extrinsic factors such quadruplex-stabilising ligands (Sun, 2010), PNAs (Onyshchenko et al., 2009, Onyshchenko et al., 2011) and/or a monovalent cation (Onyshchenko et al., 2009, Onyshchenko et al., 2011, Sun, 2010) otherwise the formation of the structures are inhibited. The initial studies conducted by Onyshchenko *et al.* (Onyshchenko et al., 2009, Onyshchenko et al., 2011) found that G-quadruplex formation of the cloned G-quadruplex forming sequence of the *Bcl-2* gene did not occur under physiological conditions which is consistent with the results presented in this thesis, however, by introducing PNAs that bind to the complementary C-rich strand they were able to induce G-quadruplex formation in the presence of either potassium or sodium cations or the quadruplex-binding ligand telomestatin. Interestingly, in a further study it was found that for the PNA to effectively invade the C-rich strand in supercoiled DNA, the G-rich strand must possess quadruplex-forming potential. This suggests that PNAs were not only sequence-specific but quadruplex-specific demonstrating a new mode of sequence-specific targeting as a means of stabilising G-quadruplex structures as an approach in gene regulation *in vivo*.

Transcriptionally-induced negative supercoiling, facilitating G-quadruplex formation was demonstrated in a recent study conducted by Zhang *et al*. (Zhang et al., 2013). They showed that G-quadruplex formation can be specifically induced upstream by negative supercoiling resulting from transcriptional events. By cloning a synthetic DNA fragment containing a $G_3(TG_3)_3$ motif into the TA cloning site of a pMD-19-T plasmid, upstream of the T7 promoter or between two divergently oriented T7 promoters and with the aid of polyethylene glycol (PEG) (Zheng et al., 2010) to stabilise and facilitate quadruplex detection, it was found that G-quadruplex formation was specifically induced by transcription and that transcription from two promoters oriented in opposite directions intensified the effects. More importantly they also found that positive supercoiling, generated in front of the moving polymerase, did not induce the formation of G-quadruplexes.

There is an abundance of PQS in regions upstream of DNA promoters (Huppert, 2006, Huppert and Balasubramanian, 2007, Huppert et al., 2008) and the insights provided by these studies have offered significant evidence to support the hypothesis that G-quadruplex structures can be formed by G-rich tracts of promoter regions and may have a functional role in gene regulation.

6.3 Final Conclusions

The studies presented in this thesis have used chemical and enzymatic probing along with two-dimensional gel electrophoresis to determine whether G-quadruplexes can form under negative supercoiled conditions. The results provide little evidence to support this hypothesis. The p(G_3T) sequences appear to form G-quadruplexes albeit independent of DNA supercoiling, whilst p(G_3T_4) sequences did not. Although the monomeric and dimeric p(G_3T_4) plasmids did not demonstrate any negative supercoiling-induced G-quadruplex formation, two-dimensional gel electrophoresis of the mutant p(G_3T_4) trimeric plasmid **pG_n,4,X_n** did reveal some supercoil-dependent structural alterations within the G-rich segment, however, neither chemical nor enzymatic probing of the mutant were able to clarify the nature of the structural perturbations. Therefore, in conclusion, the results obtained in this research suggest that negative supercoiling does not induce the formation of G-quadruplex structures.

6.4 Future Work

Despite speculation regarding the potential of PQS in promoter regions of telomeres and oncogenes, the presence and function of polypurine and polypyrimidine tracts throughout the genome is still not fully understood and many more studies are still required in both *in vitro* experiments and biological systems to understand their full implications *in vivo*. Future work studies could include:

- Repeating the cloning experiments using monomeric G-quadruplex forming oligonucleotides that terminate with different 3' flanking sequences. This is to first optimise the self-ligation experiments in order to create DNAs with longer

inserts and secondly to determine whether the flanking regions of the G-rich inserts still produce a region of melted DNA under negative supercoiling.

- Repeating the probing experiments with the previously prepared DNA clones using a variety of G-quadruplex stabilising ligands to investigate whether under negative supercoiled conditions the ligands promote quadruplex formation and stabilisation.

- Although more attention is now being directed towards the *i-motif* which is formed from the complementary C-rich strand at lower pH (~pH5), the focus still remains predominantly on the G-rich and their counterparts. Therefore in a further study; the chemical probe diethyl pyrocarbonate can be used to probe the C-rich strand under negative supercoiled conditions to determine whether the *i-motif* structure can form at physiological pH as reported by Sun and Hurley (Sun and Hurley, 2009) or whether it requires the same low pH conditions as the short linear fragments.

- Finally, the two-dimensional gel electrophoresis analysis demonstrated that a palindromic sequence at the end of the mutant plasmid **$pG_n,4,X_n$** induced conformational changes within the inserted G-rich segment. Further cloning experiments of G-quadruplex forming plasmids containing specifically placed alternative secondary structures such as cruciforms can be conducted to determine whether these structures can induce G-quadruplex formation in supercoiled DNA.

Appendices

Appendix 1 Commercial sequencing of pUC19 G-quadruplex clones

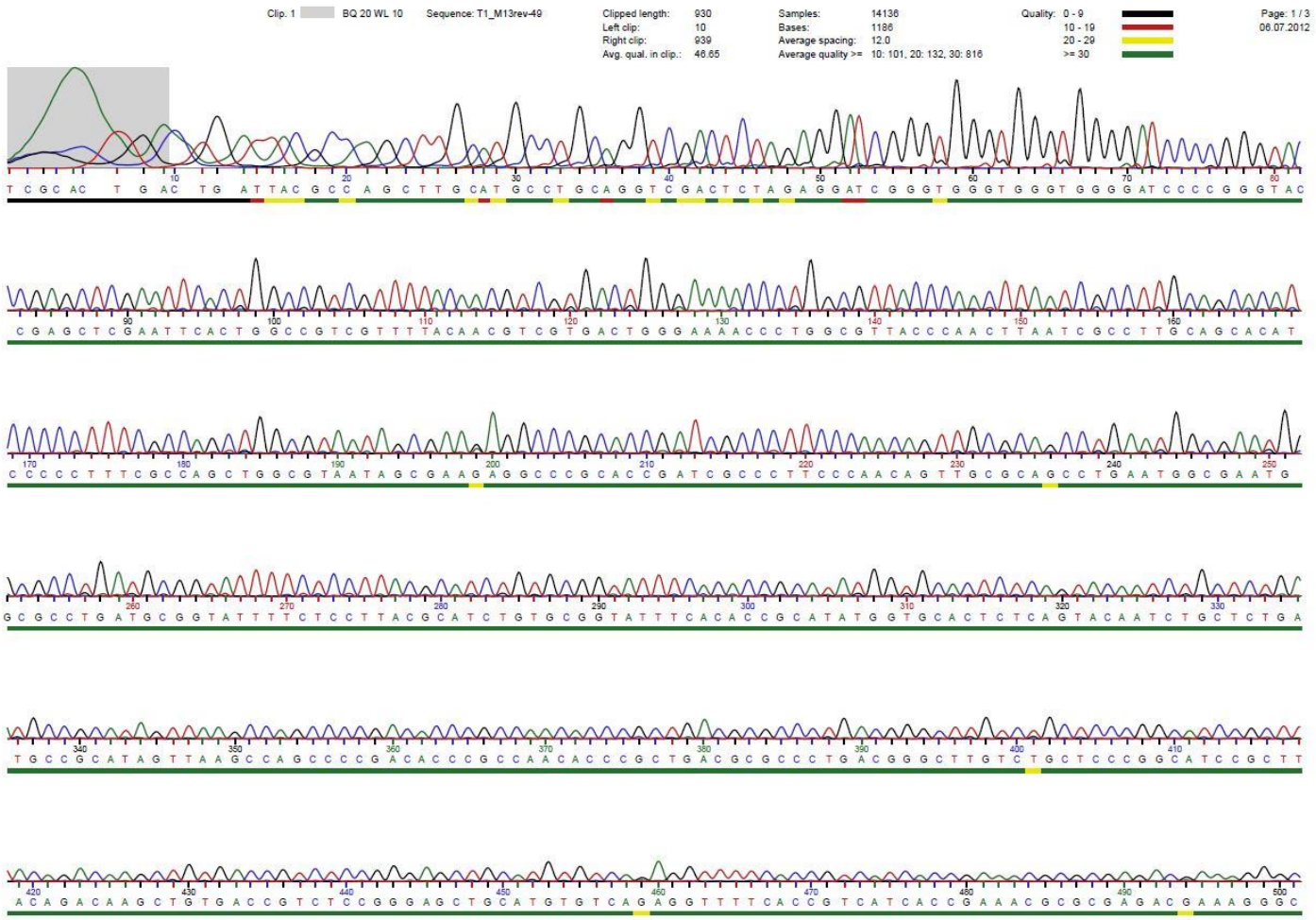


Figure A. Commercial sequencing of **pG$_4$,1,1**

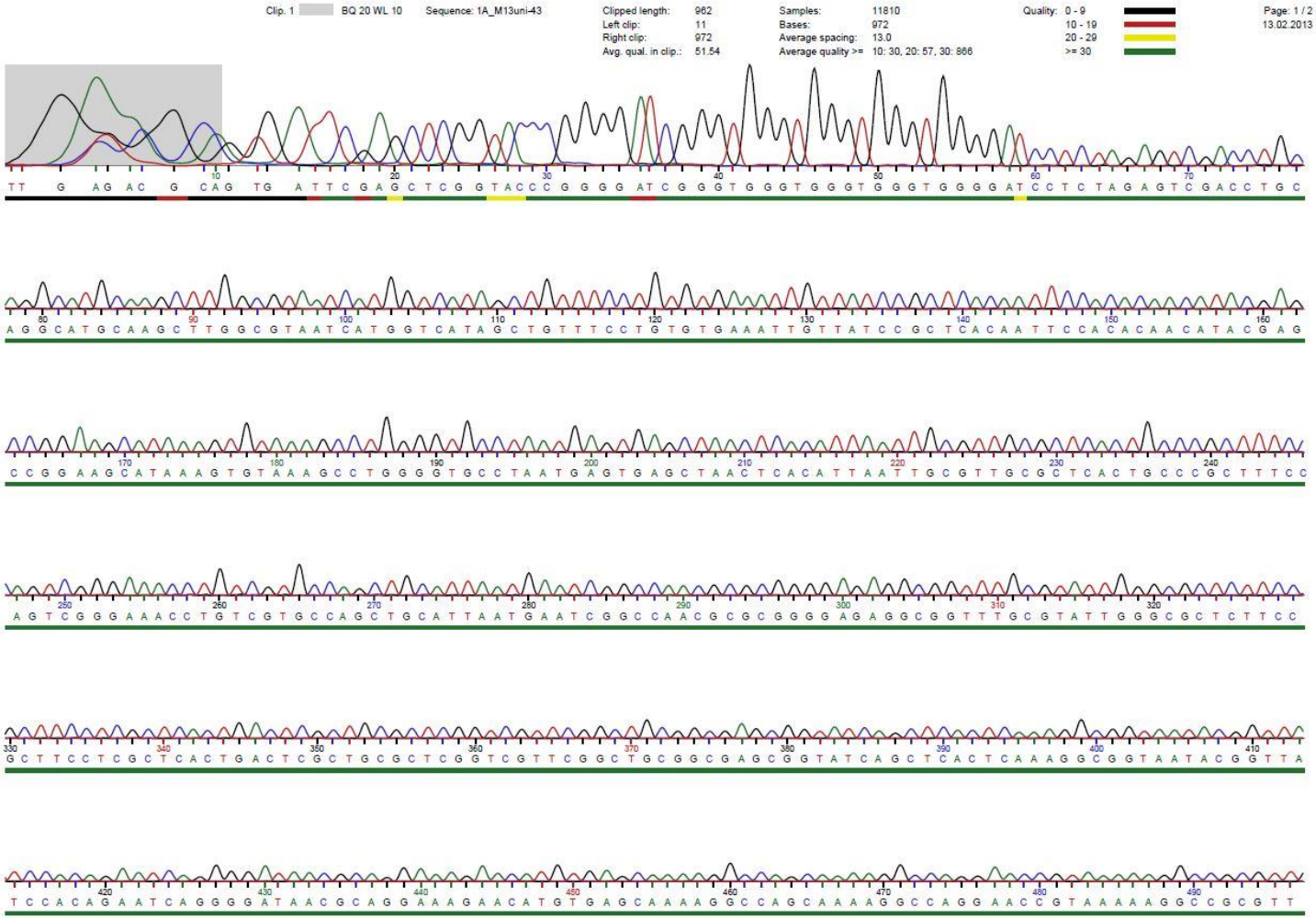


Figure B. Commercial sequencing of **pG_5,1,1**

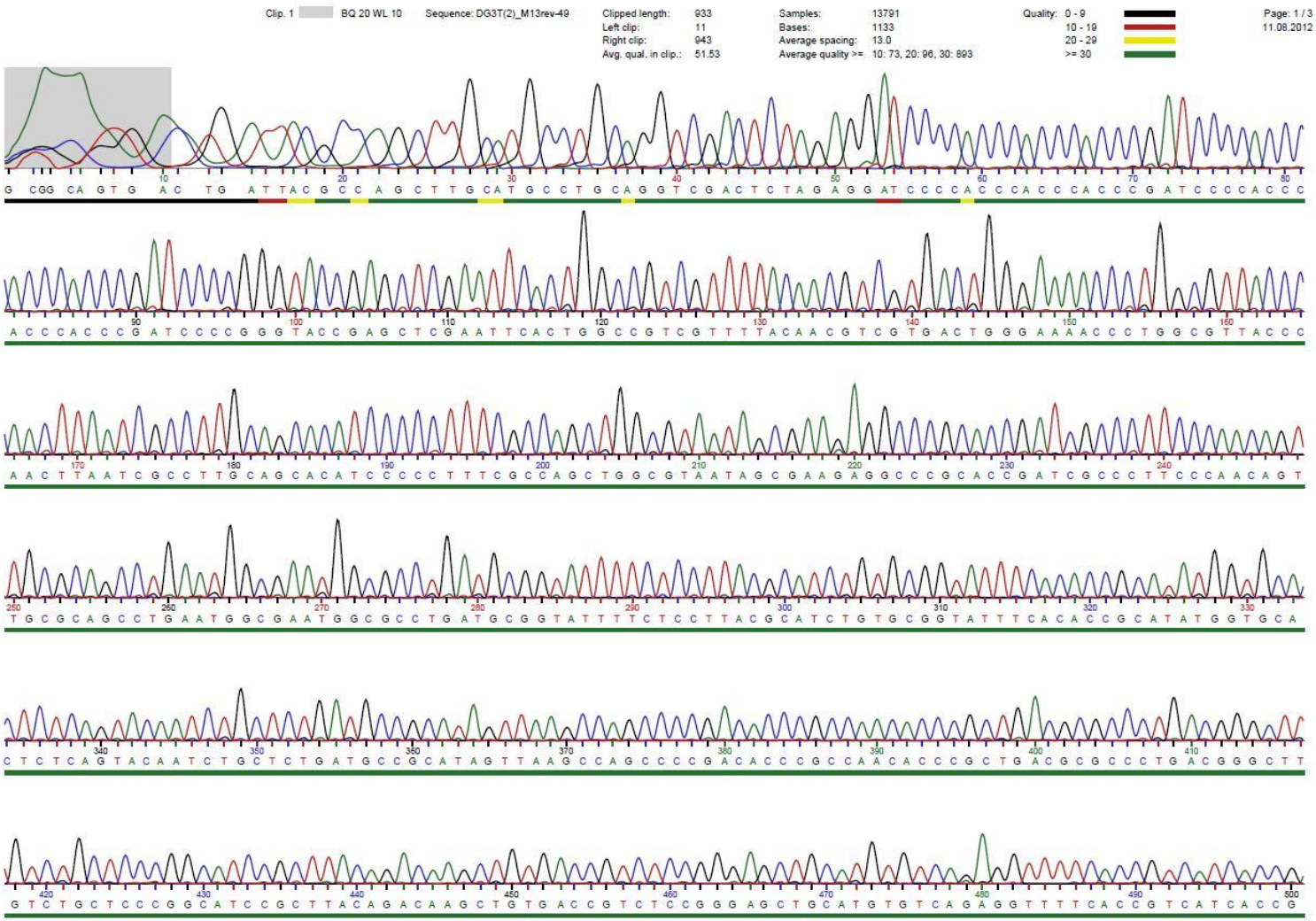


Figure C. Commercial sequencing of **pG_4,1,2**

Appendix 1

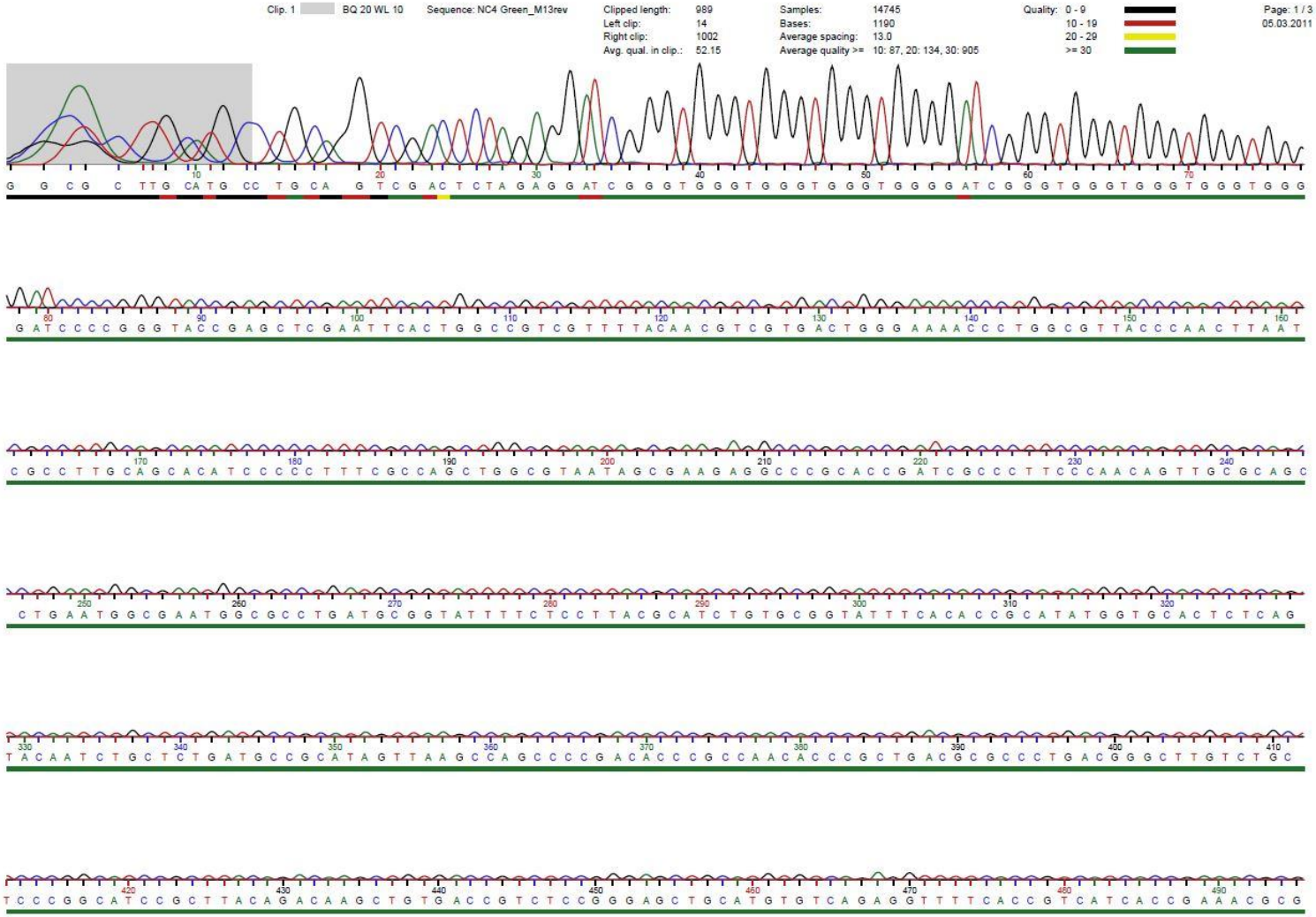


Figure D. Commercial sequencing of **pG$_4$,1,2**

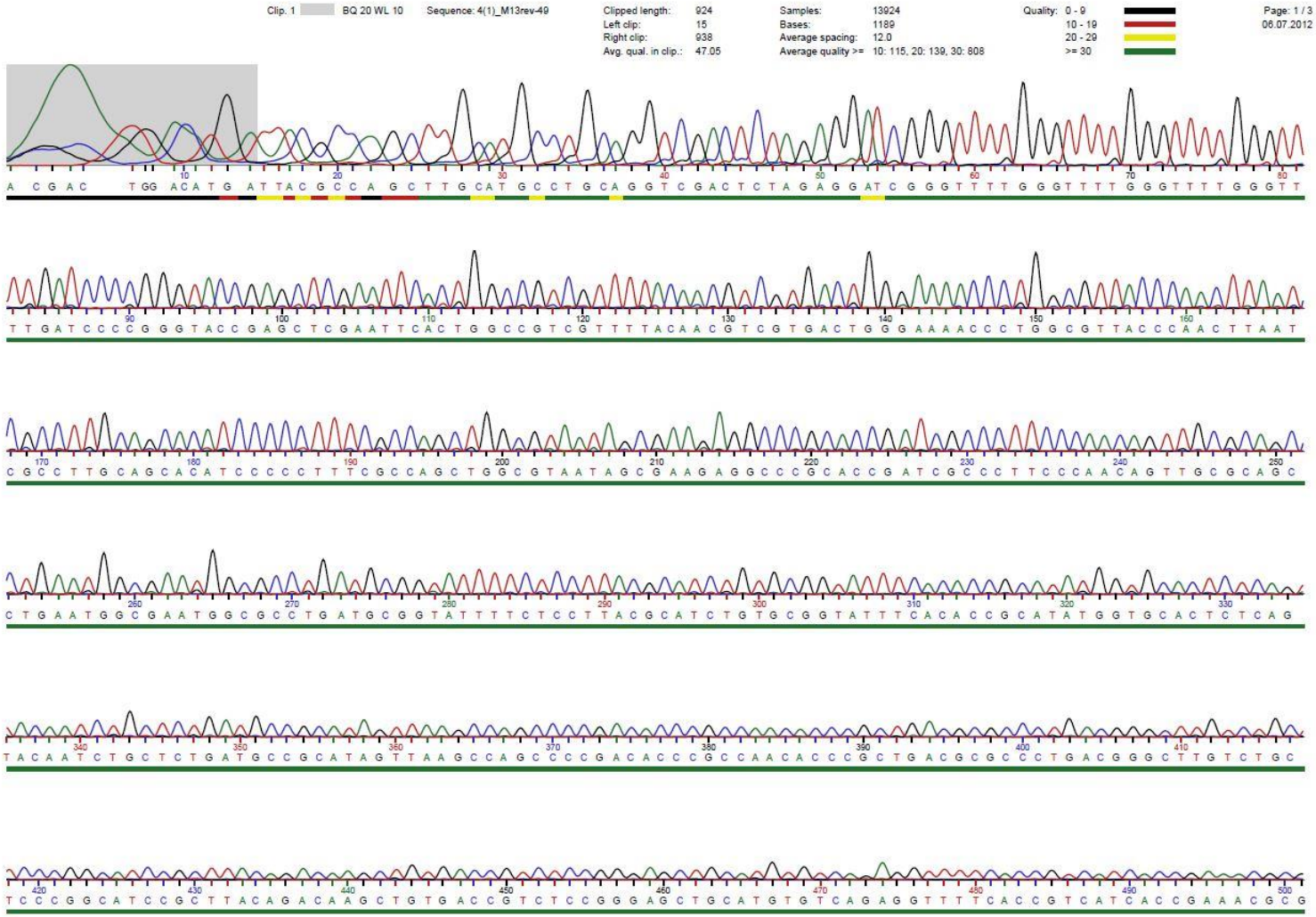


Figure E. Commercial sequencing of **pG$_4$,4,1**

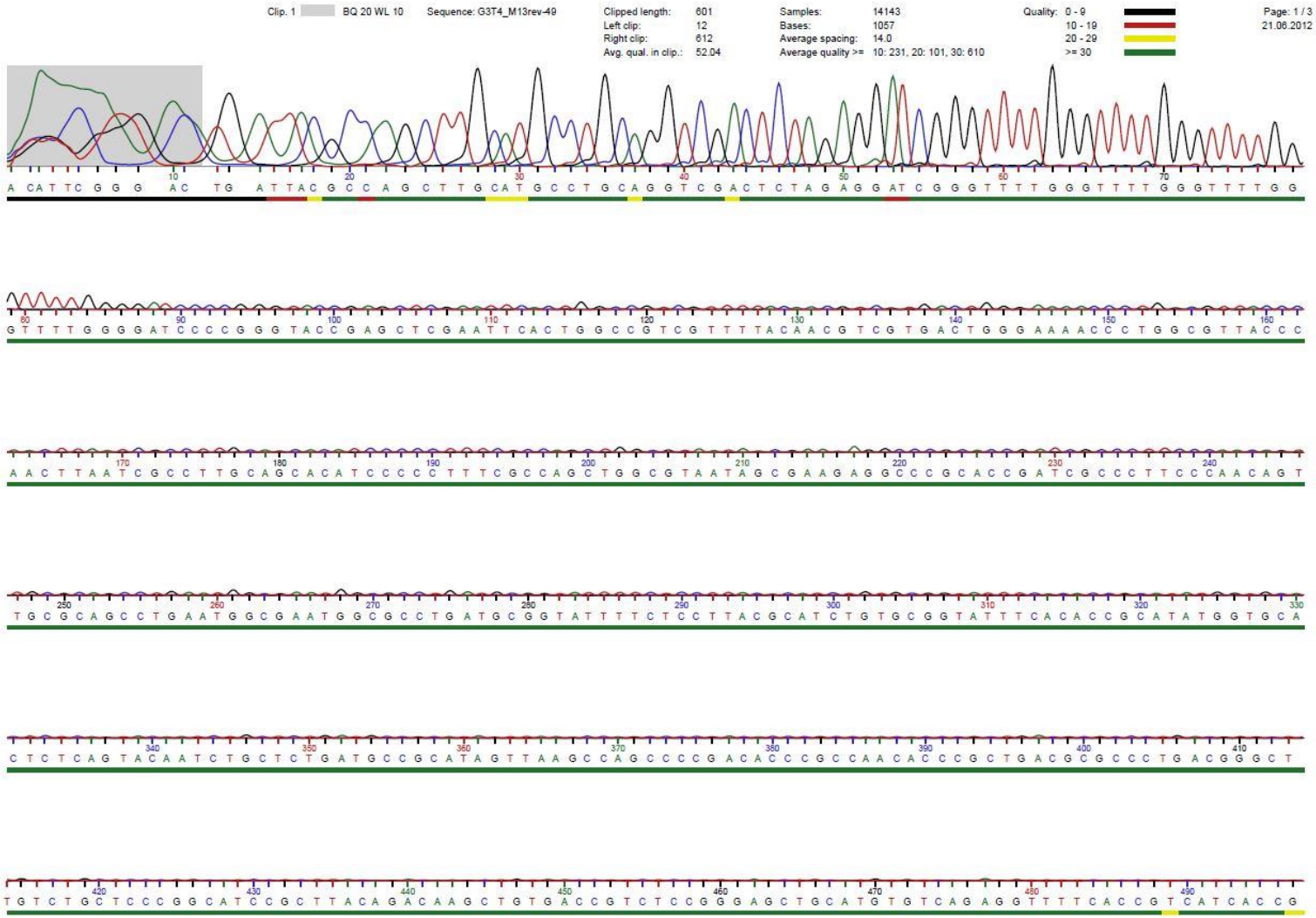


Figure F. Commercial sequencing of **pG$_5$,4,1**

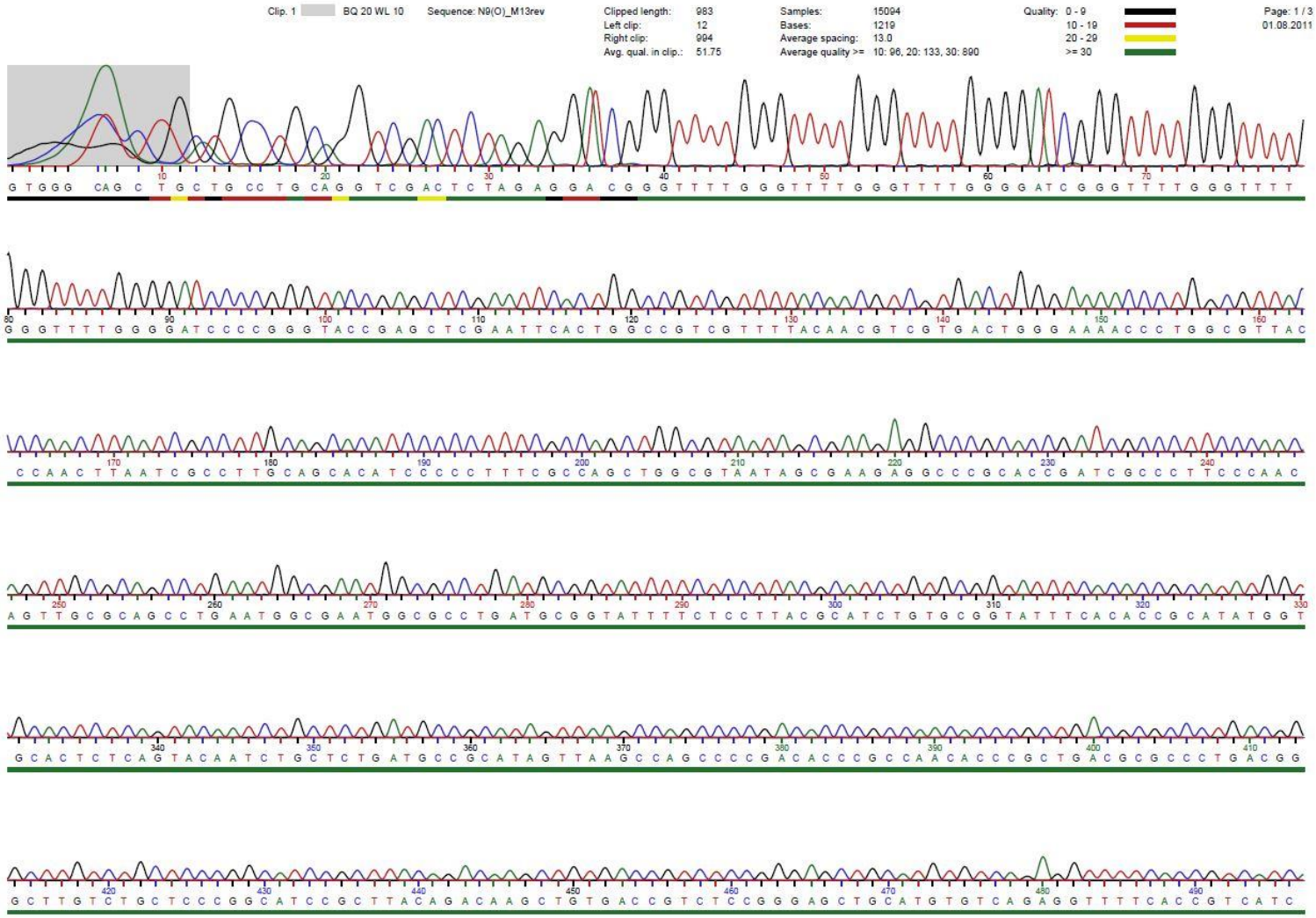


Figure G. Commercial sequencing of **$pG_4$4,2**

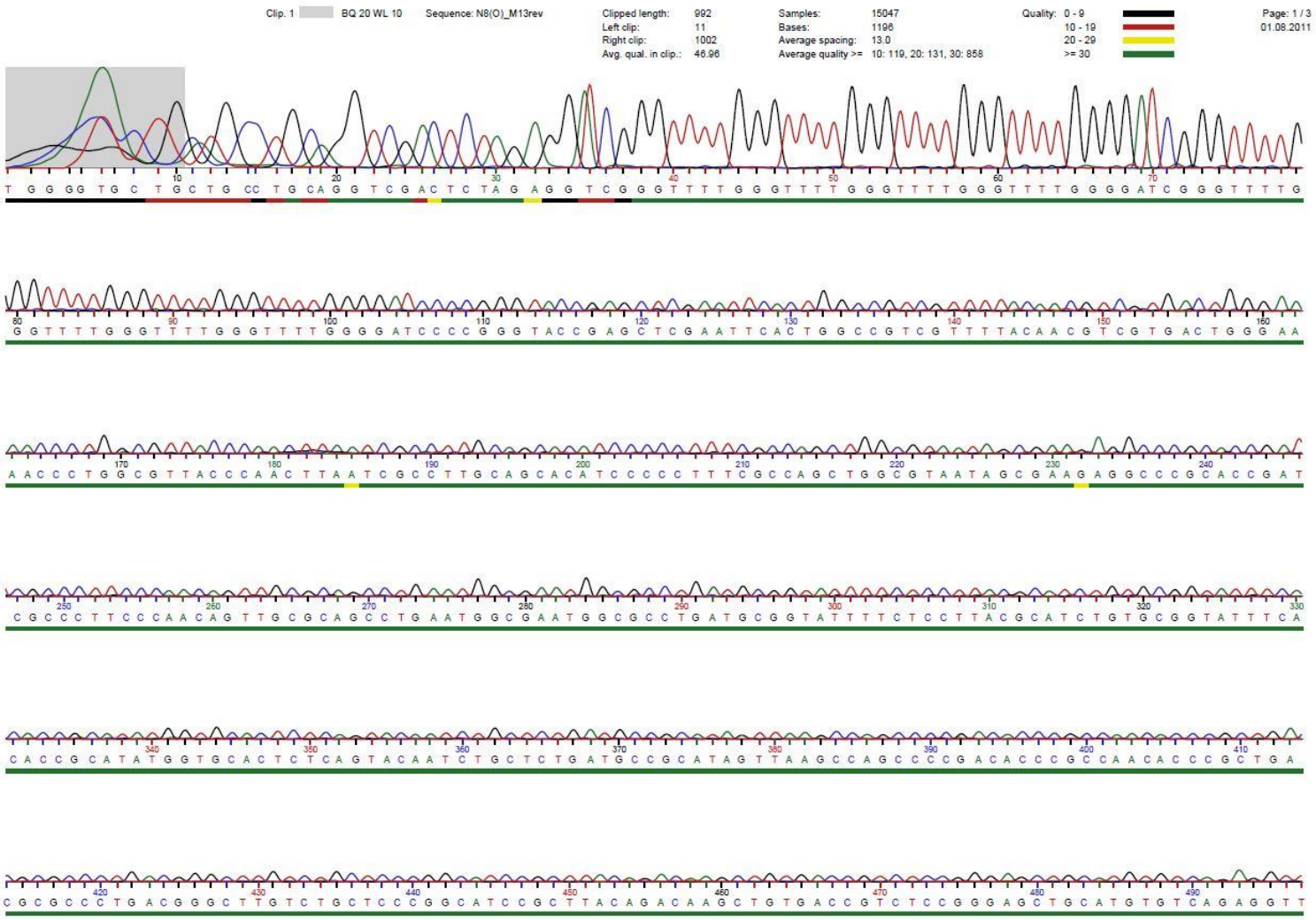


Figure H. Commercial sequencing of **pG_5,4,2**

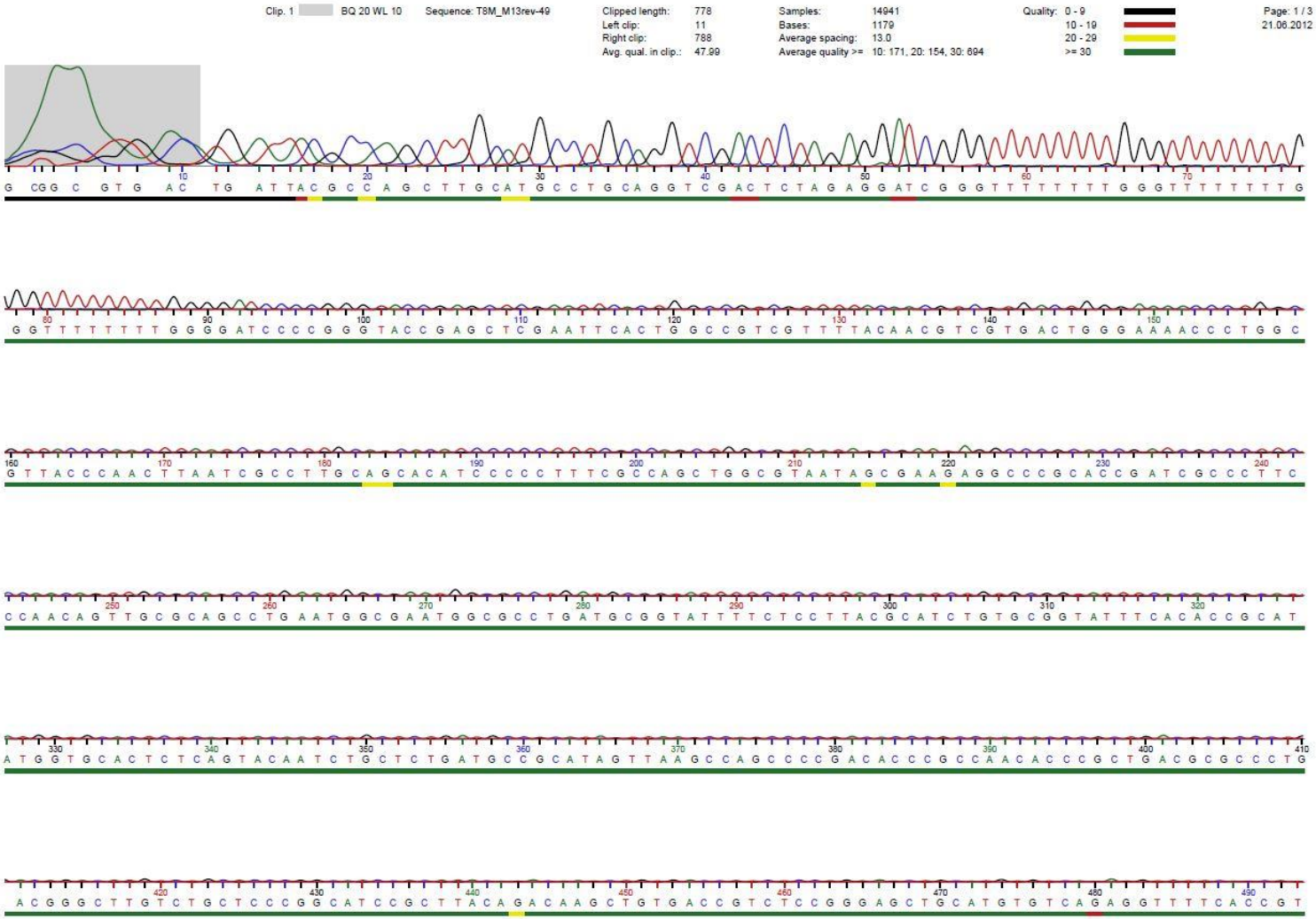


Figure E. Commercial sequencing of **pG$_4$,8,1**

List of References

ABOULELA, F., MURCHIE, A. I. H. & LILLEY, D. M. J. 1992. NMR-study of parallel-stranded tetraplex formation by the hexadeoxynucleotide d(TG4T). *Nature,* 360**,** 280-282.

ADAMS, J. M. & CORY, S. 2007. The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene,* 26**,** 1324-1337.

AGRAWAL, P., HATZAKIS, E., GUO, K., CARVER, M. & YANG, D. 2013. Solution structure of the major G-quadruplex formed in the human VEGF promoter in K+: insights into loop interactions of the parallel G-quadruplexes. *Nucleic Acids Research***,** 1-9.

AMBRUS, A., CHEN, D., DAI, J., JONES, R. & YANG, D. 2005. Solution structure of the biologically relevant G-Quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Proceedings of the American Association for Cancer Research Annual Meeting,* 46**,** 593-594.

AMBRUS, A., CHEN, D., DAI, J. X., BIALIS, T., JONES, R. A. & YANG, D. Z. 2006. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Research,* 34**,** 2723-2735.

AMRANE, S., ADRIAN, M., HEDDI, B., SERERO, A., NICOLAS, A., MERGNY, J.-L. & ANH TUAN, P. 2012. Formation of pearl-necklace monomorphic G-quadruplexes in the human CEB25 minisatellite. *Journal of the American Chemical Society,* 134**,** 5807-5816.

ANANTHA, N. V., AZAM, M. & SHEARDY, R. D. 1998. Porphyrin binding to quadruplexed T(4)G(4). *Biochemistry,* 37**,** 2709-2714.

ARNOTT, S., CHANDRASEKARAN, R. & MARTTILA, C. M. 1974. Structures for polyinosinic acid and polyguanylic acid. *The Biochemical journal,* 141**,** 537-43.

ARORA, A. & MAITI, S. 2008. Effect of loop orientation on quadruplex-TMPyP4 interaction. *Journal of Physical Chemistry B,* 112**,** 8151-8159.

AZORIN, F., NORDHEIM, A. & RICH, A. 1983. Formation of Z-DNA in negatively supercoiled plasmids is sensitive to small changes in salt concentration within the physiological range. *Embo Journal,* 2**,** 649-655.

AZZALIN, C. M., REICHENBACH, P., KHORIAULI, L., GIULOTTO, E. & LINGNER, J. 2007. Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science,* 318**,** 798-801.

BALAGURUMOORTHY, P., BRAHMACHARI, S. K., MOHANTY, D., BANSAL, M. & SASISEKHARAN, V. 1992. Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Research,* 20**,** 4061-4067.

BALASUBRAMANIAN, S., HURLEY, L. H. & NEIDLE, S. 2011. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nature Reviews Drug Discovery,* 10**,** 261-275.

BALASUBRAMANIAN, S. & NEIDLE, S. 2009. G-quadruplex nucleic acids as therapeutic targets. *Curr Opin Chem Biol,* 13**,** 345-53.

BANG, I. 1910. Untersuchungen über die guanylsäre. *Biochemisce Zeitschrift,* 26**,** 293-311.

BATES, A. D. & MAXWELL, A. 2005. *DNA topology,* Oxford ; New York, Oxford University Press.

BAUER, W. R. 1978. Structure and reactions of closed duplex DNA *Annual Review of Biophysics and Bioengineering,* 7**,** 287-313.

BEARD, P., MORROW, J. F. & BERG, P. 1973. Cleavage of circular, superhelical simian virus 40 DNA to a linear duplex by S1 nuclease. *Journal of Virology,* 12**,** 1303-1313.

BEAUDOIN, J.-D. & PERREAULT, J.-P. 2010. 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Research,* 38**,** 7022-7036.

BEAUDOIN, J. D., JODOIN, R. & PERREAULT, J. P. 2013. New scoring system to identify RNA G-quadruplex folding. *Nucleic Acids Research*.

BERBERICH, S. J. & POSTEL, E. H. 1995. PuF/NM23-H2/NDPK-B transactivates a human c-myc promoter-CAT gene via a functional nuclease hypersensitive element. *Oncogene,* 10**,** 2343-2347.

BIFFI, G., TANNAHILL, D., MCCAFFERTY, J. & BALASUBRAMANIAN, S. 2013. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chemistry,* 5**,** 182-6.

BJORNSTI, M. A. & OSHEROFF, N. 1999. DNA topoisomerase protocols Volume I: DNA topology and enzymes - Introduction. *In:* BJORNSTI, M. A. & OSHEROFF, N. (eds.) *DNA Topoisomerase Protocols, Volume 1: DNA Topology and Enzymes.*

BLACKBURN G. MICHAEL., G. M. J., LOAKES DAVID., WILLIAMS, DAVID. M. (ed.) 2006. *Nucleic acids in chemistry and biology*.

BOLES, T. C. & HOGAN, M. E. 1987. DNA structure equilibria in the human c-myc gene. *Biochemistry,* 26**,** 367-376.

BOLIVAR, F., RODRIGUEZ, R. L., GREENE, P. J., BETLACH, M. C., HEYNEKER, H. L., BOYER, H. W., CROSA, J. H. & FALKOW, S. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene,* 2**,** 95-113.

BORMAN, S. 2009. Promoter quadruplexes. *Chemical & Engineering News,* 87**,** 28-30.

BORNER, C. 2003. The Bcl-2 protein family: sensors and checkpoints for life-or-death decisions. *Molecular Immunology,* 39**,** 615-647.

BOWATER, R., ABOULELA, F. & LILLEY, D. M. J. 1992. Two-dimensional gel electrophoresis of circular DNA topoisomers. *Methods in Enzymology,* 212**,** 105-120.

BOWATER, R. P. 2005. Supercoiled DNA structure. *Encyclopedia of Life Sciences*.

BROWN, T. A. 1994. *DNA sequencing,* BIOS.

BRYAN, T. M. & BAUMANN, P. 2011. G-Quadruplexes: From guanine gels to chemotherapeutics. *Molecular Biotechnology,* 49**,** 198-208.

BUGAUT, A. & BALASUBRAMANIAN, S. 2012. 5 '-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Research,* 40**,** 4727-4741.

BUI, C. T., REES, K. & COTTON, R. G. H. 2003. Permanganate oxidation reactions of DNA: Perspective in biological studies. *Nucleosides Nucleotides & Nucleic Acids,* 22**,** 1835-1855.

BURGE, S., PARKINSON, G. N., HAZEL, P., TODD, A. K. & NEIDLE, S. 2006. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research,* 34**,** 5402-5415.

BURGER, A. M., DAI, F. P., SCHULTES, C. M., RESZKA, A. P., MOORE, M. J., DOUBLE, J. A. & NEIDLE, S. 2005. The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Research,* 65**,** 1489-1496.

CAMPBELL, N. H., PARKINSON, G. N., RESZKA, A. P. & NEIDLE, S. 2008. Structural basis of DNA quadruplex recognition by an acridine drug. *Journal of the American Chemical Society,* 130**,** 6722-+.

CASHMAN, D. J., BUSCAGLIA, R., FREYER, M. W., DETTLER, J., HURLEY, L. H. & LEWIS, E. A. 2008. Molecular modeling and biophysical analysis of the c-MYC NHE-III1 silencer element. *Journal of Molecular Modeling,* 14**,** 93-101.

CHEN, L. Q., CAI, L., ZHANG, X. H. & RICH, A. 1994. Crystal structure of a four-stranded intercalated DNA: d(C4). *Biochemistry,* 33**,** 13540-13546.

CHEN, Q., KUNTZ, I. D. & SHAFER, R. H. 1996. Spectroscopic recognition of guanine dimeric hairpin quadruplexes by a carbocyanine dye. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 2635-2639.

CHEN, Y., AGRAWAL, P., BROWN, R. V., HATZAKIS, E., HURLEY, L. & YANG, D. 2012. The major G-quadruplex formed in the human platelet-derived growth factor receptor beta promoter adopts a novel broken-strand structure in K+ solution. *Journal of the American Chemical Society,* 134**,** 13220-3.

CHUNG, W. J., HEDDI, B., TERA, M., IIDA, K., NAGASAWA, K. & PHAN, A. T. 2013. Solution structure of an intramolecular (3 + 1) human telomeric G-quadruplex bound to a telomestatin derivative. *Journal of the American Chemical Society,* 135**,** 13495-501.

COGOI, S., PARAMASIVAM, M., SPOLAORE, B. & XODO, L. E. 2008. Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins. *Nucleic Acids Research,* 36**,** 3765-3780.

COLLIE, G. W., HAIDER, S. M., NEIDLE, S. & PARKINSON, G. N. 2010. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Research,* 38**,** 5569-5580.

COLLIE, G. W., SPARAPANI, S., PARKINSON, G. N. & NEIDLE, S. 2011. Structural basis of telomeric RNA quadruplex-acridine ligand recognition. *Journal of the American Chemical Society,* 133**,** 2721-2728.

CORY, S. & ADAMS, J. M. 2002. The BCL2 family: Regulators of the cellular life-or-death switch. *Nature Reviews Cancer,* 2**,** 647-656.

CRNUGELJ, M., HUD, N. V. & PLAVEC, J. 2002. The solution structure of d(G(4)T(4)G(3))(2): a bimolecular G-quadruplex with a novel fold. *Journal of Molecular Biology,* 320**,** 911-24.

CRNUGELJ, M., SKET, P. & PLAVEC, J. 2003. Small change in a G-rich sequence, a dramatic change in topology: new dimeric G-quadruplex folding motif with unique loop orientations. *Journal of the American Chemical Society,* 125**,** 7866-71.

DAI, J. X., CARVER, M., PUNCHIHEWA, C., JONES, R. A. & YANG, D. Z. 2007a. Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Research,* 35**,** 4927-4940.

DAI, J. X., CHEN, D., JONES, R. A., HURLEY, L. H. & YANG, D. Z. 2006a. NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Research,* 34**,** 5133-5144.

DAI, J. X., DEXHEIMER, T. S., CHEN, D., CARVER, M., AMBRUS, A., JONES, R. A. & YANG, D. Z. 2006b. An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *Journal of the American Chemical Society,* 128**,** 1096-1098.

DAI, J. X., PUNCHIHEWA, C., AMBRUS, A., CHEN, D., JONES, R. A. & YANG, D. Z. 2007b. Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Research,* 35**,** 2440-2450.

DANG, C. V. 1999. c-myc target genes involved in cell growth, apoptosis, and metabolism. *Molecular and Cellular Biology,* 19**,** 1-11.

DE ARMOND, R., WOOD, S., SUN, D. Y., HURLEY, L. H. & EBBINGHAUS, S. W. 2005. Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1 alpha promoter. *Biochemistry,* 44**,** 16341-16350.

DE CIAN, A., CRISTOFARI, G., REICHENBACH, P., DE LEMOS, E., MONCHAUD, D., TEULADE-FICHOU, M. P., SHIN-YA, K., LACROIX, L., LINGNER, J. & MERGNY, J. L. 2007. Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 17347-17352.

DERECKA, K., BALKWILL, G. D., GARNER, T. P., HODGMAN, C., FLINT, A. P. F. & SEARLE, M. S. 2010. Occurrence of a quadruplex motif in a unique insert within Exon C of the bovine estrogen receptor alpha gene (ESR1). *Biochemistry,* 49**,** 7625-7633.

DEXHEIMER, T. S., FRY, M. & HURLEY, L. H. 2006a. DNA quadruplexes and gene regulation. *In:* NEIDLE, S. & BALASUBRAMANIAN, S. (eds.) *Quadruplex Nucleic Acids.*

DEXHEIMER, T. S., SUN, D. & HURLEY, L. H. 2006b. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *Journal of the American Chemical Society,* 128**,** 5404-15.

DOI, T., SHIBATA, K., YOSHIDA, M., TAKAGI, M., TERA, M., NAGASAWA, K., SHIN-YA, K. & TAKAHASHI, T. 2011. (S)-Stereoisomer of telomestatin as a potent G-quadruplex binder and telomerase inhibitor. *Organic & Biomolecular Chemistry,* 9**,** 387-393.

DOI, T., YOSHIDA, M., SHIN-YA, K. & TAKAHASHI, T. 2006. Total synthesis of (R)-telomestatin. *Organic Letters,* 8**,** 4165-4167.

DU, Z., KONG, P., GAO, Y. & LI, N. 2007. Enrichment of G4 DNA motif in transcriptional regulatory region of chicken genome. *Biochemical and Biophysical Research Communications,* 354**,** 1067-1070.

EDDY, J. & MAIZELS, N. 2006. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Research,* 34**,** 3887-3896.

ERIKSSON, M. & NORDEN, B. 2001. Linear and circular dichroism of drug-nucleic acid complexes. *Drug-Nucleic Acid Interactions,* 340**,** 68-98.

ESMAILI, N. & LEROY, J. L. 2005. i-motif solution structure and dynamics of the d(AACCCC) and d(CCCCAA) tetrahymena telomeric repeats. *Nucleic Acids Research,* 33**,** 213-224.

FERNANDO, H., RESZKA, A. P., HUPPERT, J., LADAME, S., RANKIN, S., VENKITARAMAN, A. R., NEIDLE, S. & BALASUBRAMANIAN, S. 2006. A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. *Biochemistry,* 45**,** 7854-7860.

FROELICH-AMMON, S. J., DICKINSON, B. A., BEVILACQUA, J. M., SCHULTZ, S. C. & CECH, T. R. 1998. Modulation of telomerase activity by telomere DNA-binding proteins in Oxytricha. *Genes & Development,* 12**,** 1504-1514.

FRY, M. & LOEB, L. A. 1999. Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *The Journal of biological chemistry,* 274**,** 12797-802.

GAYNUTDINOV, T. I., NEUMANN, R. D. & PANYUTIN, I. G. 2008. Structural polymorphism of intramolecular quadruplex of human telomeric DNA: effect of cations, quadruplex-binding drugs and flanking sequences. *Nucleic Acids Research,* 36**,** 4079-4087.

GEHRING, K., LEROY, J. L. & GUERON, M. 1993. A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature,* 363**,** 561-565.

GELLERT, M., LIPSETT, M. N. & DAVIES, D. R. 1962. Helix formation by guanylic acid. *Proc Natl Acad Sci U S A,* 48**,** 2013-8.

GELLERT, M., ODEA, M. H. & MIZUUCHI, K. 1983. Slow cruciform transitions in palindromic DNA. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences,* 80**,** 5545-5549.

GEORGIADES, S. N., ABD KARIM, N. H., SUNTHARALINGAM, K. & VILAR, R. 2010. Interaction of metal complexes with G-quadruplex DNA. *Angew Chem Int Ed Engl,* 49**,** 4020-34.

GIAEVER, G. N., SNYDER, L. & WANG, J. C. 1988. DNA supercoiling in vivo. *Biophysical Chemistry,* 29**,** 7-15.

GIRALDO, R., SUZUKI, M., CHAPMAN, L. & RHODES, D. 1994. Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: A circular dichroism study. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 7658-7662.

GOMEZ, D., GUEDIN, A., MERGNY, J.-L., SALLES, B., RIOU, J.-F., TEULADE-FICHOU, M.-P. & CALSOU, P. 2010. A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Research,* 38**,** 7187-7198.

GOTTSCHLING, D. E. & CECH, T. R. 1984. Chromatin structure of the molecular ends of Oxytricha macronuclear DNA: phased nucleosomes and a telomeric complex. *Cell,* 38**,** 501-510.

GOTTSCHLING, D. E. & ZAKIAN, V. A. 1986. Telomere proteins: specific recognition and protection of the natural termini of Oxytricha macronuclear DNA. *Cell,* 47**,** 195-205.

GOWAN, S. M., HARRISON, J. R., PATTERSON, L., VALENTI, M., READ, M. A., NEIDLE, S. & KELLAND, L. R. 2002. A G-quadruplex-interactive potent small-molecule inhibitor of telomerase exhibiting in vitro and in vivo antitumor activity. *Molecular Pharmacology,* 61**,** 1154-1162.

GRAND, C. L., HAN, H. Y., MUNOZ, R. M., WEITMAN, S., VON HOFF, D. D., HURLEY, L. H. & BEARSS, D. J. 2002. The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Molecular Cancer Therapeutics,* 1**,** 565-573.

GRANZHAN, A., MONCHAUD, D., SAETTEL, N., GUEDIN, A., MERGNY, J.-L. & TEULADE-FICHOU, M.-P. 2010. "One Ring to Bind Them All"-Part II: Identification of Promising G-Quadruplex Ligands by Screening of Cyclophane-Type Macrocycles. *Journal of nucleic acids,* 2010.

GRAY, D. M., WEN, J. D., GRAY, C. W., REPGES, R., REPGES, C., RAABE, G. & FLEISCHHAUER, J. 2008. Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality,* 20**,** 431-440.

GREIDER, C. W. & BLACKBURN, E. H. 1985. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell,* 43**,** 405-413.

GUO, Q., LU, M., MARKY, L. A. & KALLENBACH, N. R. 1992. Interaction of the dye ethidium bromide with DNA containing guanine repeats. *Biochemistry,* 31**,** 2451-2455.

HAIDER, S., PARKINSON, G. N. & NEIDLE, S. 2002. Crystal structure of the potassium form of an Oxytricha nova G-quadruplex. *Journal of Molecular Biology,* 320**,** 189-200.

HAIDER, S. M., NEIDLE, S. & PARKINSON, G. N. 2011. A structural analysis of G-quadruplex/ligand interactions. *Biochimie,* 93**,** 1239-51.

HALDER, K. & HARTIG, J. S. 2011. RNA quadruplexes. *In:* SIGEL, A. S. H. S. R. K. O. (ed.) *Structural and Catalytic Roles of Metal Ions in Rna.*

HALDER, K., WIELAND, M. & HARTIG, J. S. 2009. Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes. *Nucleic Acids Research,* 37**,** 6811-6817.

HAN, F. X. G., WHEELHOUSE, R. T. & HURLEY, L. H. 1999a. Interactions of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition. *Journal of the American Chemical Society,* 121**,** 3561-3570.

HAN, H., CLIFF, C. L. & HURLEY, L. H. 1999b. Accelerated assembly of G-quadruplex structures by a small molecule. *Biochemistry,* 38**,** 6981-6.

HAYATSU, H. & UKITA, T. 1967. The selective degradation of pyrimidines in nucleic acids by permanganate oxidation. *Biochem Biophys Res Commun,* 29**,** 556-561.

HAZEL, P., HUPPERT, J., BALASUBRAMANIAN, S. & NEIDLE, S. 2004. Loop-length-dependent folding of G-quadruplexes. *Journal of the American Chemical Society,* 126**,** 16405-16415.

HAZEL, P., PARKINSON, G. N. & NEIDLE, S. 2006. Topology variation and loop structural homology in crystal and simulated structures of a bimolecular DNA quadruplex. *Journal of the American Chemical Society,* 128**,** 5480-5487.

HE, Y. J., NEUMANN, R. D. & PANYUTIN, I. G. 2004. Intramolecular quadruplex conformation of human telomeric DNA assessed with I-125-radioprobing. *Nucleic Acids Research,* 32**,** 5359-5367.

HENDERSON, A., WU, Y., HUANG, Y. C., CHAVEZ, E. A., PLATT, J., JOHNSON, F. B., BROSH, R. M., JR., SEN, D. & LANSDORP, P. M. 2013. Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Research.*

HENDERSON, E., HARDIN, C. C., WALK, S. K., TINOCO, I., JR. & BLACKBURN, E. H. 1987. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell,* 51**,** 899-908.

HERBERT, A. & RICH, A. 1996. The biology of left-handed Z-DNA. *Journal of Biological Chemistry,* 271**,** 11595-11598.

HERSHMAN, S. G., CHEN, Q., LEE, J. Y., KOZAK, M. L., YUE, P., WANG, L.-S. & JOHNSON, F. B. 2008. Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. *Nucleic Acids Research,* 36**,** 144-156.

HORVATH, M. P. & SCHULTZ, S. C. 2001. DNA G-quartets in a 1.86 angstrom resolution structure of an Oxytricha nova telomeric protein-DNA complex. *Journal of Molecular Biology,* 310**,** 367-377.

HOWELL, R. M., WOODFORD, K. J., WEITZMANN, M. N. & USDIN, K. 1996. The chicken beta-globin gene promoter forms a novel "cinched" tetrahelical structure. *Journal of Biological Chemistry,* 271**,** 5208-5214.

HOWGATE, P., JONES, A. S. & TITTENSOR, J. R. 1968. The permanganate oxidation of thymidine. *Journal of the Chemical Society***,** 275 - 279.

HSU, S. T. D., VARNAI, P., BUGAUT, A., RESZKA, A. P., NEIDLE, S. & BALASUBRAMANIAN, S. 2009. A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. *Journal of the American Chemical Society,* 131**,** 13399-13409.

HTUN, H. & DAHLBERG, J. E. 1989. Topology and formation of triple-stranded H-DNA. *Science,* 243**,** 1571-6.

HUD, N. V. & PLAVEC, J. 2006. *The role of cations in determining quadruplex structure and stability*.

HUPPERT, J. 2006. Quadruplexes in the genome. *In:* NEIDLE, S. B. S. (ed.) *Quadruplex Nucleic Acids.*

HUPPERT, J. L. 2008. Hunting G-quadruplexes. *Biochimie,* 90**,** 1140-1148.

HUPPERT, J. L. & BALASUBRAMANIAN, S. 2005. Prevalence of quadruplexes in the human genome. *Nucleic Acids Research,* 33**,** 2908-2916.

HUPPERT, J. L. & BALASUBRAMANIAN, S. 2007. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Research,* 35**,** 406-413.

HUPPERT, J. L., BUGAUT, A., KUMARI, S. & BALASUBRAMANIAN, S. 2008. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Research,* 36**,** 6260-6268.

HUTCHISON, C. A. 2007. DNA sequencing: bench to bedside and beyond. *Nucleic Acids Research,* 35**,** 6227-6237.

INCLES, C. M., SCHULTES, C. M., KEMPSKI, H., KOEHLER, H., KELLAND, L. R. & NEIDLE, S. 2004. A G-quadruplex telomere targeting agent produces p16-associated senescence and chromosomal fusions in human prostate cancer cells. *Molecular Cancer Therapeutics,* 3**,** 1201-1206.

JIANG, H., JU, Z. & RUDOLPH, K. L. 2007. Telomere shortening and ageing. *Zeitschrift Fur Gerontologie Und Geriatrie,* 40**,** 314-324.

JOACHIMI, A., BENZ, A. & HARTIG, J. S. 2009. A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorganic & Medicinal Chemistry,* 17**,** 6811-6815.

JUDE, K. M., HARTLAND, A. & BERGER, J. M. 2013. Real-time detection of DNA topological changes with a fluorescently labeled cruciform. *Nucleic Acids Research,* 41.

KAHL, B. F. & PAULE, M. R. 2001. The use of diethyl pyrocarbonate and potassium permanganate as probes for strand separation and structural distortions in DNA. *Methods Mol Biol,* 148**,** 63-75.

KENDRICK, S. & HURLEY, L. H. 2010. The role of G-quadruplex/i-motif secondary structures as cis-acting regulatory elements. *Pure and Applied Chemistry,* 82**,** 1609-1621.

KENIRY, M. A. 2000. Quadruplex structures in nucleic acids. *Biopolymers,* 56**,** 123-146.

KENIRY, M. A., OWEN, E. A. & SHAFER, R. H. 1997. The contribution of thymine-thymine interactions to the stability of folded dimeric quadruplexes. *Nucleic Acids Research,* 25**,** 4389-4392.

KENIRY, M. A., STRAHAN, G. D., OWEN, E. A. & SHAFER, R. H. 1995. Solution structure of the Na+ form of the dimeric guanine quadruplex [d(G3T4G3)]2. *European journal of biochemistry / FEBS,* 233**,** 631-43.

KIM, M. Y., GLEASON-GUZMAN, M., IZBICKA, E., NISHIOKA, D. & HURLEY, L. H. 2003. The different biological effects of telomestatin and TMPyP4 can be attributed to their selectivity for interaction with intramolecular or intermolecular G-quadruplex structures. *Cancer Research,* 63**,** 3247-3256.

KIM, M. Y., VANKAYALAPATI, H., KAZUO, S., WIERZBA, K. & HURLEY, L. H. 2002. Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular G-quadruplex. *Journal of the American Chemical Society,* 124**,** 2098-2099.

KRISHNAN-GHOSH, Y., LIU, D. S. & BALASUBRAMANIAN, S. 2004. Formation of an interlocked quadruplex dimer by d(GGGT). *Journal of the American Chemical Society,* 126**,** 11009-11016.

KUMARI, S., BUGAUT, A. & BALASUBRAMANIAN, S. 2008. Position and stability are determining factors for translation repression by an RNA G-quadruplex-forming sequence within the 5 ' UTR of the NRAS proto-oncogene. *Biochemistry,* 47**,** 12664-12669.

KUMARI, S., BUGAUT, A., HUPPERT, J. L. & BALASUBRAMANIAN, S. 2007. An RNA G-quadruplex in the 5 ' UTR of the NRAS proto-oncogene modulates translation. *Nature Chemical Biology,* 3**,** 218-221.

KURYAVYI, V., PHAN, A. T. & PATEL, D. J. 2010. Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter. *Nucleic Acids Research,* 38**,** 6757-6773.

LANE, A. N., CHAIRES, J. B., GRAY, R. D. & TRENT, J. O. 2008. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Research,* 36**,** 5482-5515.

LAUGHLAN, G., MURCHIE, A. I. H., NORMAN, D. G., MOORE, M. H., MOODY, P. C. E., LILLEY, D. M. J. & LUISI, B. 1994. The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science,* 265**,** 520-524.

LEROY, J. L., GUERON, M., MERGNY, J. L. & HELENE, C. 1994. Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an *i-motif*. *Nucleic Acids Research,* 22**,** 1600-1606.

LI, J., CORREIA, J. J., WANG, L., TRENT, J. O. & CHAIRES, J. B. 2005. Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Research,* 33**,** 4649-4659.

LI, Q., XIANG, J.-F., YANG, Q.-F., SUN, H.-X., GUAN, A.-J. & TANG, Y.-L. 2013. G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Research,* 41**,** D1115-D1123.

LI, W., WU, P., OHMICHI, T. & SUGIMOTO, N. 2002. Characterization and thermodynamic properties of quadruplex/duplex competition. *Febs Letters,* 526**,** 77-81.

LI, X. M., NG, M. T. T., WANG, Y. F., ZHOU, T. Y., CHUA, S. T., YUAN, W. X. & LI, T. H. 2008. Site-specific self-cleavage of G-quadruplexes formed by human telemetric repeats. *Bioorganic & Medicinal Chemistry Letters,* 18**,** 5576-5580.

LILLEY, D. M. J. 1980. The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences,* 77**,** 6468-6472.

LILLEY, D. M. J. 1983. Structural perturbation in supercoiled DNA: hypersensitivity to modification by a single-strand-selective chemical reagent conferred by inverted repeat sequences. *Nucleic Acids Research,* 11**,** 3097-3112.

LILLEY, D. M. J. 1984. DNA: sequence, structure and supercoiling. *Biochemical Society Transactions,* 12**,** 127-140.

LILLEY, D. M. J. 1992. Probes of DNA-structure. *Methods in Enzymology,* 212**,** 133-139.

LILLEY, D. M. J. & HALLAM, L. R. 1984. Thermodynamics of the Cole1 cruciform - comparisons between probing and topological experiments using single topoisomers. *Journal of Molecular Biology,* 180**,** 179-200.

LILLEY, D. M. J. & PALECEK, E. 1984. The supercoil-stabilised cruciform of ColE1 is hyper-reactive to osmium tetroxide. *Embo Journal,* 3**,** 1187-1192.

LIM, K. W., AMRANE, S., BOUAZIZ, S., XU, W., MU, Y., PATEL, D. J., LUU, K. N. & PHAN, A. T. 2009. Structure of the human telomere in K+ solution: A stable basket-type G-quadruplex with only two G-tetrad layers. *Journal of the American Chemical Society,* 131**,** 4301-4309.

LIM, K. W., NG, V. C., MARTIN-PINTADO, N., HEDDI, B. & PHAN, A. T. 2013. Structure of the human telomere in Na+ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Research*.

LINDER, J., GARNER, T. P., WILLIAMS, H. E. L., SEARLE, M. S. & MOODY, C. J. 2011. Telomestatin: formal total synthesis and cation-mediated interaction of Its seco-derivatives with G-quadruplexes. *Journal of the American Chemical Society,* 133**,** 1044-1051.

LIPPS, H. J. & RHODES, D. 2009. G-quadruplex structures: in vivo evidence and function. *Trends in Cell Biology,* 19**,** 414-422.

LIU, L. F. & WANG, J. C. 1987. Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences of the United States of America,* 84**,** 7024-7027.

LIU, X. Q., LI, X. M., ZHOU, T. Y., WANG, Y. F., NG, M. T. T., XU, W. & LI, T. H. 2008. Site specific self-cleavage of certain assemblies of G-quadruplex. *Chemical Communications***,** 380-382.

LU, M., GUO, Q. & KALLENBACH, N. R. 1993. Thermodynamics of G-tetraplex formation by telomeric DNAs. *Biochemistry,* 32**,** 598-601.

LUEDTKE, N. W. 2009. Targeting G-quadruplex DNA with small molecules. *Chimia,* 63**,** 134-139.

LUKE, B. & LINGNER, J. 2009. TERRA: telomeric repeat-containing RNA. *Embo Journal,* 28**,** 2503-2510.

LUU, K. N., PHAN, A. T., KURYAVYI, V., LACROIX, L. & PATEL, D. J. 2006. Structure of the human telomere in K+ solution: An intramolecular (3+1) G-quadruplex scaffold. *Journal of the American Chemical Society,* 128**,** 9963-9970.

LYAMICHEV, V. I., MIRKIN, S. M. & FRANK-KAMEMETSKII, M. D. 1986. Structures of homopurine-homopyrimidine tract in superhelical DNA. *Journal of Biomolecular Structure & Dynamics,* 3**,** 667-669.

LYAMICHEV, V. I., PANYUTIN, I. G. & FRANK-KAMENETSKII, M. D. 1983. Evidence of cruciform structures in superhelical DNA provided by two-dimensional gel electrophoresis. *Febs Letters,* 153**,** 298-302.

MAIZELS, N. 2006. Quadruplexes and the biology of G-rich genomic regions. *In:* NEIDLE, S. B. S. (ed.) *Quadruplex Nucleic Acids.*

MARCU, K. B., BOSSONE, S. A. & PATEL, A. J. 1992. MYC Function and regulation. *Annual Review of Biochemistry,* 61**,** 809-860.

MARTADINATA, H. & PHAN, A. T. 2009. Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K+ solution. *Journal of the American Chemical Society,* 131**,** 2570-2578.

MATTES, W. B., HARTLEY, J. A. & KOHN, K. W. 1986. Mechanism of DNA strand breakage by piperidine at sites of N7-alkylguanines. *Biochimica Et Biophysica Acta,* 868**,** 71-76.

MAXAM, A. M. & GILBERT, W. 1977. New method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America,* 74**,** 560-564.

MCCARTHY, J. G. & HEYWOOD, S. M. 1987. A long polypyrimidine/polypurine tract induces an altered DNA conformation on the 3' coding region of the adjacent myosin heavy chain gene. *Nucleic Acids Research,* 15**,** 8069-8085.

MCEACHERN, M. J., KRAUSKOPF, A. & BLACKBURN, E. H. 2000. Telomeres and their control. *Annual Review of Genetics,* 34**,** 331-358.

MCLEAN, M. J., LEE, J. W. & WELLS, R. D. 1988. Characteristics of Z-DNA helices formed by imperfect (purine-pyrimidine) sequences in plasmids. *Journal of Biological Chemistry,* 263**,** 7378-7385.

MCLEAN, M. J. & WELLS, R. D. 1988. The role of DNA sequence in the formation of Z-DNA versus cruciforms in plasmids. *Journal of Biological Chemistry,* 263**,** 7370-7377.

MERGNY, J. L., DE CIAN, A., GHELAB, A., SACCA, B. & LACROIX, L. 2005. Kinetics of tetramolecular quadruplexes. *Nucleic Acids Research,* 33**,** 81-94.

MERGNY, J. L., LACROIX, L., HAN, X. G., LEROY, J. L. & HELENE, C. 1995. Intramolecular folding of pyrimidine oligodeoxynucleotides into an i-DNA motif. *Journal of the American Chemical Society,* 117**,** 8887-8898.

MIRKIN, S. M. 2001. DNA topology: fundamentals. *ENCYCLOPEDIA OF LIFE SCIENCES*.

MIRKIN, S. M. & FRANKKAMENETSKII, M. D. 1994. H-DNA and related structures. *Annual Review of Biophysics and Biomolecular Structure,* 23**,** 541-576.

MIRKIN, S. M., LYAMICHEV, V. I., DRUSHLYAK, K. N., DOBRYNIN, V. N., FILIPPOV, S. A. & FRANK-KAMENETSKII, M. D. 1987. DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature,* 330**,** 495-7.

MIYOSHI, D., MATSUMURA, S., NAKANO, S. & SUGIMOTO, N. 2004. Duplex dissociation of telomere DNAs induced by molecular crowding. *Journal of the American Chemical Society,* 126**,** 165-169.

MIZUUCHI, K., MIZUUCHI, M. & GELLERT, M. 1982. Cruciform structures in palindromic DNA are favored by DNA supercoiling. *Journal of Molecular Biology,* 156**,** 229-243.

MONCHAUD, D. & TEULADE-FICHOU, M. P. 2008. A hitchhiker's guide to G-quadruplex ligands. *Organic & Biomolecular Chemistry,* 6**,** 627-636.

MORRIS, M. J., NEGISHI, Y., PAZSINT, C., SCHONHOFT, J. D. & BASU, S. 2010. An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. *Journal of the American Chemical Society,* 132**,** 17831-17839.

MUKUNDAN, V. T. & PHAN, A. T. 2013. Bulges in G-quadruplexes: Broadening the definition of G-quadruplex-forming sequences. *Journal of the American Chemical Society*.

MURCHIE, A. I. H. & LILLEY, D. M. J. 1992. Supercoiled DNA and cruciform structures. *Methods in Enzymology,* 211**,** 158-180.

NEIDLE, S. 2009. The structures of quadruplex nucleic acids and their drug complexes. *Current Opinion in Structural Biology,* 19**,** 239-250.

NEIDLE, S. 2012. *Therapeutic applications of quadruplex nucleic acids*.

NELSON, H. C. M. & KLUG, A. 1988. Anomalous properties of adenine-thymine tracts. *Nature,* 332**,** 117-117.

NG, M. T. T., LI, X., WANG, Y., ZHOU, T., YANG, Z., FOO, H. Y. & LI, T. 2009. Site-specific cleavage of G-quadruplexes formed by Oxytricha telometric repeats. *Australian Journal of Chemistry,* 62**,** 1189-1193.

NORDHEIM, A., PECK, L. J., LAFER, E. M., STOLLAR, B. D., WANG, J. C. & RICH, A. 1983. Supercoiling and left-handed Z-DNA. *Cold Spring Harbor Symposia on Quantitative Biology,* 47 Pt 1**,** 93-100.

OLIVAS, W. M. & MAHER, L. J. 1995. Competitive triplex/quadruplex equilibria involving guanine-rich oligonucleotides. *Biochemistry,* 34**,** 278-284.

OLTVAI, Z. N., MILLIMAN, C. L. & KORSMEYER, S. J. 1993. Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programmed cell death. *Cell,* 74**,** 609-619.

ONYSHCHENKO, M. I., GAYNUTDINOV, T. I., ENGLUND, E. A., APPELLA, D. H., NEUMANN, R. D. & PANYUTIN, I. G. 2009. Stabilization of G-quadruplex in the BCL2 promoter region in double-stranded DNA by invading short PNAs. *Nucleic Acids Research,* 37**,** 7570-7580.

ONYSHCHENKO, M. I., GAYNUTDINOV, T. I., ENGLUND, E. A., APPELLA, D. H., NEUMANN, R. D. & PANYUTIN, I. G. 2011. Quadruplex formation is necessary for stable PNA invasion into duplex DNA of BCL2 promoter region. *Nucleic Acids Research,* 39**,** 7114-7123.

OU, T. M., LU, Y. J., TAN, J. H., HUANG, Z. S., WONG, K. Y. & GU, L. Q. 2008. G-quadruplexes: Targets in anticancer drug design. *Chemmedchem,* 3**,** 690-713.

PALECEK, E. 1991. Local supercoil-stabilized DNA structures. *Critical Reviews in Biochemistry and Molecular Biology,* 26**,** 151-226.

PANAYOTATOS, N. & FONTAINE, A. 1987. A native cruciform DNA structure probed in bacteria by recombinant T7 endonuclease. *Journal of Biological Chemistry,* 262**,** 11364-11368.

PANAYOTATOS, N. & WELLS, R. D. 1981. Cruciform structures in supercoiled DNA. *Nature,* 289**,** 466-470.

PANDEY, S., AGARWALA, P. & MAITI, S. 2013. Effect of loops and G-quartets on the stability of RNA G-quadruplexes. *Journal of Physical Chemistry B,* 117**,** 6896-6905.

PARAMASIVAN, S., RUJAN, I. & BOLTON, P. H. 2007. Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Methods,* 43**,** 324-331.

PARKINSON, G. N. 2006. Fundamentals of quadruplex structures. *Quadruplex Nucleic Acids***,** 1-30.

PARKINSON, G. N., LEE, M. P. H. & NEIDLE, S. 2002. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature,* 417**,** 876-880.

PARKINSON, G. N. & NEIDLE, S. 2007. Structural basis for porphyrin drug binding to human telomeres. *Proceedings of the American Association for Cancer Research Annual Meeting,* 48**,** 1161.

PATEL, D. J., PHAN, A. T. & KURYAVYI, V. 2007. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Research,* 35**,** 7429-55.

PECK, L. J., NORDHEIM, A., RICH, A. & WANG, J. C. 1982. FLIPPING OF CLONED D(PCPG)N.D(PCPG)N DNA-SEQUENCES FROM RIGHT-HANDED TO LEFT-HANDED HELICAL STRUCTURE BY SALT, CO(III), OR NEGATIVE SUPERCOILING. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences,* 79**,** 4560-4564.

PECK, L. J. & WANG, J. C. 1983. Energetics of B-to-Z transition in DNA. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences,* 80**,** 6206-6210.

PHAN, A. T. 2010. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *The FEBS journal,* 277**,** 1107-17.
PHAN, A. T., KURYAVYI, V., BURGE, S., NEIDLE, S. & PATEL, D. J. 2007. Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *Journal of the American Chemical Society,* 129**,** 4386-4392.
PHAN, A. T., KURYAVYI, V., GAW, H. Y. & PATEL, D. J. 2005. Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nature Chemical Biology,* 1**,** 234-234.
PHAN, A. T., LUU, K. N. & PATEL, D. J. 2006. Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K+ solution. *Nucleic Acids Research,* 34**,** 5715-5719.
PHAN, A. T. & MERGNY, J. L. 2002. Human telomeric DNA: G-quadruplex, i-motif and watson-crick double helix. *Nucleic Acids Research,* 30**,** 4618-4625.
PHAN, A. T., MODI, Y. S. & PATEL, D. J. 2004a. Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *Journal of the American Chemical Society,* 126**,** 8710-8716.
PHAN, A. T., MODI, Y. S. & PATEL, D. J. 2004b. Two-repeat Tetrahymena telomeric d(TGGGGTTGGGGT) sequence interconverts between asymmetric dimeric G-quadruplexes in solution. *Journal of Molecular Biology,* 338**,** 93-102.
PHAN, A. T. & PATEL, D. J. 2003. Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *Journal of the American Chemical Society,* 125**,** 15021-7.
POSTEL, E. H., FLINT, S. J., KESSLER, D. J. & HOGAN, M. E. 1991. Evidence that a triplex-forming oligodeoxyribonucleotide binds to the c-myc promoter in HeLa cells, thereby reducing c-myc mRNA levels. *Proceedings of the National Academy of Sciences of the United States of America,* 88**,** 8227-8231.
POTAMAN, V. N., OUSSATCHEVA, E. A., LYUBCHENKO, Y. L., SHLYAKHTENKO, L. S., BIDICHANDANI, S. I., ASHIZAWA, T. & SINDEN, R. R. 2004. Length-dependent structure formation in Friedreich ataxia (GAA)n·(TTC)n repeats at neutral pH. *Nucleic Acids Research,* 32**,** 1224-1231.
QIN, Y. & HURLEY, L. H. 2008. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie,* 90**,** 1149-1171.
QIN, Y., REZLER, E. M., GOKHALE, V., SUN, D. & HURLEY, L. H. 2007. Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Research,* 35**,** 7698-7713.
RACHWAL, P. A., BROWN, T. & FOX, K. R. 2007a. Sequence effects of single base loops in intramolecular quadruplex DNA. *Febs Letters,* 581**,** 1657-1660.
RACHWAL, P. A., FINDLOW, I. S., WERNER, J. M., BROWN, T. & FOX, K. R. 2007b. Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Research,* 35**,** 4214-4222.
RACHWAL, P. A. & FOX, K. R. 2007. Quadruplex melting. *Methods,* 43**,** 291-301.
RANDALL, A. & GRIFFITH, J. D. 2009. Structure of long telomeric RNA transcripts: the G-rich RNA forms a compact repeating structure containing G-quartets. *Journal of Biological Chemistry,* 284**,** 13980-13986.
RANDAZZO, A., SPADA, G. P. & DA SILVA, M. W. 2013. Circular dichroism of quadruplex structures. *Topics in current chemistry,* 330**,** 67-86.

RANKIN, S., RESZKA, A. P., HUPPERT, J., ZLOH, M., PARKINSON, G. N., TODD, A. K., LADAME, S., BALASUBRAMANIAN, S. & NEIDLE, S. 2005. Putative DNA quadruplex formation within the human c-kit oncogene. *Journal of the American Chemical Society,* 127**,** 10584-10589.

READ, M., HARRISON, R. J., ROMAGNOLI, B., TANIOUS, F. A., GOWAN, S. H., RESZKA, A. P., WILSON, W. D., KELLAND, L. R. & NEIDLE, S. 2001. Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 4844-4849.

REDON, S., BOMBARD, S., ELIZONDO-RIOJAS, M. A. & CHOTTARD, J. C. 2003. Platinum cross-linking of adenines and guanines on the quadruplex structures of the AG(3)(T(2)AG(3))(3) and (T(2)AG(3))(4) human telomere sequences in Na+ and K+ solutions. *Nucleic Acids Research,* 31**,** 1605-1613.

RICH, A. DNA comes in many forms. COGENE Symposium - From the Double Helix to the Human Genome: 40 Years of Molecular Genetics, Apr 21-23 1993 Paris, France. Elsevier Science Bv, 99-109.

RISITANO, A. & FOX, K. R. 2003a. The stability of intramolecular DNA quadruplexes with extended loops forming inter- and intra-loop duplexes. *Organic & Biomolecular Chemistry,* 1**,** 1852-1855.

RISITANO, A. & FOX, K. R. 2003b. Stability of intramolecular DNA quadruplexes: Comparison with DNA duplexes. *Biochemistry,* 42**,** 6507-6513.

RISITANO, A. & FOX, K. R. 2004. Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Research,* 32**,** 2598-2606.

RISITANO, A. & FOX, K. R. 2005. Inosine substitutions demonstrate that intramolecular DNA quadruplexes adopt different conformations in the presence of sodium and potassium. *Bioorganic & Medicinal Chemistry Letters,* 15**,** 2047-2050.

ROCA, J. 1995. THE MECHANISMS OF DNA TOPOISOMERASES. *Trends in Biochemical Sciences,* 20**,** 156-160.

RZUCZEK, S. G., PILCH, D. S., LAVOIE, E. J. & RICE, J. E. 2008. Lysinyl macrocyclic hexaoxazoles: Synthesis and selective G-quadruplex stabilizing properties. *Bioorganic & Medicinal Chemistry Letters,* 18**,** 913-917.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America,* 74**,** 5463-5467.

SASSE-DWIGHT, S. & GRALLA, J. D. 1989. KMnO4 as a probe for lac promoter DNA melting and mechanism in vivo. *Journal of Biological Chemistry,* 264**,** 8074-8081.

SCHOEFTNER, S. & BLASCO, M. A. 2008. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nature Cell Biology,* 10**,** 228-U106.

SCHOEFTNER, S. & BLASCO, M. A. 2009. A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. *Embo Journal,* 28**,** 2323-2336.

SCHULTZE, P., HUD, N. V., SMITH, F. W. & FEIGON, J. 1999. The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the Oxytricha nova telomere repeat oligonucleotide d(G(4)T(4)G(4)). *Nucleic Acids Research,* 27**,** 3018-3028.

SCHULTZE, P., SMITH, F. W. & FEIGON, J. 1994. Refined solution structure of the dimeric quadruplex formed from the Oxytricha telomeric oligonucleotide d(GGGGTTTTGGGG). *Structure,* 2**,** 221-233.

SEN, D. & GILBERT, W. 1988. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature,* 334**,** 364-366.

SEN, D. & GILBERT, W. 1990. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature,* 344**,** 410-414.

SHAHID, R., BUGAUT, A. & BALASUBRAMANIAN, S. 2010. The BCL-2 5 ' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry,* 49**,** 8300-8306.

SHAMMAS, M. A., REIS, R. J. S., LI, C., KOLEY, H., HURLEY, L. H., ANDERSON, K. C. & MUNSHI, N. C. 2004. Telomerase inhibition and cell growth arrest after telomestatin treatment in multiple myeloma. *Clinical Cancer Research,* 10**,** 770-776.

SHIM, J. W., TAN, Q. L. & GU, L. Q. 2009. Single-molecule detection of folding and unfolding of the G-quadruplex aptamer in a nanopore nanocavity. *Nucleic Acids Research,* 37**,** 972-982.

SHIN-YA, K., WIERZBA, K., MATSUO, K., OHTANI, T., YAMADA, Y., FURIHATA, K., HAYAKAWA, Y. & SETO, H. 2001. Telomestatin, a novel telomerase inhibitor from Streptomyces anulatus. *Journal of the American Chemical Society,* 123**,** 1262-1263.

SIDDIQUI-JAIN, A., GRAND, C. L., BEARSS, D. J. & HURLEY, L. H. 2002. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 11593-11598.

SIMONSSON, T. 2001. G-quadruplex DNA structures - Variations on a theme. *Biological Chemistry,* 382**,** 621-628.

SIMONSSON, T., PECINKA, P. & KUBISTA, M. 1998. DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Research,* 26**,** 1167-1172.

SINDEN, R. R. 1994. *DNA structure and function* London, Academic Press

SMIRNOV, I. & SHAFER, R. H. 2000. Effect of loop sequence and size on DNA aptamer stability. *Biochemistry,* 39**,** 1462-1468.

SMITH, F. W. & FEIGON, J. 1992. Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. *Nature,* 356**,** 164-168.

SMITH, F. W., LAU, F. W. & FEIGON, J. 1994. d(G3T4G3) forms an asymmetric diagonally looped dimeric quadruplex with guanosine 5'-syn-syn-anti and 5'-syn-anti-anti N-glycosidic conformations. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 10546-10550.

SPICUGLIA, S., KUMAR, S., CHASSON, L., PAYET-BORNET, D. & FERRIER, P. 2004. Potassium permanganate as a probe to map DNA-protein interactions in vivo. *Journal of Biochemical and Biophysical Methods,* 59**,** 189-194.

SUN, D. 2010. In vitro footprinting of promoter regions within supercoiled plasmid DNA. *Drug-DNA Interaction Protocols, Second Edition***,** 223-233.

SUN, D., GUO, K. & SHIN, Y.-J. 2011. Evidence of the formation of G-quadruplex structures in the promoter region of the human vascular endothelial growth factor gene. *Nucleic Acids Research,* 39**,** 1256-1265.

SUN, D. & HURLEY, L. H. 2009. The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-myc promoter: Implications for drug targeting and control of gene expression. *Journal of Medicinal Chemistry,* 52**,** 2863-2874.

SUN, D., LIU, W. J., GUO, K. X., RUSCHE, J. J., EBBINGHAUS, S., GOKHALE, V. & HURLEY, L. H. 2008. The proximal promoter region of the human vascular

endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interactive agents. *Molecular Cancer Therapeutics,* 7**,** 880-889.

SUN, D. Y., GUO, K. X., RUSCHE, J. J. & HURLEY, L. H. 2005. 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. *Nucleic Acids Research,* 33**,** 6070-6080.

SUN, D. Y., POURPAK, A., BEETZ, K. & HURLEY, L. H. 2003. Direct evidence for the formation of G-quadruplex in the proximal promoter region of the RET protooncogene and its targeting with a small molecule to repress RET protooncogene transcription. *Clinical Cancer Research,* 9**,** 6122S-6123S.

SUN, D. Y., THOMPSON, B., CATHERS, B. E., SALAZAR, M., KERWIN, S. M., TRENT, J. O., JENKINS, T. C., NEIDLE, S. & HURLEY, L. H. 1997. Inhibition of human telomerase by a G-quadruplex-interactive compound. *Journal of Medicinal Chemistry,* 40**,** 2113-2116.

SUN, H., KAROW, J. K., HICKSON, I. D. & MAIZELS, N. 1998. The Bloom's syndrome helicase unwinds G4 DNA. *Journal of Biological Chemistry,* 273**,** 27587-27592.

SUNDQUIST, W. I. & KLUG, A. 1989. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature,* 342**,** 825-829.

SUNTER, G., BUCK, K. W. & COUTTS, R. H. A. 1985. S1-sensitive sites in the supercoiled double-stranded form of tomato golden mosaic virus DNA component B: identification of regions of potential alternative secondary structure and regulatory function. *Nucleic Acids Research,* 13**,** 4645-4659.

TAETZ, S., BALDES, C., MURDTER, T. E., KLEIDEITER, E., PIOTROWSKA, K., BOCK, U., HALTNER-UKOMADU, E., MUELLER, J., HUWER, H., SCHAEFER, U. F., KLOTZ, U. & LEHR, C. M. 2006. Biopharmaceutical characterization of the telomerase inhibitor BRACO19. *Pharmaceutical research,* 23**,** 1031-7.

TAN, J.-H., GU, L.-Q. & WU, J.-Y. 2008. Design of selective G-quadruplex ligands as potential anticancer agents. *Mini-Reviews in Medicinal Chemistry,* 8**,** 1163-1178.

TODD, A. K. 2007. Bioinformatics approaches to quadruplex sequence location. *Methods,* 43**,** 246-251.

TODD, A. K., JOHNSTON, M. & NEIDLE, S. 2005. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Research,* 33**,** 2901-2907.

TODD, A. K. & NEIDLE, S. 2011. Mapping the sequences of potential guanine quadruplex motifs. *Nucleic Acids Research,* 39**,** 4917-4927.

TONG, X., LAN, W., ZHANG, X., WU, H., LIU, M. & CAO, C. 2011. Solution structure of all parallel G-quadruplex formed by the oncogene RET promoter sequence. *Nucleic Acids Research,* 39**,** 6753-6763.

TSUJIMOTO, Y. & SHIMIZU, S. 2000. Bcl-2 family: Life-or-death switch. *Febs Letters,* 466**,** 6-10.

TUVESON, D. A., WILLIS, N. A., JACKS, T., GRIFFIN, J. D., SINGER, S., FLETCHER, C. D. M., FLETCHER, J. A. & DEMETRI, G. D. 2001. STI571 inactivation of the gastrointestinal stromal tumor c-KIT oncoprotein: biological and clinical implications. *Oncogene,* 20**,** 5054-5058.

VAUX, D. L., CORY, S. & ADAMS, J. M. 1988. Bcl-2 gene promotes hemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells. *Nature,* 335**,** 440-442.

VINOGRAD, J., LEBOWITZ, J., RADLOFF, R., WATSON, R. & LAIPIS, P. 1965. The twisted circular form of polyoma viral DNA. *Proc Nat Acad Sci USA,* 53**,** 1104-1111.

VIVILLE, S. & MANTOVANI, R. 1994. S1 mapping using single-stranded DNA probes. *In:* HARWOOD, A. J. (ed.) *Methods in Molecular Biology; Protocols for gene analysis.*

VOJTISKOVA, M., MIRKIN, S., LYAMICHEV, V., VOLOSHIN, O., FRANKKAMENETSKII, M. & PALECEK, E. 1988. Chemical probing of the homopurine.homopyrimidine tract in supercoiled DNA at single-nucleotide resolution. *Febs Letters,* 234**,** 295-299.

VOLOGODSKAIA, M. Y. & VOLOGODSKII, A. V. 1999. Effect of magnesium on cruciform extrusion in supercoiled DNA. *Journal of Molecular Biology,* 289**,** 851-9.

VOLOGODSKII, A. V. & COZZARELLI, N. R. 1994. Conformational and thermodynamic properties of supercoiled DNA. *Annual Review of Biophysics and Biomolecular Structure,* 23**,** 609-43.

VORLICKOVA, M., CHLADKOVA, J., KEJNOVSKA, I., FIALOVA, M. & KYPR, J. 2005. Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic Acids Research,* 33**,** 5851-5860.

WANG, J. C. 1971. INTERACTION BETWEEN DNA AND AN ESCHERICHIA-COLI PROTEIN OMEGA. *Journal of Molecular Biology,* 55**,** 523-&.

WANG, J. C., PECK, L. J. & BECHERER, K. 1983. DNA supercoiling and its effects on DNA structure and function. *Cold Spring Harbor Symposia on Quantitative Biology,* 47 Pt 1**,** 85-91.

WANG, Y. & PATEL, D. J. 1993a. Solution structure of a parallel-stranded G-quadruplex DNA. *Journal of Molecular Biology,* 234**,** 1171-1183.

WANG, Y. & PATEL, D. J. 1993b. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure,* 1**,** 263-282.

WANG, Y. & PATEL, D. J. 1994. Solution structure of the Tetrahymena telomeric repeat d(T2G4)4 G-tetraplex. *Structure,* 2**,** 1141-1156.

WATSON, J. D. & CRICK, F. H. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature,* 171**,** 737-8.

WEI, D. G., PARKINSON, G. N., RESZKA, A. P. & NEIDLE, S. 2012. Crystal structure of a c-kit promoter quadruplex reveals the structural role of metal ions and water molecules in maintaining loop conformation. *Nucleic Acids Research,* 40**,** 4691-4700.

WIELAND, M. & HARTIG, J. S. 2007. RNA quadruplex-based modulation of gene expression. *Chemistry & Biology,* 14**,** 757-763.

WILLIAMSON, J. R. 1994. G-quartet structures in telomeric DNA. *Annual Review of Biophysics and Biomolecular Structure,* 23**,** 703-730.

WILLIAMSON, J. R., RAGHURAMAN, M. K. & CECH, T. R. 1989. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell,* 59**,** 871-880.

WOODFORD, K., WEITZMANN, M. N. & USDIN, K. 1995. The use of K(+)-free buffers eliminates a common cause of premature chain termination in PCR and PCR sequencing. *Nucleic Acids Research,* 23**,** 539-539.

WOODFORD, K. J., HOWELL, R. M. & USDIN, K. 1994. A novel K(+)-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. *Journal of Biological Chemistry,* 269**,** 27029-27035.

WRIGHT, W. E., TESMER, V. M., HUFFMAN, K. E., LEVENE, S. D. & SHAY, J. W. 1997. Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes & Development,* 11**,** 2801-2809.

WU, Y. L., MEHEW, J. W., HECKMAN, C. A., ARCINAS, M. & BOXER, L. M. 2001. Negative regulation of bcl-2 expression by p53 in hematopoietic cells. *Oncogene,* 20**,** 240-251.

XU, Y., KAMINAGA, K. & KOMIYAMA, M. 2008. G-quadruplex formation by human telomeric repeats-containing RNA in Na+ solution. *Journal of the American Chemical Society,* 130**,** 11179-84.

XU, Y., NOGUCHI, Y. & SUGIYAMA, H. 2006. The new models of the human telomere d[AGGG(TTAGGG)(3)] in K+ solution. *Bioorganic & Medicinal Chemistry,* 14**,** 5584-5591.

XU, Y., SUZUKI, Y., ITO, K. & KOMIYAMA, M. 2010. Telomeric repeat-containing RNA structure in living cells. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 14579-14584.

YANG, D. Z. & OKAMOTO, K. 2010. Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Medicinal Chemistry,* 2**,** 619-646.

YANG, E., ZHA, J. P., JOCKEL, J., BOISE, L. H., THOMPSON, C. B. & KORSMEYER, S. J. 1995. Bad, a heterodimeric partner for Bcl-XL and Bcl-2, displaces Bax and promotes cell death. *Cell,* 80**,** 285-291.

YANISCH-PERRON, C., VIEIRA, J. & MESSING, J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene,* 33**,** 103-119.

YING, L. M., GREEN, J. J., LI, H. T., KLENERMAN, D. & BALASUBRAMANIAN, S. 2003. Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 14629-14634.

YOON, J., KANG, H.-J., SUNG, J., PARK, H.-J. & HOHNG, S. 2010. Highly Polymorphic G-quadruplexes in the c-MYC Promoter. *Bulletin of the Korean Chemical Society,* 31**,** 1025-1028.

YOUNG, R. L. & KORSMEYER, S. J. 1993. A negative regulatory element in the bcl-2 5'-untranslated region inhibits expression from an upstream promoter. *Molecular and Cellular Biology,* 13**,** 3686-3697.

YU, H., GU, X., NAKANO, S.-I., MIYOSHI, D. & SUGIMOTO, N. 2012. Beads-on-a-string structure of long telomeric DNAs under molecular crowding conditions. *Journal of the American Chemical Society,* 134**,** 20060-20069.

YU, H. J., ZHAO, C. Q., CHEN, Y., FU, M. L., REN, J. S. & QU, X. G. 2010. DNA loop sequence as the determinant for chiral supramolecular compound G-quadruplex selectivity. *Journal of Medicinal Chemistry,* 53**,** 492-498.

YUE, D. J. E., LIM, K. W. & PHAN, A. T. 2011. Formation of (3+1) G-quadruplexes with a long loop by human telomeric DNA spanning five or more repeats. *Journal of the American Chemical Society,* 133**,** 11462-11465.

ZAHLER, A. M., WILLIAMSON, J. R., CECH, T. R. & PRESCOTT, D. M. 1991. Inhibition of telomerase by G-quartet DNA structures. *Nature,* 350**,** 718-720.

ZAUG, A. J., PODELL, E. R. & CECH, T. R. 2005. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 10864-10869.

ZHANG, C., LIU, H. H., ZHENG, K. W., HAO, Y. H. & TAN, Z. 2013. DNA G-quadruplex formation in response to remote downstream transcription activity: long-range sensing and signal transducing in DNA double helix. *Nucleic Acids Research.*

ZHANG, D.-H., FUJIMOTO, T., SAXENA, S., YU, H.-Q., MIYOSHI, D. & SUGIMOTO, N. 2010. Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. *Biochemistry,* 49**,** 4554-4563.

ZHANG, D.-H. & ZHI, G.-Y. 2010. Structure monomorphism of RNA G-quadruplex that is independent of surrounding condition. *Journal of Biotechnology,* 150**,** 6-10.

ZHANG, L. & GRALLA, J. D. 1989. In situ nucleoprotein structure at the SV40 major late promoter: melted and wrapped DNA flank the start site. *Genes and Development,* 3**,** 1814-1822.

ZHANG, N., PHAN, A. T. & PATEL, D. J. 2005. (3+1) assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. *Journal of the American Chemical Society,* 127**,** 17277-17285.

ZHENG, K. W., CHEN, Z., HAO, Y. H. & TAN, Z. 2010. Molecular crowding creates an essential environment for the formation of stable G-quadruplexes in long double-stranded DNA. *Nucleic Acids Research,* 38**,** 327-38.