IN VIVO DETECTION OF G-QUADRUPLEX NUCLEIC ACIDS USING MONOCLONAL ANTIBODIES

by

Alexander Henderson

B.Sc., Simon Fraser University, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Experimental Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

November 2011

Abstract

G-quadruplex nucleic acids are a group of nucleic acids formed from the non-Watson-Crick base pairing of guanine nucleic acids. They can readily form at physiological pH and physiological temperatures within sufficiently long stretches of guanine-rich oligonucleotides. Although, the existence of the G-quartet (the fundamental unit of a G-quadruplex) in a Petri dish has been recognized since the early 60's, the existence of G-quadruplex nucleic acids in mammalian cells remains unclear. Yet while unequivocal evidence of the existence of G-quadruplex nucleic acids in live cells remains unclear, interest in these potentially important biological structures continues to intensify. G-quadruplex nucleic acids have been suggested to play key roles in essential human molecular pathways including telomere biology, transcriptional regulation and disease development. One of the major obstacles in G-quadruplex nucleic acid research is a lack of tools for the *in vivo* detection of these structures.

In our work, we have harnessed hybridoma technology to produce the first monoclonal antibodies to these unique nucleic acid structures. To our knowledge, these are the first hybridomas secreting monoclonal antibodies obtained through the immunization of mice with purified and validated G-quadruplex structures. Monoclonal antibodies have been approved for use in diagnostic tests and for therapeutic treatments in both cancer and autoimmune diseases, and continue to be very effective laboratory research tools.

Using monoclonal antibodies to different G-quadruplex nucleic acids we have explored the existence of G-quadruplex nucleic acids in mammalian cells. One of our antibodies, termed

1H6, forms discrete nuclear foci in human and murine cells and strong nuclear staining in most cells of human tissues. Based on the specificity of the antibodies for defined G-quadruplex structures *in vitro,* these foci could represent the detection of G-quadruplex nucleic acid structures in mammalian cells. If so, the work presented here provides the first direct evidence for the existence of G-quadruplex nucleic acid structures in human cells.

Preface

Chapter 2 was a collaborative effort whereby most of the G-quadruplex nucleic acid structure formation was done at Simon Fraser University in the laboratory of Dr. Dipankar Sen. Oligonucleotides were initially purchased from the University of Calgary DNA Synthesis Lab and folded into various G-quadruplex structures. I performed most of the experiments including design and synthesis with some guidance from Dr. Sen and his post-doctoral fellow Dr. Yu Chuan Huang. Dr. Huang was instrumental in teaching me the basic properties of guanine-rich nucleic acids.

All animal studies in Chapter 2 required animal ethics approval by UBC and the Canadian Council on Animal Care (CCAC). All animal immunizations were conducted at the British Columbia Research Centre (BCCRC) under standard guidelines outlined by the CCAC and UBC animal care committee. I performed all animal experiments and wrote all animal care applications. I have been properly trained in animal husbandry, humane end-point analysis/euthanasia and surgery. All of the animal experiments were conducted under the UBC animal care protocol number A06-0145 and I maintain an animal training certification number from UBC: 2145.

Chapter 2 and chapter 3 was collaboration with Dr. Robert Brosh Jr. and his post-doctoral fellow Dr. Yuliang Wu at the National Institutes of Health. Dr. Yuliang Wu performed the experiments outlined in appendix (A1, B6, B7 and B8). All other experiments and analysis in Chapter 2 and Chapter 3 were performed by me. Some experiments highlighted in the appendix (B1-5) were done by Liz Chavez from the BCCRC.

Chapter 4 was a partial collaboration with the Genome Sciences Centre (GSC) and specifically their next-generation sequencing centre run by Dr. Martin Hirst. I performed all of the G4 chromatin immunoprecipitation (G4-ChIP) experiments, DNA size selection and purification, and built the DNA sequencing libraries. The GSC loaded the libraries onto a HiSeq flow cell and gathered the data into a file for us to analyze. I analyzed the sequence data with the help of Dr. Olivia Alder and Dr. Mark Hills.

Chapter 4 was a partial collaboration with Dr. Gregg Morin from the GSC. Dr. Grace Cheng under the supervision of Dr. Gregg Morin performed the in gel trypsinization, alkylation, reduction and mass spectrometry analysis of proteins that I had isolated after G4-ChIP. I designed, performed and analyzed all other experiments in Chapter 4.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost I would like to thank my PhD supervisor Dr. Peter M. Lansdorp. I am very grateful for the learning environment in which he provided and the challenges they presented along the way. I sincerely believe the freedom that Peter has provided in his lab has allowed me to develop my own personal scientific method (with its own quirks) and allowed me to become the "Master of my own destiny".

I would also like to thank my committee members for their supportive guidance throughout my degree. Dr. Dipankar Sen has been instrumental in my academic development and has constantly challenged my understanding of G-quartet biochemistry. Dr. Anne Rose has always been very enthusiastic about my work and she has continuously provided a supportive voice for future ventures using these novel reagents. I owe particular thanks to Dr. Robert Kay for playing the "devil's advocate" which has provided me with some of the most interesting perspectives on this project. Rob has always provided insightful alternative explanations for experimental results and has challenged me to understand that absolute proof is hard to come by.

I offer my enduring gratitude to my fellow Lansdorp lab members who have inspired me throughout my training and who continue to provide a great training environment for all students in the lab. I have good friends in the lab and truly believe that all of my skills and abilities are a direct reflection of their leadership and support. I am especially grateful to all the postdoctoral fellows who have been actively involved in my development as a scientist.

Special thanks are owed to my parents, who have supported me throughout my many years of education, both morally and financially. Last but not least my academic pursuits could have never been fully achieved without the unwavering support of my (soon to be) wife, Michelle. I am extremely grateful for all of the sacrifices (including listening to me rant about crazy science things) she has made, that have allowed me to realize this part of my life.

Chapter 1: Introduction

1.1   Nucleic acids and the B-form

DNA and RNA molecules are long linear chains of copolymers composed of many

nucleotides (1). These molecules contain the genetic information of all known organisms

and can be generally built from five bases; the purines adenine (A) and guanine (G), and the

pyrimidines cytosine (C), thymine (T) and uracil (U) (1). The classical complementary base

pairing of nucleotides from antiparallel strands, also known as Watson-Crick base pairing,

occurs through hydrogen bonds formed between complementary bases; adenine to thymine

or uracil (two hydrogen bonds) and guanine to cytosine (three hydrogen bonds). The right-

handed canonical double helix (B-form) secondary structure of DNA arising from this

complementary base pairing was first proposed by James Watson and Francis Crick in 1953

(2) and it was for this accomplishment they shared the Nobel Prize for medicine in 1962.

Although once considered an inert molecule used only for the storage and passage of the

genetic code, DNA is now seen as an extremely dynamic set of molecules that can undertake

significant structural modifications. It is quite easy to forget that the double helix has

considerable dynamic flexibility and that there are an abundance of distinct structural

conformations of nucleic acids (other than the B-form) that exist in active equilibrium with

each other (3). Despite the fact that DNA spends much of its time in the relatively simple

structural conformation B-form, wrapped around histones and packaged as chromatin, it does

exhibit the ability to form complex alternative structures. We know that complex structures

form in (single-stranded) RNA and that single-stranded DNA is formed during several

biological processes; including, replication, transcription, recombination and repair (4). Like

RNA, single-stranded DNA may form alternative structures that are important in the regulation of gene expression and other biological processes such as meiotic recombination. Alternative structures of RNA and DNA such as G-quadruplex structures not only provide new avenues for research but also provide novel targets for artificial intervention and therapies (5).

1.1.1   Z-form DNA

Discovered by accident through the crystallization of duplex DNA, Z-form DNA is like B-form DNA in that it is a double helix formed by two Watson-Crick complementary strands in an antiparallel conformation (3). The major structural difference arises from the handedness of the helix; B-form DNA is right-handed while Z-form DNA is left-handed (Figure 2). In addition, all bases in B-DNA are in the same conformation; the bases in left-handed Z-DNA alternate in a *syn* and *anti* conformation. Due to the natural stability of purines in *syn* over pyrimidines, Z-form DNA is favored in sequences that alternate purines and pyrimidines and is thought to be stabilized by negative supercoiling (unwinding) (3). Although some scientific papers have proposed biological roles of Z-from DNA including transcriptional regulation, there is limited evidence to support a major role of Z-form DNA in human cells. It is important to note that Z-form DNA is much less stable than B-form DNA at physiological conditions, due in part to the increased electrostatic repulsion by the negatively charged phosphates of each strand being closer together in Z-form DNA than B-from DNA (3). The lack of stability of Z-form DNA and the limited indications of any

significant biological function has made targeting Z-DNA for therapeutic intervention a difficult task.

1.1.2   Triplex nucleic acids

The canonically-formed double helix held together by complementary Watson-Crick base pairing can bind a third oligonucleotide strand, producing triplex oligonucleotides. Triple helix structures are formed when pyrimidine or purine bases occupy the major groove of an already formed double helix (6). Although the previously formed double helix is held together by Watson-Crick base pairing the third strand is bound by Hoogsteen pairs (6) and arises through the bonding of pyrimidine or purine motifs (7). Unlike Watson-Crick base pairing where there are specific hydrogen bond donors and acceptors between complementary bases, Hoogsteen pairs (denoted as such by their discovery in 1963 by Karl Hoogsteen) arise when the purine base acts as both a hydrogen bond acceptor (N_7 position) and hydrogen bond donor (C_6 amino group) in turn binding the Watson-Crick bound pyrimidine base. Triplex forming oligonucleotides (TFO) consisting of pyrimidines (thymine or cytosine) bind to parallel purine-rich A and G duplex oligonucleotides via Hoogsteen base pairing. However, protonation at the N_3 position is required for proper bonding of cytosine rich TFO's to G:C duplexes (7). TFOs consisting of purines (adenine or guanine) provoke triplex formation via Hoogsteen bonding to A:T and G:C duplex oligonucleotides (Figure 2). TFO's have limited biological importance due in part because of the limited *in vivo* evidence of their existence. They have however, found renewed importance as powerful tools for the sequence specific manipulation of DNA. TFO's can be

used as "molecular scissors," and possess therapeutic potential by their potential for mediating targeted disruption of gene transcription (8). For a full review of triplex nucleic acids see references (6, 7).

1.1.3   Quadruplex nucleic acids

Pre-dating the discovery of DNA, it was established that high concentrations of guanine nucleobases behaved very differently than other nucleobases and could spontaneously form gels (9). It wasn't until 52 years later using fiber diffraction studies that the structure responsible for guanine gel formation was discovered, the guanine tetrad (or G-quartet) (10). At the core of guanine gels is the formation of alternative base pairing, notably guanine bases bonding to other guanine bases. It is not surprising that guanines can self-associate due in part to the presence of complementary hydrogen bond donors and acceptors on the edges of guanine. G-quartets form from the bonding of four guanines in a square planar array, associated by two hydrogen bonds on each side of a guanine (Figure 1) (11, 12). G-quartets can be further stabilized by the presence of monovalent cations (usually K^+or Na^+) in the centre of the quartet, interacting with the lone pair on O_6. The self-assembly of guanine is enhanced further by the polarizable aromatic surface of guanine that, when bound in a square planar array, creates large π surfaces with a high propensity to stack (13). Stacking of G-quartets allows for improved stabilization between quartets via π-π stacking interactions and cation intercalation between adjacent stacks. Stacked quartets forming larger stable structures are called G-quadruplex (G4) structures.
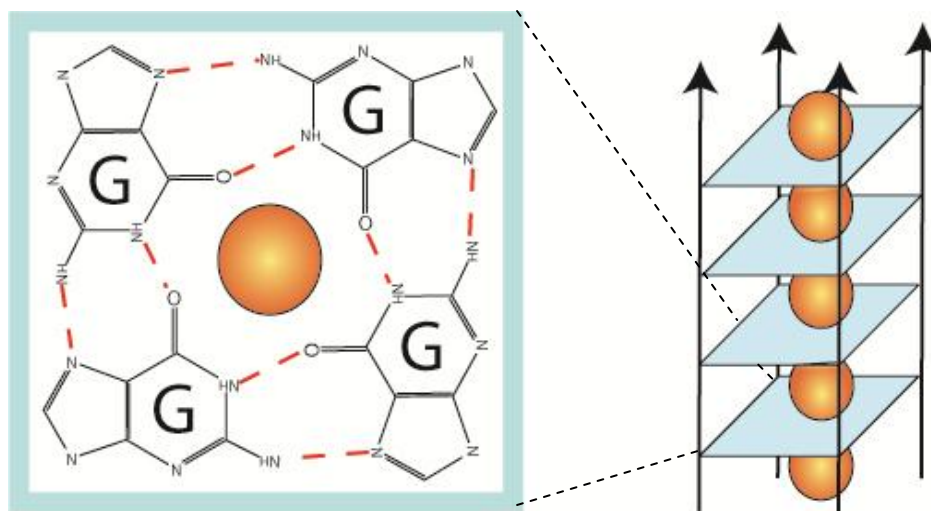
Figure 1. From G-quartet to G-quadruplex

Four guanines (G) can self-assemble into stable square planar arrays known as the G-quartet (left panel). Guanines are held together by four hydrogen bonds on each guanine (shown by the dashed red lines) and may be further stabilized by the presence of monovalent cations (orange ball). Stacking of G-quartets generates a G-quadruplex (right panel).

Similar to the conventional double helix where complementary oligonucleotides will self-assemble into a stable B-form structure, G4 structures are stable, self-assembled structures generated from guanine-rich oligonucleotides. However unlike any other alternative conformation, G4 structures are arguably the most promising alternative nucleic acid structure for therapeutic intervention and as molecular targets due in part to their extreme structural polymorphisms and their stability in physiological conditions.

B-form   Z-form   Triplex   Quadruplex
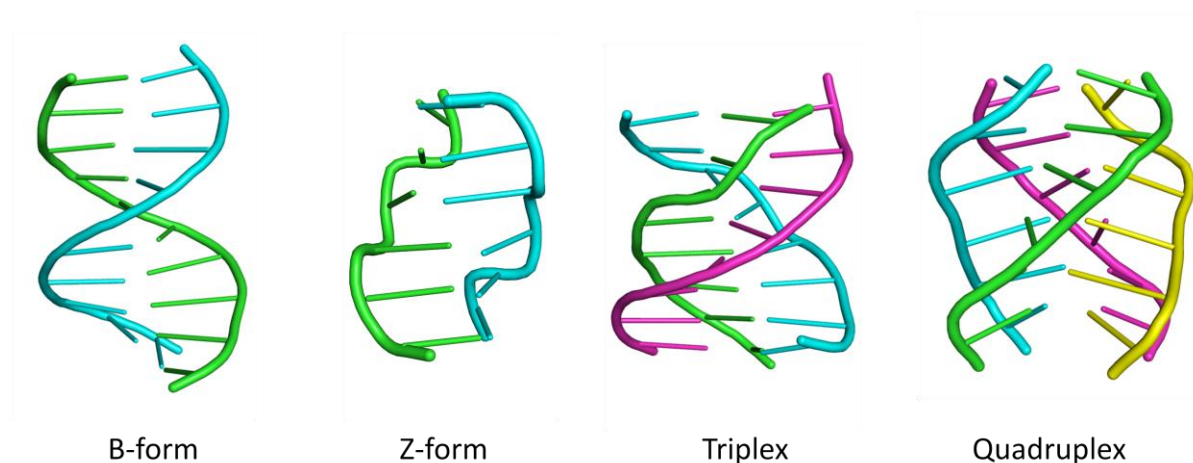
Figure 2. Alternative nucleic acid structures.
Nucleic acids are now viewed as dynamic entities that can fold into numerous structural variants. The B-form structure was once considered the only biologically relevant structure. Different strands have been artificially colored for simplicity.   3-D images were generated using Polyview 3-d and PDB files from the RCSB protein data bank PBD ID #: 3r86 (B-form), 3fqb (Z-form), 1d3x (Triplex) and 1rau (Quadruplex).

1.2   The structural polymorphisms of quadruplex nucleic acids

The family of structures known as G4 structures can easily be created from DNA and RNA. G4 structures are incredibly complex and vary in six major attributes: the number of oligonucleotide strands, the nucleotide sequence composition, the directionality of strands, the glycosidic bond angles, the type of intervening loops, and the interacting alkali metal present.

A primitive classification of G4 structures can be made by the number of guanine-rich strands used to form a G4 structure. Tetramolecular structures arise from the multimerization of four single guanine-rich strands. Bimolecular G4 structures generally arise from the pairing of two hairpin oligonucleotides generated from sequences containing

6

two guanine runs and unimolecular G4 structures develop when a single oligonucleotide contains at least four runs of guanines.
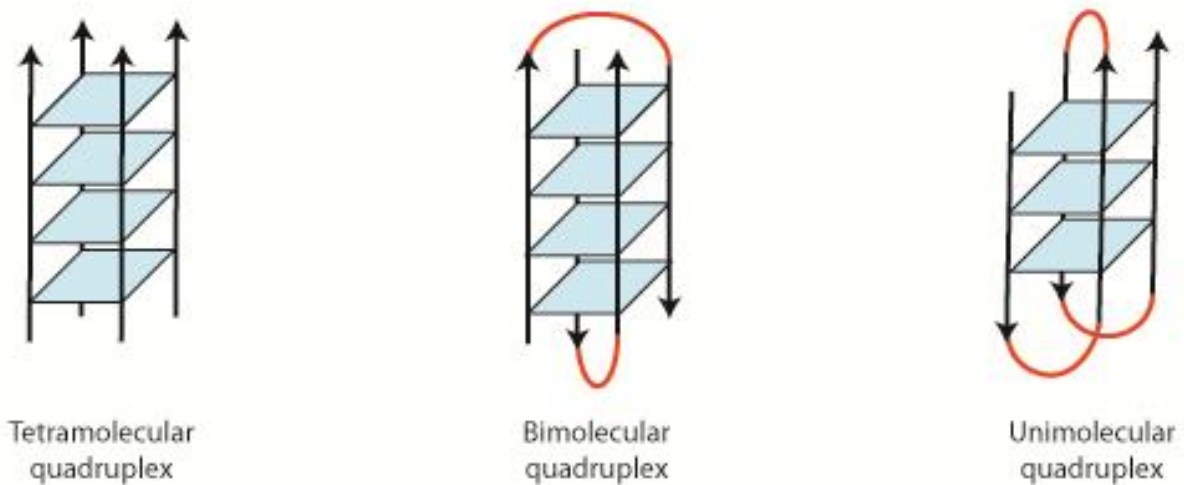
Figure 3. Molecularity of G-quadruplex structures.
A primitive classification of G-quadruplex structures by the number of oligonucleotide strands used to form a quadruplex structure. Tetramolecular structures are generated from four separate strands; bimolecular structures are generated from two separate strands while unimolecular structures are folded structures derived from a single guanine-rich strand. Looping sequences are highlighted in red and strand polarity is indicated by arrow directions.

Although the number of oligonucleotide sequences used to form a G4 structure can be used to roughly classify G4 structures, the 5′ to 3′ directionality of the strands will further distinguish the G-quartet core. Directionality of strands can be classified as parallel, antiparallel or mixed. The directionality will have direct impact on two features of a quadruplex: the glycosidic bond angle and the loop types between G-quartets.

Syn   Anti
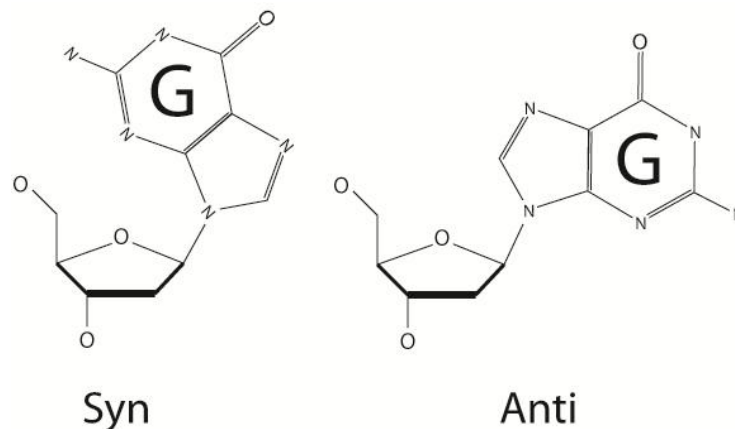
Figure 4. Glycosidic bond angle of guanines.
The glycosidic bond angle of G-quartets will change depending of the polarity of strands in a G-quadruplex.

The glycosidic bond, which is found between the guanine base and its sugar backbone can be either in a *syn* or *anti* conformation (Figure 3). To properly form hydrogen bonds between guanines in the G-quartet core the glycosidic bond will alternate between *syn* and *anti* to accommodate for the directionality of the strands. The all parallel strand orientation found in all tetramolecular G4 structures results from having all guanines in an *anti* conformation. When any strand in a G4 structure is in an antiparallel direction the bases must be in *syn* to properly form hydrogen bonds (11, 12). Mixed hybrid structures occur when one or more strands are parallel and one or more are antiparallel. In these structures glycosidic bonds will adjust to accommodate the directionality of the sequence.

The directionality of strands also plays an important role in determining the adjoining loops between guanines. There have been three types of loop orientations described for G4 structures: diagonal loops, lateral loops and double chain reversal loops (Figure 4). There are no loops in tetramolecular G4 structures (where four individual strands form a G4 structure); loops only occur in bimolecular and unimolecular G4 structures. Diagonal loops or lateral

loops appear when strand directionality is antiparallel. Double chain reversals develop when the directionality of the strands are parallel. When the directionality of the strands is a mixed hybrid the loop generated will also become a mixture of conformations (11).
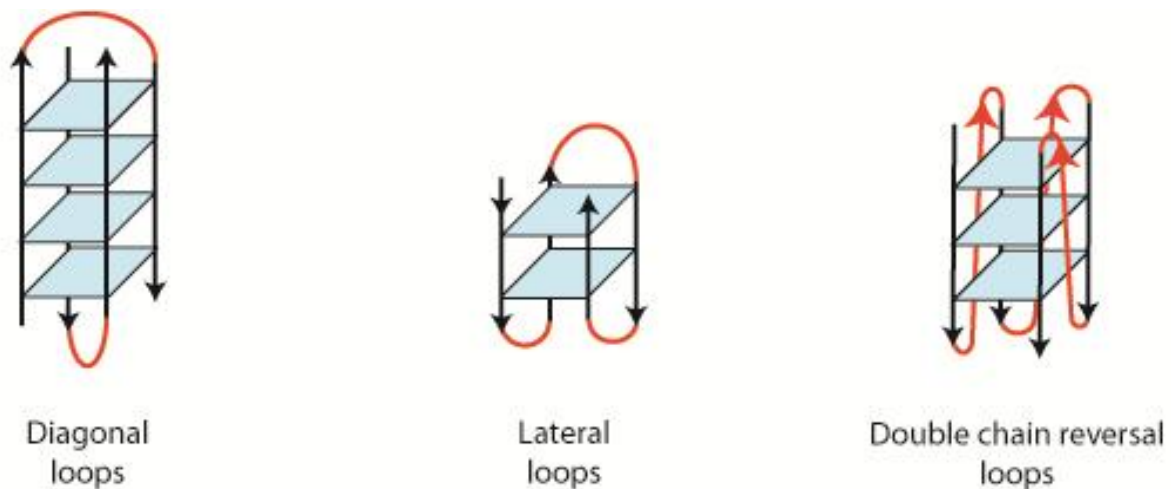
Figure 5. Possible loop arrangements in G-quadruplex nucleic acids
There are three different types of loop arrangements that have been described for joining G-quartets in a G-quadruplex structure; diagonal, lateral or double chain reversal loops. Loop sequences can only arise in bimolecular and unimolecular structures and have been highlighted in red. Strand polarity is highlighted by arrow directionality.

The self-assembly of G4 structures is cation dependent and stabilized by alkali and alkali earth metals (14). Given that the intracellular space and extracellular spaces in most organisms are dominated by K^+ and Na^+ respectively, these two metals typically get the most attention (15). The stabilization of G-quartets by cation binding is understandable, since the G-quartet has four oxygen atoms clustered in its centre, without a bound ion this cyclic arrangement would be electronically unfavorable (16). In spite of this, Sessler *et al*., have described the formation of a G-quartet without a central cation (16). In all other cases to date, cations are required for the stabilization of self-assembled G4 structures. In comparison

9

to other alternative nucleic acid structures, there are numerous topological variations of G4 structures many of which have been folded and studied *in vitro*.

1.2.1 Methods for the interrogation of G-quadruplex nucleic acids

A number of techniques have been developed for detecting and measuring G4 nucleic acid structures that differentiate it from B-form nucleic acids. One of the simplest yet powerful tools for analyzing G4 nucleic acids is by polyacrylamide gel electrophoresis (PAGE). The electrophoretic mobility of molecules through acrylamide is determined by the size, shape and charge of a molecule (17). Under non-denaturing conditions (native PAGE), oligonucleotides that can adopt different structures will migrate through the gel differently then single-stranded oligonucleotides of the same length. Oligonucleotides that assume compact unimolecular G4 structures have increased electrophoretic mobilities in comparison to their unstructured single-stranded counterparts (18). Conversely, oligonucleotides that assume larger oligomeric G4 structures, like tetramolecular G4, have exceedingly retarded electrophoretic mobilities in native PAGE (19). Native gels are therefore especially useful for detections of alternate conformations of oligonucleotides based on the distinct mobilities of each conformer.

Another method useful for the interrogation of G4 structures is circular dichroism spectropolarimetry (CD). CD works by measuring the differential absorption of left-handed and right-handed circularly polarized light (20). The circular polarization of light occurs when an electric field rotates about its propagating direction maintaining its magnitude, thus generating a helix in space (Figure 5).
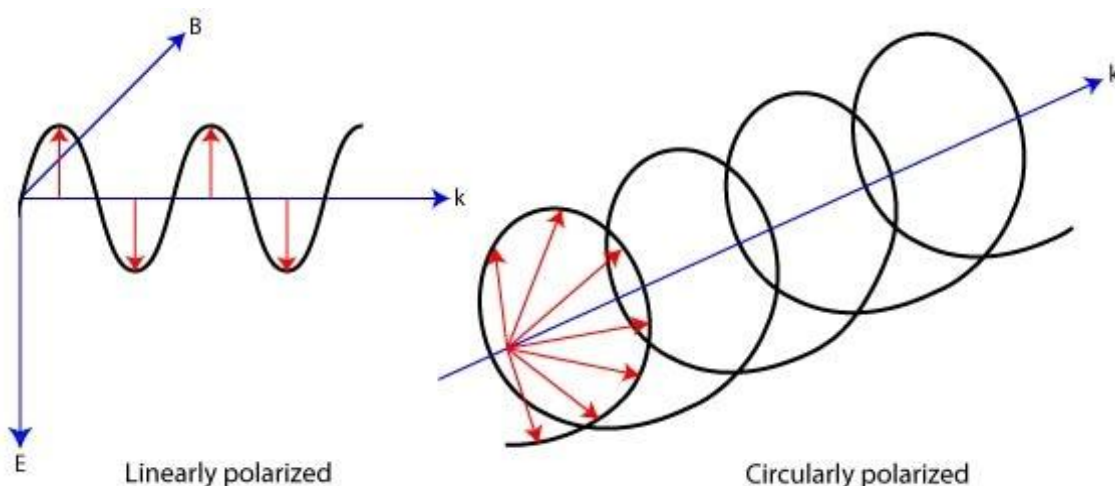
Figure 6. The polarization of electromagnetic waves.
Visualizing the propagation of electromagnetic waves, where E represents the electric field vector, B represents the magnetic field vector and k is the propagating direction of light. In comparison to linearly polarized light, circularly polarized light maintains is magnitude but the electric field vector rotates about its propagating direction.

If the electric field vector rotates counterclockwise towards the observer, this is termed left-handed circularly polarized light. If the electric field vector rotates clockwise towards the observer, this is termed right-handed circularly polarized light. When circularly polarized light passes through an optically active medium the right-handed and left-handed circularly polarized light will be differentially absorbed (21). The differential absorption of left and right-handed circularly polarized light can be measured using a circular dichroism spectropolarimeter. During a CD experiment, equal amounts of left-handed and right-handed circularly polarized light of a selected wavelength are alternately radiated into a sample. The two polarizations of light will be absorbed to different extents; the wavelength dependent difference of absorption is measured in a quantity called ellipticity, θ (21).

CD has been extensively used for interpreting many fundamental physical properties of quadruplexes, including structure (22, 23), kinetics of formation (24, 25), the effects of

chemical modifications (26, 27), thermal melting profiles (24, 25, 28), the effects of varying cations and ligand binding (23-25, 28-38). For a complete review of methods used for interrogating G4 structures see (17, 39, 40).

1.2.2   Why study G4 structures?

G4 structures were long considered an *in vitro* phenomenon as they were predominately generated in the laboratory under non-physiological conditions outside the realm of biological context. They did not attract the attention of biologists (40) until the discovery that short guanine-rich sequences located at the ends of eukaryotic chromosomes could associate together to form four-stranded structures in DNA, and thus might be biologically relevant (18, 19, 41, 42). Since then there has been increasing amounts of circumstantial evidence that indicate G4 structures play key physiological roles, including the discovery of evolutionarily conserved potentially G4 forming sequences throughout the human genome and the identification of key proteins that stabilize, modify or resolve G4 structures (45, 59-61). It has now been suggested that G4 structures function in numerous biological phenomenon including; replication, transcription, translation, recombination, telomere biology and disease (45, 59, 62, 203). However, the debate continues regarding all functional roles of G4 structures in living cells, because there is no direct evidence of their existence in mammalian cells. In addition, under normal physiological conditions, transitions from B-form DNA to G4 DNA have not been shown to be energetically favorable. Many scientific reports pertaining to G4 nucleic acids are moderated by the disclosure that the existence of G4 structures in mammalian cells *in vivo* have yet to be proven (43).

1.2.3   Putative quadruplex forming regions in the genome

While no *in vivo* data exists regarding the presence of G4 structures in human cells, *in vitro* and *in silico* data continues to accumulate. Now that much of the human genome has been sequenced and annotated (although large gaps remain unsequenced like repetitive GC-rich regions where quadruplexes are most likely to form) computers can be used to search for putative quadruplex sequences (PQS) that may form G4 structures. In theory any sequence containing sufficiently long stretches of guanine nucleotides can form a unimolecular G4 structure upon opening of double-stranded DNA during transcription, replication or repair. However, in reality not all linear single-stranded DNA sequences will form G4 under physiological conditions, and the stability of a G4 structure formed from a particular sequence cannot yet be predicted. It has been assumed that the chance of generating multimeric G4 structures from the combination of multiple single strands in a biological context is relatively low. For this reason the greater part of today's research has focused on the formation of unimolecular G4 structures (13).

Using similar search algorithms, multiple surveys have looked for PQS in the human genome (44, 45). Although different statistical and analytical approaches were employed, the overall number of PQS that was found is similar. Although G4 structures could be formed from PQS with medium to long intervening sequences as large loop lengths, most studies have eliminated such variables in the search algorithms for practical reasons. Although in theory limiting loop lengths to <8 nucleotides could result in arbitrarily missing many PQS, it is assumed that the general stability of G4 structures decreases with increasing

13

loop size. A general nucleotide algorithm for the search of PQS has therefore been restricted to:

$$G_{3\text{-}5}\ N_1\ G_{3\text{-}5}\ N_2\ G_{3\text{-}5}\ N_3\ G_{3\text{-}5}$$
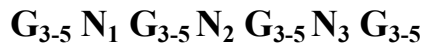
Where $N_{1\text{-}3}$ are loops of unknown length with any nucleotide identity and limits in nucleotide length $1 \leq N_{1\text{-}3} \leq 7$. Using this approach PQS have been discovered throughout the genome including: in genes (exons, introns, promoter sequences) in gene deserts, immunoglobulin switch regions, 5' untranslated regions, recombination hot spots and telomeres. When every possible combination is considered (from the annotated human genome) and analysis is restricted to only distinct non-overlapping PQS there are ~400, 000 sequences in the genome that could form G4 structures (12). However, when analysis includes possible overlapping sequences there are close to 6 million possible PQS in the human genome. Although the analysis can give us a general idea of how many PQS exist, it does not inform us of the likelihood or stability of each PQS to exist as a G4 structure in cells. The stability of G-quadruplex structures *in vivo* continues to be controversial however; unimolecular quadruplexes may be less stable than once suggested (207). In addition, knowing that most PQS exist in the presence of cytosine-rich complementary strands, there may be a dynamic equilibrium between G4 structures at PQS and double-stranded DNA. The presence of a cytosine-rich complementary strand reduces the probability that a PQS would fold into a G4 structure because it is assumed, under normal physiological conditions the B-form is favored. However, the cytosine-rich complementary strand may form an alternative structure of its own termed the i-motif (11).

Not surprisingly, a high proportion of PQS are located in areas of high GC-richness including within gene promoters. However, the discovery of PQS enrichment in proto-oncogenes and reduced PQS found in tumor suppressor genes is suggestive of a correlation between PQS and specific gene function (46). Therefore, many groups have focused on the characterization of G4 structures within promoter regions that have been suggested to reflect a mechanism for global regulation of gene expression (46).

1.3   Biological implications of G4 structures in the genome

Studies exploring the hypothesis that G4 structures formed in promoter sequences effectively act as transcriptional regulatory signals have gained a lot of support from the examination of multiple prokaryotic species and the occurrence of PQS in their genomes. PQS are statistically over-represented in positions proximal to the transcription start site and could therefore participate in global gene regulation (47). If PQS sequences and subsequent G4 structures are involved in the regulation of gene expression, the polymorphic nature of G4 structures will provide specific targeting potential for small-molecule therapies (48).

The most studied promoter region containing multiple PQS is located in the *c-myc* oncogene. The *c-myc* oncongene encodes a transcription factor involved in the expression of many genes involved in cell proliferation and helps to regulate global chromatin structure. Mutant versions of C-MYC have been found in multiple cancers where there is ubiquitous expression of *c-myc* and unregulated expression of many genes resulting in the formation of cancer. Interestingly, one PQS is located in the nuclease-hypersensitivity element III of the *c-myc* promoter, which is responsible for 80-90% of its transcriptional activity. Initially

discovered by Simonsson *et al*. in 1998, it wasn't until a few years later that the relationship between G4 structural stabilization and *c-myc* gene transcription was discovered (48, 49). The expression of *c-myc* was shown to be significantly suppressed upon the drug-mediated stabilization of the G4 structure. Due to the presence of more than four guanine-rich tract repeats in the PQS of *c-myc* there are multiple G4 species that can be formed with this PQS. However, synthetic oligonucleotides containing short regions of the *c-myc* gene have been shown to form very stable parallel-stranded unimolecular G4 structures in solution (50).

PQS have also been discovered in the promoter region of the *c-kit* kinase gene, which has important roles in cell survival, proliferation and differentiation. Mutations in the *c-kit* gene are associated with gastrointestinal tumors, acute myeloid leukemia and other cancers and therefore *c-kit* has also been suggested as an important therapeutic target (51, 52). There are multiple PQS located within the promoter of the *c-kit* gene that are highly conserved across vertebrate species and therefore suggestive of a functional role (12). Similar to *c-myc*, one PQS is located at the site of core promoter activity of *c-kit* (52). Structural studies continue but some truncated single species sequences of *c-kit* do form parallel-stranded G4 structures (53).

Similarly, PQS have been found in a number of other oncogenic promoter regions including: *HIF-1α, VEGF, RET, KRAS, BCL-2* and *PDGF*. Hypoxia inducible factor 1α (*hif-1α*) is another transcription factor, that responds to changes in the oxygen levels in a cellular setting and is over-expressed and unregulated in many human tumors. A parallel-unimolecular structure formed from the PQS in the promoter of *hif-1α* has been proposed to regulate its expression (54). The vascular endothelial growth factor (*VEGF*) gene is responsible for the generation of a protein that stimulates vasculogenesis and angiogenesis.

vegf has been implicated in tumor angiogenesis and as such has influenced an increased interest in *vegf* as a potential target for cancer therapy (55). Within the promoter region responsible for basal promoter activity of *VEGF* is a PQS. Analogous to *c-myc,* studies have assigned a parallel-stranded unimolecular G4 structure to the drug-stabilized PQS (56).

Stabilizing PQS may not always lead to the suppression of transcription and in many cases it may be required for the activation of transcription. For instance, mutations that disrupt G4 folding in a guanine-rich repeat located upstream of the human insulin gene disrupt insulin gene transcription, but presence of K^+ ions appears to activate transcription (57). The upregulation of genes caused by the stabilization of G4 structures has been suggested as part of a mechanism by which nucleosomes and other proteins that might otherwise repress transcription can be excluded from native B-form duplex DNA (58).

The potential transcriptional regulation of oncogenes by means of folding or unfolding G4 structures will continue to propel the structural characterization of PQS in hopes of manufacturing small molecules that can bind to specific G4 structures. Of particular note, a significant proportion of PQS found in oncogenic promoters have been suggested to form G4 with all strands in a parallel conformation (12). Tools designed to bind and stabilize parallel G4 conformations may therefore act at multiple gene sites.

1.3.1   Replication, recombination and G4 unfolding

One of the fundamental functions of DNA is the passing on of genetic information from mother cell to daughter cell through DNA replication. The process of replication involves the unwinding of double-stranded DNA into two single strands such that the newly synthesized DNA can be generated from the addition of complementary bases. When

double-stranded DNA containing guanine-rich sequences are unwound, the single-stranded

guanine-rich sequences will have a high propensity to form G4 structures especially when the

molecules that typically bind to single-stranded DNA such as replication protein A (RPA)

become limiting during replication. This could be the case upon stalling of the replication

fork and single-stranded DNA not coated by RPA could be involved in DNA damage

signaling (204). With or without fork stalling, G4 structures formed at replication forks

could impair natural replication and lead to DNA repair reactions and as a consequence to

genomic instability (59). Similarly, during recombination reactions single-stranded DNA is

transiently present and may allow for the formation of G4 structures where guanine-rich

repeats persist (60, 61). The presence of stable G4 DNA structures during recombination

could be detrimental and cause errors.

The adverse effects of alternative structures formed during replication and

recombination are presumed to be countered by specialized proteins called helicases.

Interestingly, a number of helicase proteins involved in the unwinding of DNA, have been

described to specifically catalyze the ATP dependent unwinding of G4 structures (Table 1)

(59, 62). RecQ helicase family members WRN, BLM and Sgs1 have all been shown to

unwind numerous G4 structures *in vitro* (60, 61, 63, 64). In addition to the 3′ to 5′ helicases

of the RecQ family there are also 5′ to 3′ helicases, including human FANCJ, Pif1 and DNA2

that can unwind G4 DNA (62). The discovery of G4 capable helicases (some of which have

higher specificity for G4 DNA over B-form DNA) implies a mechanism for G4 structural

unfolding under normal cellular conditions and provides support that G4 structures naturally

exist in mammalian cells.

Evidence supporting an important role for G4 specific helicases derives from the clinical ramifications of defects in these genes. Bloom's syndrome which is caused by defects in the *blm* gene is characterized by increased genomic instability, high rates of leukemia and other forms of cancer and the formation of quadriradial chromosomes, that appear to be aberrant recombination structures (60, 65, 66). Similarly, Werner's syndrome which is associated with early onset aging, genomic instability, increased gene deletions and chromosomal abnormalities is caused by defects in the *wrn* gene (65, 66). Mutations in the related gene, Sgs1 found in yeast, cause errors in replication, transcription, increased levels of recombination and the rapid onset of aging (analogously seen in Werner's) (61, 64, 67). Remarkably, the premature aging phenotype in yeast is rescued by BLM but not WRN (68). This finding suggests two things 1) many helicases could have redundant roles in organisms, and 2) different helicases may have G4 structure specificity.

Similarly, Fanconi Anemia (FA) is a genetic disorder resulting from mutations in one or more of 13 complementation genes (62). Patients with FA show clinical symptoms such as bone marrow failure, chromosome instability and again, predisposition to many types of cancers (69, 70). Mutations in the FA complementation group J gene which encode the FANCJ helicase, result in defective enzymatic activities of the FANCJ protein. *In vitro* assays have established that normal FANCJ proteins catalytically unwind G4 structures in a 5' to 3' manner (62). In addition, mutations in this gene are more often found in breast cancer patients than in normal individuals suggesting a role for FANCJ in genome stability, homologous recombination repair and/or normal G4 metabolism. The latter is supported by genetic experiments in *Caenorhabditis elegans* (*C.elegans*) where the FANCJ ortholog *dog-1* (for deletion of guanine-rich DNA) was found to be required to maintain the genomic

stability of guanine-rich tracts (22 or more consecutive guanines) (71). Researchers found

that mutations in *dog-1* resulted in deletions that started around the 3' end of polyguanine

tracts and terminated at variable 5' positions from these tracts in 50% of the polyguanine

tracts examined (71). The DOG-1 protein in *C.elegans* has therefore been proposed to play a

key role in the maintenance of genomic integrity by reducing the accumulation of irregular

DNA structures that occasionally form during lagging strand replication (71). In related

research the mammalian gene *Rtel1* (for regulator of telomere length 1)*,* encoding a protein

with sequence homology to DOG-1, was identified by genomic mapping of loci responsible

for the control of telomere length in mice (72). Murine cells deficient for the RTEL1

helicase show genomic instability due to in part, to telomere loss (72). Interestingly,

mutations in the *Rtel1* gene in the *C.elegans* model did not result in the deletion of guanine-

rich sequences as seen in the DOG-1 mutants and suggests that these two proteins act in

different molecular pathways (73). Although closely related to the human FANCJ helicase,

this evidence remains somewhat circumstantial as neither RTEL nor DOG-1 helicases have

yet been shown to resolve G4 structures.

Another interesting 3' to 5' DNA helicase that has been shown to unwind G4

structures in an ATP dependent manner is Dna2 (74). In addition to its helicase activity

Dna2 has 3' to 5' endo/exonuclease functions and has been implicated in single-strand

annealing and strand exchange events (75). Dna2 has been suggested to play multiple roles

including coordinating with flap endonuclease I (FEN1) during Okazaki fragment maturation

(76, 77) for DNA replication and been shown to be important for DNA repair (78). It was

proposed that Dna2 resolves G4 DNA structures during replication of guanine-rich strands

and works in conjunction with RPA to prepare the substrate for FEN1. Interestingly, in yeast

Dna2 localizes to the telomeres (79) and has been shown to be involved in the generation of the guanine-rich overhangs by helicase/nuclease processing of telomeres (80). In addition, the stability of ribosomal DNA (which contains PQS) appears to be dependent on the activities of Dna2 (74). Somewhat surprisingly, there is no human disease associated with mutations in Dna2 and may indicate that Dna2 is an essential gene (62).

In a recent study Law *et al*., have provided good evidence that another helicase protein ATRX (alpha thalassemia/mental retardation syndrome X-linked) may associate with G4 DNA *in vivo* (195). Mutations in the *ATRX* gene, which encodes a large helicase, have been associated with severe mental retardation, developmental defects and blood diseases. However, how mutations in *ATRX* alter gene expression is not well understood. Law *et al*., have shown that ATRX preferentially binds G4 DNA over B-form DNA *in vitro* and that over 50% of ATRX target sites are predicted to adopt G4 conformations *in vivo* (195). Interestingly, many of the ATRX binding sites are guanine-rich repetitive elements located near genes with altered gene expression in patients with ATR-X syndrome. ATRX may therefore be regulating global gene transcription through localization to specific sites in the genome by associating with G4 DNA structures (195).

Further evidence that G4 DNA structures can act as specific genomic recognition sites can be derived from studies done in the human pathogen, *Neisseria gonorrhoeae*. *N.gonorrhoea* and other pathogens use DNA recombination to promote antigenic variation of surface structures to avoid immune detection (208). Cahoon *et al*., identified a cis-acting DNA sequence near the antigenically variable pilin locus of *N.gonorrhoeae* that could fold into a G-quadruplex structure *in vitro* and was required for pilin antigenic variation (208). Mutations that disrupted the G4 structure blocked pilin antigenic variation and prevented the

nicks in the G-rich strand that were required for recombination within the G4 region.

Supplementary stabilization of the G4 structure by the addition of G4 stabilizing drugs also

impaired antigenic variation and prevented the required nicks (208). The authors proposed

that the specific location, sequence and structure formed at this region act as a recombination

initiation hot spot that directs gene conversion to a specific chromosomal locus (208).

While the science of G4-protein interactions is a developing field, understanding the

specificity of each helicase/G4 interacting protein and its preferred substrate may ultimately

provide insight into future therapeutic interventions. For excellent reviews on G4 associated

proteins and helicases see (62, 81, 82).

Name	Organism	Direction	Phenotype
WRN	Human	3'to5'	Premature aging, genomic instability
BLM	Human	3'to5'	Cancer, elevated sister-chromatid exchange, genomic instability
FANCJ	Human	5'to3'	Fanconi anemia, breast cancer, defective interstrand X-link repair
G4R1/RHAU	Human	Yes	Abnormal mRNA deadenylation and decay, unwinds G4 RNA
Pif1	Human	unknown	Possible telomere defect
	Yeast	5'to3'	Mitochondrial DNA and G-rich instability, longer telomeres, defective Okazaki fragment metabolism
Dna2	Human	5'to3'	Aberrant cell division, genomic and mitochondrial instability
	Yeast	5'to3'	Defects in Okazaki fragment metabolism
Sgs1	Yeast	3'to5'	Defective double-strand break repair, stalled replication forks inappropriate processing of meiotic recombination intermediates Telomere instability

Table 1. G4 unwinding helicase proteins.
Specialized proteins have been discovered that have the capability to unwind alternative
structures. Several helicase proteins have been shown to preferentially unwind G4 structures
in vitro. The above table describes the name of the protein, organism of origin, the
directionality of unwinding and the corresponding clinical and biochemical phenotypes
associated with mutant versions of the protein.

1.3.2   G4 structures and genomic instability

Guanine-rich sequences are located throughout the genome of all organisms and therefore cannot be inherently deleterious. The clinical manifestation of mutations in helicase proteins with the ability to unfold G4 structures suggests that maintenance of proteins which properly fold/unfold G4 structures is mandatory to maintain genomic stability. (40)

Unless repaired faithfully errors during DNA replication will lead to abnormal cellular phenotypes. Guanine-rich sequences have been associated with genomic instability caused by abnormal DNA breaks and subsequent translocations. The chromosomal breakpoint associated with the lymphoma-associated *bcl-2* gene translocating from human chromosome 14 to chromosome 18 has been linked to guanine-rich repeats with a propensity to form G4 structures (83). Similar guanine-rich repeats with a propensity to fold into G4 structures have been mapped in the breakpoint location of the *shank3* gene (84). An inability to unfold G4 structures during DNA replication may lead to DNA breaks at G4 structures and create regions inadvertently repaired by DNA translocation events.

Genomic instability has also been associated with a number of human diseases relating to an unusual expansion of triplet repeat sequences (85, 86). The most common form of human mental retardation Fragile X syndrome, is caused by the rapid expansion of d(CGG) triplet repeats. The d(CGG) triplet repeat expansion occurs adjacent to the promoter of the *fmr1* gene and suppresses its expression leading to a diseased state (87). Under physiological conditions d(CGG) repeats have been shown to form stable G4 structures *in vitro* and it has therefore been proposed that looped-out secondary structures of G4 during replication of the *fmr1* promoter may lead to DNA slippage and subsequent expansion of

repeats with each round of replication (59). Confounding the analysis of this disease is a complete understanding of the FMR1 protein. The FMR1 protein has been implicated in the regulation of multiple messenger RNAs and mutations in FMR1 proteins may result in an inability to bind certain mRNAs (88) . Of significant importance was the discovery that the FMR1 protein binds strongly to G4 RNA structures (89). As a result, it has been suggested that mutations in FMR1 may cause the loss of G4 RNA binding and the deregulation of mRNA molecules. Continued characterization of the folding and unfolding mechanisms of G4 structures may help to elucidate how loss or gain of G4 structures may lead to genomic instability.

1.3.3   Telomeres

Predictably, telomeres have always and continue to occupy the vast majority of the scientific literature pertaining to G4 structures in the genome. Telomeres are specialized complexes located at the ends of all linear chromosomes and consist of simple DNA repeats and an array of associated proteins (90). All vertebrates share the same conserved repeat sequence of $d(TTAGGG)_n$ (91, 92). In addition, telomeres feature a highly conserved single-stranded overhang at the extreme 3′-terminal end of the guanine-rich strand, which in humans is 75-300 nucleotides long (93-95). The tandem DNA repeats and associated proteins of telomeres are essential parts of the genome and help to protect the chromosomal ends from end-to-end fusions, non-homologous end joining and aberrant recombination (96).

Although the extreme 3' end of some telomeric DNA sequences is fixed, like that of the 16 nucleotide *Oxytrica nova* telomeric G-overhang of exactly

d(TTTTGGGGTTTTGGGG) (97), most other organisms possess much longer telomeres up to several kilobases in length (98, 99). Also, unlike the precise number of repeats found at the terminal ends of *Oxytrica nova*, the number of tandem telomeric repeats in most other species varies from organism to organism and from telomere to telomere.

Human telomeric repeats interact with a large number of proteins including some that appear to be telomere specific like TRF1 and TRF2 (telomere repeat factor 1 and 2) and others that can associate with multiple DNA sequences like BLM, FANCJ and Ku (59). The combination of tandem DNA repeats and associated proteins act as protective caps at the ends of chromosomes. In addition, mammalian telomeres have been shown to form multiple secondary structures including a loop (T-loop) structure proposed to form by the invasion of duplex DNA by the 3' guanine-rich overhang (100). The joint of the T-loop may involve a G4 structure of DNA at the point of strand invasion and the exact structure at this junction (D-loop) is not known. However, there are many nuclear proteins that specifically bind the complementary cytosine-rich strand that may help to stabilize the telomere when a G4 structure is formed (101).

In vivo evidence for the presence of G4 structures at human telomeres remains controversial. However, human telomeric sequences have been found to spontaneously assemble, under physiological conditions, into a number of G4 structures *in vitro* (102-106). Some of the best *in vivo* evidence for the formation of G4 structures at telomeres comes from research on model organisms. The ciliated protozoa *Stylonychia* has been an invaluable model organism for the study of telomeres due in part by the high number of telomeres ($<10^9$) in the transcriptionally active macronucleus (107). It has been shown that antibodies developed to antiparallel-stranded G4 structures bind to the telomeres of *Stylonychia* as

visualized by immunostaining (108). In addition, it was discovered that the expression of two telomere end-binding proteins (TEBP) α and β were required for the binding of this antibody (109). It was shown that TEBPβ could catalyze the formation of G4 structures *in vitro*, and that TEBPα was required for the recruitment of TEBPβ to the telomere. Importantly, the constraint of TEBPs for the successful detection of G4 structures showed that visualization of telomeric G4 structures in *Stylonychia* was not due to the stabilization of G4 by the antibodies but rather depended on the natural existence of G4 structures at the telomeres *in vivo*.

In a recent study Smith *et al*., provided further *in vivo* evidence for the existence of G4 structures at telomeres in the yeast model organism *Saccharomyces cerevisiae* *(S.cerevisiae)* (110). Telomere protection and capping in yeast requires the function of one protein, Cdc13 (111). By examining temperature sensitive Cdc13-1 mutants under conditions in which G4 structures were stabilized, including: the over-expression of G4 DNA binding proteins, loss of Sgs1 helicase and the addition of G4 stabilizing ligands, Smith *et al*., noticed that telomere capping was enhanced and defects associated with Cdc13-1 telomere uncapping were suppressed. Consistent with this, telomere uncapping occurred under conditions that did not allow for G4 structure formation at the telomere. Including targeted mutations in the G-rich strand there by disrupting the G-quadruplex forming potential at telomeres. It was proposed that G4 formation at the telomeres in *S.cerevisiae* may enhance the capping function by allowing G4 DNA binding proteins to bind G4 structures and help stabilize the telomeric ends, thereby facilitating the action of Cdc13 which normally blocks the 5' to 3' resection of telomeric ends (110).

The insinuation of telomeric G4 structures *in vivo* in both *Stylonychia* and

S.cerevisiae does suggest that G4 structures may be a general mechanism for maintaining

telomere integrity. These two model organisms have provided credible evidence that G4

DNA structures possibly exist and that they may function similarly in more complex

organisms, such as mammals. However the existence of *in vivo* evidence for the presence of

G4 structures in mammals remains elusive.
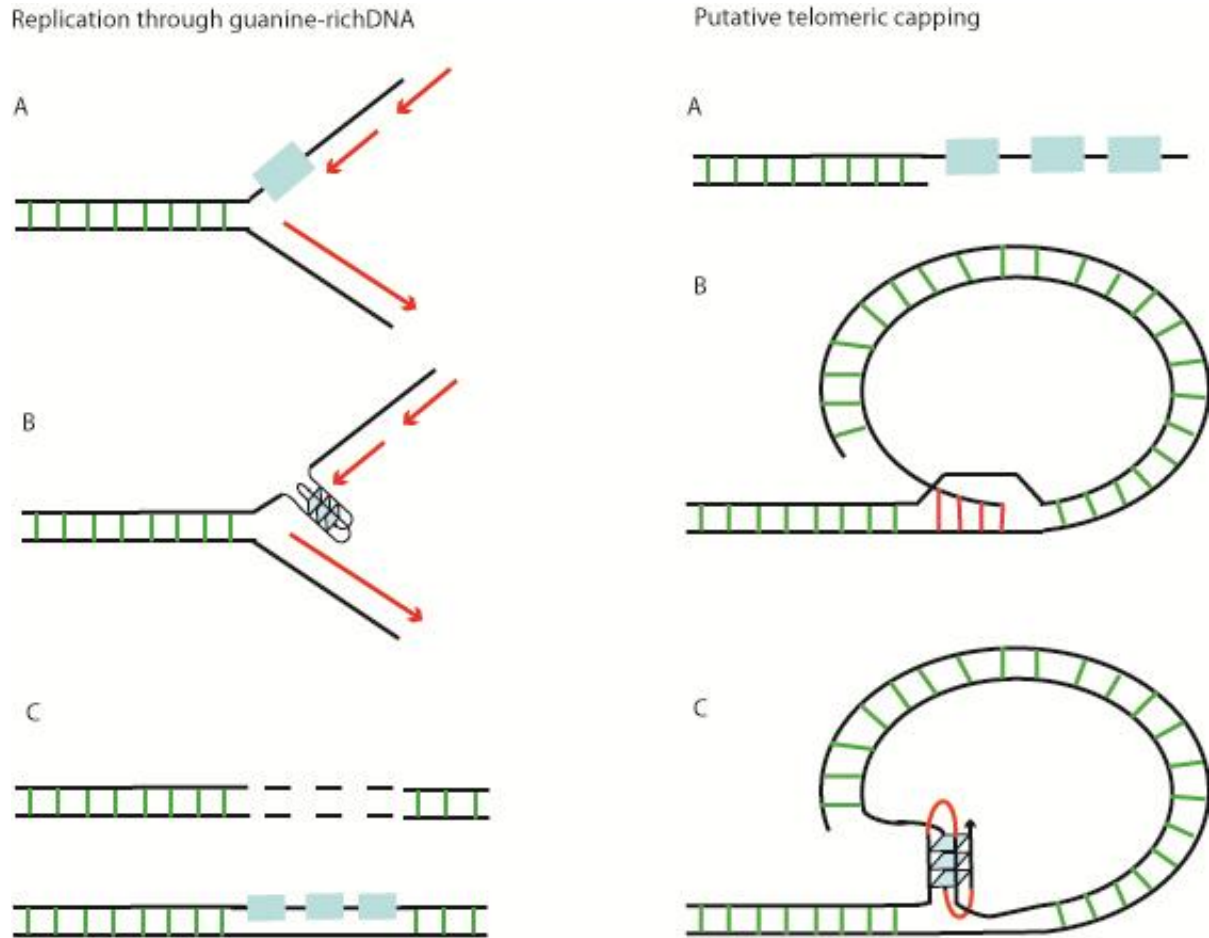
Figure 7. Biological implications of G4 structures in the genome.

Some of the possible biological consequences of having G4 structures forming in the genome are depicted. In the left panel, during replication (A) blocks of guanine-rich sequences (blue blocks) become naturally single stranded and may form G4 structures (B). Two possible outcomes include deletion of guanine sequences (like those seen in *dog-1* mutants in *C.elegans* (71)), or triplet repeat expansion (C) (as seen in Fragile X syndrome (88) . G4 structures could also arise at guanine-rich telomeric ends (right panel A). The single stranded invasion of the guanine-rich strand in the T-loop (B) could also be explained by the formation of G4 structures and may result in alternative modes of telomere capping.

1.3.4   Inhibition of telomerase

Mammalian telomeric DNA is very heterogeneous in length and contains anywhere from 5 to 30 kilobases of d(TTAGGG) repeats at the end of each chromosome in humans to over 150 kilobases at the ends of some chromosomes in mice (112-115) . The very terminal ends of telomeres cannot be fully replicated and cause the slow erosion of 50 to 200 base pairs of DNA during each round of replication (the so called end replication problem) (116-118). The gradual shortening of telomeres results in the eventual uncapping of telomeres and activation of a DNA damage response and cell cycle arrest or apoptosis. It is possible that a decrease in telomere length in cells over time contributes to normal aging.

An enzyme called telomerase is used to maintain telomere length during cellular division in some cells of the body. Telomerase is a reverse transcriptase that contains an RNA template for the *de novo* addition of guanine-rich telomeric repeats to the ends of chromosomes. Telomerase is highly regulated in healthy human cells and only actively expressed in a subset of cells. However, it has been shown that 85-90% of all cancers show high levels of telomerase activity allowing tumor cells to evade the end replication problem and divide indefinitely (119-121). The only other known mechanism of telomere maintenance in tumor cells is termed ALT (alternative lengthening of telomeres) which is a recombination based mechanism that allows the maintenance of telomeres by recombination events between the ends of chromosomes and perhaps circular telomeric DNA (122). Regardless of the mechanism, the maintenance of telomere length in tumor cells is critical to their ability to proliferate indefinitely. A great deal of research has therefore focused on the manipulation of tumor cell fate by telomerase inhibition (123-127). As a side note, the

activation of telomerase for the purpose of slowing cellular aging is fraught with risk, including the adverse activation of oncogenes (128).

The template for the enzymatic elongation of telomeres by telomerase is the 3′ single-stranded guanine-rich over-hang. Telomerase requires the guanine-rich overhang as a site of binding and de novo addition of repeats. Remarkably, some G4 DNA structures have been shown to inhibit the enzymatic activity of telomerase (126, 129). Zahler *et al*., have shown that telomerase activity decreases significantly in the presence of K^+ ions and conclude that telomeric folding into G4 structures inhibited the enzymatic activities of telomerase *in vivo* (129). It would appear that conversion of the single-stranded guanine-rich overhang into a G4 structure interferes with the direct base-pairing interaction of the RNA template of telomerase and the guanine-rich strand (17). This discovery has led to the rapid production of small-molecule ligands that selectively stabilize G4 DNA in an attempt to convert the telomeric ends of chromosomes into structures no longer capable of undergoing elongation by telomerase. However, in most studies a distinction between the effect of compounds on the substrate for telomerase (a guanine-rich oligonucleotide) and the reaction product of telomerase (single-stranded guanine-rich DNA) was not made. As a result it cannot be excluded that many of the compounds that are claimed to inhibit telomerase activity by blocking its substrate in fact have no such activity.

1.3.5   G4 stabilizing ligands

In theory, if the only function of telomerase was the maintenance of telomeric repeats then the inhibition of telomerase would not be toxic to cells until several rounds of

replication have eroded away the telomeric repeats until they became critically short, causing telomere uncapping, senescence and cell death. However, it has been shown that some telomerase inhibitors have much quicker cytotoxicity to cells and therefore might act on mechanisms other than just telomere length maintenance.

Some G4 structures have been shown to inhibit the enzymatic addition of telomeric repeats by telomerase (126, 129). Not only have these results made G4 DNA a promising target for the therapeutic intervention of cancer cells it has spurred an explosion of research into the development of small-molecule stabilizers of G4 structures (123, 130-134). The first small molecule that was shown to bind G4 DNA and inhibit telomerase was an anthraquinone derivative created in 1997 (135). Since then hundreds of compounds have been either created or found to bind G4 DNA and have the potential to be toxic to cancer cells (123, 130-134).

One example of a small-molecule that was designed is a drug called PIPER (136). Using different modeling approaches PIPER was designed and synthesized by the Hurley lab and found to be a potent telomerase inhibitor. In addition, PIPER has been shown to accelerate (by 100 fold) the formation of bimolecular G4 structures and is therefore known as a G4 molecular chaperone (137). Another famous small-molecule binder of G4 structures is the porphyrin molecule TMPyP4 (136, 138, 139). TMPyP4 has also been shown to inhibit telomerase, stabilize anaphase bridges and be very toxic to human cancer cells (140, 141). Interestingly TMPyP4 has also been shown to suppress the proliferation of ALT cells as well as telomerase positive cells (141). Nevertheless, the most potent and selective telomerase inhibitor to date is actually a naturally occurring compound called telomestatin (142).

Telomestatin is isolated from *Streptomyces anulatus* and is the only telomerase inhibitor that can stabilize G4 structures in the absence of monovalent cations (143).

Most structural formulas of telomerase inhibitors indicate that aromatic compounds with large electron deficient rings are best suited for the end stacking of the electron rich G-quartets. Detailed information about small-molecule complexes with G4 structures will help drug design and increase our understanding of G4 molecular recognition (43). One of the major challenges of getting G4 stabilizing small-molecule therapies into clinical trials will be the identification of nontoxic small-molecules that truly select distinct G4 structures (96).

1.4   Experimental hypothesis

One of the foremost concerns regarding the potential applications and biological roles of G4 structures in cells is a lack of *in vivo* proof that G4 structures naturally occur in human cells. To date there have been limited tools for the examination of G4 structures in human cells, as such much of the proposed biological roles have not been scientifically investigated. Research into the existence of G4 structures in human cells requires better tools for the *in vivo* detection of these structures.

Although *in vivo* evidence does exist in other organisms, notably *Stylonychia* and *Saccharomyces cerevisiae*, no group to date has shown any *in vivo* evidence that G4 structures occur in vertebrates. Understandably there are key differences that may preclude similar experiments done in model organism from effectively discovering G4 structures in vertebrates. For instance, the visual detection of G4 structures in *Stylonychia* through the use of specific antibodies is facilitated by the extreme abundance of nanochromosomes in the

macronucleus of this organism. By contrast, most healthy human cells only have 46 chromosomes and therefore only 92 telomeres. The detection of only 92 fluorescently-labeled antibodies (assuming a minimum of one G4 structure per telomere) requires much more extreme resolution than required to visualize the hundreds of thousands found in *Stylonychia*. In addition, the *in vivo* evidence that telomeres are capped by G4 structures in cdc13 mutants of *Saccharomyces cerevisiae* cannot be directly compared to the capping mechanisms found in human cells. *Saccharomyces cerevisiae* cells do not fold into structures like the T-loops and healthy human telomeres are capped by a multitude of proteins not found in *Saccharomyces cerevisiae*. Nevertheless, these model organisms do provide *in vivo* evidence that may lead to the eventual discovery of G4 structures in vertebrate systems.

In addition, as previously stated many diseases have been associated with defects in genes and proteins associated with potential G4 structures. To fully understand and potentially therapeutically intervene in these diseases we require new tools to explore G4 structures in human cells.

Given that G4 structures exist *in vivo* in other organisms we (like many other groups) hypothesize that G4 structures do naturally exist in human cells. If our hypothesis is correct we believe that new tools that can discriminate between single-stranded, double-stranded and quadruplex structures could provide insight into the *in vivo* existence of G4 structures in human cells. Monoclonal antibodies that can distinguish between G4 structures may help elucidate the natural existence of specific G4 nucleic acid structures in human cells. The main goals of our research are thus:

1) Develop novel monoclonal antibodies to different G-quadruplex structures, and
2) Using these novel reagents, attempt to ascertain the *in vivo* existence of G-quadruplex nucleic acids in human cells

Chapter 2: Constructing validated quadruplex structures and the biochemical analysis of purified monoclonal antibodies

While chemical and *in vitro* analyses of quadruplex structures have been extensive, the study of G-quadruplex DNA *in vivo* has been limited, partly due to the lack of reagents designed to target such secondary structures of DNA. To date there has been no direct detection of G-quadruplex DNA in human cells. To facilitate biological studies on the role of G-quadruplex DNA in human cells we have generated monoclonal antibodies to multiple G-quadruplex DNA structures. Here we outline the experimental methodologies used for generating different structural polymorphisms of G-quadruplex DNA, the subsequent immunization strategy used for producing antibodies directed against defined G-quadruplex DNA structures and the biochemical characterization of the novel monoclonal antibodies that were obtained. One antibody, termed 1H6 appears to bind several different quadruplex structures.

2.1 Introduction

Hybridoma technology has been around since the 70's and has provided a valuable tool for the development of monoclonal antibodies. Monoclonal antibodies have been approved for use in diagnostic tests and for therapeutic treatments in both cancer and autoimmune diseases. They continue to be very valuable laboratory research tools. As such, the inventors of hybridoma technology, Cesar Milstein and Georges Köhler were award the Nobel Prize for Medicine in 1984 for their breakthrough. Although numerous other techniques for generating antibodies have been developed since, including phage display and

plant based antibody production, hybridoma production continues to be a popular approach for generating antibodies in the research laboratory.

Several groups have developed antibodies and antibody fragments that can bind G4 DNA structures *in vitro* (108, 171, 172, 176). However, only the antibodies made by Schaffitzel *et al.*, that bind the telomeric G4 DNA in *Stylonychia* have been shown to detect naturally occurring G4 structures *in vivo* (109). These antibodies have not been shown to bind anything in vertebrates and appear to be specific to the folded telomeric structures of *Stylonychia* (109). The telomeric repeats from *Stylonychia* (T_4G_4) are different from vertebrates (T_2AGGG) and therefore may preclude these antibodies from detecting similar structures of a different sequence variant in vertebrates.

In an attempt to achieve direct *in vivo* evidence and explore the biology of G4 DNA in human cells we have developed a set of antibodies to G4 DNA through the immunization of mice with defined G4 DNA structures. Our initial strategy for generating these antibodies was to first create distinct G4 structures that were stable enough for immunization. Tetramolecular G4 structures are extremely stable once formed and can easily be purified from unfolded strands. In this regard we chose to begin our experiments with tetramolecular G4 structures. In later experiments we illustrate our attempts at generating antibodies to bimolecular and unimolecular structures.

In this chapter we discuss the development and characterization of different G4 DNA structures used for immunization or antibody analysis. Native polyacrylamide gel mobility assays and circular dichroism spectroscopy were methods employed to identify and characterize purified G4 DNA structures. Subsequently, we describe how these purified G4 DNA structures are exploited for immunization and screening of antibody titers. Following

good titer levels, we detail how hybridomas were generated, screened and cloned. Finally,

we describe the initial characterization of the purified monoclonal antibodies to G4 DNA,

including their specificity, affinity and CDR amino acid composition.

2.2   Experimental methodology

2.2.1   Oligonucleotide secondary structure formation

Oligonucleotides were synthesized using an ABI 394 DNA/RNA synthesizer and underwent standard desalting prior to shipment from the University of Calgary DNA Synthesis lab. All sequences were desalted and lyophilized prior to shipment. Both 5'-amino and 3'-biotin modifications were incorporated into DNA oligonucleotides at the time of synthesis. A list of oligonucleotides used in this study is depicted in Table 3. For unimolecular G4 structures, oligonucleotides were resuspended at a final concentration of 0.1 mM in 10 mM Tris-HCl, 1mM EDTA (TE). Solutions were heat denatured for 4 minutes at 95 °C before the addition of 100-150 mM (final) NaCl or KCl. Samples were allowed to slowly equilibrate to room temperature for 5-8 hours. For tetramolecular G4 structures, oligonucleotides were resuspended at a final concentration of 1mM in TE. Samples were heat denatured at 100 °C for 2 minutes and subsequently placed on ice prior to the addition of 1M (final) NaCl. Cooled samples were sealed in microcentrifuge tubes and incubated at 50 °C for 72 hours. Duplex DNA and the previously designed GC hairpin (148) was generated by 2 minute heat denaturation at 95 °C and annealing complementary sequences at room temperature. Single stranded oligonucleotides were maintained in TE and underwent 2 minute heat denaturation at 95 °C prior to use. The previously designed triplex structure (149) was generated by first annealing the complementary sequences (TC30W + TC30C) into a duplex structure. To prepare a stable triplex the inverse 30-mer strand was annealed to the 30 bp duplex in a 1:2 molar ratio in the presence of 33 mM Tris-acetate (pH 5.5) and100

mM NaCl. Oligonucleotide complexes were separated from unbound oligonucleotides by native polyacrylamide electrophoresis and gel purification.

2.2.2   Polyacrylamide gel analysis and quadruplex purification

DNA secondary structure was assessed by native polyacrylamide gel electrophoresis (PAGE) with 12-15% polyacrylamide gels (19:1 ratio of acrylamide to bisacrylamide) containing 10 mM NaCl or 10 mM KCl. Running buffer consisted of 1X TBE with 10 mM NaCl or 10 mM KCl. Samples were diluted 1:2 in non-denaturing loading buffer (30% glycerol, 1 mM EDTA with xylene cyanol and bromophenol blue). Oligonucleotides consisting of known stretches of thymines were used as molecular weight markers in conjunction with corresponding heat denatured monomers. Gels were run at room temperature at 5-10 watts until xylene cyanol marker had migrated roughly 20 cm (5-10 hours). Samples were subsequently visualized by UV shadow, ethidium bromide or Gel Red staining. For UV shadowing, gels were wrapped in saran wrap and placed on a TLC plate. A hand held Mineralight lamp emitting short wave UV light (254 nm) was used to detect bands of interest. In this way, bands corresponding to both structured DNA and unstructured monomers could be visualized and excised. Each excised gel fragment was crushed into elution buffer (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 25mM NaCl or KCl) and shaken overnight (17-20 hours) at 4 $^{\circ}$C. Eluted DNA was syringe-filtered from gel pieces and concentrated by repeated rounds of 2-butanol extraction. Butanol extraction was performed by the addition of an equal volume of 2-butanol to an equal volume of aqueous DNA sample, vigorously shaken and centrifuged for 30 seconds at 6000 rpm. After removal of the upper organic layer, the procedure was repeated until the lower aqueous layer had decreased to less

than 400 µl. Once DNA samples had been concentrated to 400 µl or less, ethanol precipitation was performed by adding 1/10th the volume of 3 M NaOAC followed by the addition of 1 µl of LPA (linear polyacrylamide) as a co-precipitant and 2.5x the volume of ice-cold 100% ethanol. Samples were then vortexed for 1 min and placed at -20 °C for 30 minutes. Following incubation, samples were centrifuged for 30 minutes at 4°C at 13,000 rpm. The sample pellet was washed one time in 70% ethanol followed by a 5 minute centrifuge at 13, 000 rpm. Purified G4 samples were resuspended in phosphate-buffered saline (PBS) and monomers were resuspended in TE. Both were stored at -20 °C until further use.

2.2.3 Nanodrop quantification

Nucleic acids were quantified spectrophotometrically on a Nanodrop ND-1000 using a calculated 1 cm path length. Molar concentrations of nucleic acids and their structures was calculated using molar extinction coefficients (ε) calculated per strand using the IDT Biophysics calculator (http://biophysics.idtdna.com/) and the following equation:

A/ε = molar concentration

Where A represents the absorbance measured at 260 nm.

2.2.4 Circular dichroism spectropolarimetry

Samples prepared for circular dichroism (CD) spectroscopy were diluted in PBS or TE to a final concentration of 5-10 µM (70-95 µg/ml). CD spectra were measured using a Jasco-810 spectropolarimeter at 25 °C. Readings were recorded over a wavelength range of

200 to 320 nm in a quartz cuvette with a 1 mm path length. Measurements were averaged

between three accumulations with an instrument scanning speed of 200 nm/min and readings

were blank-corrected with buffer alone.   Measurements are presented in molar ellipticity and

nanometers.

2.2.5   Antigen formulation and immunization strategy

Reactive sulfhydryl groups were generated on keyhole limpet hemocyanin (KLH) by

the addition of a 20-fold molar excess of Traut's reagent. The heterobifunctional crosslinker

sulfo-MBS was used to conjugate the amino modified G4 DNA and reactive KLH to produce

KLH-G4 conjugates. KLH-G4 conjugates were emulsified (1:1) in complete Freund's

adjuvant (CFA) and used in immunizations of C57BL/6J mice, obtained from the Jackson

Laboratories. Groups of four 6-8 week old mice were immunized 4-5 times at 2-week

intervals with 50 µg of KLH-G4 DNA. Primary immunizations consisted of KLH-G4 DNA

emulsified in CFA and administered by intraperitoneal injection.   Subsequent immunizations

consisted of KLH-G4 DNA emulsified in incomplete Freund's adjuvant (IFA) also

administered via intraperitoneal injection. Blood samples were regularly collected 2 weeks

post immunization, centrifuged at 10,000 rpm for 5 minutes at 4 °C. Serum was extracted

from blood samples and stored at -20 ºC until use.

2.2.6   ELISA serum screening

For screening of anti-G4 DNA antibodies, 96-well enzyme-linked immunosorbant

assay (ELISA) plates were pre-coated with streptavidin (1.5 µg/mL) in PBS over night at 4

°C. Plates were subsequently blocked for 1 hour at 22 ºC with 2% BSAP (2% Bovine Serum

Albumin (BSA), 0.1% Tween-20 in PBS). After 3 washes with PBS/0.1% Tween-20

(PBST), the plates were incubated at 22 ºC for 1 hour with 0.13 µg/ml of biotinylated-G4

DNA diluted in PBS. After 3 washes with PBST, serial dilutions of mouse sera diluted in

1% BSAP were incubated at 22 ºC for 2 hrs on previously coated plates. Upon 3 washes

with PBST, bound anti-DNA antibodies were detected with HRP labeled goat-anti-mouse

IgG and identified using 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) (ABTS) as

the substrate. Antibody binding was quantified by spectrophotometry at 405 nm.

2.2.7   Hybridoma generation

Splenocyte/myeloma fusions were performed with highest serum responders

according to the method described by Köhler and Milstein (21). In brief, SP2//0 0 Ag 14

myeloma cells (205) grown at log phase in Iscove's Modified Dulbecco's Medium (IMDM)

supplemented with 10% fetal calf serum (FCS), 2 mM glutamine, 50 units/ml penicillin and

50 µg/ml of streptomycin (IMDM/10/PS). 3 days prior to fusion, mice were boosted with 50

µg of KLH-G4 in PBS without adjuvant. Mice were sacrificed under CO_2 and spleens were

removed. Spleens were dissociated into a single cell suspension by mincing with frosted

slides and resuspended in 10 ml of IMDM/10/PS. Splenocytes were transferred to a T-75

tissue culture flask with an additional 15 ml of IMDM/10/PS. Cells were incubated at 37 °C

in a 5% CO_2 incubator for 30 minutes to removed adherent cells including macrophages.

After centrifugation at 1,2000 rpm for 5 minutes at 4 °C, splenocytes were counted using 3%

acetic acid in methylene blue while myeloma cells were counted using a hemocytometer and

trypan blue. FCS was cleared from solutions with splenocytes and myeloma cells by

washing cells twice with warm IMDM. Splenocytes and myeloma cells were combined in a

3:1 ratio and washed one more time in IMDM kept at 37 °C.   After the final wash, cells were

resuspended in 1 ml warm IMDM containing 50% polyethyleneglycol 1500 (PEG, Sigma

P7181). 9 mls of warm IMDM was added over a period of 5 minutes. Fused cells were then

pelleted by centrifugation at 1,200 rpm for 10 minutes and slowly resuspended in selective

HAT medium (IMDM/10/PS supplemented with hypoxanthine-aminopterin-thymidine).

Cells were cultured at a minimum of 125,000 fused spleen cells per well of a 96-well plate.

Fused cells were cultured in HAT medium for 7 days followed by removal of ½ the

supernatant and replacement with HT media (IMDM/10/PS supplemented with

hypoxanthine-thymidine). After 10 days of culture hybridoma supernatants were screened by

standard ELISA assays using 96-well plates coated with 1μg/35 μl biotinylated-G4 DNA.

Positive colonies were expanded in HT media only.

2.2.8   Cloning analysis by ELISA screening

Upon positive identification by ELISA, cells were harvested and transferred to 1.5 cm

Petri dish plates containing semi-solid medium: 3 ml of IMDM/10/PS supplemented with

0.9% methylcellulose and 75 units of IL-6. Hybridomas were cultured until colonies were

visually identifiable. Clonal populations were picked from methylcellulose plates and

transferred to individual wells of a 96-well plate in liquid IMDM/10/PS supplemented with

50 units of IL-6. Clones were cultured for four days and supernatants from selected clones

were screened by ELISA. Three rounds of single cell cloning were performed before

hybridomas were considered monoclonal populations.

2.2.9 Antibody purification using FPLC

Antibodies were purified from hybridoma supernatants by affinity chromatography using HiTrap protein G HP columns on an ÄKTA Purifier instrument according to standard protocols. In brief, 500 ml spinner flasks containing overgrown culture medium from clonal hybridoma cultures were centrifuged at 7,000 rpm for 30 minutes at 4 °C. Supernatant was transferred and the pH was adjusted with $1/20^{th}$ the volume of 1 M Tris-HCl (pH 8) and the addition of 0.02% sodium azide. 1 ml protein columns were first equilibrated with loading buffer consisting of 20 mM sodium phosphate buffer (pH 7.4) and subsequently used to capture monoclonal antibodies by running 300 – 500 ml of equilibrated supernatant through each column at a rate of 1 ml/minute. After re-equilibration with loading buffer, antibodies were eluted from the column with elution buffer consisting of 100 mM Glycine (pH 2.9). Eluted antibodies were re-equilibrated to physiological pH by the addition of 1M Tris-HCl (pH 9.0) in the capture microcentrifuge tubes. Purified antibodies were then buffer-exchanged into PBS by gel filtration chromatography with a Fast Desalting column HR10/10 or Microcon centrifugal filter tubes. Sample purity was confirmed by SDS-PAGE, Coomassie staining and quantified by Nanodrop ND-1000.

2.2.10 Isotyping

Antibody isotyping was performed on 96-well plates first coated with streptavidin at 1.5 µg/ml and subsequently blocked with 2% BSAT. 1µg/35µl of biotinylated-G4 was added to the wells and incubated at 37 °C for 1 hour. Upon washing, wells were blocked with 50 µl of normal rabbit serum and a single specific rabbit-anti-mouse antibody: anti-mouse IgG_1,

anti-mouse IgG$_{2a}$, anti-mouse IgG$_{2b}$, anti-mouse IgG$_3$, anti-mouse IgA, anti-mouse IgM, anti-mouse Kappa light chain or anti-mouse Lambda light chain to individual wells on a plate. Bound antibodies were detected using goat-anti-rabbit-HRP and revealed with the ABTS substrate, and were quantified by spectrophotometry at 405 nm.

2.2.11 Apparent affinity by ELISA

Apparent antibody affinity was determined by binding curve analysis (150, 151). Biotinylated versions of immunizing G4 DNA was captured on SA coated ELISA plates and probed with various concentrations of antibodies to generate a binding curve for each antibody G4 combination. Bound antibodies were detected using goat-anti-rabbit-HRP and revealed with the ABTS substrate, and were quantified by spectrophotometry at 405 nm using a microplate reader. Apparent affinities were determined from the resulting binding curves by nonlinear regression and the antibody concentration at half-maximal binding using GraphPad Prism software.

2.2.12 Antibody specificity analysis

Antibody specificity was initially screened by a direct ELISA assay where biotinylated G4 structures were captured by SA coated plates while double-stranded DNA, single-stranded DNA and various proteins were dried to the plates. Purified antibodies were subsequently added at a ½ maximal binding concentration and incubated on pre-coated plates for 1 hour at room temperature. Bound antibodies were detected using goat-anti-mouse-HRP and revealed with the ABTS substrate, and were quantified by spectrophotometry at 405 nm

using a microplate reader. Antibody 1H6 also underwent another round of specificity testing by competition ELISA. The relative binding specificity by the 1H6 antibody was determined by triplicate measurements by pre-incubation with different competitors at room temperature for 1 hour.   1H6 antibody concentration was below ½ maximal binding and constant in each competitive reaction. All competing agents were in 1000-fold molar excess over 1H6 antibody. Plates coated with immobilized G4-DNA were used to evaluate percentage competition in comparison to incubated antibodies without competitor. Soluble competitors included: *Oxytricha* intermolecular G4 (Oxy-2), vertebrate intermolecular G4 (Ver-3), poly d(G), previously designed GC hairpin (148), G5 intermolecular parallel stranded quadruplex (G5 G4), *Oxytricha* intramolecular G4 (Oxy-4), vertebrate intramolecular G4 (Ver-4), *Oxytricha* unfolded monomer, vertebrate unfolded monomer, *Oxytricha* duplex, vertebrate duplex, and biotin. A minimum of 4 µg of synthetic DNA or protein was loaded per competition. All competition reactions were incubated in 1% BSAP and monomeric samples were initially denatured for 2 minutes at 95 °C to maintain single stranded form. Anti-DNA antibodies were detected with HRP labeled goat-anti-mouse IgG and revealed with ABTS substrate. % competition = [(OD control- OD with competitor)/OD control] x 100.

2.2.13   Antibody variable heavy and light-chain sequencing

Variable heavy and light chains of each hybridoma were sequenced following Wang *et al.*, (152). In brief, isotyped hybridomas were grown in 24-well tissue culture plates and harvested for RNA isolation using a Qiagen RNeasy kit. The resultant RNA was dissolved in 60 µl of RNAse free water and stored at -20 °C prior to use. cDNA was synthesized using reverse transcriptase included in the High Capacity RNA-to-cDNA kit by Applied

Biosystems. Up to 2 µg of RNA was converted to cDNA using a PCR cycling setup of 37 °C

for 60 minutes followed by 95 °C for 5 minutes and a final step of 4 °C indefinitely. The

resultant cDNA was amplified using the high fidelity enzyme Extensor mix. Each PCR

reaction contained: 2 µl of cDNA, 20 pmol 5' and 3' primers, 0.4 µl 25 mM dNTPs, 10 µl of

10x PCR buffer and 1 µl Extensor enzymes mix. Modified primers used for specific variable

heavy and variable light chain amplification are listed in table 2 bellow. PCR cycling

conditions were: initial melt at 94 °C for 3 minutes followed by 30 cycles of a three step

program 94 °C for 1 minute, 45 °C for 1 minute and 72 °C for 2 minutes. The reactions were

subsequently held at 72 °C for 10 minutes and cooled to 4 °C. Amplicons were visualized

using a 1.2% agarose gel stained with ethidium bromide. PCR amplicons were then cut from

agarose gels and purified using QIAquick PCR purification spin columns. Purified

amplicons were sequenced by the NAPS unit (Nucleic Acid Protein Service Unit, University

of British Columbia) using M13 forward and reverse primers and analyzed using IMGT/V-

Quest program (International ImMunoGeneTics Information System) (http://imgt.cines.fr/).

Sequence 5' to 3'	Length	Target
CAGGAAACAGCTATGACCATAGACAGATGGGGGTGTCGTTTTGGC	36	3' IgG1
CAGGAAACAGCTATGACCCTTGACCAGGCATCCTAGAGTCA	32	3' IgG2a
CAGGAAACAGCTATGACCGGCCAGTGGATAGACTGATGG	33	3' IgG2b
TGTAAAACGACGGCCAGTSARGTNMAGCTGSAGSAGTC	32	5' MH1
TGTAAAACGACGGCCAGTSARGTNMAGCTGSAGSAGTCWGG	35	5' MH2
CAGGAAACAGCTATGACCGGATACAGTTGGTGCAGCATC	30	3' Kc
TGTAAAACGACGGCCAGTGAYATTGTGMTSACMCARWCTMCA	32	5' Mk

Table 2. Table of modified primers used for variable heavy and light chains of antibodies.
Primers used for the sequencing of variable chains were modified with the addition of M13
sequences for ease of use. The degenerate primers are modified from Wang *et al.,* 2000
(152).

2.3   Results

2.3.1   Native polyacrylamide gel migration patterns discern distinct quadruplex species

Telomeric DNA from most species comprises regions of guanine-rich repeats (90). In addition, the 3′ distal portion of most telomeres is naturally single-stranded (albeit in association with proteins). Due to the inherent single-stranded guanine-richness of most telomeric sequences (Table 1, Chapter 1) we chose to generate stable G-quadruplex structures from telomeric oligonucleotides. The immunizing sequences were derived from vertebrates and ciliates (*Oxytricha* and *Tetrahymena*) (as initiated by Dr. Melita Irving). Our first objective was to isolate stable and distinct quadruplex structures for immunization and screening. To analyze the structures that were obtained, all oligonucleotides were visualized by native polyacrylamide gel electrophoresis (PAGE) (see Experimental Methodology). PAGE allows structures to be visually separated by migratory speeds. Small changes in secondary structure will alter the rate of migration upon acrylamide gel electrophoresis and can be detected by UV shadowing. Although the sensitivity of UV shadowing is relatively low (10-50 ng DNA/RNA) it does allow the visual detection of DNA without the use of intercalating dyes that might otherwise affect G4 structure or interfere with immunization attempts later. However in part due to the low sensitivity of UV shadowing other techniques to characterize G4 structures (like circular dichroism) are employed in conjunction to native PAGE to improve structural characterization. Figure 8 illustrates how high sensitivity dyes

like Gel Red appear to detect multiple species in a single incubation condition (Lane 1 and 3) that may only be detected as a single species by lower sensitivity UV shadowing.

Tetramolecular G4 DNA structures were generated from telomeric oligonucleotides and visually separated from monomers using PAGE. Tetramolecular G4 DNA structures are easy to identify over duplex DNA or unstructured single stranded DNA by their migratory retardation in PAGE (Lane 1, Figure 8). In addition, tetramolecular G4 DNA structures are readily formed at high oligonucleotide concentrations and high temperatures and therefore have discrete conditions for folding for easy purification and immunization (15, 19, 41).

Formation of tetramolecular G4 DNA using *Oxytricha* telomeric sequences is relatively favorable, whereas under similar incubation conditions the vertebrate telomeric sequence does not efficiently assemble into an equivalent four-stranded quadruplex (communication with Dr. Melita Irving). To increase the attractive forces between vertebrate sequence oligonucleotides, the telomeric sequence was modified by adding a stretch of five guanine nucleotides to the 5' end of the vertebrate sequence (15) improving the efficiency of tetramolecular G4 DNA formation. However, the modification of the telomeric sequence could give rise to antibodies specific for the modification and not the unmodified telomeric sequence. Antibodies generated to the modified sequence required cross-reactivity testing to the natural telomeric sequence. Slow moving bands corresponding to higher molecular structures were excised and purified prior to circular dichroism analysis (Lane 1 Figure 8).

In contrast, unimolecular G4 DNA structures frequently exist in dynamic equilibrium between different structural polymorphs. For proper immune recognition and immunization we sought to isolate and stabilize a single species. As such, we used UV cross-linking to stabilize unimolecular G4 DNA structures, which in turn were used for immunization. UV

cross-linking of thymines in G-quadruplex structures has been shown to stabilize a single

species in a mixture of G4 DNA structures (42). By contrast to the slow moving

tetramolecular structures, compact folded unimolecular G4 DNA structures can be

distinguished from duplex DNA and unfolded monomers by their increased migratory speeds

(42, 153). Using telomeric sequences derived from vertebrates and ciliates (*Oxytricha*,

Tetrahymena) we induced the formation of unimolecular G4 DNA structures (See

Experimental methodology and lane 3 of Figure 8).

All G4 DNA structures used in this study were initially identified by gel

electrophoresis and visualized by UV shadowing, Figure 8 is an example of such a PAGE

experiment stained with Gel Red for descriptive purposes. Under UV shadowing bands of

interest were excised and eluted following protocols listed in the experimental methodology.
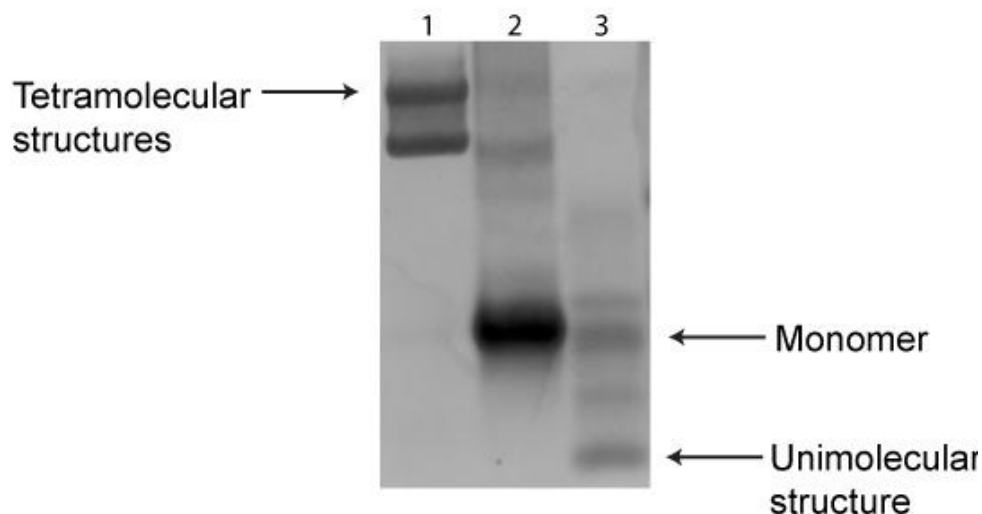
Figure 8. Native polyacrylamide gel electrophoresis can resolve different nucleic acid structures. Purified nucleic acids are separated by native PAGE and visualized with UV shadowing. The migration of different nucleic acids can be visualized and appear as distinct species through gels. Tetramolecular G4 DNA structures are slow migrating species and generally consist of only one or two species (Lane 1 Ver-3) in contrast, unimolecular G4 DNA structures are fast moving structures and may persists as multiple species in a gel (Lane 3 Ver-4). Lane 2 is an example of an unstructured monomeric oligonucleotide sample, Ver-3. Multiple bands appear often in gels and require other investigative techniques (like CD) to help characterize these differently migrating structures.

Sequence 5' to 3'	Molecularity	Cation	ε L/(mol*cm)	Name	Reference
d [TGGGGG(TTAGGG)$_2$T]	Tetramer	1 M NaCl	190100	Ver-3	(17)
d (TTTTGGGG)$_2$	Tetramer	1 M NaCl	147000	Oxy-2	(15)
d (CGCGCGCGTTTCGCGCGCG)	GC hairpin	none	161000	GC Hair	(148)
d (TTTTGGGG)$_4$	Basket	100 mM NaCl	293800	Oxy-4	(42)
d (TTGGGG)$_4$	Mixture	100 mM NaCl	229000	Tet-4	(42)
d [AGGG(TTAGGG)$_3$]	Basket	100 mM NaCl	228500	Ver-4	(106)
d (TTAGGGGGTTA)	Tetramer	1 M NaCl	112800	G5 G4	(42)
d (GGTTGGTGTGGTTGG)	Monomeric chair	100 mM KCl	143300	M.Chair	(154)
d [AGGG(TTAGGG)$_3$]	Monomeric basket	100 mM NaCl	228500	M.Bask	(106)
d [AGGG(TTAGGG)$_3$]	Propeller	150 mM KCl	228500	Prop	(103)
d (GGGGTTTTGGGG)	Dimeric basket	100 mM NaCl	115200	D.Bask	(155)
r (UGGGGU)	RNA tetramer	1 M NaCl	628000	RNA G4	(156)
d (TTTCTTTTTTCTTCTTTTCTTTCTTTTTCT)	Triplex	100 mM NaCl	453188	TC30W	(149)
d (AGAAAAAGAAAGAAAAGAAGAAAAAAGAAA)	Triplex	100 mM NaCl	359200	TC30C	(149)
d (TCTTTTTCTTTCTTTTCTTCTTTTTTCTTT)	Triplex	100 mM NaCl	238200	TC30	(149)

Table 3. Oligonucleotides used in this study to synthesize DNA or RNA G4 structures.
Sequences used for generating DNA or RNA secondary structures and their corresponding
extinction coefficients. The cation concentration used to generate specific structures is listed;
following purification structures were stored in PBS and maintained at -20 °C until use.
Assigned molecularity is derived from native PAGE, CD experiments and referenced
material.

2.3.2   Circular dichroism spectropolarimetry determines G4 strand polarity

To help distinguish between different nucleic acid structures, which might not readily

be distinguished by UV shadowing and/or native PAGE alone, we measured all purified

nucleic acids by circular dichroism spectropolarimetry (CD). CD has been extensively used

to characterize the structure of proteins and nucleic acids (19, 157). The majority of

publications have used as a reference two characteristic spectral forms, distinguishing

between parallel and antiparallel quadruplexes based solely on different wavelength maxima and minima. Classical antiparallel quadruplexes have a positive ellipticity maximum around 290 nm and a negative ellipticity minimum around 260 nm. Quadruplexes with these wavelengths maximum and minimum are proposed to be folded unimolecular and bimolecular quadruplexes. Classical parallel quadruplexes have a positive ellipticity maximum around 260 nm and a negative ellipticity minimum around 240 nm (12, 39, 153, 158, 159). CD was used in conjunction with native PAGE migration patterns to gain insight into the oligonucleotide folding patterns. For this purpose we compared the patterns that were observed with our structures to the known reference spectrum of specific, well defined structures in the literature.

Parallel vs. anti-parallel spectrum
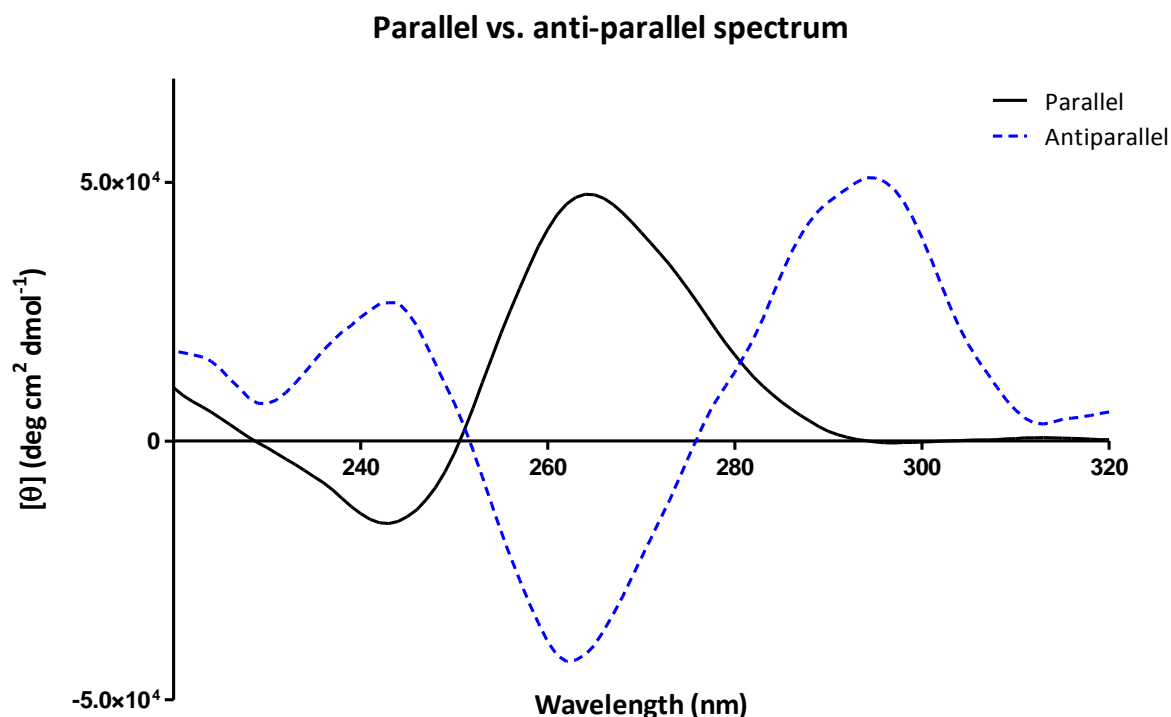
Figure 9. CD spectrum of parallel G4 DNA have different ellipticity maximum and minimum than antiparallel G4 DNA structures.
An example of the differences between characteristic parallel G4 DNA and antiparallel G4 DNA. Characteristic parallel CD spectrum contains a molar ellipticity maximum at ~260 nm and a minimum at ~240 (Purified tetramolecular Ver-3 is shown as classical parallel example). While, characteristic antiparallel CD spectrum contain a molar ellipticity maximum at ~290 nm and a minimum at ~260 nm (Purified D.Bask is shown as classical antiparallel example).

The generation of discrete, stable structure was very important for immunization and specificity testing. One of the challenges when working with G-quadruplex structures (especially unimolecular structures) is their dynamic flexibility. Often a single guanine-rich sequence can fold into many different polymorphic structures, depending on the conditions in which it is developed. Notably, d(TTAGGG)$_4$ and its variants can adopt different topologies in Na$^+$ vs K$^+$ (36, 160-163). As such, we sought to generate discrete quadruplex structures

for immunization so that any immune response that would subsequently be mounted would generate antibodies specific to one particular G-quadruplex structure. Initial attempts at generating antibodies focused on the immunization of animals with stable tetramolecular G4 DNA structures. These structures are easy to purify in large quantities and are among the best resolved of all quadruplex structures (144, 164-167). The polarities of all 4 strands in any tetramolecular G4 structure are always parallel. Therefore the CD spectrum of any tetramolecular G4 structure fits the classical characteristic ellipticity maximum and minimum of all parallel structures. In addition, tetramolecular G4 DNA structures are very stable once formed and easy to visualize by their slow migratory patterns in PAGE. All purified and assembled tetramolecular G4 structures from Oxy-2, Ver-3, G5 G4 and RNA G4 telomeric sequences displayed characteristic ellipticity maximums around 260 nm and ellipticity minimums around 240 nm (example Figure 10).

Parallel Tetraplex Structures
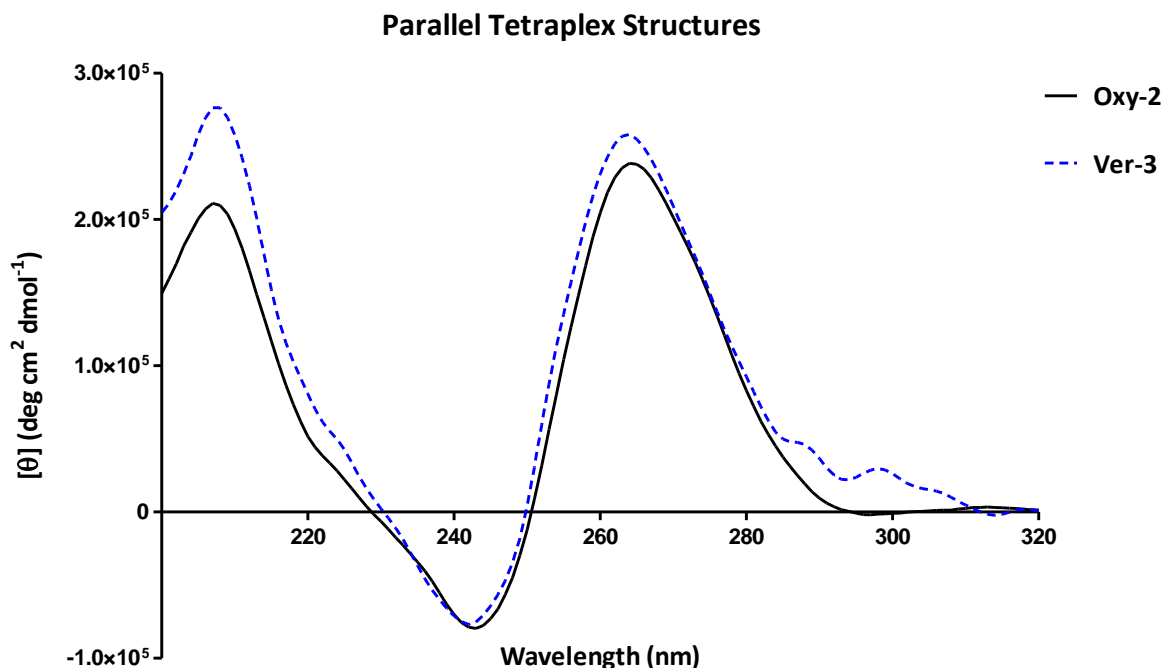
Figure 10. Circular dichroism of parallel stranded tetramolecular G4 DNA have classical ellipticity spectrum.

Circular dichroism spectroscopy of immunizing tetramolecular G4-DNA structures show characteristic ellipticity maximum around 260 nm and minimum around 240 nm. Telomeric oligonucleotide Oxy-2 and Ver-3 were incubated in appropriate conditions (See Experimental Methodology) for the formation of all parallel-stranded tetramolecular G-quadruplex DNA.

After the successful creation of parallel tetramolecular G4 DNA structures, we attempted to generate the more dynamic antiparallel unimolecular and bimolecular structures. Oligonucleotides containing at least four repeats of dG_n sequences tend to fold unimolecularly into single stranded G-quadruplexes while double dG_n repeats tend to assemble into bimolecular quadruplex structures (two hairpins coming together). Using the Oxy-4, Tet-4 and Vert-4 sequences we incubated oligonucleotides at low oligonucleotide concentration and low salt conditions, appropriate for antiparallel G4 DNA folding (See Experimental Methodology). Following visualization through PAGE all purified samples

were analyzed by CD to assign topology. Both Oxy-4 and Ver-4 purified sequences fold into

characteristic antiparallel G4 DNA structures as shown by the classical CD spectrum

ellipticity maximum around 290 nm and ellipticity minimum around 260 nm (Figure 11).
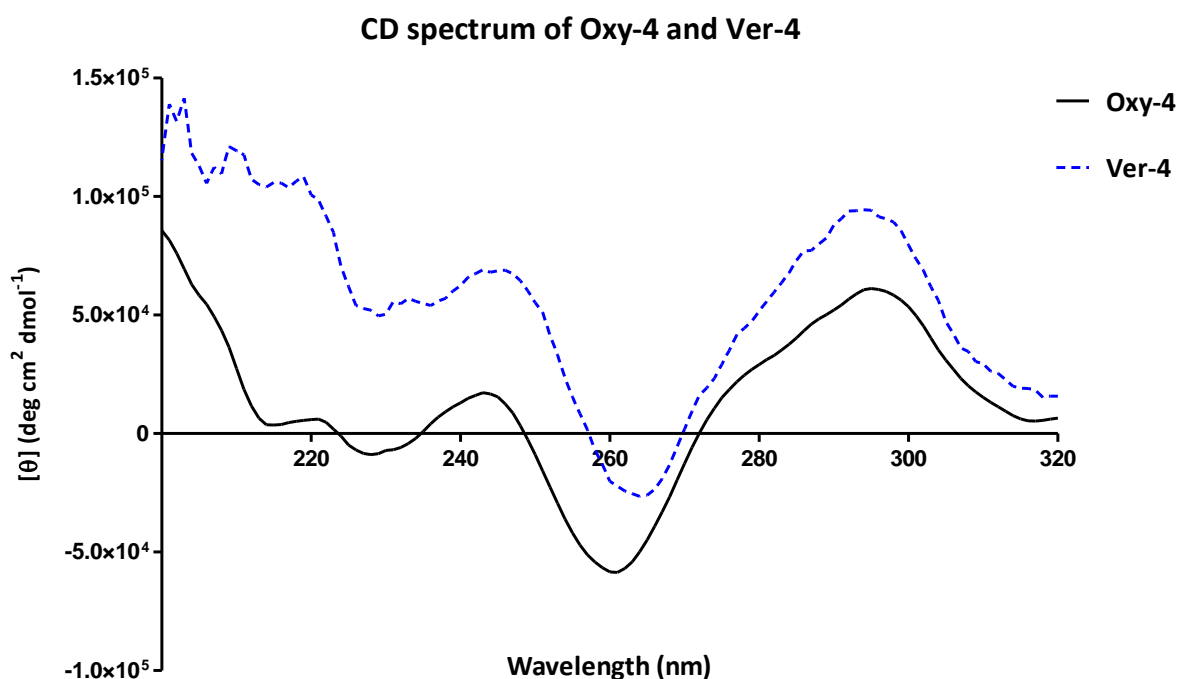
Figure 11. Oxy-4 and Ver-4 sequences have CD spectra of classical antiparallel G4 structures.
Circular dichroism spectrum for Oxy-4 and Ver-4 sequences show characteristic pattern of
unimolecular antiparallel G4 DNA structures. Both spectra display a major molar ellipticity
maximum around 290 nm and a major molar ellipticity minimum around 260 nm.

Interestingly, the incubation of a third sequence Tet-4, in conditions appropriate for

unimolecular folding did not generate a classical spectrum easily assignable to parallel or

antiparallel topology based solely on CD spectrum analysis (Figure 12). Many other groups

have used the Tet-4 sequence to generate quadruplex structures and it has been noted to form

a 3+1 mixed hybrid parallel and antiparallel structure under these conditions (164). The Tet-

4 CD spectrum possesses characteristics of both parallel and antiparallel spectrum and thus represents a good example of the difficulty of assigning G4 topology based solely on CD. Publication of different CD spectrum for the same sequence incubated under similar conditions outlines the limitations of assigning complex topology to G4 structures and may indicate that samples contain a mixture of species.

CD spectrum of Tet-4
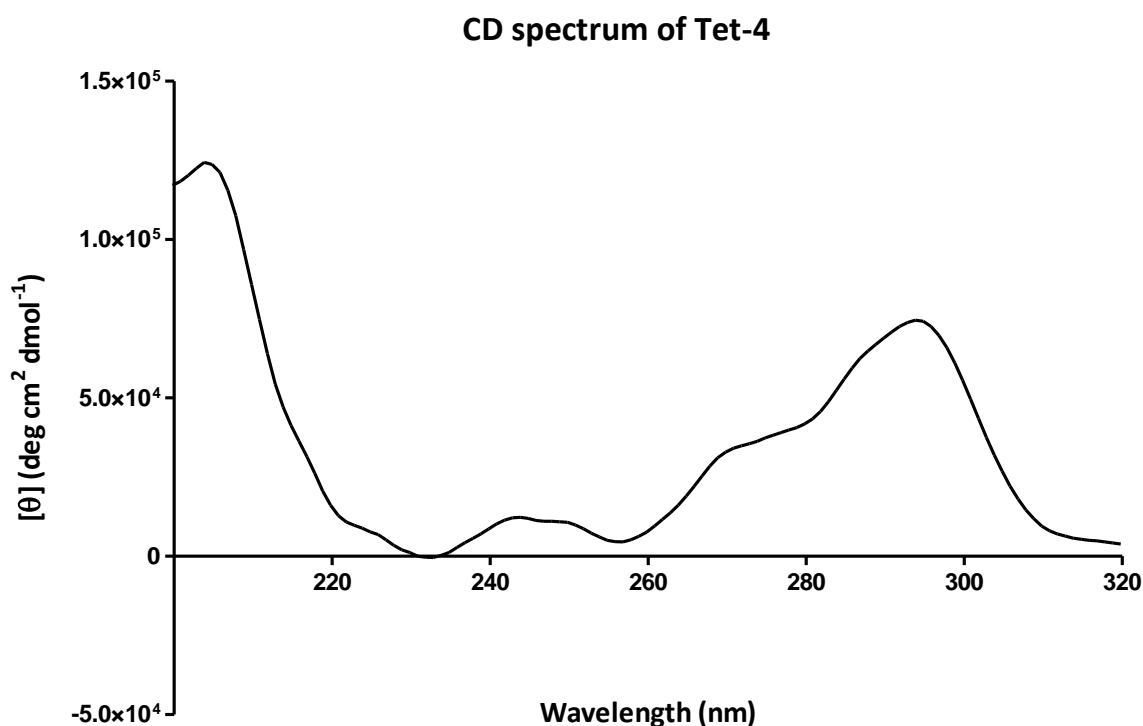
Figure 12. CD spectra from a four-repeat telomeric sequence from Tetrahymena (Tet-4) suggest antiparallel/parallel hybrid mixture is formed.

CD spectra from Tet-4 displays a spectral pattern with both parallel and antiparallel characteristics. A major ellipticity maximum is detected ~290 nm characteristic of antiparallel structures, while two minor troughs are detected at ~260 nm and ~230 nm. This spectrum indicates the presence of multiple topologies and illustrates the weakness of assigning G4 topology based solely on CD.

Initial experiments with unmodified oligonucleotides resulted in the appropriate incubation conditions for the formation of several different G4 structures. However, there was cause to fold G4 structures using modified oligonucleotides for downstream applications. Two types of modifications were required for future applications including biotinylated oligonucleotides and amino modified oligonucleotides. Biotinylated oligonucleotides would serve as a means of attaching G4 structures to streptavidin (SA) coated ELISA plates for screening, while terminally modified amino oligonucleotides would act as a means of attaching G4 structures to protein carriers for immunization of mice.

Given that terminal modification may alter the G4 topology formed in comparison to unmodified oligonucleotides, we sought to examine structures formed with terminally modified sequences. To test whether or not terminal modification of an oligonucleotide perturbed secondary structure formation, we examined the CD spectra of both terminally modified sequences and unmodified sequences under the same folding conditions. Figure 13 illustrates that generating G4 DNA structures using terminally modified sequences had no effect on the formation of G4 DNA with these sequences. In comparison to unmodified oligonucleotides, terminally modified sequences with biotin or amino groups allowed for the formation of similar G4 structures (Figure 13). All sequences used in this study that had been terminally modified were compared in this way to unmodified sequences. We did not detect any differences in CD spectra between those sequences that were terminally modified and those that were unmodified.

Effect of terminal modifications
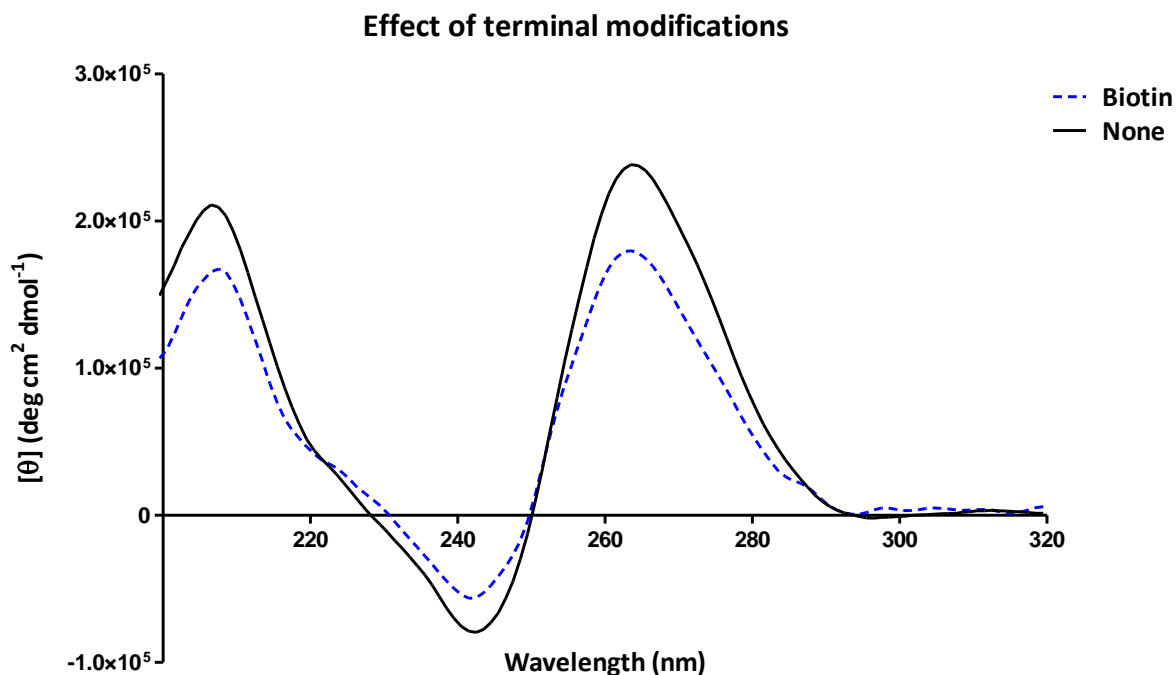
Figure 13. Terminal modifications of oligonucleotides do not affect G4 formation.
CD of purified tetramolecular G4 DNA shows that the characteristic ellipticity maxima and minima of an all parallel stranded tetramolecular structure remain after formation using terminally modified oligonucleotides. All terminally modified oligonucleotides were monitored by CD and compared to the unmodified oligonucleotides to confirm G4 DNA formation.
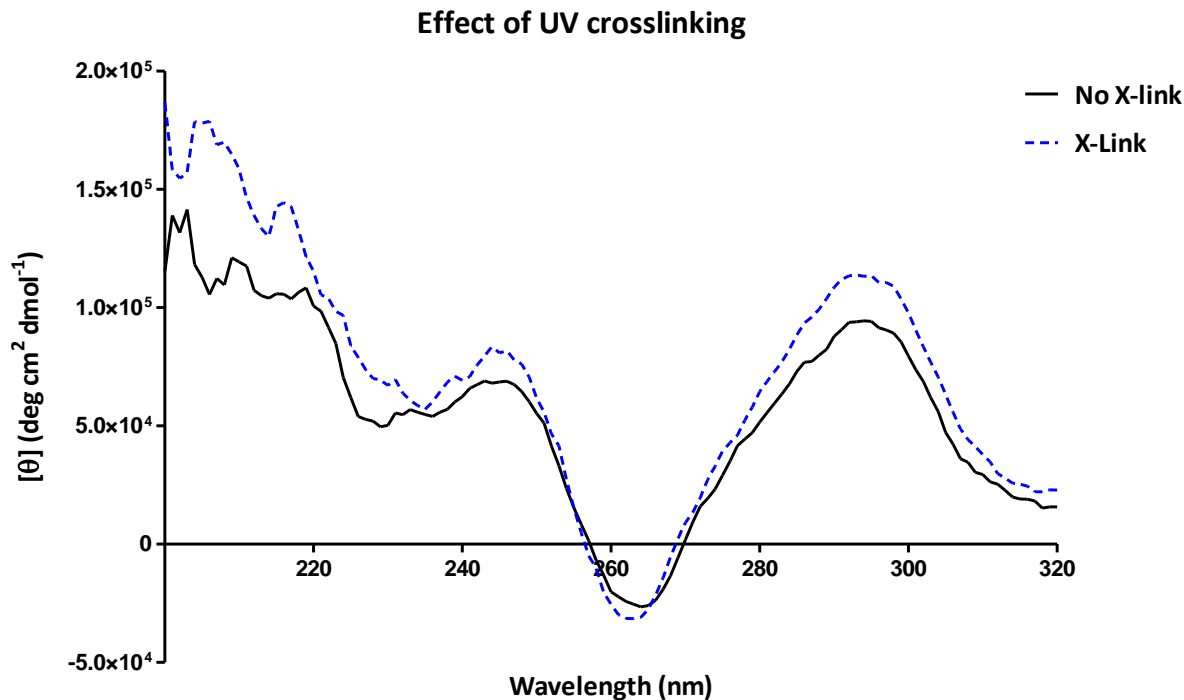
Figure 14. The effect of UV cross-linking antiparallel G4 DNA structures.
CD spectrum of antiparallel G4 DNA structures with and without UV cross-linking. UV cross-linking did not change the circular dichroism spectrum obtained from Oxy-4, Tet-4 or Ver-4 and suggests that the G4 DNA structure was maintained after cross-linking. UV cross-linking was used to stabilize one conformational structure of flexible unimolecular G4 DNAs used for immunizations.

In a similar experiment we compared the CD spectra obtained from oligonucleotides that had been folded and stabilized by UV cross-linking to those folded oligonucleotides that had not been UV cross-linked. Figure 14 illustrates that stabilizing some unimolecular G4 DNA sequences (Oxy-4, Ver-4 and Tet-4) with UV cross-linking did not affect the expected folding topology of the structure. Only unimolecular structures used for immunization (Oxy-4, Ver-4 and Tet-4) were UV cross-linked prior to immunization.

Although immunization and subsequent serum screening required the production of a small group of quadruplex structures, testing the specificity of subsequent antibodies required a larger panel of G4 polymorphs to be generated. As G-quadruplex structural research has evolved, the folding topologies of many guanine-rich telomeric and non-telomeric sequences has been well characterized (15, 19, 50, 53, 103, 104, 157, 163, 164, 169, 170). Following guidance from the literature we incubated numerous telomeric and non-telomeric sequences in condition appropriate for G-quadruplex folding (See Experimental Methodology). Different sequences and different incubation conditions produced a small library of structural polymorphs of G-quadruplex DNA and RNA structures (Table 3 and Figure 15). Using sequences shown in Table 3, we induced the formation of different polymorphs of G4 DNA and RNA structures for specificity testing. Figure 15 illustrates the variety of structures that were produced. The folding topology assigned to each structure was based on the referenced literature (Table 3), their experimental migratory speeds in PAGE and CD spectrum. Figure 15 depicts some of the polymorphic topologies of G4 structures not already presented.
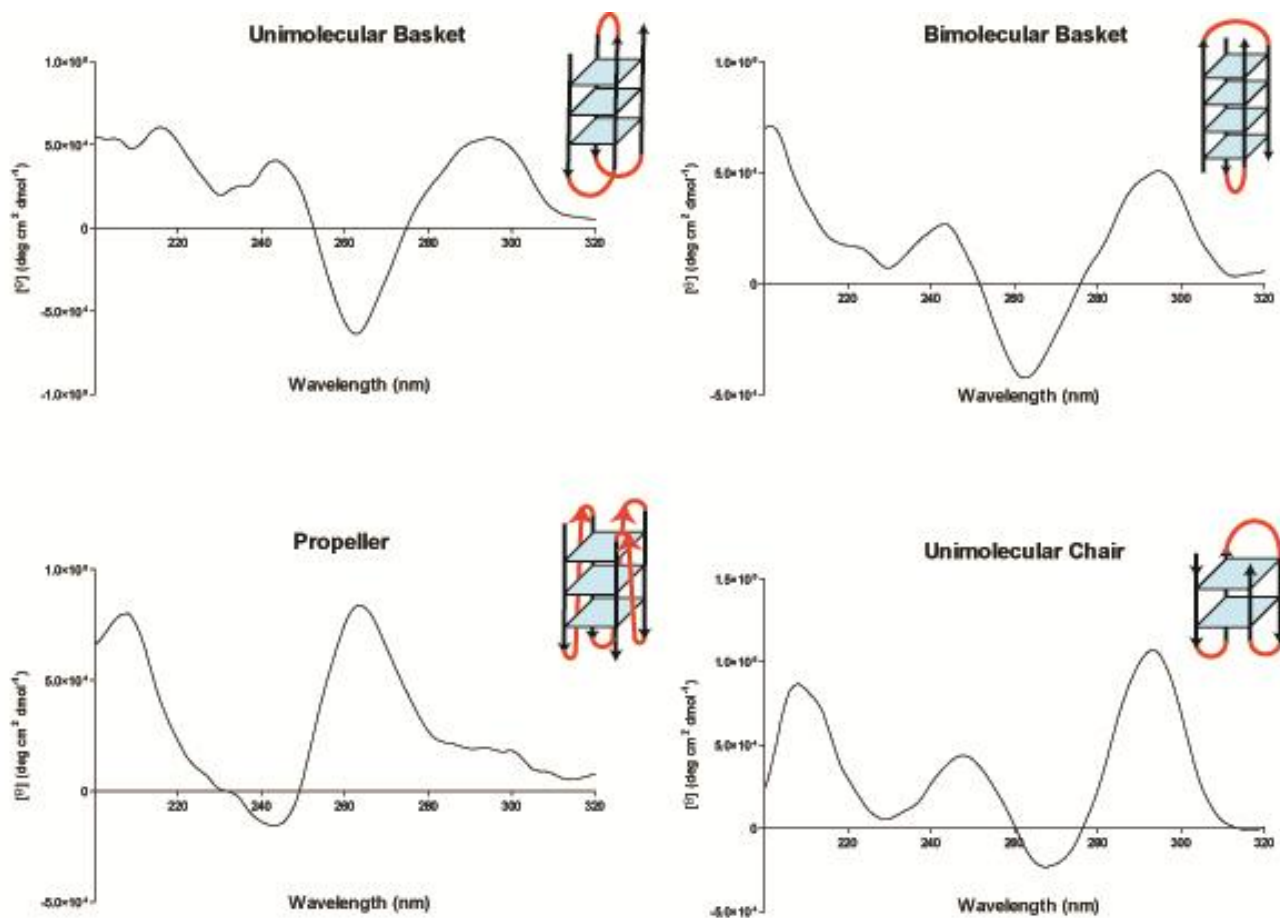
Figure 15. CD spectra from folded unimolecular and bimolecular G4 DNAs used in this study.
CD spectra of a variety of guanine-rich sequences from Table 3 prepared under conditions for G4 DNA folding. Cartoons in the upper right hand corner of each spectrum illustrate the assigned folding topologies based on the literature and experimental migratory speeds and ellipticity spectra. Each CD spectrum has been labeled with the common molecularity associated with that spectrum detailed in Table 3.

Taken together, native gel mobility assays and CD spectroscopy confirm that the assembled oligonucleotides formed desired G4 DNA structures.

2.3.3 ELISA analysis of serum verifies the generation of anti-DNA antibodies

Confirmation of structural topologies allowed for the immunization of animals with distinct G4 DNA structures. Animals were divided into small groups and immunized with repeated rounds of one structural variant of G4 DNA. All G4 DNAs used in immunizations were prepared from amino modified sequences. Terminal amino modification of G4 DNA structures allowed for the conjugation of G4 DNA molecules to key hole limpet hemocyanin (KLH) carrier proteins. Presenting G4 DNA structures on the surface of KLH was used to produce a more robust immune response in an attempt at increasing the immunogenicity of the nucleic acid structures. Following immunization, blood from immunized mice was analyzed for antibody titers to specific G4 DNA structures. A simple ELISA method was used for screening large numbers of mice for the generation of antibodies to G4 DNA. The immunizing sequence terminally modified with biotin was used to coat the wells of a streptavidin coated ELISA plate. Primary antibodies in the sera of mice that reacted with the bound G4 DNA molecules in a well were detected using goat-anti-mouse antibodies conjugated to HRP. Positive wells were scored with green color changes using ABTS and an optical density plate reader. Following repeated rounds of immunization most mice were producing antibodies to G4 DNA detected in this system.

Serum Titration
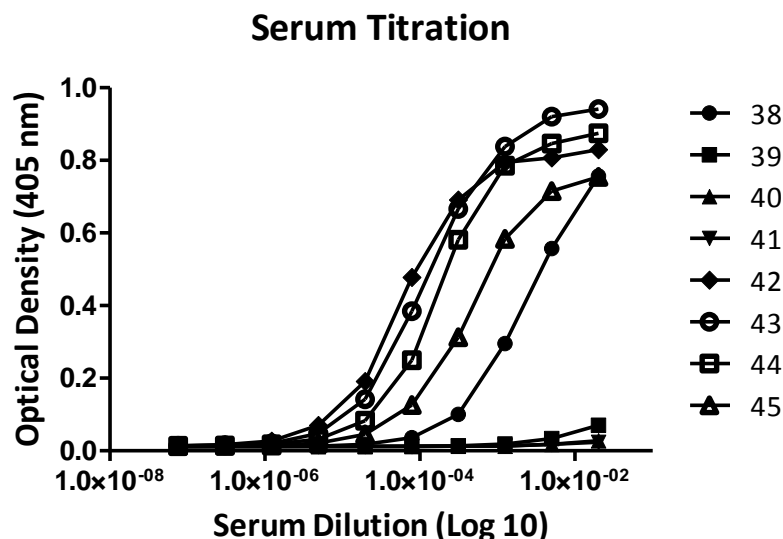
Figure 16. Mice immunized with G4 DNA produce antibodies to G4 DNA.
ELISA screening of mice immunized with tetramolecular G4 DNA display classical immune response by generating antibodies reactive to tetramolecular G4 DNA *in vitro*. Individual mouse numbers are labeled in the legend, not all mice respond equally to immune challenge with tetramolecular G4 DNA Oxy-2.

Figure 16 is an example of a titration of serum titers in mice after 6 rounds of immunization. Animals with the highest titers by G4 ELISA were euthanized and spleens removed. Following standard hybridoma techniques, spleen cells were fused to myeloma cells to produce hybridoma cells lines producing anti-G4 DNA antibodies. Hybridoma supernatants were screened for the presence of anti-G4 antibodies by G4 ELISA using the same protocol as described for serum screening. Hybridoma cells, selected by ELISA, were cloned in semi-solid medium. Individual hybridoma clones were picked from semi-solid medium and re-plated into individual wells of liquid culture for re-screening. This process was repeated for a minimum of three rounds to obtain monoclonal populations. Monoclonal populations of hybridoma cells were expanded in 1 L spinner flasks. Monoclonal antibodies

were purified using a fast protein liquid chromatography (FPLC) system with Protein G capture columns. Purified antibodies were evaluated and characterized in various ways. A total of 11 different monoclonal antibodies were selected for further study which can be subdivided into three groups: antibodies to Oxy-2, Ver-3 and Tet-4 (Table 4).

2.3.4   Variable isotype and affinity of distinct monoclonal antibodies to G4 DNA

Due to the inherent variability in G4 structures it is very unlikely that only one particular G4 topology is present *in vivo*. Without knowing which G4 DNA topology to focus on, we set out to generate structure specific antibodies to distinct G4 DNA structures. Our goal was to generate monoclonal antibodies that could distinguish between single-stranded DNA, double-stranded DNA, parallel and antiparallel G4 DNA structures. Upon purification of monoclonal antibodies, we performed isotype characterization and affinity determination of all 11 antibodies (see Experimental Methodology). Isotyping analysis shows that purified antibodies display a characteristic antibody isotype pattern consistent with multiple immunizations and affinity maturation of an intact immune response (Table 4). To characterize the apparent affinities of the purified antibodies a titration was performed for all antibodies to their respective immunogen. Biotinylated G4 DNA structures were captured on streptavidin coated ELISA plates. Primary antibodies were serially diluted and corresponding binding curves generated. There were three groups of antibodies tested in this system, purified antibodies to parallel tetramolecular Oxy-2 or Vert-3 G4 DNA and purified antibodies to mixed parallel antiparallel unimolecular Tet-4 G4 DNA. Figure 16 outlines the binding curves of individual groups to their corresponding G4 DNA immunogen. Most

antibodies were determined to have affinities in the low nanomolar range, suggesting a strong interaction with their binding partner in the ELISA system. Table 4 outlines the specific antibody names, calculated affinities and the immunizing sequence tested for apparent affinity.

Name	Isotype	Kd (nM) *	Immunizing Sequence
8H2	IgG 2b	7.22+/-0.80	d[TGGGGG(TTAGGG)$_2$T]
5E11	IgG 1	13.1+/-0.92	d[TGGGGG(TTAGGG)$_2$T]
9A1	IgG 1	77.8+/-7.1	d[TGGGGG(TTAGGG)$_2$T]
4E11	IgG 2a	3.64+/-0.28	d(TTTTGGGG)$_2$
4C9	IgG 1	undetermined	d(TTTTGGGG)$_2$
5C10	IgG 1	24.4+/-1.2	d(TTTTGGGG)$_2$
1H6	IgG 2b	0.33+/-0.03	d(TTTTGGGG)$_2$
5G5	IgG 1	0.18+/-0.02	d(TTGGGG)$_4$
4G2	IgG 1	0.13+/-0.01	d(TTGGGG)$_4$
7F8	IgG 1	0.60+/-0.03	d(TTGGGG)$_4$
8H4	IgG 1	0.23+/-0.01	d(TTGGGG)$_4$

Kd Standard deviation based on triplicate measurements by ELISA

Table 4. The majority of purified monoclonal antibodies are IgG$_1$ isotype and have low nanomolar apparent affinities.

The majority of purified monoclonal antibodies that bind G4 DNA *in vitro* are IgG1 and have low nanomolar apparent affinities by ELISA. Immunizing sequences used to generate the antibody is indicated under immunizing sequence. Three different sequences were used Ver-3, Oxy-2 and Tet-4.
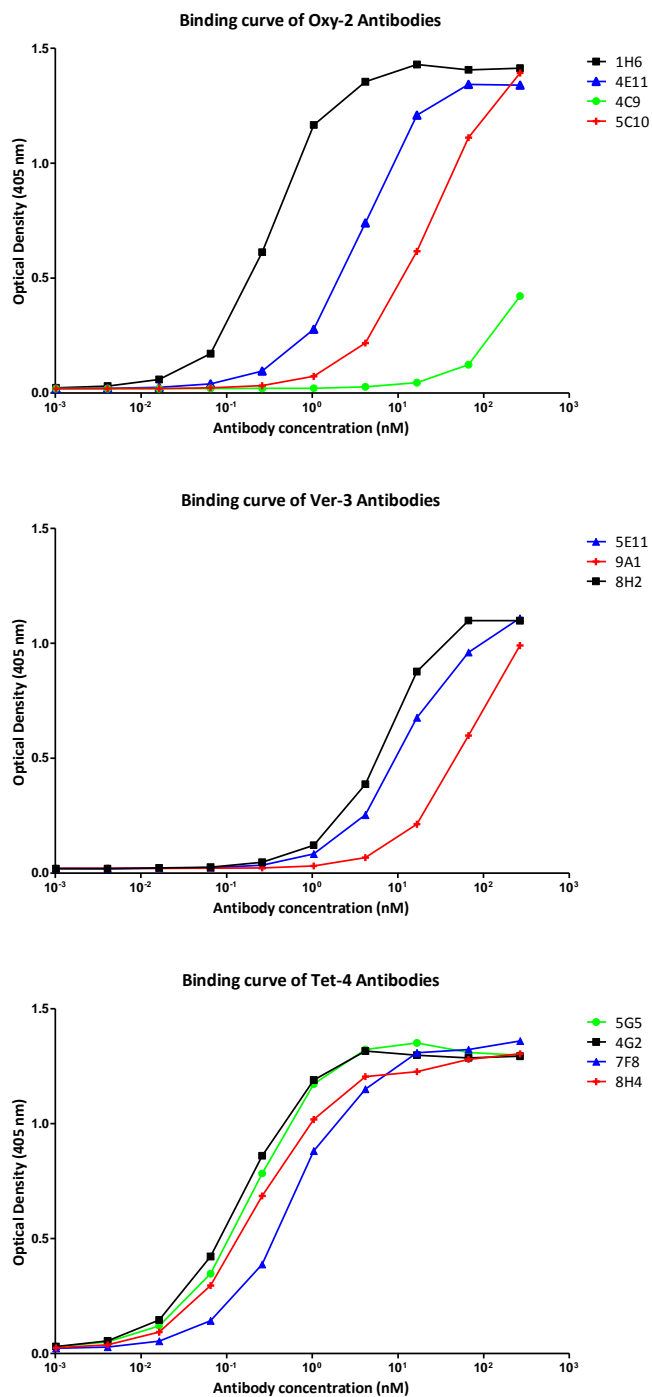
Figure 17. Binding curve analysis of antibodies to different G4 DNA immunogens.
Biotinylated G4 DNA structures were coated on ELISA plates and used to screen purified monoclonal antibodies. Serially diluted antibodies give rise to binding curves for the determination of apparent affinity. Legends indicate the antibody name and corresponding color coded binding curve.

2.3.5   Purified monoclonal antibodies have varying specificities to G4 DNA structures

To examine the specificity of purified antibodies we performed multiple direct ELISA assays and examined the direct binding of purified antibodies to simple structures bound to the plate or dried to the plate (See Experimental Methodology).

Although variations among antibody specificities were observed, antibodies raised against the tetramolecular Vet-3 G4 DNA structure designated 5E11, 8H2 and 9A1 demonstrate high specificity for Vet-3 G4 DNA. Interestingly, 5E11 displayed significant binding to the G5 G4 tetramolecular competitor (data not shown) and also bound the other tetramolecular structure generated from a different sequence Oxy-2. None of the Vet-3 specific antibodies 5E11, 8H2 or 9A1 significantly bound double stranded or single stranded DNA in this assay (Figure 18).

Antibody 5C10 (α Oxy-2)

Antibody 4E11 (α Oxy-2)

Antibody 1H6 (α Oxy-2)

Antibody 4C9 (α Oxy-2)

Antibody 5E11 (α Ver-3)

Antibody 9A1 (α Ver-3)

Antibody 8H2 (α Ver-3)
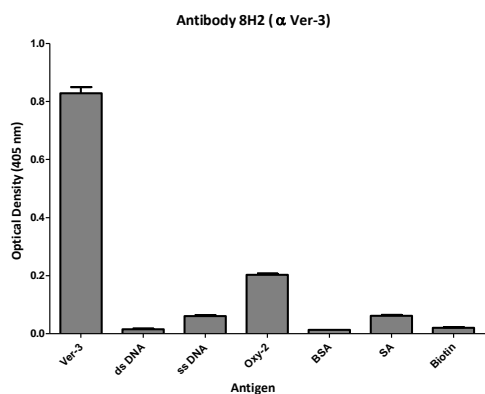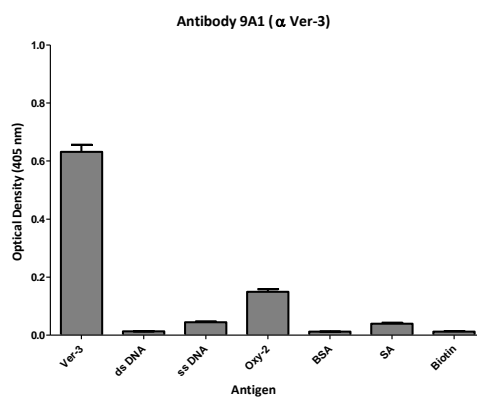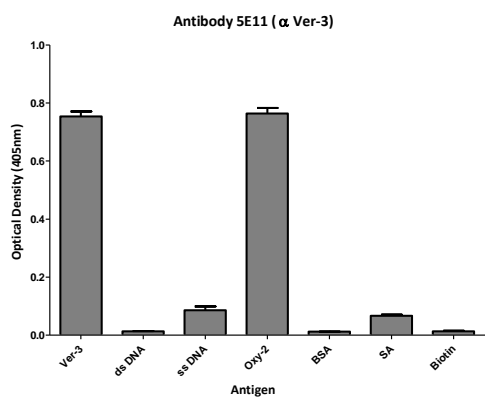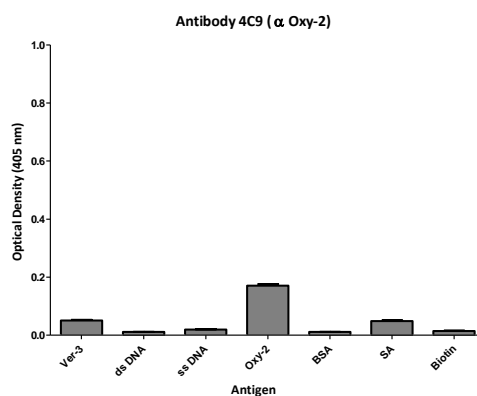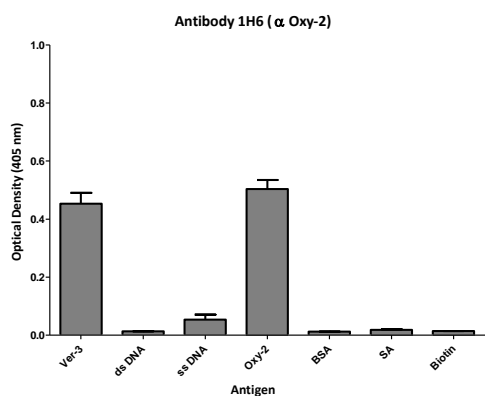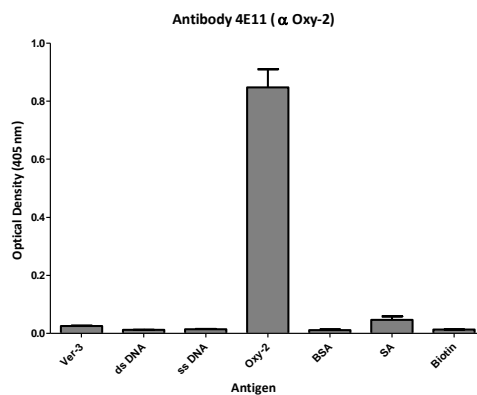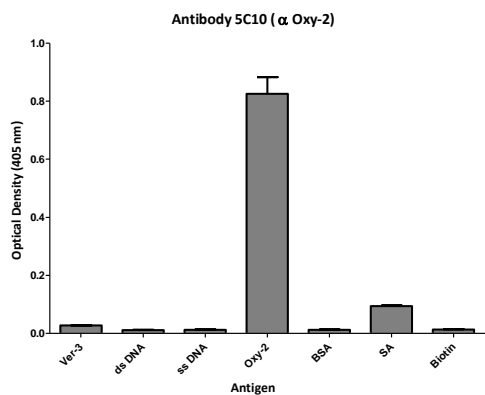
Figure 18. Direct binding ELISA for specificity of antibodies to tetramolecular G4 DNA.
Initially, antibodies raised against tetramolecular G4 DNA were subject to specificity testing by coating competitors to ELISA wells. Positive binding is detected in wells with increased optical density at 405 nm. No purified antibody significantly bound single-stranded or double-stranded DNA. All antibodies were tested for direct binding to G4 DNA (Oxy-2 or Ver-3), dsDNA (Oxy-2 or Ver-3 with complementary strand), ssDNA (denatured Oxy-2 or Ver-3) and proteins (BSA, SA and biotin).

A similar ELISA was performed to examine the specificity of antibodies raised against tetramolecular Oxy-2 G4 DNA. Antibodies designated 4E11 and 5C10 demonstrate high selectivity for the tetramolecular Oxy-2 G4 DNA structure (Figure 18). Somewhat surprisingly a third antibody raised against tetramolecular Oxy-2 G4 DNA designated 1H6 bound both tetramolecular structures generated from Oxy-2 or Ver-3 sequences (Figure 18). The final antibody tested by direct ELISA designated 4C9 had low specificity to all competitors and may represent a non-specific low affinity antibody. None of the antibodies raised against tetramolecular Oxy-2 G4 DNA significantly bound to double-stranded or single-stranded DNA, supporting the fact that these purified antibodies can distinguish between DNA secondary structure.

To further characterize the antibodies that bound multiple tetramolecular topologies (5E11 and 1H6) we performed competition ELISA assays with an increased pool of G4 DNA topologies. Competition ELISA permits equilibrium to be reached between antibodies that bind to competing structures and those that don't in a separate chamber prior to antibody capture by ELISA. In this way, reactions happen in solution and more appropriately mimic cellular conditions.

Although the 5E11 antibody bound all tetramolecular structures it did not bind any unimolecular or bimolecular samples tested (no change from direct ELISA). Interestingly,

the 1H6 antibody that had been previously shown to bind to both tetramolecular structures from Oxy-2 and Ver-3 also bound to the unimolecular structures from Oxy-4 and Tet-4 (Figure 19). However, the 1H6 antibody did not seem to bind the unimolecular Vet-4 structure. It appears that the 1H6 antibody is somewhat promiscuous in its binding and does not appear to be completely sequence specific or completely structure specific. Expanding our pool of competitors to include an RNA tetramolecular structure and a triplex DNA structure we sought to determine if the 1H6 antibody could bind RNA or if the binding interaction was purely charge density based (triplexes have increased charge density). In our assay, the 1H6 antibody did not significantly bind a tetramolecular RNA structure or a triplex structure (Figure 19).
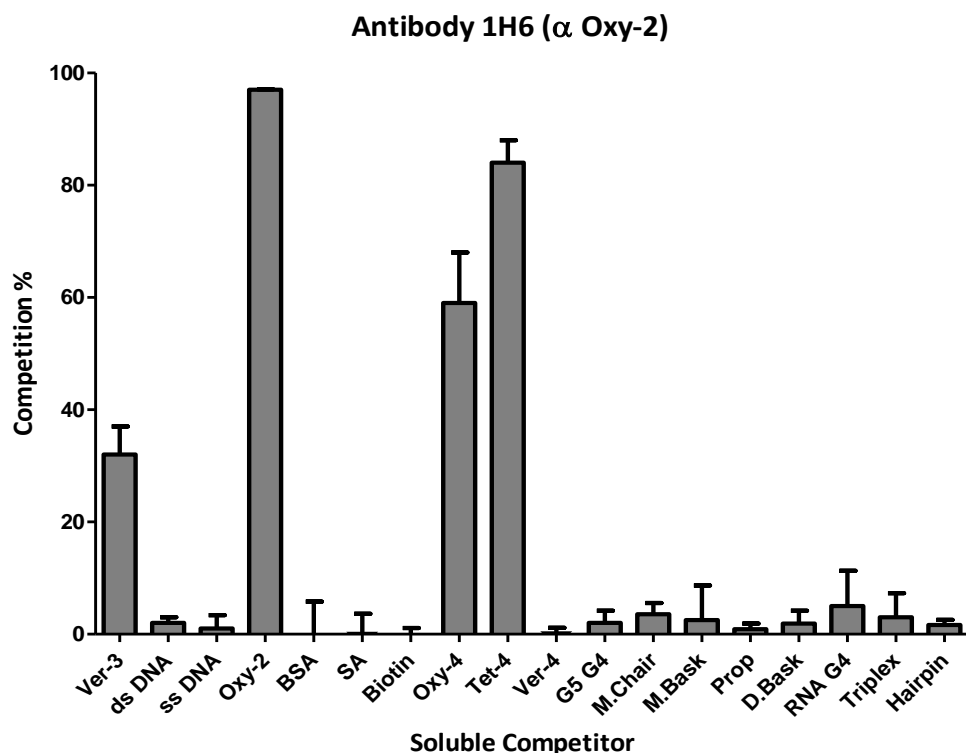
Figure 19. Expanded specificity testing of monoclonal antibody 1H6 displays promiscuous binding.
Competition ELISA for examining the specificity of antibody 1H6 to various soluble competitors. See experimental methodology for setup, positive competition suggest binding to elements in solution. 1H6 binds to tetramolecular structures and unimolecular structures without sequence specificity. Error bars represent the standard deviation of triplicate experiments and abbreviated soluble competitor structures are outline in Table 3 (Page 52). Soluble competitors that compete for binding the 1H6 antibody include tetramolecular G4 DNA from Ver-3 and Oxy-2 and unimolecular G4 DNA from Oxy-4 and Tet-4.

2.3.6 Insight into the mechanism of binding between antibodies and DNA structures by variable heavy and light-chain sequence analysis

The most important part of any antibody is the region that makes contact with the antigen. The antigen binding region of an antibody is derived from the variable heavy (V_H)

and variable light (V_L) chain genes of a cell. There are three antigen contact regions referred to as complementarity determining regions (CDRs) in both heavy and light chains. To elucidate the molecular binding characteristics between the purified monoclonal antibodies and G4 DNA structures we sequenced the V_H and V_L chain genes for purified hybridomas of interest. Sequence data is available online from GenBank, accession numbers FJ652787-FJ652794. Alignment of the primary amino acid sequences was performed using IMGT/V-Quest program (http://imgt.cines.fr/, see Experimental Methodology. There are thought to be distinguishing amino acids molecules in complementarity determining regions (CDR) of antibodies that influence the affinity and specificity of antibodies (171).

Heavy Chain

	CDRH-1				CDRH-2				CDRH-3													
	27	30	35	38	56	59	62	65	105													
8H2	G Y T F	T S Y G			I	Y P R	S G N	T	A	R V R G G Y Y G S S S H W Y												
4E11	G Y T F	T D Y Y			I	F P G	S G S	T	A	R G G E L W S S Y Y A M D Y												
5E11	G F A F	S S F G			I	T S G	G T Y	T	A	R H W A Y Y S N Y L F A Y												
5C10	G Y T F	T S Y W			I	D P N	S G G	T	A	R S P E I Y Y P A W F A Y												
1H6	G F T F	R N Y W			I	R L K S	D N Y A	T	T	N W Y Y F D Y												
Brown	G F T F	S S F G			I	S S G	S S T	L	T	S V T V S S A K T T P												

Light Chain

	CDRL-1				CDRL-2		CDRL-3									
	27	32	34	38	56	65	105									
8H2	E S V D N	Y G I S	F		A A	S	Q Q S K E V P A R T									
4E11	S S V S	S	S H		S T	S	Q Q Y S G Y P L N T									
5E11	Q S L V H S	N G N T	Y		K V	S	S Q S T H V P P T									
5C10	K R I	S K	Y		S G	S	Q Q H N E F P L T									
1H6	Q S L L Y S	N G K T	Y		L V	S	V Q G T H F P L T									
Brown	K S V S T	S G Y S	Y		L V	S	Q H I R E L Y T									

Table 5. Variable heavy and light chain amino acid composition of purified monoclonal antibodies. Amino acid composition of CDRs for purified monoclonal antibodies illustrates variability of amino acids. CDRs contain the contact points between antibodies and target ligands. Light grey shading highlights conserved amino acids in at minimum four antibodies. Red highlighted amino acids (basic residues) have been previously shown to directly interact with nucleic acid molecules.

Anti-G4 DNA antibodies would benefit from an amino acid composition in their CDR's capable of binding nucleic acid bases or the phosphate backbone. Histidine, arginine and lysine (H, R, K) are all capable of forming charge-charge interactions with the phosphate back bone of nucleic acids. Asparagine and tyrosine (N, Y) can engage in hydrogen bonding with thymidine bases, while serine has also been shown to bind DNA. All five hybridomas sequenced contain basic amino acids (H, R, K) throughout the V_H and V_L chains. Interestingly, hybridoma 8H2 contains 3 basic residues in its CDRH-3 region, in comparison to two in 5E11 and only a single basic residue in both 4E11 and 5C10 (Table 5). This trend does not correlate to the affinity data (Table 4).

All of these potential DNA binding amino acids have been labelled in bold red letters in Table 3. For comparative analysis, table 3 also shows an antibody labelled "Brown" which was the first anti-quadruplex antibody isolated from non-immunized autoimmune "viable motheaten" mice (172).

In addition to distinct amino acids capable of binding DNA, conservation of amino acids across CDRs may suggest important G4 DNA-amino acid scaffold interactions. Sequenced V_H and V_L chains have several conserved amino acid residues throughout their CDRs (Table 5 grey shading). The sequence data presented here may help guide future experiments aimed at generating synthetic antibodies, similarly capable of bind G4 DNA structures.

2.3.7 1H6 antibody does not induce structural formation of G-quadruplex DNA

Although antibodies are considered well validated tools for use in targeting a multitude of molecules, they have also been unintentionally used to propagate accurate evidence for the existence of a target. Unfortunately, due in part to the nature of antibodies and the dynamic flexibility of most targets, antibodies can also induce and stabilize structure/target that may or may not naturally exist without the presence of the antibody. Historically speaking there is no better example than the case of antibodies to Z-DNA (See Chapter 1). During the period of preliminary research into the possible existence of Z-DNA, antibodies to Z-DNA were developed to target such structures in cells (173). The binding of antibodies in cells was inappropriately used as substantiation for the biological existence of Z-DNA (174). Regrettably, later on it was discovered that anti-Z-DNA antibodies were in actuality inducing a structural transition from B-DNA to Z-DNA (175) that may not happen in biology without the use of antibodies. The case of Z-DNA in addition to the mass production of substandard antibodies on the open market has inadvertently increased the level of scrutiny for novel antibodies (especially to nucleic acids).

The antibody designated 1H6 was interesting to us as it had been shown to interact with multiple G4 DNA topologies (Figure 19) including some unimolecular G4 DNA structures. This promiscuous binding suggested to us that the 1H6 antibody may be inducing a G4 structure formation upon binding. In this regard we sought to investigate the binding of the 1H6 antibody to single-stranded and G4 DNA folded structures.

Using CD experiments we tested if the addition of antibodies to single-stranded sequences would induce conformation changes in topology. Following similar experiments

that have previously been published regarding the induction effects of antibodies (176)

Figure 20 shows that adding the 1H6 antibody to the guanine-rich single-stranded oligonucleotide Oxy-2 does not induce a topological transition to G4 DNA. Figure 19 also illustrates the point that adding the 1H6 antibody to a previously formed tetramolecular G4 DNA structure does not alter the classical spectrum associated with parallel quadruplexes. This experiment suggests that the binding of the 1H6 antibody and parallel tetramolecular G4 DNA does not induce a structural transition nor can the 1H6 antibody induce a tetramolecular structure to form from guanine-rich oligonucleotides.

ssDNA and 1H6 Ab

Oxy-2 G4 and 1H6 Ab
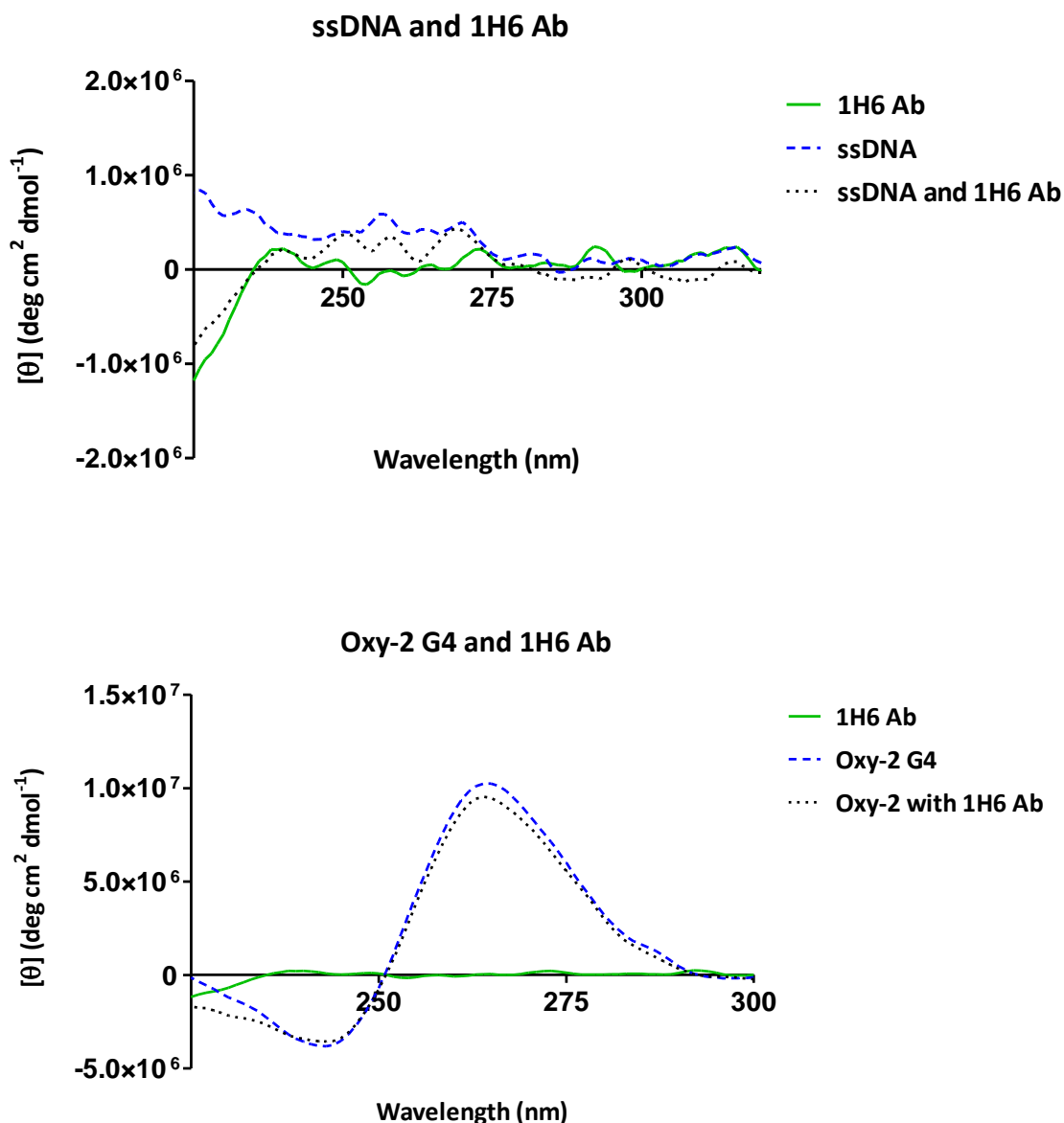
Figure 20. Anti-G-quadruplex antibody 1H6 does not induce the formation of quadruplex structures. CD analysis of tetramolecular G4 DNA in the presence and absence of the 1H6 antibody shows that addition of the antibody has no effect on the classical parallel CD spectrum obtained. In addition, single-stranded guanine-rich oligonucleotides capable of forming tetramolecular G4 DNA are not induced to form G4 structures in the presences of the 1H6 antibody. CD analysis was performed at room temperature over the course of 15 minutes. Nucleic acid structures or sequences (PBS) were incubated at 10-fold molar excess over 1H6.

2.4   Discussion

The generation of monoclonal antibodies specific for various G4 conformational isomers could help elucidate their existence in the genome. Many different G4 DNA isomers have been chemically derived, but their existence *in vivo* and biological relevance have yet to be demonstrated. This study has set out to develop biologically relevant G-quadruplex specific probes. Our initial immunization strategy was focused on the immunization of mice with tetramolecular and unimolecular G4 DNA structures from telomeric oligonucleotides derived from Oxy-2, Ver-3 and Tet-4 (Table 4).   We chose to focus on stable tetramolecular structures for initial immunizations for their classical gel mobility pattern and well defined CD spectrum. Although we included the development of unimolecular and bimolecular structures in our specificity assays, antibodies derived in animals that were immunized with unimolecular structures are still under investigation. Telomeric oligonucleotides are very polymorphic in that they can form many structural conformations. Understanding that G4 structures are frequently reported as single species and later on reported as a mixture of species (42, 177-179) we have only included unimolecular and bimolecular structures that we could confidently synthesize and identify.

Upon purification of these antibodies, specificity ELISAs described two antibodies specific for tetramolecular G4 DNA derived from Vet-3 and two antibodies specific for tetramolecular G4 DNA derived from Oxy-2 (Figure 18). Antibodies specific to Tet-4 are currently under investigation and we hope to explore the development of other antiparallel specific antibodies. Two other antibodies designated 5E11 and 1H6 were derived from mice immunized with only one sequence specific tetramolecular G4 DNA yet bound both tetramolecular immunogens from Oxy-2 and Vet-3 sequences (Figure 18). Upon further

testing the 1H6 antibody seemed to bind multiple topologies of G4 DNA including

tetramolecular and unimolecular structures (Figure 19). Importantly, these antibodies have

been shown to discriminate between single stranded, double stranded and quadruplex DNA

(Figure 18 and 19). In addition to our own competition experiments, the direct binding of the

1H6 antibody to different DNA structures including single-stranded, double-stranded and

quadruplex DNA was studied using gel-shift assays in collaboration with Dr. Brosh

(Appendix A). The gel-shift results are in accordance with our competition ELISA results

and suggest that the 1H6 antibody can discriminated between single or double-stranded DNA

and folded G4 DNA structures.

By producing highly specific probes we hope to visualize G-quadruplex DNA in a

vertebrate system. We have shown that our antibodies are of nanomolar affinity (Table 4)

which may prove important for visualization of an antigen with depleted abundance or for the

future development of single-chain antibody fragments that offer the possibility of live-cell

imaging from these selected immunoglobulins.

The CDRs of antibodies are important contact portions between antibodies and target

ligands. In the future the amino acid composition of the CDRs could be exploited for the

development of antibodies with increased affinity or structural specificity to G4 structures.

Determining amino-acids important in nucleic acid recognition and specifically G4 structure

recognition would be an asset for such studies. Other than the highlighted amino acid

residues in Table 5 that have been shown to bind nucleic acids, Isalen *et al*., have shown that

Zn^{2+} finger proteins bind G4 structures with high specificity and contain high amounts of

glutamic acid (E) and aspartic acid (D) (180). Notably several E and D residues are present

in the purified antibodies specific to G4 DNA (Table 5) and may increase the ability to hydrogen bond to the N_2 amino group of guanosine in the G-quartets.

It has been previously reported that arginine residues in the amino terminus of the CDRH-3 region may play a role in DNA binding (181, 182). Four hybridomas (designated 8H2, 4E11, 5E11 and 5C10) include at least one position-specific arginine residue in the CDRH-3 region (Table 5). However, neither the originally purified G4 DNA binding antibody ("Brown") nor the antibody designated1H6 contain any arginine residues in the CDRH-3 region. The characterized amino acid conservation and the dispersion of basic residues throughout the CDRs are indicative of DNA binding antibodies. We hope to exploit this sequence analysis for future generations of reagents that may be able to discriminate between, not only, the G-tetrad cores of quadruplexes but also the topology, loop sequences and strand orientation of these interesting molecular signals. Nonetheless, these first line monoclonal antibodies will provide invaluable tools for the biological understanding of G4 DNA.

By examining the amino acid composition of these purified antibodies, an interesting concept develops pertaining to the method of molecular recognition between proteins and different G4 structures. There have been four ways described in which a small molecule may recognize a G4 structure: 1) by G-quartet face recognition, 2) by edge recognition, 3) by loop recognition and 4) by simultaneous face and adjacent loop/groove recognition ((43). Most small-molecule ligands that induce the formation of G4 structures (telomestatin and PIPER for example) have been shown to be end-stacking ligands that bind the G-quartet. However, because of their size, it is more likely that specific antibodies that could discriminate between different G4 structures would recognize different types of loop arrangements and/or strand

polarities. Future molecular studies of the molecular recognition between these antibodies and different G4 structures may help us to understand how different proteins may be created to increase specificity interactions.

The potential biological importance of G4 structures can be seen in two major mechanisms, transcriptional regulation and the inhibition of telomerase (183, 184). Telomerase is a holo-enzyme complex which acts to prevent telomere shortening. Telomerase is expressed in the great majority of cancerous cells. G-quadruplexes are indirectly implicated in cellular aging and cancer if these structures inhibit telomerase *in vivo*. Anti-G-quadruplex antibodies could be used to study the natural existence of these structures in human cells (possibly by fluorescence microscopy) and correlative studies could further associate their presence with telomerase inhibition and cellular mortality rates or senescence.

The interconversion between double and single stranded DNA is presumed to be critical for the formation of G-quadruplex structures. Previous studies in ciliates have shown that G-quadruplex DNA in cells is dependent on chaperone proteins, such as the β subunit of the heterodimeric telomere end-binding protein complex (TEBPα/β) from *Oxytricha* (185). TEBPα/β has been shown to facilitate G-quadruplex formation, whereas other protein helicases such as Sgs1 from yeast (61) resolve these structures. Taken together, the simplicity of interconversion at physiological conditions from double stranded to single to quadruplex, and the prevalence of quadruplex binding proteins suggests that G-quadruplex DNA is an excellent candidate as a biological signaling molecule (59). The many possible isoforms of G-quadruplex DNA could bestow unique molecular recognition properties to G-quadruplex DNA in comparison to duplex DNA. Monoclonal antibodies specific for

different isoform could possibly assist to resolve the possible complex signaling nature of G-quadruplex DNA.

Moreover, such antibodies could be used to examine previous theories regarding the biological roles for tetramolecular G-quadruplexes. For example, Sen and Gilbert (1988) have proposed that tetramolecular G-quadruplex structures could be created by the association of 3' overhangs of telomeres of two chromosomes and that four homologous sister chromatids may be stabilized by G-quadruplex formation, aligning chromatids side by side during meiosis (19). Both 9A1 and 8H2 antibodies are specific for Ver-3 tetramolecular G-quadruplex DNA (Figure 11) and could be used to shed insight into this model.

Some of the discrepancies in the literature regarding G-quadruplex biology stem from studies conducted on different structural polymorphs of quadruplex DNA (59). Chemists continue to debate as to the exact chemical conformation of certain G-quadruplexes under physiological conditions. For example the vertebrate quadruplex sequence seems more structurally heterogeneous than once thought (103, 186). During the development of these antibodies we considered this dilemma and initially focused our efforts on the development of antibodies to stable single species of tetramolecular G4 DNA structures. In doing so, we have developed unique monoclonal antibodies with a high degree of specificity and affinity to specific tetramolecular G4 structures. Together with future monoclonal antibodies directed against unimolecular G4 DNA structures we hope to exploit these reagents for detection of G4 structures in mammalian cells by microscopy.

Chapter 3: Cellular localization of anti-G-quadruplex monoclonal antibody 1H6

The detection of G-quadruplex nucleic acids in cells has been limited in part by a lack of reagents that specifically target different G-quadruplex structures. Using antibodies to parallel stranded and antiparallel G-quadruplex DNA Schaffitzel *et al.,* were the first group to visually detect G-quadruplex DNA in the macronuclei of *Stylonychia lemnae* (108). It was proposed that the telomeric ends of *Stylonychia* associate through antiparallel G4 DNA formation. Unfortunately, similar detection of G-quadruplex DNA in a mammalian system has yet to be achieved. Exploiting our novel antibodies to various G-quadruplex structures we have identified a single antibody (designated 1H6) that generates distinct nuclear foci in mammalian cells. Using fluorescence microscopy and flow cytometry analysis we have analyzed the cellular localization of the 1H6 antibody under conditions in which the existence of native G-quadruplex nucleic acids in mammalian cells may be examined. In this chapter we describe our hypothesis testing of possible G-quadruplex nucleic acids in mammalian cells through the visualization of the 1H6 antibody. Fluorescence microscopy and flow cytometry are used to examine G-quadruplex nucleic acids in cells using the 1H6 antibody.

3.1 Introduction

Although the study of G-quadruplex nucleic acids has had a long history pre-dating the discovery of classical B-form DNA, visualization of G4 nucleic acids in mammalian systems has never been achieved. As a consequence the study of G4 nucleic acids remains occupied by a relatively small field of chemists and biologists. Nevertheless, interest in G4

nucleic acids was initiated by the discovery that telomeric oligonucleotides could form G4

DNA structures under physiological conditions (19) and has increased since the discovery

that some G4 structures could inhibit telomerase (129).

Interestingly, several groups have published reports on the development of antibodies

and antibody fragments specific to G4 structures (108, 171, 172, 176). However, no group

has shown the subcellular localization of these antibodies in mammalian systems. The best

in cell visualization of G4 structures is from Schaffitzel *et al*., who have shown the *in vivo*

localization of telomeric G4 structures in the ciliate *Stylonychia* using novel antibodies.

Unfortunately, these antibodies have not been reported to perform similarly in mammalian

cells. Given the extreme abundance of telomeric ends in the macronuclei of *Stylonychia*, the

overall fluorescent staining pattern in mammalian cells would be expected to be much less.

However, telomeres are not the only putative quadruplex forming regions in the human

genome and may only represent a small fraction of G4 structures. PQS are located

throughout the genome and have been estimated to be located at ~400,000 non over-lapping

regions and in excess of 6 million over-lapping regions (45). Given that these numbers are

estimates based on conservative putative quadruplex folding algorithms and rationalized on

the basis of the current human genome builds (which lack large portions of repetitive

sequences capable of forming G4 structures) it could be suggested that these numbers are

gross under-estimates. Conversely uncertainty remains regarding the stability of many G4

structures and therefore this may suggest a gross overestimation of G4 structures in the

genome. Better tools for the direct detection of G4 structures could help elucidate these

issues.

The *in vivo* evidence of G4 structures in *Styloyicha* and *Saccharomyces cerevissiae* (108, 110) suggest that G4 structures may be a general structure found in many organisms, including humans. Using the novel antibodies created in Chapter 2 we have studied the subcellular localization of these antibodies in human and mouse cells and tissues. One antibody, termed 1H6, produced clear punctuate nuclear foci in all cells as detected by fluorescence microscopy. In this chapter we outline the examination of 1H6 antibody staining patterns and the experimental conditions in which we attempt to influence G4 folding and stabilization. In conjunction with flow cytometry experiments examining hundreds of thousands of cells we focus on the subcellular localization of this antibody through fluorescence microscopy.

3.2   Experimental methodology

3.2.1   Immunofluorescence microscopy

Cells were grown on flamed coverslips in 12-well tissue culture plates and incubated at 37 ºC until 70% confluence. Cells were subsequently washed and fixed with either 1-4% paraformaldehyde in PBS for 5-15 minutes, ice-cold methanol for 5 minutes or methanol:acetic acid (3:1) for 10 minutes. Fixed cells were subsequently washed with PBS to remove excess fixative. Paraformaldehyde fixed samples were permeabilized with 0.5% triton-X 100 for 5 minutes. All washed and permeabilized cells were then blocked with 10% goat serum/2% BSA/0.1% tween-20 in PBS for 12 hrs at 4 ºC. Cells were then incubated with purified G4 antibodies either over-night at 4 ºC or 1 hour at room temperature. Cells were washed with PBST and then incubated with either: PE goat anti-mouse IgG (Jackson Immuno Research), Cy3 donkey anti-mouse IgG (Jackson Immuno Research), Alexa Fluor

596 goat anti-mouse IgG (Invitrogen) followed by staining for DNA with 20 ng/ml Dapi (Sigma). Coverslips were washed, mounted in Vectashield (Vector Laboratories, Inc.) or Prolong gold antifade (Invitrogen). Where indicated, cells were incubated with varying concentrations of NMM (N-methyl mesoporphyrin IX) (Frontier Scientific), TMPyP4 (Mesotetra(N-methyl-4-pyridyl)porphine) (Calbiochem) for the indicated amount of time. Where stated some fixed and permeabilized cells were treated with 10 µg/500 µl or 20 µg/500 µl RNAse A or RNAse T1 (Invitrogen). Fluorescence images were acquired using a digital camera mounted on a fluorescence microscope with either the Deltavision RT (Applied Precision, Issaquah, Washington) microscope or Zeiss microscope. Three-dimentional reconstruction and deconvolution were performed using Softworx Suite (Applied Precision, Issaquah, Washington).

3.2.2   Cell culture conditions

Cell lines used in this study include: primary human foreskin fibroblast cell line BJ (ATCC #CRL-2522), cervical adenocarcinoma cell line HeLa (ATCC# CCL-2) the ALT cell line GM847 and normal mouse embryonic stem cells from (72). All cell lines were maintained in a humidified chamber at 37 °C and 5% CO_2. Cells were cultured in tissue culture treated flasks and maintained in standard growth medium including IMDM/10/PS.

3.2.3   Fluorescence *in situ* hybridization

Cells used for *in situ* hybridization were first grown to 70% confluence in standard conditions. Prior to harvesting 1:1000 dilution of colecemid was added to cells for 3 hrs at

37 ˚C. Cells were subsequently washed with warm PBS and treated with trypsin/citrate to create a clean cellular suspension. Cells are then centrifuged for 5 minutes at 1,200 rpm to pellet cells and remove trypsin. Cells were then resuspended in 75 mM KCl (Hypo) for 10 minutes at 37 ˚C. Again, cells are centrifuged for 5 minutes at 1,200 rpm to pellet and remove Hypo. Cells are then resuspended in a small volume of PBS ~200 µl and dispersed into fresh methanol:acetic acid solution (greater than 2 mls) to avoid clumping. After 10 minutes of fixation cells are re-centrifuged for 5 minutes at 1,200 rpm to pellet and remove methanol:acetic acid. Fresh methanol:acetic acid (3:1) is applied to cellular pellet and one to two drops of fixed cells are dropped onto to clean glass slides and allowed to warm in the palm for 1 minute. Slides are air-dried over-night and examined under a light-microscope for good metaphase spreading. Fixed metaphase preparations were rehydrated in PBS for 15 minutes at room temperature. In a fume hood, metaphase preparations are then fixed in 4% formaldehyde solution for 2 minutes followed by 3 changes of fresh PBS for 5 minutes to remove fixative. Slides are then digested with pepsin at 1 µg/µl in 0.01N HCl (pH 2.0) for 10 minutes at 37 ˚C. Digested slides are again washed two times in PBS for 5 minutes. Metaphases were then subjected to RNAse A treatment at 0.5 µg/ml for 10 minutes at 37 ˚C. After RNAse A treatment, metaphases are washed in PBS three times for 5 minutes each. The previous formaldehyde fixation and wash step is then repeated prior to blocking. Metaphases are blocked with 10% goat serum/2%BSA/PBST for 1 hour at room temperature. Primary antibody (1H6 or IgG$_{2b}$ control) is diluted in block buffer and incubated with metaphase preparations at 4 ºC over night. Slides were washed three times with PBST for 10 minutes each prior to incubation with goat-anti-mouse-AF594 at a 1:2000 dilution in block buffer in the dark for 1 hour at room temperature after which slides were washed three times

with PBST for 15 minutes each. Fixation and washing was again repeated with an extended

fixation length of 10 minutes with 1% formaldehyde. After washing, slides were dehydrated

with increasing ethanol concentrations (70%, 90% and 100%) for 5 minutes at each

concentration. In some cases telomeric PNA probes are hybridized as follows, after slides

have air dried, 20 µl of hybridization mixture containing 70% deionized formamide, 0.25%

NEN, 10 mM Tris-HCl pH 7.4, 0.5 µg/ml PNA Tel Cy3 and 25 mM $MgCl_2$ is placed on a

coverslip. The coverslip is gently placed on the dried slide and chromosomal DNA is

denatured at 80 ˚C for 3 minutes. Slides are subsequently removed from the oven and

allowed to hybridize at room temperature in a damp dark box for 2 hours. Slides are then

washed twice for 15 minutes with wash solution 1 (70% formamide, 0.1% BSA and 10 mM

Tris-HCl pH 7.4) removing the cover slip. Washing continues with three washes with wash

2 (0.1 M Tris-HCl pH 7.4, 0.15 M NaCl and 0.1% tween-20) for 5 minutes each. Slides are

dehydrated as above and counter stained/mounted as stated in immunofluorescence

microscopy.

3.2.4   Immunohistochemistry

Normal human tissues were obtained from the Centre for Translational and Applied

Genomics (CTAG) at the BC Cancer Agency. Tissues were fixed, embedded in paraffin,

sectioned and stained on site at CTAG using a Tissue-Tek Xpress continuous rapid tissue

processor and the Discovery XT for detection. Images were obtained using a Zeiss

Axioplan microscope and OpenLab software (Improvision).

3.2.5   Flow cytometry

Cells were grown to 70% confluence in 12 well tissue culture dishes under standard conditions. Where indicated, cells were incubated with varying concentrations of NMM (N-methyl mesoporphyrin IX), TMPyP4 (Mesotetra(N-methyl-4-pyridyl)porphine) or PIPER for the indicated amount of time. Cells were then trypsinized for 5 minutes at 37 ^{o}C and centrifuged for 5 minutes at 1,200 rpm. Resuspended cell pellets were fixed in 4% formaldehyde, and permeabilized with 0.5% Tween 20 in PBS. Fixed and permeabilized cells were incubated with 5% normal goat serum for 4 hrs at 25 ^{o}C to block protein binding sites. Subsequently, suspended cells were incubated with 1 µg/ml 1H6 antibody for 12 hrs at 4 ^{o}C. After washing cells were resuspended in PBS containing 0.5 µg/ml dilution of secondary antibody conjugated with PE or FITC for 1 hr at 25 ^{o}C. Stained cells were analyzed using a FACSCalibur (BD Biosciences). Data were recorded and analyzed using CellQuest (BD Biosciences) and FlowJo (Tree Star, Inc.) in succession. For DNA content determination we harvested cells into ice-cold 70% ethanol (1×10^6 cells/mL) for 30 minutes at 4 ˚C. Fixed cells were spun down at 2,000 rpm for 5 minutes. Cells were subsequently resuspended in PBS and washed twice more. 100 µl or RNAseA (100 µg/ml) was added to the cells and left at room temperature for 5 minutes. 1H6 antibody staining was performed as above and finally, 400 µl of propidium iodide (50 µg/ml) was added to the cells prior to analysis.

3.3 Results

3.3.1 Immunofluorescent staining of mammalian cell lines with monoclonal antibody 1H6 generates distinct nuclear foci

In the past, antibodies and antibody fragments to various G4 DNA structures have been produced, however there are no reports of the visualization of G4 DNA structures in vertebrates (108, 171, 172, 176). The most promising candidate was a single-chain antibody fragment generated via ribosome display technology that visually detected the resolution and formation of G4 DNA at the telomeres in the macronucleus of the ciliated protozoa, *Stylonychia* (108). Unfortunately, this antibody has not been shown to bind the telomeres of vertebrates, possibly due in part to the few telomeres present in vertebrates (96 ends in humans) in comparison to the extreme abundance ($>10^9$) telomeres found in *Stylonychia*. Additionally, the *Stylonychia* specific antibodies have not been shown to resolve individual telomere foci. Alternatively, the anti-Stylonychia antibodies may not bind mammalian telomeric DNA structures or may have insufficient affinity required for the detection of mammalian G4 DNA. However, based on these studies, we hypothesized that if an antibody could detect G4 DNA in vertebrates, based solely on the decreased number of telomeric sites, staining patterns would be significantly decreased in comparison to these earlier studies in *Stylonychia*. Interestingly, one antibody (designated 1H6) showed discrete nuclear foci that was clearly above that of matched isotype controls. Figure 21 illustrates the dramatic nuclear localization of foci present when human cell lines are fixed and stained with 1H6 and detected by fluorescently coupled secondary antibodies. Inverted black and white images clarify the localization of foci where the majority of staining is detected in perfect overlap with Dapi nuclear staining.
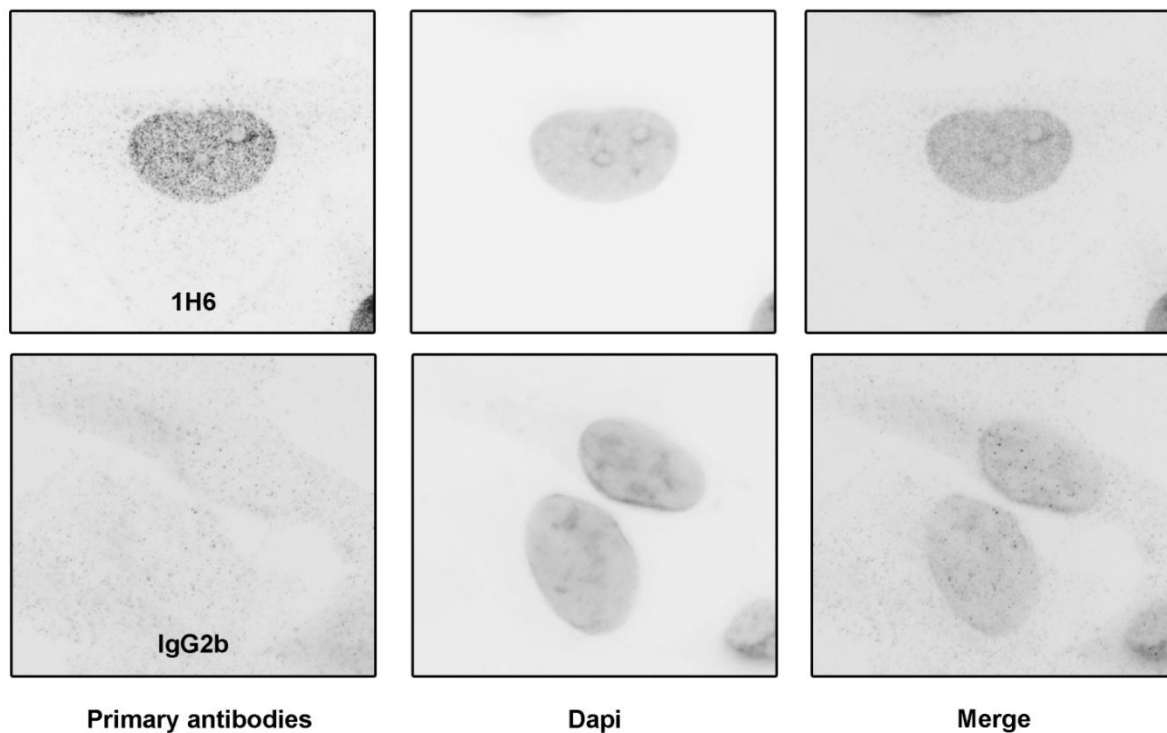
Figure 21. Anti-G-quadruplex antibody 1H6 produces distinct nuclear staining pattern above isotype controls.

Human HeLa cells have been fixed and stained with anti-G-quadruplex monoclonal antibody 1H6. 1H6 antibody staining was fluorescently detected using a goat-anti-mouse secondary and appears to be significantly localized to the nucleus. Black and white images have been inverted for clarity. 1H6 antibody staining co-localizes with Dapi nuclear stain.

The status of telomerase in cells may influence the detection of G4 DNA at telomeres. Lacking knowledge of what impact telomerase status might have on these foci we chose to examine BJ (non-cancerous primary cell line), GM847 (ALT positive cancer cell line) and HeLa (telomerase positive cancer cell line) to examine differences in staining patterns. Thus far we have not detected differences in the location and gross florescence patterns in these cells. All cell lines tested to date exhibit this nuclear staining pattern. Using CY3-coupled secondary anti-mouse antibody, fluorescent staining is nuclear, and is restricted

93

to bright, punctuate foci.   Representative images (Figure 22) have been reconstructed as projections of three-dimensional nuclei from stacks of two-dimensional sections to generate overall staining patterns in these cell lines. This staining pattern is consistent within a cell population and between different cell lines although the number of foci and therefore overall mean fluorescence appears to vary from cell line to cell line (Figure 22). Upon three-dimensional reconstruction minimal foci are detected outside of the nucleus, but are present in some preparations.
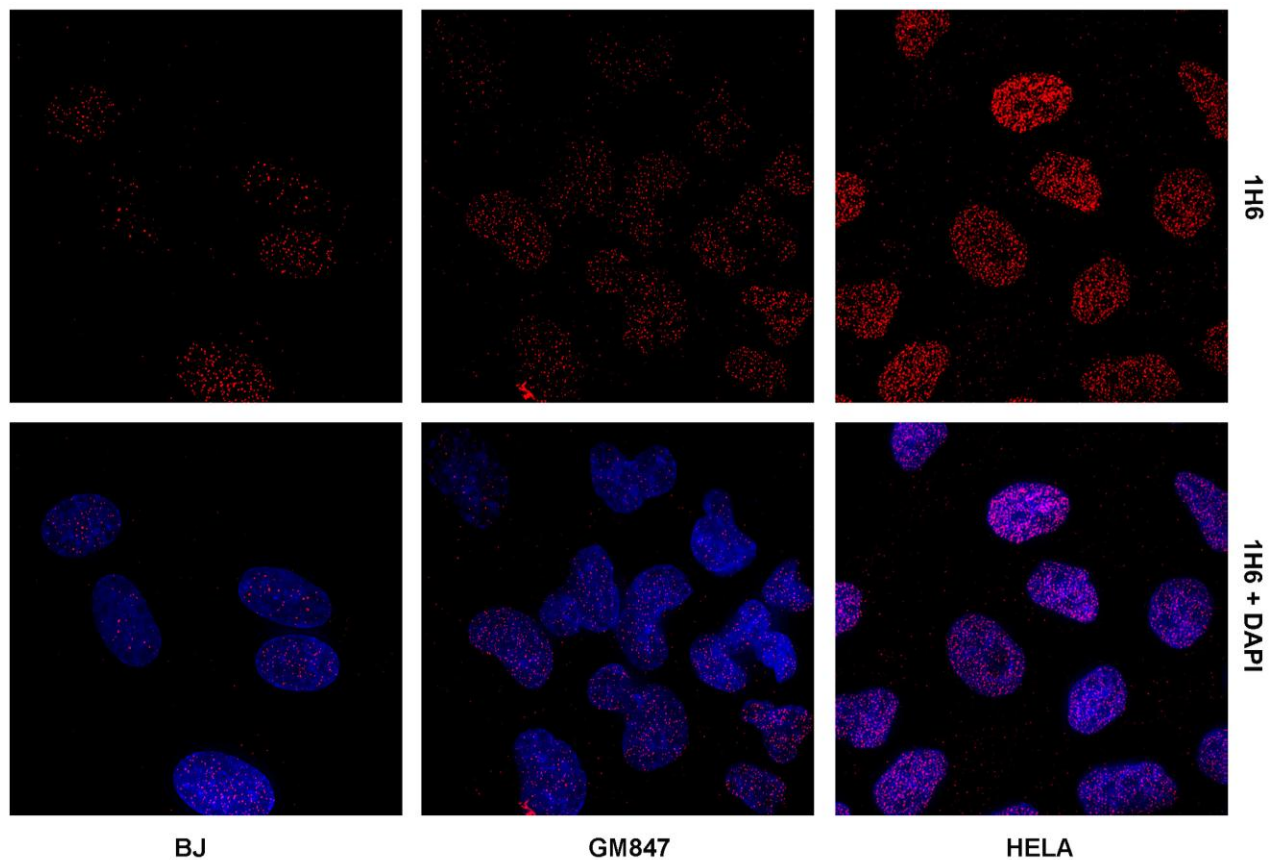
Figure 22. 1H6 antibody appears to generate discrete nuclear foci in multiple human cell lines. Monoclonal antibody 1H6 generates discrete punctuate nuclear foci in primary (BJ), ALT positive (GM847) and telomerase positive (HeLa) human cell lines. Although the number of foci varies from cell line to cell line, all cells tested to date (mouse or human) positively stain with the 1H6 antibody.

3.3.2   Localization and intensity of nuclear 1H6 antibody staining varies

Understanding the staining patterns observed using the 1H6 antibody might help elucidate where G4 DNA forms most readily and how these structures might be influenced by cell cycle progression. Remarkably the staining patterns observed within cell populations varied between focused nucleolar localization and an absence of nucleoli staining all together (Figure 23). Most common was an absence of staining in the nucleoli. However, significantly increased localization of staining at nucleoli was also observed in many preparations and may be an indication of increased levels G4 RNA accumulating in the nucleoli during different phases of the cell cycle. To further explore this hypothesis we examined nuclei chemically blocked at metaphase. Figure 23 illustrates the increased staining intensity observed when fixed metaphase cells are detected with the 1H6 antibody. In comparison to interphase cells, the intensity is increased and staining appears to be directly localized to chromosomes.

| Nucleoli staining | Absence of nucleoli staining | Metaphase staining |
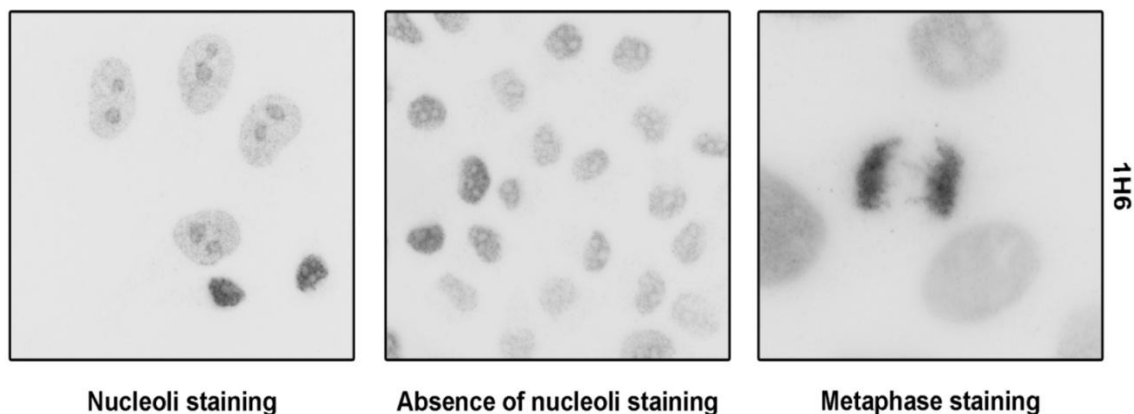
Figure 23. Variation of anti-G-quadruplex antibody 1H6 staining pattern across the cell cycle. The antibody staining pattern generated using fixed HeLa cells in the presence of anti-G-quadruplex antibody 1H6 shows variable localization and intensity. Significant portions of cellular populations show an absence of nucleoli staining signal while other portions of the population appear to have increased localization of signal in the nucleoli. Metaphase chromosomes have increased intensity and staining with the 1H6 antibody. Black and white images have been inverted for clarity.

3.3.3 Different fixation techniques have limited effect on 1H6 foci.

Our preliminary microscopy results generated discrete nuclear foci, which suggested to us that if G4 was being detected such structures are clustered in nuclei. Alternatively, the antibody could accumulate in foci by means of some poorly understood experimental artifact. To further examine these possibilities, we investigated other fixation techniques. Paraformaldehyde and other cross-linking fixatives function by cross-linking proteins and nucleic acids while maintaining cellular structure, but may eliminate some specific epitopes. Conversely, organic solvent fixation techniques precipitate proteins and destabilize membranes during fixation. If the 1H6 antibody is detecting nucleic acid structures, organic solvent fixation should not dramatically alter the discrete nuclear foci pattern detected. To this end we investigated methanol and ethanol fixation techniques in comparison to our

standard formaldehyde fixation technique. Organic solvent fixation did not dramatically alter

the discrete nuclear foci detected (Figure 24). An antibody that can be successfully utilized

across cross-linking and precipitating fixatives is rare and suggests that the antibody target is

durable during these preparations. By observing similar staining patterns across varying

fixatives we have not eliminated the possibility that the 1H6 antibody is binding proteins

specific to the nucleus. However, varying fixation technique while observing similar staining

patterns in comparison to unfixed cells (Appendix B.2) provides support that the 1H6

antibody is not aggregating due to an unforeseen artefact.

Methanol   **Methanol:Acetic acid**   **Paraformaldehyde**
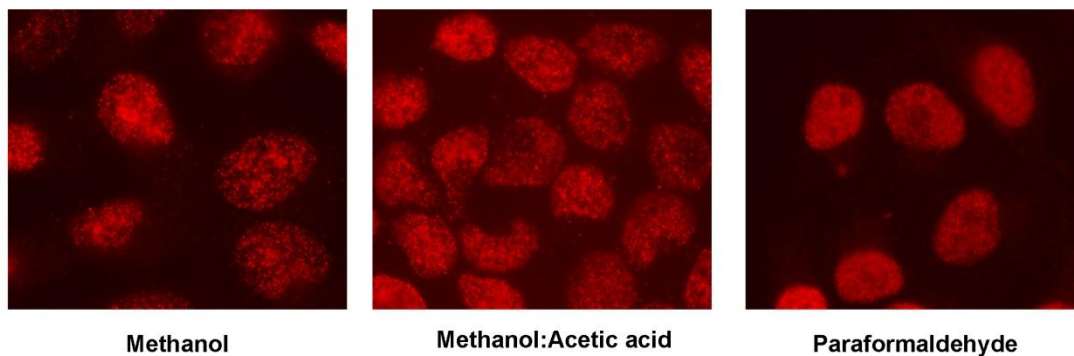
Figure 24. Different fixation techniques have limited effect on nuclear localization of 1H6 antibody.
Human cell line HeLa was fixed using precipitating fixatives methanol or methanol and
acetic acid, and compared to the cross-linking fixative paraformaldehyde (See Experimental
methodologies). Variations in fixation did not alter the nuclear localization of anti-G-
quadruplex antibody 1H6 (Red).

3.3.4   Mean fluorescence signal increases with DNA content

If the 1H6 antibody was detecting nucleic acid structures that may be folded and

unfolded during different stages of the cell cycle, different DNA contents may influence the

number of folded structures. In this regard the mean fluorescence signal would be different

between stages of the cell cycle with 2n as opposed to 4n DNA content. Using flow

cytometry we stained cells with propodium iodide in conjunction with 1H6 antibody staining

to visualized DNA content in relationship to antibody staining. Our results showed almost a

two-fold mean fluorescence signal increase when comparing G1 cells to cells in S and G2/M

(Figure 25). These results suggest that the greater the DNA content the greater the 1H6

antibody staining and in turn, possibly more folded G4 structures.
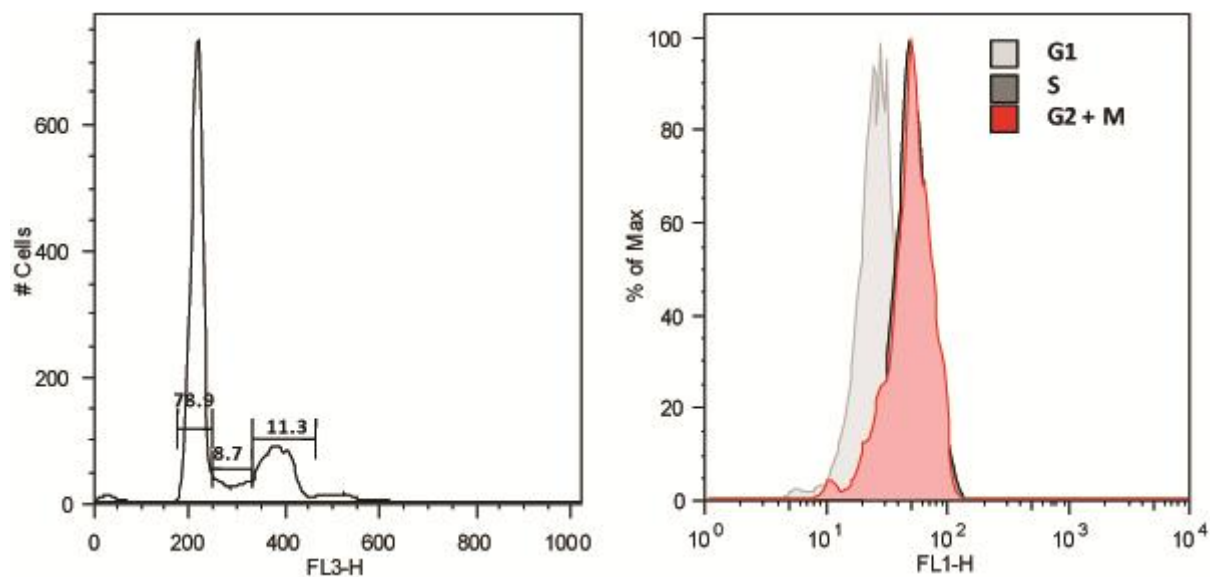
Figure 25. Cells stained with 1H6 antibody have increased mean fluorescence in late stages of cell cycle. HeLa cells were fixed in ethanol and stained with propidium iodide (detected in FL3, left panel) in conjunction with the 1H6 antibody (detected in FL1, right panel). Overall mean fluorescent signal is greater in S-phase (black, exact overlap by red) and G2/M-phases (red) over G1-phase (light grey).

3.3.5 Nuclear staining pattern of 1H6 antibody is unaffected by RNAse but decreases in the presence of synthetic G4 DNA.

Another possibility that has been described in previous studies suggests that G4 structures can be folded from RNA molecules (187-190). G4 RNA could arise from the transcription of genomic sequence containing long stretches of cytosines, which may subsequently fold into alternative secondary structures such as G-quadruplexes. RNA molecules are transcribed from DNA in single-stranded form and therefore may represent more propensity than DNA (usually bound by a complementary strand) to fold into alternative structures. To examine the prospect that the foci being detected were RNA quadruplexes, we pre-incubated paraformaldehyde fixed cells with increasing concentrations of RNAse to digest RNA oligonucleotides before antibody staining. RNAse treatment with RNAse A or T1 did not have an effect on the overall nuclear staining pattern detected (Figure 26). However, we are unaware of any G4 RNA specific RNAse enzymes at this time and cannot eliminate the possibility that RNAse A and T1 cannot digest G4 RNA. Nevertheless, the binding of 1H6 to G4 DNA in ELISAs (Chapter 1), the comparable staining observed across fixed or unfixed cells and a lack of effect observed with RNAse enzymes suggest that the nuclear foci being detected are predominantly of DNA origin.

To further investigate the nuclear foci generated by the 1H6 antibody, we hypothesized that incubation with synthetic G4 DNA would compete with native G4 DNA in cells for binding of the antibody. Competition would lead to a decrease in fluorescent signal detected. In this regard we incubated increasing concentrations of synthetic G4 DNA (tetramolecular Oxy-2) with constant amounts of 1H6 antibody (Figure 26). At equimolar concentrations of antibody to synthetic G4 DNA, nuclear foci were reduced to background

levels. This result suggests that the native epitope recognized by the 1H6 antibody in human

cells could be G4 DNA.
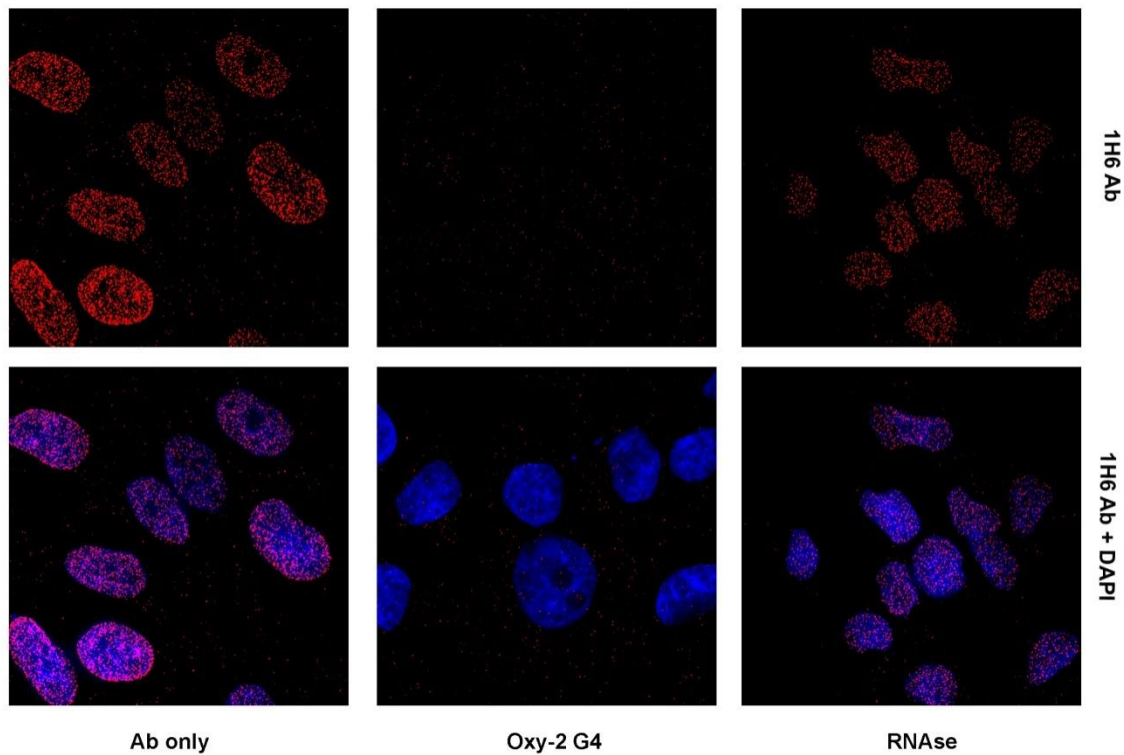
Figure 26. Nuclear staining pattern of 1H6 antibody is decreased in the presence of synthetic tetramolecular G4 DNA whereas treatment with RNAse had no effect.
Fluorescent microscopy analysis of HeLa cells stained with anti-G-quadruplex antibody 1H6 (Red) and Dapi DNA dye (Blue). Nuclear staining is significantly reduced in the presence of synthetic tetramolecular G4 DNA (Oxy-2). Treating HeLa cells with RNAse (A and T1) prior to staining with 1H6 had no effect on the staining pattern observed (RNAse).

In addition to visualizing these experiments using fluorescence microscopy we

wanted to explore similar experiments by flowcytometry. Unlike fluorescence microscopy,

whereby one examines a small number of cells per image field, flow cytometry can be used

to examine thousands of cells. In a similar fashion to our fluorescence microscopy

experiments with synthetic G4 DNA, we found that the mean fluorescence decreased

significantly upon incubation of cells with synthetic tetramolecular Oxy-2 G4 structures

(Figure 27). All cell lines examined exhibited this characteristic relationship. These

experiments suggest that the 1H6 antibody is detecting DNA structures in human cells. The

competition between *in vivo* DNA and synthetic G4 DNA provides support that the 1H6

antibody is detecting G4 DNA in human cells. However, this finding could also be explained

by an in cell epitope with lower affinity to the 1H6 antibody.
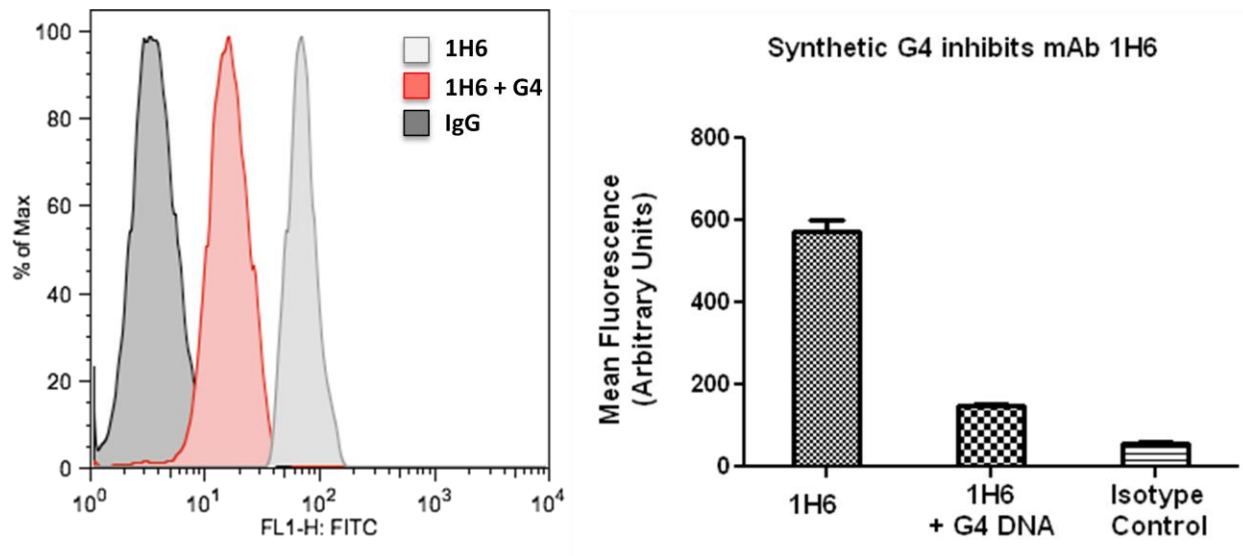
Figure 27. Synthetic tetramolecular G4 DNA decreases 1H6 nuclear foci as detected by flow cytometry.
Human HeLa cells were prepared and stained with anti-G-quadruplex antibody 1H6 with and
without the presence of synthetic tetramolecular G4 DNA. Mean fluorescence was measured
by flow cytometry and plotted above. Incubating the 1H6 antibody in the presence of
tetramolecular DNA decreases the total mean fluorescence.

3.3.6 Quadruplex stabilizing drugs cause increased mean fluorescence of 1H6 nuclear foci

Although G4 DNA has been hypothesized to form throughout the human genome if these structures are formed then they must also be resolved, to preserve genomic integrity through multiple rounds of DNA replication and transcription. G4 DNA could potentially cause stalled replication forks by physically blocking DNA access for the replication/transcription machinery (17). If the 1H6 antibody is detecting G4 DNA structures in cells there should be a dynamic equilibrium between G4 DNA—double-stranded DNA and single stranded DNA. Altering the equilibrium may cause variations in staining patterns observed. Fortunately, there are several synthetic and natural small molecules that have been discovered which bind and stabilize G4 DNA structures (123, 130-134). We hypothesized that treatment of cells with small molecules that stabilize G4 DNA would increase the number of nuclear foci detected by antibodies specific to G4 DNA. We therefore treated human cell lines HeLa, GM847 and BJ with increasing concentrations of G4 stabilizing drugs TMPyP4 or NMM (N-methyl mesoporphyrin IX) (Figure 28). Following standard immunofluorescent staining we detected an increase in fluorescent signal with increased concentrations of G4 stabilizing drugs. Visually it appeared that there were more nuclear foci and brighter nuclear foci present in cells treated with high concentrations of TMPyP4 or NMM. Even at low concentrations of TMPyP4 or NMM foci were brighter than foci in control experiments. This finding suggested that more secondary antibody was binding at these regions and that this may be due to an increase in G4 DNA stability and 1H6 antibody binding.
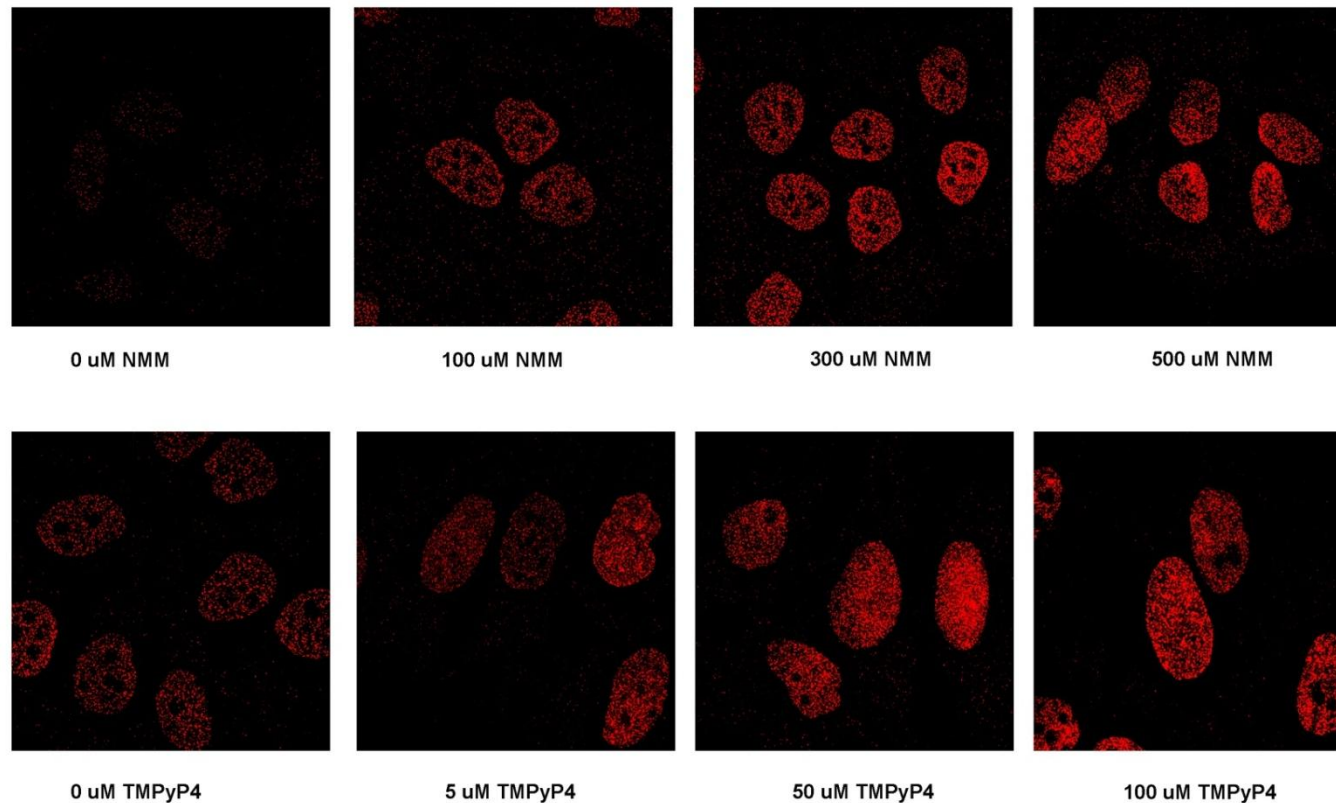
Figure 28. Cellines treated with G4 stabilizing drugs show increased mean fluorescence by microscopy.
Cells stained with 1H6 antibody (Red) after incubation with the G4 stabilizing drug NMM or TMPyP4 show increased mean fluorescence in fluorescent microscopy experiments. GM847 cells (upper panel) and HeLa cells (lower panel) are treated with increasing concentrations of G4 stabilizing drugs. Cells were subsequently fixed and visualized with the same exposure settings. Increased staining was observed across all cell lines with either TMPyP4 or NMM (two examples shown for clarity).

TMPyP4 treatment of HeLa cells

Legend: IgG, 1H6

X-axis: Drug treatment — untreated, TMP 5, TMP 50, untreated, TMP 5, TMP 50

Y-axis: Mean Fluorescence (Arbitrary Units)
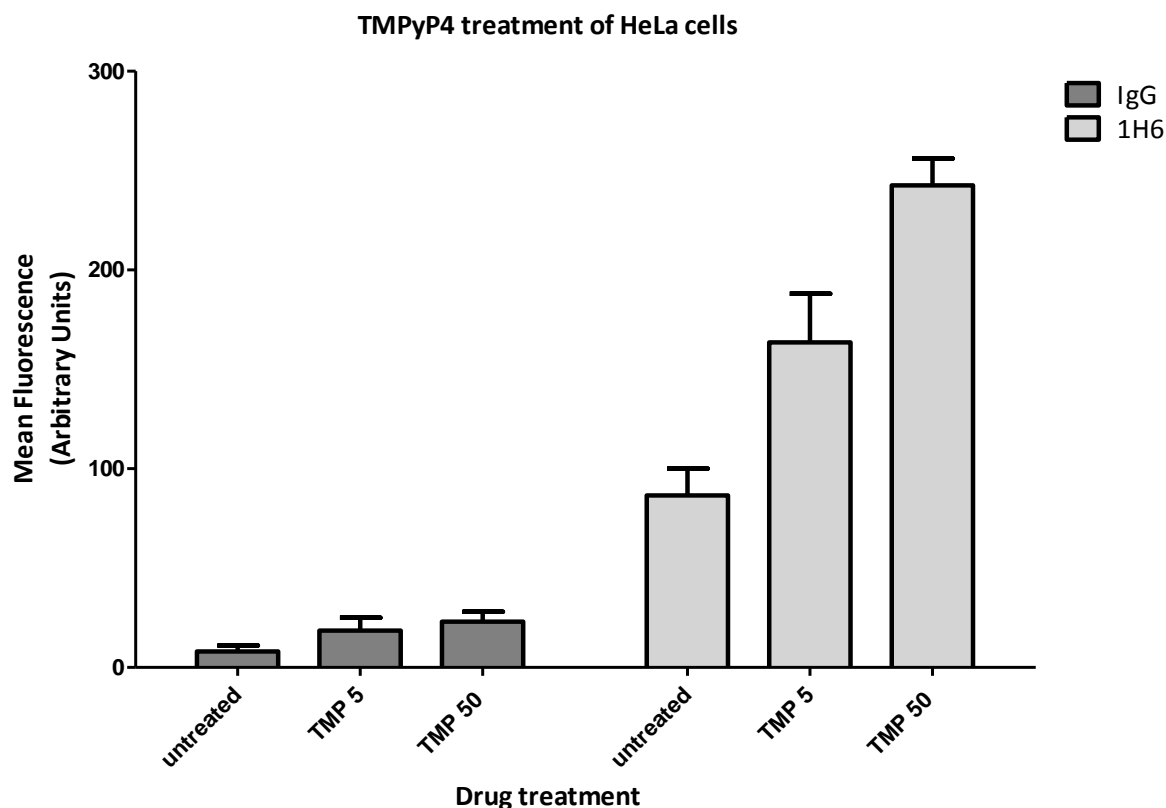
Figure 29. Flow cytometry analysis of HeLa cells show increased mean fluorescence upon treatment with TMPyP4.
HeLa cells were treated with increasing concentrations of G-quadruplex stabilizing drug TMPyP4 and subsequently stained with the 1H6 antibody. Increased TMPyP4 treatment resulted in increased mean fluorescence detected. TMPyP4 concentrations (T) in μM are listed in the graph legend.

In conjunction with our fluorescent microscopy experiments we sought to gather an overall quantitative analysis of the fluorescent patterns observed upon treatment of cells with these small molecules. Therefore we examined the mean fluorescent patterns of these cell populations via flow cytometry (Figure 29). First of all, consistent with microscopy experiments we detected an increased mean fluorescent signal in accordance with increasing amounts of TMPyP4 and NMM. This increase in mean florescence was independent of the

increased florescence signal due to the back-ground fluorescence of these drugs. Fluorophores were initially tested and chosen such that the emission spectrum of the fluorophores tested were sufficiently distant from the excitation and emission spectrum of TMPyP4 or NMM as not to influence the experiment. Also, the increased mean fluorescent signal was not detected in our isotype control and suggests an overall increased number and intensity of nuclear foci detected in concurrence with increased concentrations of G4 stabilizing drugs. Taken together, the immunofluorescence and flow cytometry data suggest that G4 stabilizing drugs TMPyP4 and NMM cause an increase in the intensity of nuclear foci detected by anti-G-quadruplex antibody 1H6.

3.3.7   1H6 antibody foci exhibit limited co-localization with telomeric and helicase proteins.

Much of the current G-quadruplex literature has focused on the possible existence of G4 DNA at human telomeres. Human telomeres represent a likely source of biologically relevant G4 structures because of the conservation across species of guanine-richness and the naturally occurring single-stranded 3' overhang end.

To examine the possible existence of human telomeric G4 DNA we co-stained cells with anti-G-quadruplex antibody 1H6 in addition to antibodies targeting proteins found at telomeres. Figure 30 represents a small subset of those experiments using antibodies to POT 1 (Protection of telomere 1) and PTOP (POT1- and TIN2-organizing protein) also known as TPP1 or PIP1 or TINT1. Although some co-localization of spots was observed, the patterns were highly variable. In addition, given the vast numbers of 1H6 foci we could not

conclusively say that the 1H6 antibody foci were co-localizing with any telomeric protein

foci at significantly greater numbers than expected by chance.
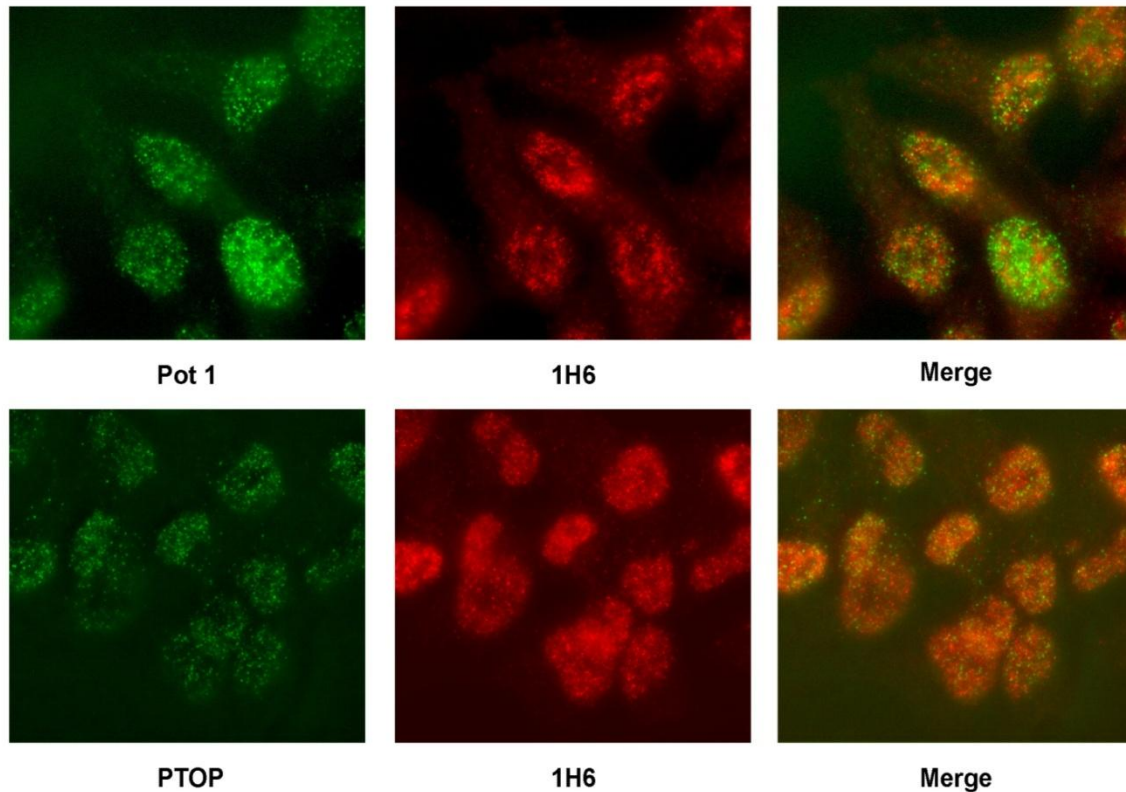
Figure 30. Telomeric specific proteins POT1 and PTOP show limited co-localization with anti-G-quadruplex antibody 1H6.
HeLa cells were fixed with paraformaldehyde and stained by standard techniques (See Experimental Methodology). Green fluorescence represents telomeric proteins while red fluorescence represents 1H6 antibody staining. Limited co-localization was observed in our standard preparations.

Similarly, we attempted to visualize the co-localization of the WRN helicase protein

with specific 1H6 antibody foci. Helicase proteins have often been cited in the literature as

unwinding different G4 structures *in vitro* which suggest they would also mark locations of

G4 structures in immunofluorescent experiments. Analogous to our co-localization

experiments with telomeric foci we had difficulty convincing ourselves that the small

106

percentage of co-localization events were statistically greater than one would expect by chance (Figure 31). Most 1H6 foci did not co-stain with telomeric proteins. We believe these experiments were unclear due in part to the nature of fluorescence microscopy and the numerous foci present using the 1H6 antibody.

Interestingly, during several microscopy experiments we visualized long DNA/protein fibers between two cells. Although DNA signal was relatively weak in the DAPI channel, we could easily detect the fibers using the 1H6 antibody. Figure 31 reveals one such fiber between two cells co-stained with 1H6, anti-WRN and Dapi. Magnification of the fiber shows a long thin strand of DNA coated with both 1H6 foci and WRN foci. It is appealing to postulate that these fibers could have been the result of improper unfolding of G4 DNA structures during the cell cycle that have persisted after division. Nonetheless, the co-staining of these fibers with the 1H6 antibody and anti-WRN antibodies could be an unforeseen artifact of the procedure and requires further control experimentation.
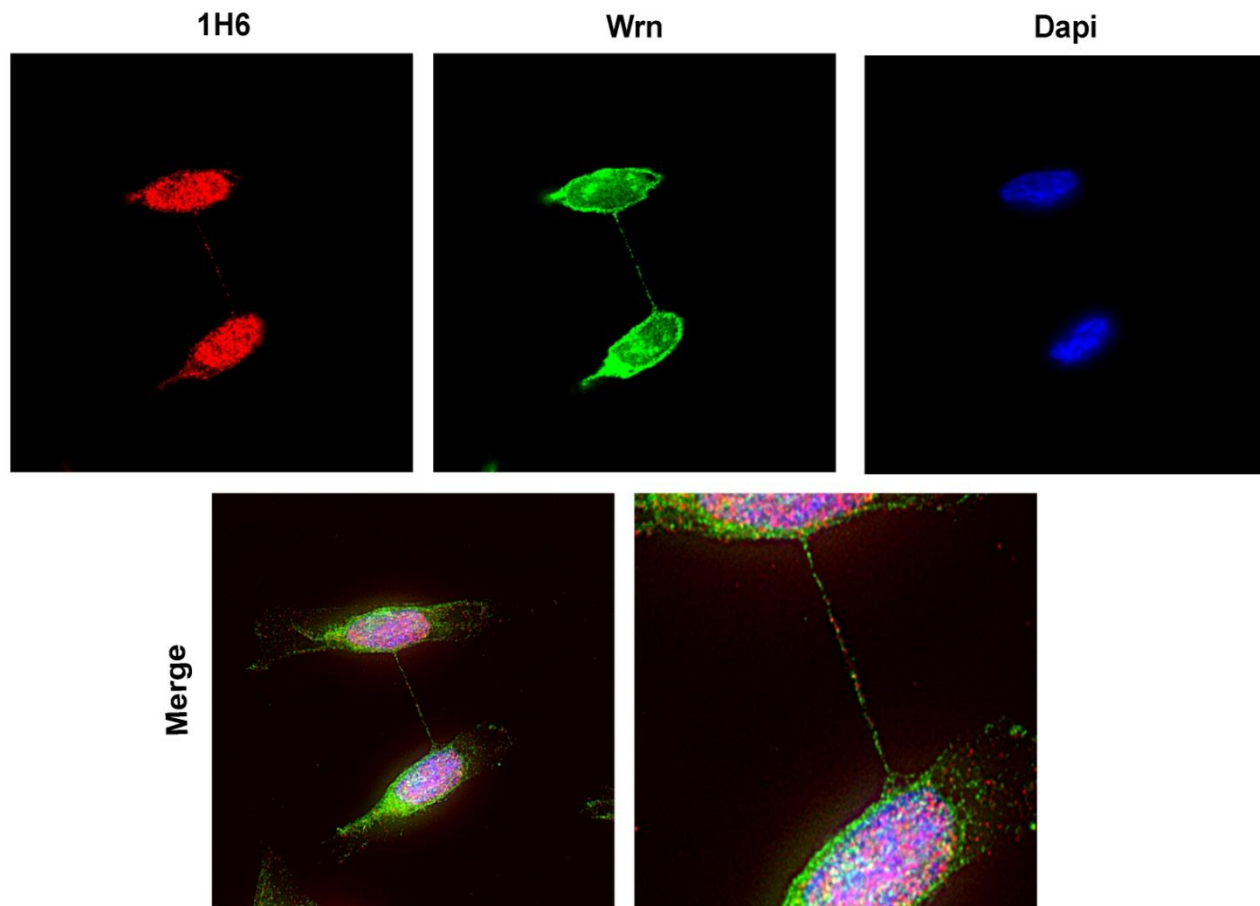
Figure 31. DNA fibers between cells contains both 1H6 foci and WRN protein foci.
Fluorescence microscopy of HeLa cells using the 1H6 antibody (Red) in co-localization
experiments with WRN (Green) reveal fibers between cells. Dapi staining is faint but
magnification of the fibers indicates the presence of both 1H6 foci and WRN foci.

3.3.8   Immunofluorescence *in situ* hybridization of telomeric probes and fixed 1H6 antibody foci reveal chromosomal location of 1H6 antibody

Given the difficulty of quantifying co-localization experiments involving an
overwhelming number of 1H6 antibody foci we sought to clarify the physical location of 1H6
antibody foci on chromosomes using other methods. Generating metaphase preparations of
cells increases the clarity of chromosomes and facilitates the visual localization of antibody

binding patterns due in part by the flattening and spreading of chromosomes. Fortunately, the fixation required for generating good metaphase preparations (a combination of methanol and acetic acid, see Experimental Methodology) has little effect on the overall nuclear staining using anti-G-quadruplex antibody 1H6 (See Figure 24).

Standard mouse ES cell metaphase preparations were generated and subsequently stained with primary 1H6 antibodies and detected with goat-anti-mouse antibodies conjugated to Alexafluor 594. Figure 32 illustrates the localization of 1H6 antibody foci in comparison to Dapi DNA dye. 1H6 antibody foci are discrete and isolated to chromosomes. Although it appears that significant portions of 1H6 antibody foci are randomly distributed across all chromosomes, some foci are found at the terminal ends of chromosomes. These terminal foci may indicate binding to telomeres.

DAPI   **1H6**
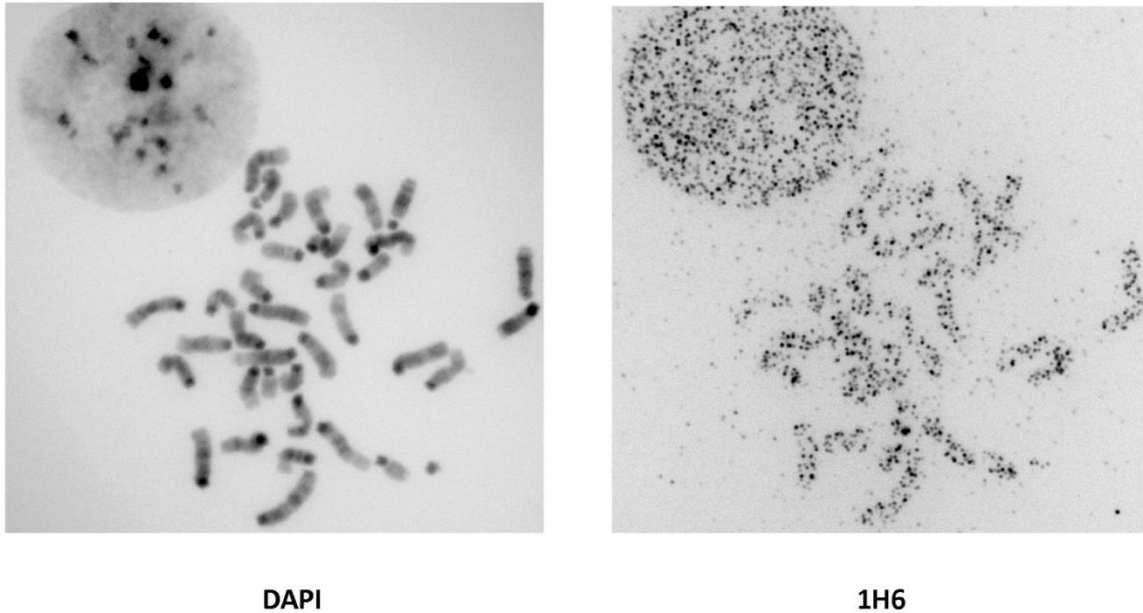
Figure 32. 1H6 antibody shows discrete foci on chromosomes of standard mouse metaphase preparations.

Standard metaphase preparations of normal mouse ES cells were stained with anti-G-quadruplex antibody 1H6 (right panel). Significant staining is detected on chromosomes as confirmed by localization of 1H6 staining and Dapi dye (left panel). 1H6 antibody foci are revealed throughout the chromosomes including at the terminal ends. Images are in grey scale and inverted for clarity.

Given that some 1H6 antibody foci appear to be located at the terminal ends of chromosomes of metaphase preparations we sought to identify telomeres of metaphase preparations. Combining other antibodies for co-localization experiments using metaphase preparations was problematic, as most antibodies to proteins do not perform properly under these conditions. However, modification of standard immunofluorescent techniques allows for the combination of immunofluorescence and fluorescence *in situ* hybridization (FISH). Using standard metaphase preparations, FISH permits the localization of telomeres by annealing fluorescently labeled PNA probes directly to telomeric repeats (See Experimental Methodology). Exploiting the immuno-FISH technique we combined 1H6 antibody staining

and telomeric PNA probes to examine the localization of 1H6 foci. Figure 33 reveals that it is possible to visually detect 1H6 antibody foci at the same location as telomeric PNA probes. Interestingly, not all telomeric foci co-localize with 1H6 antibody staining and again most 1H6 antibody staining appears to be non-telomeric. Future experiments could attempt to quantify this observation and as of now we are only speculating based on a small number of observed chromosomes. Nevertheless, the staining pattern observed suggests that not all mouse telomeres have folded quadruplex structures. In addition, because not all 1H6 spots are at every telomere this may indicate that folded quadruplex structures may play a role in telomere maintenance on some ends and not others (possibly due to the presence of T-loops or other capping mechanisms). Additionally, we cannot eliminate the possibility that the observed co-localization was coincidental, without proper quantification experiments. Outside of telomeres, magnifications of mouse chromosomes in Figure 33 reveal that a significant portion of 1H6 antibody foci are distributed throughout the chromosome and appear to bind at similar locations on sister chromatids. This could be the result of similar DNA sequences on sister chromatids consequentially forming similarly folded structures. Although intriguing, this observation also requires quantification and further experimentation to substantiate this idea.

Red = 1H6 antibody
Green = C-tel-probe
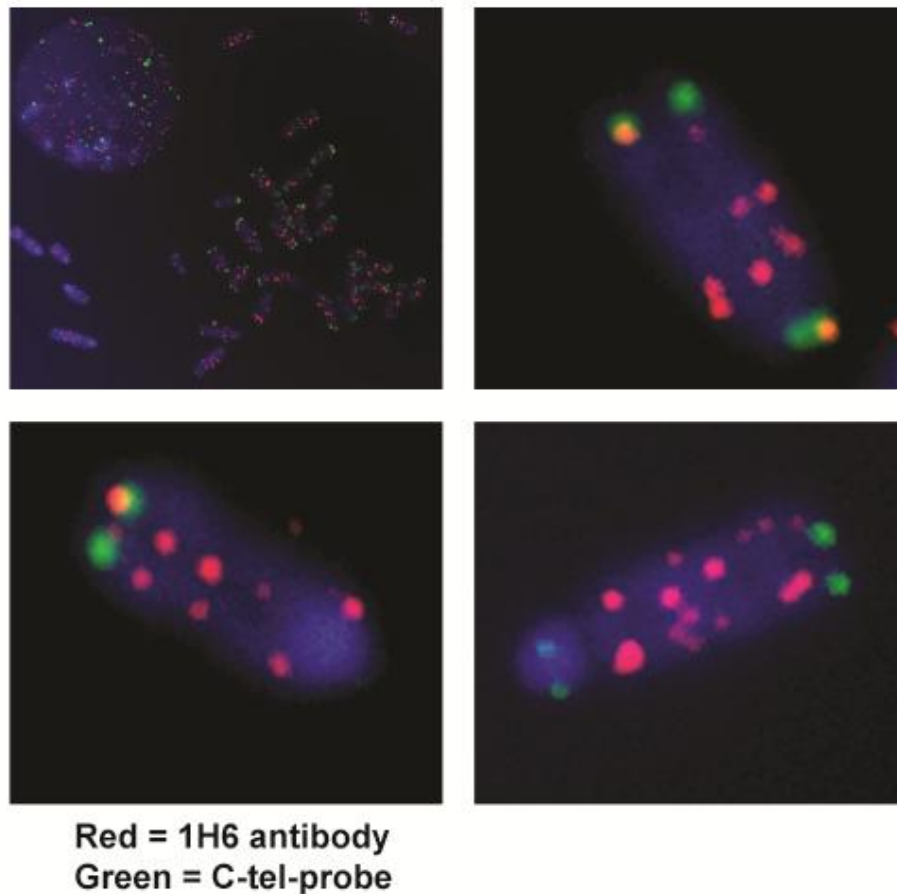
Figure 33. Immuno-FISH staining illustrates co-localization of 1H6 foci and telomeric repeats.
Normal mouse ES cells were used to prepare metaphase preparations and stained with the
1H6 antibody (Red). After fixation of the antibody, FISH was performed with a telomeric
probe (Green). Three panels have been magnified to give a clear representation of
chromosomal staining patterns. Red fluorescence indicates the 1H6 antibody and green
represents the telomeric probe, yellow is over-lapping signals.

**3.3.9   Immunohistochemical staining of normal human tissues with 1H6 antibody
shows antibody staining in all human cell types except a subpopulation of cells in the
human testis.**

Although cell lines are valuable tools for exploring the use of these novel reagents,
observed staining patterns may be caused by unknown artifacts found only in laboratory cell

lines grown in tissue culture. In addition to examining the staining patterns of the 1H6 antibody in multiple human and mouse cell lines, we therefore explored the staining of normal human tissues. In collaboration with the Centre for Translational and Applied Genomics (CTAG) we stained various human tissues with anti-G-quadruplex antibody 1H6. Normal human tissues are continually collected during tumor resection and stored as fixed paraffin embedded samples for analysis at CTAG. 1H6 antibody staining was detected with a secondary anti-mouse antibody conjugated to HRP in the presence of the chromogenic substrate DAB (3, 3′-Diaminobenzidine). Figure 34 (and Appendix B9) confirms the presence of 1H6 antibody staining in all nucleated cells (as present by dark-brown staining). Interestingly, no staining was observed in a subset of testis cells including the spermatogonial cells. Given that guanine-rich DNA is present throughout the human genome it is likely that anti-G-quadruplex antibodies would stain any cell with DNA (note lack of staining in red blood cells in the liver, Appendix B.9). However, the lack of staining in the spermatogonial cells of the testis (where DNA is present) suggests that there are limited epitopes for 1H6 antibody staining and may indicate the presence of highly expressed G4 unfolding proteins. Additionally, the global condensation of chromatin in spermatogonial cells may prohibit the access of the 1H6 antibody to G4 DNA.
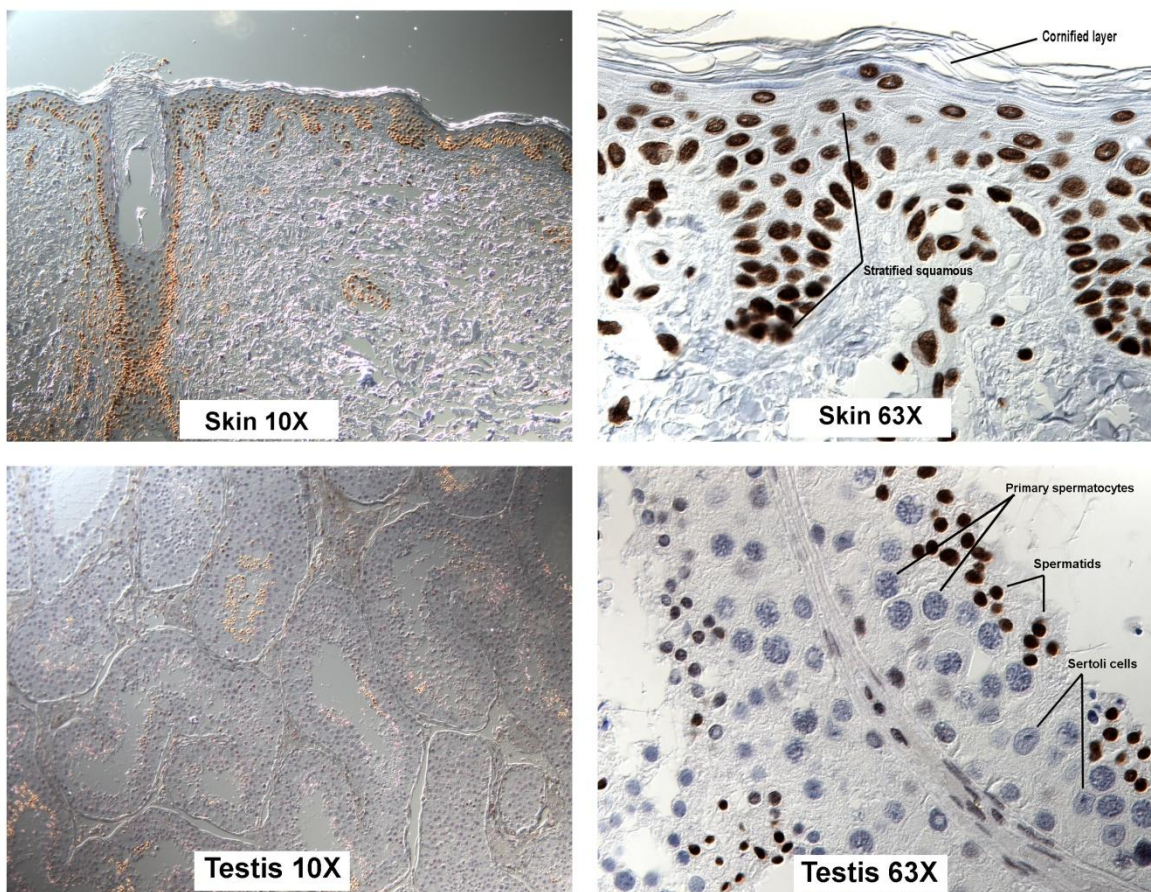
Figure 34. Histological preparations of normal human tissues confirms staining of all nucleated cells except a subset of cells found in the testis.

Normal human tissues were fixed and embedded in paraffin prior to staining with anti-G-quadruplex antibody1H6.   All nucleated cells stain positive (dark-brown color) while non-nucleated (red blood cells) and some cell populations of the testis are not detected.

3.4   Discussion

Although numerous studies have implied the biological existence of G-quadruplex nucleic acids in vertebrate systems no group to date has conclusively demonstrated the presence of this type of alternatively folded structure. Many researchers continue to explore the vast experimental analysis of folded guanine-rich oligonucleotides but have yet to prove that any of these structures occur in the biological context of human cells.

We have generated several antibodies that can target G4 nucleic acids over single-stranded and double-stranded structures (Chapter 2). Remarkably one of these antibodies, 1H6 appears to bind at discrete nuclear locations (Figures 21, 22 and 23).

Using fluorescence microscopy, it is clear that incubation of the 1H6 antibody shows distinct nuclear foci above that of isotype controls (Figure 21). All other antibodies generated in this thesis did not show signal above isotype controls (data not shown). Although characterization experiments suggest that other antibodies generated in this thesis may be more specific and bind G4 DNA with higher affinities, specific antigenic targets for these antibodies may exist at such low levels they may not be detectable without signal amplification. We believe that the promiscuous nature of the 1H6 antibody binding across different G4 structures allows for the recognition of several topologically different G4 structures. Other G4 specific antibodies may be highly specialized binders for only a single G4 structure and as such bind at such low abundance in the cell that detection thresholds are never reached. Additionally, the 1H6 antibody could be detecting a common epitope in cells.

Primary incubations with the 1H6 antibody reveal numerous isolated nuclear foci across telomerase positive, ALT positive, and non-cancerous primary cell lines. The *in vivo* antigen being detected is therefore not a primary cell line specific phenomenon or an ALT/telomerase positive cell line phenomenon. Also, after three-dimensional reconstruction, the number of nuclear foci represent more than the number of human telomeres present. In turn, this suggests that these antigens are not only present at a few telomeres but located throughout the chromosome. This data is in accordance with bioinformatic studies of G4 DNA in the human genome (44, 45). It has been previously reported that thousands of putative G4 DNA forming sequences exist throughout the human genome and likely that

these putative G4 DNA forming sequences fold into multiple G4 DNA conformational

isomers and thus require a less specific probe for identification. The antibody designated

1H6 may profit from binding multiple conformational isomers of G4 DNA and subsequently

be detected above background.

During our initial examination of 1H6 antibody localization, we noticed that most of

the signal detected was localized to the nucleus of all cells. We have seen a proportion of

preparations retaining some staining in the cytoplasm (Appendix B10). However, there is

significantly less signal in the cytoplasm and this staining appeared to be dependent on

individual experimental preparation and the exposure time during microscopy. Faint signal

detected in the cytoplasm were also detected in the isotype controls and may indicate non-

specific binding. If 1H6 is detecting G4 structures it is likely that a high signal concentration

would be detected in the nucleus as both RNA and DNA molecules amass therein. However,

signal detected outside the nucleus could be the result of RNA G4 structures that are located

in the cytoplasm. Interestingly, we often observed two different nuclear staining patterns

which were mutually exclusive, either staining concentrated inside the nucleoli or absence of

staining in the nucleoli (Figure 21). Increased nucleoli staining could arise during different

stages of the cell cycle where increased levels of RNA (with a propensity to form G4 RNA)

are built up in these centers. On the other hand we did see significantly more cells stained

with the latter pattern where no staining was observed in the nucleoli. To further examine

the nuclear foci detected by this antibody we attempted to eliminate the possibility that G4

RNA structures were present. Incubation with RNAses did not significantly alter the nuclear

foci detected (Figure 26 and Appendix B3). Given that G4 structures tend to be resistant to

cleavage by many types of nucleases it is likely that RNA G4 structures may be resistant to

RNAse enzymes tested in our experiments. Nevertheless, in addition to the competition

ELISA experiment in Chapter 2 involving a tetramolecular G4 RNA structure these

experiments suggest that the 1H6 antibody does not detect G4 RNA. These experiments do

not preclude the possibility that G4 RNA structures exist but suggest G4 RNA may not

represent a significant number of foci detected in our assay. Future experiments looking at

nucleolar proteins and RNA staining dyes may help elucidate these findings further.

Additionally, incubation with DNAse enzymes that could be shown to digest G4 DNA could

be used to help validate the 1H6 antibody binding to G4 DNA. Moreover, we cannot

completely eliminate the possibility that the 1H6 antibody is binding a protein epitope. We

might test this issue further by exploring western blots of nuclear and cytoplasmic cell

fractions. Quantification of non-specific binding of proteins by an isotype control in

comparison to the 1H6 antibody could help elucidate the different binding preferences of the

1H6 antibody to protein vs. DNA.

Experimental fixation techniques were altered to explore the possibility that fixation

was somehow influencing staining patterns. Although mean fluorescence signals were lower

in precipitating fixatives (such as methanol, see appendix B1), we did not observe significant

localization pattern changes often observed using antibodies to proteins. Since we observed

similar staining patterns in unfixed (Appendix B2), methanol fixed (Figure 24) and

formaldehyde fixed cells (Figure 24) we propose that the epitope targeted by 1H6 is most

likely not of protein origin. The greater part of all antibodies available today are targeted to

protein components and do not work well across different fixative due to the modification of

the epitopes by different fixation techniques. 1H6 antibody detection across a spectrum of

fixatives suggests that the epitope is not significantly modified during fixation perhaps in

agreement with nucleic acids as its origin. Moreover a onetime experimental over digestion

of paraformaldehyde fixed cells with proteinase K resulted in 1H6 foci detected on the DNA

fibers of exposed cells (Appendix B.11).   In addition, pepsin digestion of metaphase

preparations also resulted in the detection of 1H6 antibody foci (Figure 33). Further

experimentation is required however, taken all together it is unlikely that the epitope

recognized by the 1H6 antibody is of protein origin.

Once stable detection of 1H6 was established we sought to explore methods that

would modify the putative structures such that we could in turn detect changes in 1H6

antibody foci. Initial experiments incubating 1H6 in the presence of synthetic G4 DNA was

shown to completely abolish the detection of foci in cells (Figure 26 and 27). Although not

conclusive evidence for the presence of naturally occurring G4 DNA in cells this experiment

indicates that the presence of synthetic G4 can out compete the 1H6 epitope present in cells.

In parallel experiments we examine the converse effect when we incubated cells with

different drugs thought to stabilize G4 structures. In the presence of TMPyP4 or NMM

(Figure 28 and 29) increasing the concentration of these drugs increased the mean

fluorescence of the cell populations detected with the 1H6 antibody. Using other G4

stabilizing drugs such as PIPER and telomestatin produce similar results (Appendix B4 and

B5). However, in our hands, PIPER did not consistently generate an increased mean

fluorescent signal and was only observed to produce an increase in fluorescent signal

detected in a subset of cells in the total population (Appendix B4). Several of these

experiments reported here were also performed by Dr. Brosh from the NIH. His group also

found that increasing the exposure time to G4 stabilizing drugs increased the number of 1H6

foci present overtime (although future experiments will be required to quantify this observed

finding) at a fixed of antibody concentration (Appendix B6). Many of the drugs used here might be locking G4 structures into a constant fold such that helicase proteins cannot unwind them resulting in accumulation of G4 structures. The dynamic equilibrium that might otherwise persist with the folding and unfolding of these structures may be pushed to stabilization of folded forms. The increased 1H6 antibody staining could result from the accumulation of stabilized G4 structures over time.

After the visualization of telomeric G4 structures in ciliates was reported by Schaeffer *et al*., this group reported that two telomeric proteins TEBPα and TEBPβ were required for the proper folding of telomeres into G4 structures (109). While the telomeres of ciliates are noticeably less complex than the telomeres of vertebrates we sought to examine if any of the 1H6 foci were co-localizing with any telomere specific proteins. The fluorescence microscopy experiments involving the co-localization of telomeric proteins and the 1H6 antibody was problematic and lacked statistical power because of the sheer number of 1H6 foci. We did not see clear co-localization with telomeric proteins using our staining procedures and microscopy instruments. However, our research partner Dr. Brosh did indicate co-localization of 1H6 antibody foci and 15-27% of 53bp1 DNA damage foci upon treatment with telomestatin (Appendix B8). Based on these observations further co-localization experiments are warranted. While co-localization of the WRN helicase was observed in atypical fibers joining cells (Figure 31), Dr. Brosh showed that cells lacking a related helicase protein FANCJ showed increased 1H6 antibody staining when treated with telomestatin (Appendix B7). Both FANCJ and WRN have been shown to unwind G4 structures *in vitro* (62, 63). Decreased levels of helicase proteins specific for the unfolding of G4 structures may allow G4 structures to accumulate overtime and therefore increase the

levels of 1H6 antibody detected. Many human proteins have redundancy such that deletion of one might not completely destroy a cellular mechanism however, patients with helicase mutations have been shown to have increased propensity to develop acute diseases such as cancer (62). Further studies involving different helicase mutant models and 1H6 antibody staining may help expose the molecular associations between G4 structures and clinical disease.

Although traditional fluorescence microscopy experiments did not show significant co-localization of proteins located at the telomere and 1H6 foci we wanted to explore the possibility of 1H6 foci being at the ends of chromosome, with increased clarity. Methanol acetic acid preparations of chromosomes gave detailed information on the localization of 1H6 foci across chromosomes (Figure 32). We did find 1H6 foci located throughout the chromosomes and at the very terminal ends of chromosomes in all metaphase preparations observed (not yet quantified). To further explore co-localization of 1H6 foci and mouse telomeres, we combined immunofluorescent staining of 1H6 with PNA probe binding to telomeres (Figure 33). Similar to the original detection of T-loops at telomeres (100) we did not find 1H6 foci at all telomeres. The absence of 1H6 foci at some telomeres may indicate a dynamic equilibrium between different telomere structures or may possibly an artifact due to coincidental overlap. G4 structures may be detected more readily at the telomeres of mice given that their telomere length can span upwards of 150 kb and therefore have increased guanine sequences capable of folding into many G4 structures for antibody binding. Future quantification of the foci located at the telomeres in human cells and mouse cells is warranted.

In a generalized approach at examining the staining across many human tissues we stained various normal human tissues with the 1H6 antibody. Strikingly, 1H6 staining was widespread throughout nuclei of almost all cell types tested, in line with nucleic acid being the target of 1H6 (Figure 34 and appendix B9). However, a subset of cells located in the testis showed little to no staining. We are uncertain if this was a technical/experimental abnormality or if this subset of cells is lacking G4 structures all together. We would estimate that the latter is not probable and that these experiments should be repeated to formally exclude the possibility of an error.

One of the most significant putative biological roles of G4 DNA structures *in vivo* is based on the findings that quadruplexes can inhibit the function of telomerase *in vitro* (183, 184). In normal stem cells, telomerase normally adds de novo telomere repeats to the ends of chromosomes in order to prevent DNA replication-based telomere shortening and senescence. In most cancers, telomerase is reactivated and contributes to the immortalization of tumor cells. Telomerase-catalyzed addition of telomere repeats to oligonucleotides *in vitro* is prevented by the presence of stable G4 structures, likely preventing access to the oligonucleotide 3' end. Thus, by inhibiting telomerase, G4 DNA may have a role in cellular aging and cancer. To this end, several small-molecules have been developed that bind and stabilize G4 DNA *in vivo*, their mechanism of action is thought to be partially through the inhibition of telomerase (131, 191, 192). The 1H6 antibody described here could be used to investigate this phenomenon and provide direct insight into this proposed mechanism of action.

In guanine-rich regions of the genome, conversion between double-stranded and single-stranded DNA during replication and transcription may be required for the formation

of G4 DNA structures. The formation and resolution of G4 DNA is certainly facilitated by numerous proteins, some of which have been reviewed here (62). Notably, the β subunit of the heterodimeric telomere end-binding protein complex (TEBPα/β) has been shown to facilitate G4 DNA formation at the telomeres of ciliates (185). Conversely helicase proteins such as Sgs1 from yeast (61) resolve these structures. The existence of G4 DNA forming and resolving proteins, suggests that G4 DNA may be an important biological signaling molecule (59). If G4 DNA operates as a signaling molecule, the many G4 DNA isoforms are all potentially unique molecular recognition sites that can be exploited by diverse biological pathways. Monoclonal antibodies specific for other isoforms of G4 DNA will be able to resolve the complex biological signaling of G4 DNA. However, other than the unrestricted binding of 1H6, the monoclonal antibodies generated here are specific for intermolecular G4 DNA. Previous reports have hypothesized that intermolecular G4 structures could exist through the association of 3' overhangs at telomeres of two chromosomes (19). Four homologous sister chromatids may be stabilized by G4 DNA formation, precisely aligning chromatids side by side during meiosis. The antibodies presented here may be used to examine this possibility.

Prior to this study, many different G4 DNA isomers have been assembled *in vitro*, but there was no *in vivo* evidence of their existence in human cells. Our data suggests that *in vivo* G4 DNA may exist in human and mouse cells as detected visually by the binding of the 1H6 antibody. We have provided the first direct evidence that these structures may exist in cells and believe that these novel probes will be invaluable tools for understanding the biological roles of G4 DNA *in vivo*. Subsequent experiments using the 1H6 antibody in immunoprecipitation (IP) experiments may lead to the *in vivo* detection of G4 structures in

human cells. If we could show that the 1H6 antibody can pull-down G-rich DNA sequences in the context of human cells, capable of folding into G4 DNA structures, then future experiments could explore these sequences in detailed experiments. Enriching for PQS may help to reveal the existence of G4 DNA structures *in vivo* through the examination of mutant sequences and interacting protein partners. Immunoprecipitation experiments combined with downstream experiments such as sequencing the IP DNA would provide detailed insight into the biologically relevant PQS sequences forming *in vivo*.

Chapter 4: *In vivo* detection of G-quadruplex nucleic acids and G-quadruplex associated proteins

The fundamental issue concerning the therapeutic potential and biological role of G-quadruplex nucleic acids continues to be the shortage of *in vivo* verification of these structures in mammalian cells. However, given that there is mounting *in vivo* verification of G-quadruplex structures in other organisms (*Stylonychia* and *S.Cerevisiae*) it is to be expected that the verification of G-quadruplex structures in other cells including human cells is not far away. Antibodies to G-quadruplex structures are one way by which the *in vivo* verification of these structures could be accomplished.

Although fluorophore conjugated antibodies help to describe the location of where a particular antigen may be at any one time, antibodies can also be exploited for their high affinity to antigens for examining biological associations. Novel antibodies that may detect G-quadruplex nucleic acids in human cells could in theory be used to probe the mammalian genome for regions that by inference of binding the antibody fold into G4 structures.

In this chapter we describe our attempts to identify genomic sequences bound by the 1H6 antibody *in vivo*. Using a novel technique (G4-ChIP) the 1H6 antibody is used to isolate DNA/protein complexes from human HeLa cells. Through the use of traditional and quantitative PCR methods these DNA/protein complexes are interrogated for the enrichment of putative quadruplex forming sequences. In addition, the isolation of DNA from G4-ChIP experiments is used to create genomic libraries of putative quadruplex forming sequences and interrogated by next-generation sequencing. Finally, proteins bound to G4 structures *in vivo* that have been isolated from G4-ChIP experiments are investigated by mass spectrometry analysis.

124

4.1   Introduction

In vivo detection of an antigen with as much published biological potential as G-quadruplex nucleic acids would lift the research field of G-quadruplex nucleic acids from the fringe into the mainstream. This is not to say that the research field of G-quadruplex nucleic acids is not filled with brilliant and exciting research but rather a reduced significance due in part by the lack of frank evidence for the existence of G-quadruplex nucleic acids in human cells.

Antibodies to different antigens have been exploited throughout the literature to ascertain that a distinct target(s) exists in cellular contexts. Unfortunately, a common misconception regarding antibodies, if an antibody exists therefore the antigen must exist, is not accurate. This misunderstanding cultivates from a lack of understanding for a particular antigen and the mechanisms in which antibodies bind their targets. Although this is not necessarily true, antibodies are valuable tools, which can in addition to other techniques and experimental approaches help elucidate the presence of a particular antigen in cells.

Many groups have developed antibodies and antibody fragments to different structural polymorphs of G4 DNA (108, 171, 172, 176). In spite of this no group has shown any cellular images of antibody binding using these antibodies in human cells. If the anti-G-quadruplex antibody 1H6 is in fact detecting G-quadruplex nucleic acids in cells we should be able to exploit this fact to explore the sequences in the genome possibly forming these structures and any proteins that might be associated to those structures. Although a powerful tool, the 1H6 antibody alone does not prove the existence of G4 DNA.

One of the largest collaborative endeavors in science has been the Human Genome Project (HGP). The HGP is an international collaboration that began in 1989 which primary

goal was to determine the sequence of the 3 billion base pairs that make up human DNA and identity all of the 20-25,000 genes of the human genome, (http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml). The publicly funded HGP has been completed at a cost of 3 billion dollars and took approximately 15 years. During the evolution of the HGP DNA sequencing technologies were rapidly changing from slow expensive procedures to accelerated high-throughput methods. The modernization of DNA sequencing technologies was instrumental in sequencing the human genome.

New techniques for combining antibody specificities and DNA sequencing technologies can be used to investigate the interaction between proteins and DNA in the cell. Chromatin immunoprecipitation (ChIP) has historically employed the use of antibodies specific to DNA binding proteins (such as transcription factors) to purify DNA-protein complexes of interest *in vivo*. Isolated DNA-protein complexes can be further characterized by sequencing technologies to decipher the *in vivo* DNA sequences bound by the proteins.

Identifying the DNA or RNA nucleic acid sequences to which the anti-G-quadruplex antibody 1H6 is binding *in vivo* could elucidate the nature of G4 structures *in vivo*. Our current data demonstrates that 1H6 can binds nuclear DNA. However, it is not known whether or not the nucleic acid sequences that are bound by 1H6 in cells are guanine-rich and fold into G-quadruplex structures nor the nature of any proteins that might be associated with these sequences.

Modification of the traditional ChIP method can be used to probe the *in vivo* existence of G-quadruplex nucleic acids using antibodies specific to G-quadruplex structures. We modified ChIP based methodology to target nucleic acids in order to investigate the *in vivo* existence of naturally occurring G-quadruplex nucleic acids and to define their genomic

distribution in cells. To our knowledge, no scientific group has used structure specific anti-nucleic acid antibodies in this manner.

Combining G4-ChIP with PCR technologies we have set out to validate the 1H6 antibody binding to putative quadruplex forming regions in the genome. We would expect that G4-ChIP eluted DNA is enriched for putative quadruplex forming sequences, and therefore should find discrepancies between the DNA amplified by primers designed to genomic location capable of forming G4 structures and the DNA amplified by primers to regions devoid of putative quadruplex sequences. In this regard we have combined G4-ChIP with traditional PCR and quantitative PCR based methods to validate the *in vivo* binding of putative quadruplex sequences by 1H6.

We have endeavored to describe the *in vivo* genomic location of putative quadruplex forming sequences by combining next-generation sequencing with our G4-ChIP assays. We isolated DNA from 1H6 antibody immunoprecipitations, built genomic libraries of putative quadruplex regions and used Illumina based sequencing methods to decipher these sequences and mapped them back to the referenced human genome.

In addition, antibodies that bind G4 structures *in vivo* could also be utilized to investigate any associated proteins bound to G4 structures. Other scientific groups have done traditional ChIP based assays to isolate proteins that have been proposed to be associated with G4 structures *in vivo* (193-195). Conversely, we have attempted to directly isolate putative quadruplex forming regions by interaction with the 1H6 antibody. Proteins associated with these nucleic acid structures have been isolated from G4-ChIP experiments and probed by mass spectrometry.

In this Chapter we employ modified ChIP techniques using the anti-G-quadruplex antibody 1H6 to isolate *in vivo* G-quadruplex DNA from human HeLa cells. Following successful isolation, DNA was purified and manipulated into genomic libraries suitable for next-generation DNA sequencing. In addition to DNA sequencing, modified ChIP techniques using the 1H6 antibody were employed to isolate G4 nucleic acid associated proteins. Unknown proteins were analyzed by mass spectrometry. Our first experiments using these techniques are discussed.

4.2   Experimental methodology

4.2.1   G-quadruplex chromatin immunoprecipitation (G4-ChIP)

HeLa cells were grown to 80% confluence under standard conditions. Cells were then lifted from tissue culture plates by treatment with trypsin-EDTA (0.25%) for 5 minutes and pelleted by centrifugation for 5 minutes at 1,200 rpm. Cells were pooled into groups of 2×10^7 total cells in 10 ml of PBS prior to fixation. 270 µl of 37% formaldehyde solution (1% final) was added to pooled cells and incubated for 10 minutes at room temperature rocking. The fixation was immediately quenched with the addition of 500 µl of 2.5 M glycine pH 8.0 (0.125 M final) and allowed to incubate for 5 minutes rocking. Fixed cells were then pelleted at 1,200 rpm for 5 minutes at 4 °C and washed 3X with freshly prepared, pre-chilled PBS and phenylmethylsulfonyl fluoride (PMSF) (500 µM final). Washed and fixed cell pellets subsequently resuspended in 2 ml of lysis buffer (50 mM Tris-HCl pH 8.0, 1% sodium dodecyl sulfate (SDS), 10 mM ethylenediaminetetraacetic acid (EDTA), 1X protease inhibitor cocktail (PIC) and 1X PMSF). Lysis took place on ice for 30 minutes with

occasional inversion. Cells were subsequently sonicated to shear chromatin into fragments less than 1 kb. 2 ml samples were either sonicated using a Biologics ultrasonic homogenizer model 150 V/T with 30 rounds of 12X 0.4 second bursts at ~35% power with 1 minute rest between rounds with a 3.9 mm microtip or using a Misonix sonicator 3000 using a power setting 8 watts. During sonication all samples were maintained in ice-cold water and in between rounds samples were vigorous mixed. After sonication sheared chromatin was cleared by centrifugation at 13,000 rpm at 4 °C for 10 minutes. Soluble chromatin was separated into fresh tubes and diluted 1:10 with IP buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 1% triton X-100, 0.1% sodium deocycholate, 1X PIC and 1X PMSF). 10 µg of 1H6 antibody, isotype control (IgG$_{2b}$) or positive control anti-histone H3 (Ab cam ChIP grade) was mixed with chromatin and rotated over night at 4 °C. Antibody-chromatin complexes were captured using protein G Dynal beads by first incubating 50 µl of magnetic beads with antibody-chromatin complexes for 90 minutes rolling at 4 °C. Captured complexes were subsequently washed 3X with cold PBST. Chromatin was eluted from magnetic beads by the addition of 100 µl of freshly made elution buffer (100 mM Na$_2$CO$_3$, 1% SDS) and 5 µg of RNAse A. All samples were eluted, RNAse digested and reverse cross-linked at 65 °C for 8-14 hrs. After elution samples were digested with 20 µg of proteinase K at 37 °C for 90 minutes. DNA was then purified using Qiaquick PCR purification kit and resuspended in 20-30 µl of EB buffer (10 mM Tris-HCl pH 8.5)

4.2.2   ChIP-PCR for G4 promoters

10 ng of purified DNA from G4-ChIP was used as template input into traditional PCR reactions containing: 200 µM dNTPs, 0.5 µM forward and reverse primers, 3% DMSO, 1X

Phusion GC buffer and 0.02 U/µl Phusion Hot Start polymerase. Genomic primers used for

conventional PCR were *Myc* 5′-CGGAGATTAGCGAGAGAGGA-3′ and 5′-

GAGCAGCAGAGAAAGGGAGA-3′ which gave a 239 bp product, *c-Kit* 5′-

GGCATTAACACGTCGAAAGA-3′ and 5′-GATCCGAGCTCTGGTCCAC-3′ which gave a

219 bp product, and *VEGF* 5′-TTTTCAGGCTGTGAACCTTG-3′ and 5′-

GATCCTCCCCGCTACCAG-3′ giving a 227 bp product. PCR was performed as follows:

95 °C for 1 minute followed by 32 cycles of 95 °C for 30 seconds, 63 °C for 40 seconds, 70

°C for 90 seconds. Amplification products were separated on a 2% agarose gel in TBE and

visualized using ethidium bromide under UV exposure on a Gel-DOC apparatus. Semi-

quantitative real-time-PCR was performed following following standard protocols. An ABI

7900h real-time PCR system (Applied Biosystems) and the QuantiFast SYBR Green

supermix (Quiagen) were used for qPCR experimental setup and analysis. 10 ng of purified

DNA from G4-ChIP experiments was used as template for qPCR reactions and cycling

conditions as set by the manufacturer were: 95 °C for 5 minutes followed by 40 rounds of 95

°C for 10 seconds and 60 °C for 30 seconds. Genomic primers used for qPCR are listed in

Table 6 below.

Sequence 5' to 3'	Length	Target
AGCTCCTCCTTGCTTCACAG	20	Kit NON F
GTGAGGACCCAGGTTGATGT	20	Kit NON R
CTGGAGCAGGAACTCGAATC	20	Rb NON F
GGACTCACCAGAAGCTGAGG	20	Rb NON R
TGCTTGAGACAGTGACACCA	20	GAPDH NON F
TCTTGGAAAAGCCTGGAGAA	20	GAPDH NON R
TCTGCCTAGGCGTTCAACTT	20	P53 G4 F
TCTCGGCTCCGTGTATTTTC	20	P53 G4 R
GCCAGGCTGTAAATGTCACC	20	GAPDH G4 F
GGGAGCACAGGTAAGTGCAT	20	GAPDH G4 R
AAAGAGCAGGGGCCAGAC	18	KIT G4 F
GGTCCACGTTCCAGCTCTC	19	KIT G4 R
GGTTTTTGAGGGTGAGGGTGAGGGTGAGGGTGAGGGT	37	TELO F
TCCCGACTATCCCTATCCCTATCCCTATCCCTATCC	36	TELO R

 **NON indicates products over 1000 bp from any putative quadruplex forming sequence

Table 6. qPCR primers used in this study.

ChIP-qPCR analysis was performed following standard fold enrichment protocols where data have been normalized relative to the mock antibody. Sometimes referred to as the $2^{-\Delta\Delta Ct}$ method.

4.2.3   DNA size selection

Purified DNA from G4-ChIP underwent size selection using pre-cast Novex 8% TBE gels. Samples were diluted in Novex Hi-Density TBE sample buffer and loaded along with 10 µl of 100 bp DNA ladder (10 ng/µl) as a reference. TBE gels were run at 200V for ~35 minutes. Gels were post stained with Sybr Green I for 1 min to visualize reference ladder and corresponding smear for excision. Gel fragments used for ChIP-PCR were excised from the 50 bp to 800 bp range whereas, gel fragments used for building libraries for sequencing were excised from the 100 bp to 300 bp range. Gel slices were sheared by placing gel fragments into a 0.5 ml microsentrifuge tube with a pin hole in the bottom of the tube, in a 2 ml microsentrifuge tube. Samples were centrifuged at 12,000 rpm at room temperature for 3

minutes. Sheared gel slices were eluted over-night at 4 °C from gel fragments with 200 µl of

5:1 elution buffer (10 mM Tris-HCl, 0.1 EDTA):7.5 M Ammonium acetate). After

incubation at 65 °C for 15 minutes gel slurries were centrifuged through Spin-X filter tubes

at 12,000 rpm at 4 °C for 3 minutes, to remove gel fragments. Following Spin-X filtration

DNA was purified and concentrated following standard ethanol precipitation.

4.2.4   Library construction for next generation sequencing

G4-ChIP samples that had undergone size selection (100 bp-300 bp) and purification were

used as input into library construction for next generation sequencing using standard

procedures. In brief, DNA end repair and phosphorylation was done by incubating 40 µl of

each sample in the presence of 1X phosphorylation buffer, 1 mM dNTPs, 0.5 µl T4 DNA

polymerase, 0.1 µl of Klenow DNA polymerase and 0.5 µl of T4 PNK for 30 minutes at

room temperature. After incubation was complete, samples were phenol:chloroform

extracted in MaxTract tubes and precipitated by ethanol using standard procedures. A-tailing

was performed by combining purified samples with A-tailing master mix (6.6X NEB buffer

2, 1.7 mM dATP and 1 µl of Klenow fragment (3' to 5' exonuclease) and incubating at 37 °C

for 30 minutes. After incubation was complete, samples were phenol:chloroform extracted in

MaxTract tubes and precipitated by ethanol using standard procedures. Forked PCR adapters

were subsequently ligated to each sample by incubation of samples with adaptor ligation mix

(1X Quick ligase buffer and 1 µl of PE adapters, in the presence of 4 µl of Quick Ligase) for

15 minutes at room temperature. Adapter ligated DNA samples were purified using

AmpureXP magnetic beads following manufacturer's instructions. Purified adapter ligated

DNA libraries were amplified using separate indexing primers for each sample to be pooled

in the library. 13.5 µl of PCR master mix (2X Phusion HF master mix, 1 µM primer PE 1.0

and 1 µM PE 2.0 indexed primer) was added to 10.5 µl of adaptor-ligated template.

Libraries were amplified following a PCR program of 98 ºC for 30 sec followed by 15 cycles

of 98 ºC for 10 seconds, 65 ºC for 30 seconds and 72 ºC for 30 seconds with a final extension

of 72 ºC for 5 minutes. PCR amplified libraries are finally cleaned up using size selection of

AmpureXP magnetic beads and manufacturer's instructions. Prior to sequencing, libraries

are analyzed by Agilent high sensitivity chips.

4.2.5 ChIP-sequencing and bioinformatic analysis

G4-ChIP libraries were sequenced using Illumina sequencing technology powered by clonal

array formation and proprietary reversible terminator technology as outlined at

http://www.illumina.com/ . Paired-end sequencing was performed at the Genome Sciences

Centre using a HiSeq 2000 analyzer. Illumina data collection software was used to generate

files of sequence reads. Reads were aligned to the human genome build 18. The subsequent

BAM files created contained IP sequences aligned to hg18 and were compared to input

control sequences by first converting BAM files to BED files using BEDTools (available at

http://code.google.com/p/bedtools/). BED files and BAM files were subsequently analyzed

using ChIP-Seq software Cistrome accessible at: http://cistrome.dfci.harvard.edu/ . Quality

control was computed using summary statistics on each dataset and subsequently peaks were

called using MACS two sample peak calling with the SeqPeak algorithm in Cisgenome

(model-based analysis for ChIP-Seq data). Input G4 ChIP peaks were compared to 1H6

antibody peaks in this way to generate a called peaks file for quality control. Threshold

method using (P-values) was assigned using the default parameters in the MAC algorithm.

4.2.6   ChIP-mass spectrometry and analysis

Chromatin immunoprecipitation was carried out as above with the following modifications: lysis buffer was changed to (50 mM Tris-HCl pH 8.0, 0.5% nonyl phenoxypolyethoxylethanol (NP-40), 0.1% sodium deoxycholate, 10 mM EDTA, 150 mM NaCl, 1X PIC and 1X PMSF) and IP buffer was changed to (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 1X PIC and 1X PMSF). Dilution of chromatin in IP buffer was modified to 1:5 and elution buffer was changed to 100 mM Glycine pH 2.9. After cross-link reversal the pH was adjusted to 7.5 with the addition of 20 µl of 1 M Tris-HCl pH 7.5 and did not undergo proteinase K digestion. DNA was digested with 15 µg of DNAse 1 in DNAse 1 buffer (20 mM Tris-HCl pH 7.4, 2.5 mM $MgCl_2$ and 0.1 mM $CaCl_2$) for 45 minutes at 37 °C. DNAse 1 was arrested by the addition of 20 mM EGTA and protein samples were concentrated by centrifugation at 37 °C in a SpeedVac centrifuge (~1 hour). Proteins were then run through a Nupage Bis-Tris 12% SDS-PAGE gel using MES running buffer to separate proteins prior to mass spectrometry. Proteins were visualized using SimplyBlue safe stain containing Coomassie G-250. Bands of interest were excised and in-gel trypsin, alkylation, reduction were done on a ProGest digestion robot (Digilab Genomic Solutions). Mass spectrometry was performed using a 4000 QTrap mass spectrometer (Applied Biosystems/Sciex). The MS/MS spectra from the gel slices for each lane were concatenated and searched against the Uniprot human database using the Mascot (Matrix Science) search engine.

4.3 Results

4.3.1 1H6 antibody isolates DNA quantities in excess over isotype controls in a modified chromatin immunoprecipitation experiment with human cells

Chromatin immunoprecipitation (ChIP) has been traditionally used to isolate *in vivo* DNA sequences bound by proteins targeted by antibody and antibody fragments. We have modified the standard anti-protein ChIP procedure to exploit the targeting ability of the 1H6 antibody to G4 nucleic acids. To our knowledge the only other group to employ a similar strategy based on an anti-DNA ChIP procedure was created by Weber *et al.*, termed MeDIP (methylated DNA immunoprecipitation) whereby anti-5-methylcytosine antibodies were exploited for the immunoprecipitation of denatured methylated DNA (201). In our novel approach (termed G4-ChIP) we have used the anti-G-quadruplex antibody 1H6 to isolate G4 nucleic acids from human cells. We believe that the novel approach of directly binding of G4 structures using this antibody can provide insight into the proposed existence of these structures *in vivo*. Our first question regarding G4-ChIP was whether or not the 1H6 antibody could in fact pull-down DNA in the context of chromatin in excess of the non-specific pull-down of a matched isotype control. Following immunoprecipitation reactions with the 1H6 antibody or an IgG_{2b} control we compared the quantity of DNA recovered. In our hands, the total amount of DNA isolated using the 1H6 antibody was consistently greater than the amount of DNA isolated from the negative control antibody. Figure 35 illustrates the absolute quantities of DNA isolated using the anti-G-quadruplex antibody 1H6 from triplicate G4-ChIP experiments in comparison to an isotype control. While this suggests that the 1H6 antibody targets a chromatin component, these experiments cannot determine if the 1H6 antibody is specific for any particular DNA, RNA or protein component of chromatin.

G4-ChIP absolute DNA pull-down
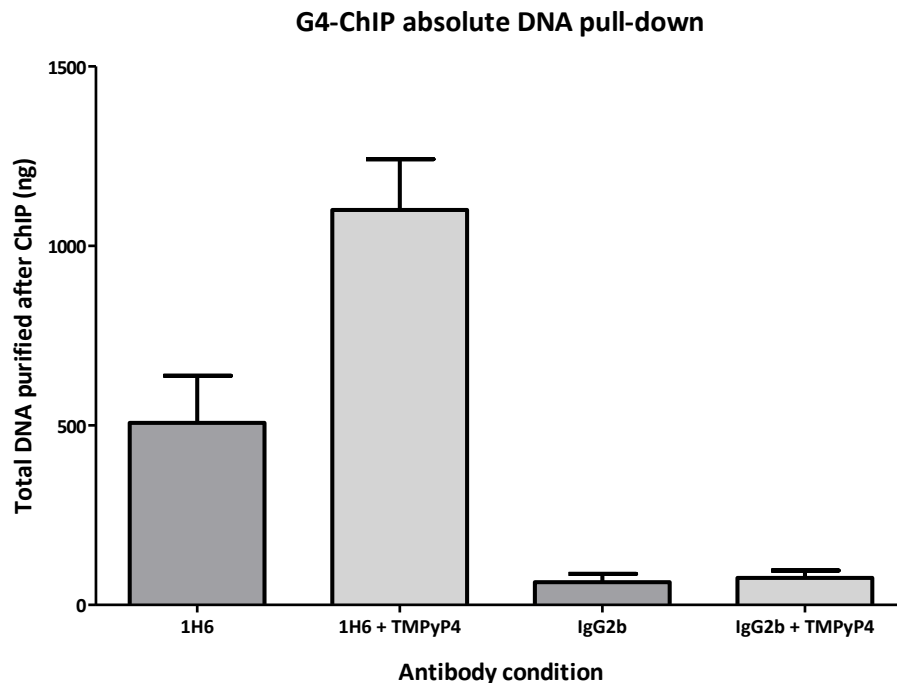
Figure 35. Anti-G-quadruplex antibody 1H6 pulls-down DNA in excess of isotype controls.
A novel technique to isolate G4 DNA structures from human cells (G4-ChIP) was employed
to quantify the amount of DNA bound by the 1H6 antibody in comparison to an isotype
control. Dark grey bars represent DNA eluted from G4-ChIP while light-grey bars represent
TMPyP4 treated G4-ChIP. Error bars represent standard deviation of total DNA between
triplicate experiments. In this assay the 1H6 antibody bound more DNA than the IgG_{2b}
control. Treating the cells with TMPyP4 increased the amount of DNA bound by 1H6 by a
factor of ~2.

To test whether stabilization of dynamic G4 structures in cells would affect the

quantity of DNA isolated in the G4-ChIP experiments we treated HeLa cells prior to fixation

with 20 μM TMPyP4 for 12-16 hours prior to G4-ChIP. Analogous to our fluorescence

microscopy experiments in Chapter 3, treating cells with TMPyP4 prior to G4-ChIP resulted

in an increase in the absolute amount of DNA isolated (Figure 35). This enrichment of DNA

was not recapitulated in the isotype control. This suggests that if TMPyP4 is in fact

stabilizing dynamic G4 structures in cells, then conceivably the stabilization allows for 1H6 antibody binding more targets and therefore a greater DNA yield after immunoprecipitation. If the effects of TMPyP4 were specifically targeted to G4 DNA then we could conclusively say that the increased quantity of DNA purified after immunoprecipitation is directly linked to the stabilization of dynamic G4 structures and the direct binding of the 1H6 antibody to these molecules. However, in reality there are several unknown effects of TMPyP4 and therefore we do not know what component of chromatin (DNA, RNA or protein) that the 1H6 antibody is binding after treatment with TMPyP4.

4.3.2   Conventional PCR suggests enrichment of select quadruplex forming promoters in immunoprecipitated 1H6 antibody fraction over isotype control

While DNA isolation in excess of an isotype control suggests that the 1H6 antibody is binding more DNA than the isotype control it does not disclose any information regarding the sequence of DNA being eluted. To test if the eluted DNA after G4-ChIP with 1H6 is in fact enriching for putative G4 DNA sequences we attempted to examine putative G4 promoter regions in the genome. Several groups have proposed that putative G4 sequences in the genome may be located in the promoter regions of critical genes including: *myc, c-kit, VEGF* and *bcl-2* (224, 264, 265, 303) as these sequences can form G4 structures *in vitro*. To explore the possibility that putative G4 sequences were being enriched in our G4-ChIP we designed primers to regions of the genome with putative G4 potential. Proceeding with conventional PCR we utilized eluted DNA from G4-ChIP as input template for reactions with primers to putative quadruplex sequences (PQS) in the human genome. Initial experiments included a negative control IgG antibody, an input control and a positive control

anti-histone 3 (H3) antibody for comparison to the 1H6 antibody. Figure 36 demonstrates that there appears to be an enrichment of these PQS in the 1H6 immunoprecipitation reactions over the isotype control. After PCR and agarose gel electrophoresis the visualization of bands from different genes of interest containing PQS regions appears to be present in the 1H6 antibody G4-ChIP fractions, the input fraction and the anti-H3 fraction but not in the isotype control. Not surprisingly, the IgG isotype control PCR fragments are barely detectable by gel electrophoresis (Lanes 10, 11, 12) and given the small amount of DNA eluted in the G4-ChIP with this isotype control the amplification of any product is limited in part by starting template quantity. Nevertheless, PCR fragments containing guanine-rich regions capable of forming G4 DNA were amplified from the eluted DNA G4-ChIP assays using the 1H6 antibody. This proof of principle experiment suggests that eluted DNA from 1H6 immunoprecipitation reactions do contain some guanine-rich sequences. The primer used for the promoter region of *myc* did not produce a detectable band in the input or anti-histone H3 (positive control for ChIP conditions) lanes and as such may indicate a problem with the primer set or relatively few *myc* DNA sequence fragments after immunoprecipitation with anti-histone H3 (Lanes 1 and 4). This well characterized quadruplex forming region was detected in the 1H6 immunoprecipitation and may indicate enrichment after ChIP (Lane 7). However, as previously shown by anti-Z-DNA antibodies (175) the 1H6 antibody could be inducing a G4 structure in these promoter regions that in turn are enriched during IP. Alternatively, we cannot eliminate the possibility that these results can be explained by random variation observed in IP output. We did observe some variation in these experiments that have not yet been fully explained. However, future

experiments examining the reproducibility of these results may eliminate the possibility that the observations were due to random coincidence.
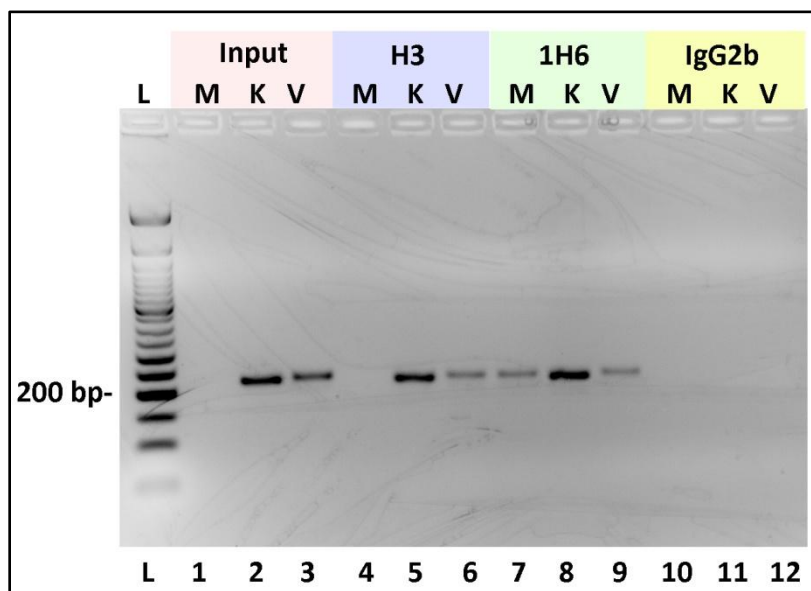
Figure 36. Eluted G4-ChIP DNA sequences examined by conventional PCR show possible enrichment of potential G4 forming sequences in 1H6 antibody immunoprecipitation.
Conventional PCR of potential G4 forming sequences was completed to test the eluted DNA fractions after G4-ChIP. M, K and V bands represent guanine-rich promoters of human *Myc*, *Kit* and *VEGF* respectively.

4.3.3 Quantitative real-time PCR of 1H6 antibody ChIP does not provide evidence for enrichment of predicted G4 regions

Conventional PCR does not easily lend to the quantification of PCR products for comparative analysis between G4-ChIP templates. Based on the possibility that the 1H6 antibody was enriching for guanine-rich genomic sequences with G4 folding potential over non-specific DNA we sought to compare enrichment between samples. In this sense we switched to the real-time quantification of PCR products by genomic qPCR. It is fairly simple to find putative G4 forming sequences throughout the human genome with user

friendly free software packages including Quadfinder and Quadparser (accessible at

http://miracle.igib.res.in/quadfinder/ and http://www.quadruplex.org/?view=quadparser_web

respectively). Using Quadparser (default parameters) and assisted by literature searches we

designed qPCR primers for potential G4 forming sequences throughout the genome.

Additionally, we chose genes that had been well characterized to form G4 structures *in vitro*

and some that have not been characterized to fold into G4 structures (as of yet) but contain

PQS (i.e. *Gapdh* G4). In an attempt to determine the amount of non-specific binding to DNA

in general we also designed primers for regions that did not screen as putative G4 forming

sequences (on the template or reverse template strand) that were at least 1000 bp away from a

putative G4 forming sequence. This distance allowed us to exclude ChIPed DNA that may

be in close proximity to our negative control sequences by selecting all of the

immunoprecipitated DNA below 1000 bp as templates into qPCR reactions.

Interestingly, qPCR reactions that contained size fractionated immunoprecipitated

material from G4-ChIP experiments did not result in significant enrichment of predicted

quadruplex forming sequences over non-quadruplex forming sequences (Figure 37). We

expected to see the enrichment of genomic sequences containing predicted G4 sequences

several fold over IgG, relative to genomic sequences containing no predicted G4 sequences.

Figure 37 illustrates that one of the non-G4 forming regions from the retinoblastoma gene

(indicated by Rb NON) was also enriched in our 1H6 G4-ChIP assays over the isotype

control. Tumor suppressor genes akin to *Rb* and *p53* have been reported to have less PQS

than oncogenes such as *kit* (46). Kit G4 was heavily enriched over IgG after

immunoprecipitation with 1H6 however Rb NON which did not contain any PQS was

enriched as well. This result suggests that the 1H6 antibody may be binding DNA in a

manner that isn't exclusively dependent on G4. Furthermore we know from the specificity ELISAs in Chapter 2 that the 1H6 antibody binds different G4 structures with different affinities. The guanine-rich sequences found in the genes tested here are not the same and therefore may not fold into the exact same G4 structure. The different levels of enrichment we observed for each gene may also be explained by different G4 structures and the affinity of the 1H6 antibody to those structures.

In addition, while at least one G4 forming region was enriched in the 1H6 immunoprecipitations there was large variability between specific G4 regions. This is not unexpected as some *in vivo* G4 structures may be more stable than others and therefore be enriched in our assay using the 1H6 antibody. Furthermore no enrichment of telomeres was detected after immunoprecipitation with 1H6. However, the amplification of telomeric DNA in both the IgG control and the 1H6 antibody immunoprecipitated template was very quick (low cycle threshold value). This suggests that both IgG and 1H6 IP samples contain high amounts of telomeric DNA sequences. As such there may not be enough specificity in our primer sets to distinguish between the two samples or may indicate difficulty associated with quantitative-PCR for highly repetitive regions on the genome. Furthermore, the chromatin state of telomeric DNA may impair 1H6 antibody binding as similarly observed in our *in situ* staining of spermatogonial cells (Figure 34, Chapter 3).

Our results in qPCR may also be unclear due to fundamental issues regarding PCR through repetitive regions of high guanine-richness. PCR guidelines indicate that high GC richness and an ability to form DNA secondary structures tends to impede the abilities of most polymerase enzymes. While we have attempted to compensate for such issues by modifying our PCR conditions (such as including DMSO in the reactions to abrogate

secondary structure formation) we do not know to what extent different types of secondary structures in these PQS regions influence our PCR reactions. Based on these experiments there is no direct evidence for the selective IP of predicted G4 sequences following IP with the 1H6 antibody.

Currently, there are no positive controls for the binding of the 1H6 antibody to G4 DNA structures in the context of chromatin during IP. Additionally, there are no positive controls for the efficient amplification of G4 forming sequences during PCR. In our future experiments will have to consider these difficulties and explore different techniques that may help clarify our lack of positive controls.

qPCR validation of 1H6 antibody
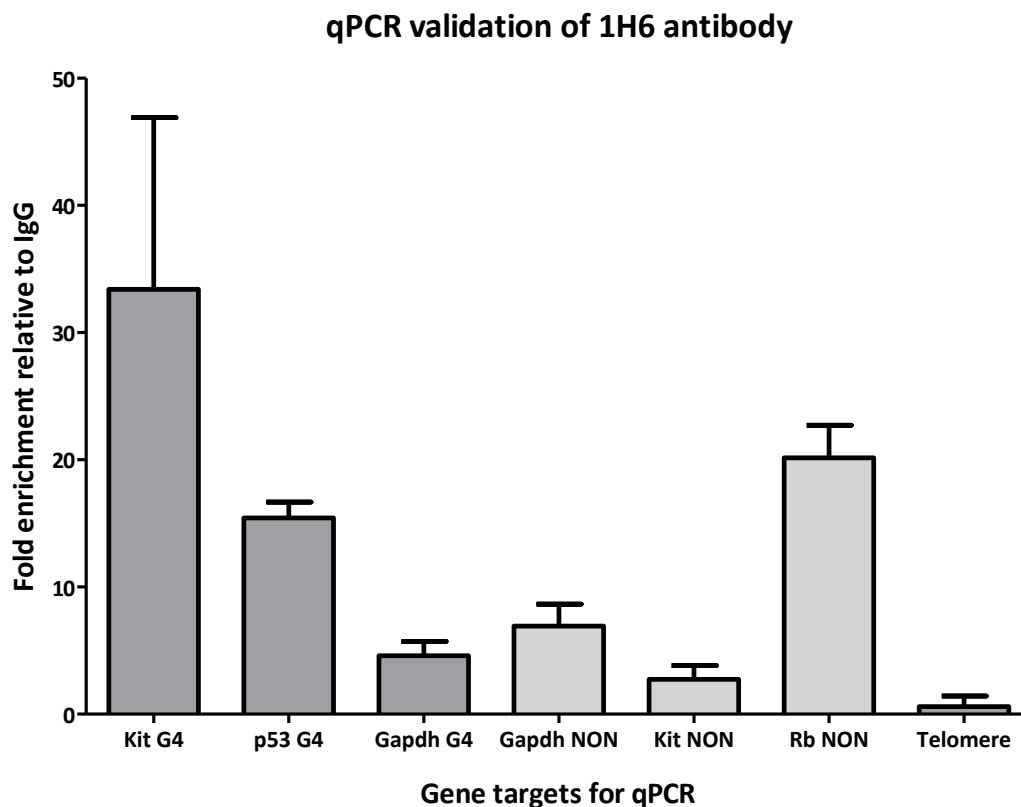
Figure 37. qPCR does not validate the binding of 1H6 antibody to putative quadruplex forming sequences.

qPCR was used to test various genomic sites for enrichment in the 1H6 antibody immunoprecipitated DNA relative to an IgG control. Putative G4 forming regions (indicated by gene name followed by G4) are enriched (kit G4 and p53 G4) in addition to a non-G4 forming sequence (Rb NON). There was no detectable enrichment of telomeric DNA in 1H6 immunoprecipitations.

4.3.4 Agilent profiling of constructed libraries validates high-quality adaptor ligation and size selection technique prior to Illumina sequencing

While qPCR amplification of selected putative G4 forming sequences in the genome might indicate that G4-ChIP with the anti-G-quadruplex antibody 1H6 is enriching for sites with G4 potential, it relies heavily on the selection of appropriate primers and finding

conditions suitable for PCR through GC rich secondary structures. Although the current literature is bursting with papers detailing the location of G4 motifs in selected human DNA sequences, there are no papers describing the absolute existence of any folded motif *in vivo*. In this regard, we don't know which putative G4 forming sequences actually form G4form in the cell.

To examine the actual nucleotide sequences eluted after G4-ChIP we sequenced the eluted DNA by next-generation sequencing technologies. Instead of making educated guesses to which putative G4 forming sequences form *in vivo*, we endeavored to analyze all of the eluted DNA fragments from immunoprecipitation reactions with anti-G-quadruplex antibody 1H6.

Following G4-ChIP with 1H6, eluted DNA was size selected and modified for next-generation sequencing following standard Illumina preparation protocols (See Experimental Methodologies). The 100-300 bp fractions from G4-ChIPs were isolated and four libraries of fragmented DNA were prepared. Library 1 was created from the DNA purified after 1H6 antibody immunoprecipitations while library 2 (a negative control) was generated from the DNA of the input chromatin. In a similar set of experiments we built library 3 from the eluted DNA after G4-ChIP with 1H6 from cells that had been treated with TMPyP4 and library 4 (a negative control) was generated from the DNA from the TMPyP4 treated input chromatin. In this regard, sequence analysis could be conducted by comparing the random DNA fragments expected in either input fraction to the 1H6 antibody specific immunoprecipitations. At this time we are unaware of any positive control antibody that has been shown to bind putative quadruplex forming sequences in the human genome. However, it would be possible to compare the conditions for immunoprecipitation with another DNA-

binding antibody, anti-5-methylcytosine, to the 1H6 antibody to help elucidate immunoprecipitation conditions for 1H6.

Initial testing of the libraries suggested that size selection and amplification had been properly achieved (Figure 38 and 39). After library construction all samples were analyzed using high sensitivity DNA chips on an Agilent bioanalyzer. Computer generated gel electrophoresis images (Figure 38) show a distinct band slightly less than the 300 bp marker. These bands are the direct result of proper adaptor ligation during library construction from the initial size selected range of 100-300 bp. In addition, relative quantification of libraries was achieved by examining the relative fluorescence of each library Figure 39. A faint smear was detected around the 900 bp range of unknown origin on the high sensitivity DNA chip (Figure 38 and 39). This band was not considered detrimental to down-stream sequencing (As discussed with the GSC). All four libraries were pooled prior to sequencing.
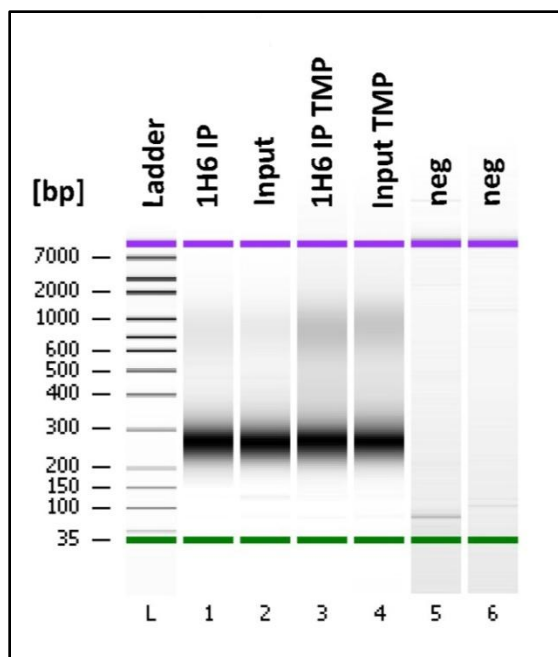
Figure 38. Agilent bioanalyzer shows discrete fragments of adaptor ligated libraries prepared for next-generation sequencing.

G4-ChIP using anti-G-quadruplex antibody 1H6 was performed on cells with or without TMPyP4 treatment (TMP). Input controls and eluted DNA from immunoprecipitations were subject to end-repair, A-tailing, adaptor ligation and amplification prior to Illumina sequencing. Agilent bioanalyzer profiles suggest that all four libraries successfully amplified and maintain discrete fragment sizes. Lane 1 is IP with 1H6 Ab, Lane 2 is input control, Lane 3 is IP using 1H6 with TMPyP4 treated cells, Lane 4 is TMPyP4 input control while Lanes 5 and 6 are negative controls. Libraries were generated from the isolation of a 100 – 300 bp fraction of DNA isolated after IP. Adaptors add an additional 120 bp and the resulting smear should be from 220 bp – 420 bp.
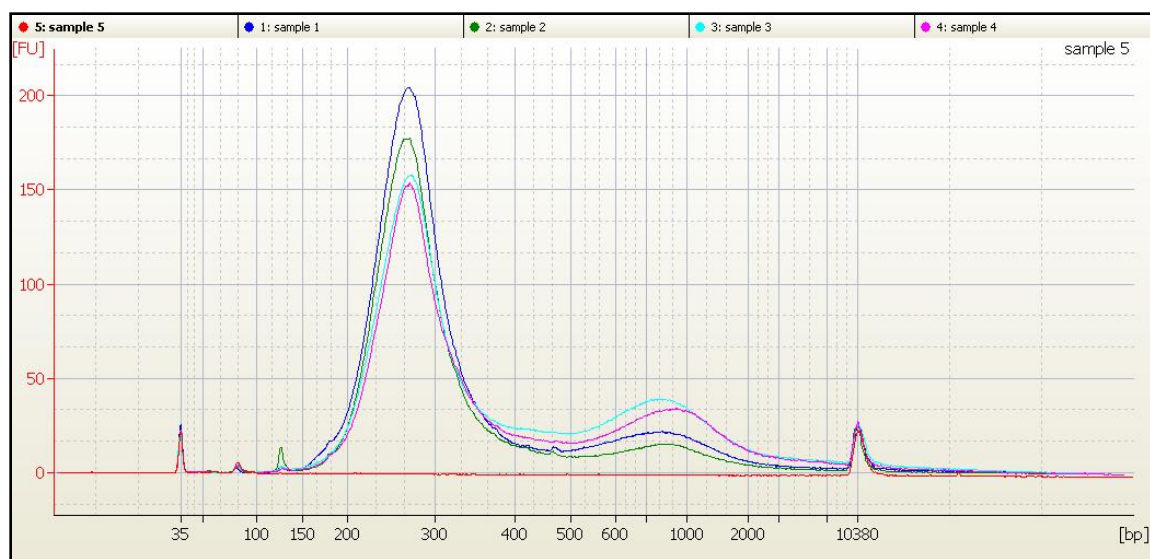
Figure 39. Computer generated profile of constructed libraries suggests proper adaptor ligation in all libraries.

Upon elution of DNA from G4-ChIP samples were size selected through PAGE. The 100-300 bp fractions was purified and used as input into library construction. Agilent bioanlysis suggests that proper adaptor ligation of fragments occurred as a major peak of fragments is located at the expected size range from 220 bp to 420 bp. The Y-axis is a measure of fluorescence units (FU) and helps to quantify total DNA per library.

4.3.5 Bioinformatic analysis of G4-ChIP indicates no enrichment in 1H6 antibody libraries over input control libraries

Upon the successful completion of Illumina sequencing all libraries undergo quality control (QC) evaluation and mapping to the reference genome by the GSC. For unknown reasons the G4-ChIP sequencing libraries failed GSC QC. We can suggest that after discussions with the GSC that there were no problems recorded on the sequencing instrument that day and that the most probable explanations for errors were located further upstream (Personal communication with Dr. Martin Hirst). The initial QC statistics of the libraries show that the sequenced libraries have similar statistics in terms of total reads, duplicates and

147

reads mapping to multiple chromosome (MMDC) (Table 7). These statistics do not tell us about the sequence reads themselves, just that there are no gross abnormalities between libraries using these statistics. Our statistics matched similar statistics from other ChIP-seq libraries and suggest that there were no obvious errors during the Illumina run. While the total number of reads for each library was approximately 40 million only ~50% of those reads were uniquely mapped back to the human genome (hg18). 50% is the standard mappability of ChIP-seq results (personal communication with Dr. Martin Hirst, GSC).

The most likely explanation for failing QC at the GSC is a lack of specific signal detected in the 1H6 antibody immunoprecipitated libraries. Standard ChIP-sequencing practice is to compare the read pile-ups (regions of the genome where many sequenced reads over-lap) between experimental libraries (1H6) and control libraries (Input). It is assumed that the libraries generated from input controls would have randomly distributed sequence reads throughout the reference genome with little to no pile-up. Any pile-up of sequence reads can be traced back to highly repetitive regions and poorly annotated reference sequence. The average sequence pile-up in the input controls is used to statistically determine the background threshold signal. Therefore, any pile-up above this threshold could be considered a binding/enrichment event. Unfortunately, the input libraries did not show an even distribution of reads expected for an input, non-enriched library. When the experimental libraries were compared to control libraries no detectable signal was observed above background. For unknown reasons, there was high background signal detected in the controls that masked the signal in the experimental libraries. Therefore we cannot interpret any of these ChIP-sequencing libraries due to a lack of specific signal above background in the experimental libraries.

Library Name	Total reads	QC Failure %	Duplicates %	Mapped %	Properly Paired %	Singletons %	MMDC %
1H6 IP	42237298	24.7	6.6	51.3	36.8	6.3	7.3
Input	46090208	25.6	1.8	53.4	36.5	8.4	7.7
1H6 IP TMP	36777234	22.7	2.1	55.5	37.4	8.3	8.8
Input TMP	41297314	25.8	4.2	49.5	34.4	6.9	7.2

Table 7. Initial quality control of sequenced libraries.

Sequenced libraries were initially analyzed following standard operating procedure examining key characteristics that can indicate failed sequenced runs. All libraries show similar QC results indicative of typical ChIP-seq experiments. QC failure refers to reads that are fragmented and/or reads that contain too many N bases (unknown bases). Duplicates are reads that are exact sequence duplicates and most likely created during PCR. Singletons are single-reads that are missing their paired-end partner read while MMDC is a record of when a mate read maps to different chromosomes.

4.3.6   Mass spectrometry interrogation of proteins after G4-ChIP cannot be interpreted due to low quality scores

In addition to examining the DNA sequences to which the anti-G-quadruplex antibody 1H6 is binding, we also wanted to investigate the proteins bound to these sequences. It has been shown that specific proteins interact with G4 structures *in vitro* and these proteins have been implied as circumstantial evidence for the existence of naturally occurring G4 structures *in vivo* (58, 62, 194, 195). To investigate the protein interactions between nucleic acid sequences capable of forming G4 structures we have combined G4-ChIP with mass spectrometry. Chromatin immunoprecipitation experiments combined with mass spectrometry have been used to identify unknown protein components of protein:

nucleic acid complexes (196, 197). However, we are unaware of any other group targeting

G4 DNA structures with antibodies to investigate the interacting protein components.

At the outset we separated all protein components through typical SDS-PAGE gels.

In two different experiments a low number of proteins were visualized on SimplyBlue

stained gels. Although prominent bands appeared in the 1H6 IP lanes, they also appear in the

IgG control lanes (Figure 41) and are most likely contaminant proteins and antibody

heavy/light chains. However, small faint bands also appeared to be present in 1H6 IP lanes

and not in the IgG controls, and vice versa.
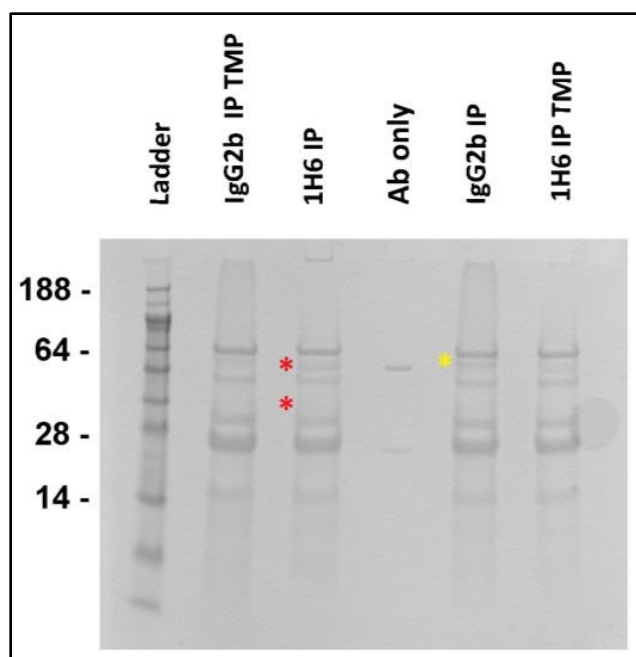
Figure 40. SDS PAGE separation of proteins eluted after G4-ChIP illustrates limited differences between anti-G-quadruplex antibody 1H6 and a matched isotype control.

G4-ChIP was performed with untreated HeLa cells or cells treated with TMPyP4 (TMP). Purified proteins from IP reactions with anti-G-quadruplex antibody 1H6 or IgG$_{2b}$ controls were subsequently separated through SDS-PAGE gels and stained with SimplyBlue. Purified 1H6 antibody (1H6 Ab) was run in the center of the gel (lane 3) to visualize the antibody protein components in other lanes. Seeblue 2 protein ladder is indicated in Ladder lane. Red asterisks highlight faint bands not detected in isotype control, while yellow asterisk highlights band found in isotype control but not in the experimental lanes.

To identify specific proteins unique to only the 1H6 antibody IP reactions (with or without TMPyP4) whole gel lanes from 1H6 antibody IPs were excised and interrogated by mass spectrometry (See Experimental Methodology). After mass spectrometry analysis, proteins detected in lanes from IgG controls were eliminated from the total proteins detected in lanes from 1H6 IP experiments. Proteins detected in IgG controls are assumed to be non-specific proteins interacting with the isotype control. Table 8 outlines the proteins detected

in 1H6 antibody IPs (both TMPyP4 treated and untreated) with non-specific proteins found in IgG controls (both TMPyP4 treated and untreated) removed.

Under similar experimental conditions to G4-ChIP-sequencing, experimental results from G4-ChIP-mass spectrometry were limited in part to quality control. While many proteins were identified in the 1H6 antibody IP experiments that were not present in the IgG control IP experiments (Table 8) the overall ion score associated with each protein is low. Ion scores are used to determine the level of confidence for which a protein identity is associated with the observed mass spectral data. In our experiment a cut-off ion score of 50 was calculated as the 5% confidence threshold. Proteins with ion scores of < 50 indicate matches that are outside the 5% confidence threshold. In addition, proteins with only 1 uniquely mapped peptide are likely random (personal communication Dr. Gregg Morin)

Although overall protein abundance was low and improved IP conditions could change these results, we did detect some interesting proteins in these first G4-ChIP-mass spectrometry experiments with ion scores above 50 (Table 8, green highlights). We identified 12 of 23 proteins (with ion scores >50) in the 1H6 antibody IP experiments, that were not present in the IgG control IP experiments, that generally reside in the nucleus. This finding supports the idea that the 1H6 antibody is binding epitopes predominantly located in the nucleus. However we did not find a significant difference between the total proportion of nuclear proteins detected in the IPs of the 1H6 antibody versus the control IgG. This observation will require further exploration in subsequent experiments and therefore cannot conclusively state that the 1H6 antibody epitope is nuclear. However, in addition we identified four RNA specific proteins that suggest the 1H6 antibody may be binding G4 RNA structures. Although it is unclear due to the overall low abundance of proteins whether or not

this is significant, it remains intriguing to think that anti-G-quadruplex antibody 1H6 could also be binding G4 RNA structures *in vivo*. There has been much less reported on the stability and characterization of G4 RNA structures, however, since RNA molecules tend to be partially single-stranded they may have a higher likelihood of folding into quadruplex structures in comparison to double-stranded DNA (187-190). Future G4-ChIP experiments with modified IP conditions may elucidate if G4 RNA binding proteins are a significant proportion of protein components isolated or if this was just an incongruity due to low numbers of proteins eluted in the reaction.

1H6 IP (with and without TMPyP4) minus IgG control IP (with and without TMPyP4)

Accession	IonScore	Mass	Unique Peptides	Description
Q9Y5G6-2	43.8	90	1	Isoform Short of Protocadherin gamma-A7 OS=Homo sapiens GN=PCDHGA7
Q9Y490	53.3	271.8	2	Talin-1 OS=Homo sapiens GN=TLN1 PE=1 SV=3
Q9Y2I1	35	168.3	2	Nischarin OS=Homo sapiens GN=NISCH PE=1 SV=3
Q9UPA5	33.5	418.3	1	Protein bassoon OS=Homo sapiens GN=BSN PE=1 SV=4
Q9UMX5	49.5	18.8	2	Neudesin OS=Homo sapiens GN=NENF PE=1 SV=1
Q9UDY2-2	29.1	117.8	1	Isoform A2 of Tight junction protein ZO-2 OS=Homo sapiens GN=TJP2
Q9NZC7-3	32.5	21.7	1	Isoform WOX3 of WW domain-containing oxidoreductase OS=Homo sapiens GN=WWOX
Q9NS39	34.2	81.3	2	Double-stranded RNA-specific editase B2 OS=Homo sapiens GN=ADARB2 PE=1 SV=1
Q9HCM1	28.6	196.6	1	Uncharacterized protein C12orf35 OS=Homo sapiens GN=C12orf35 PE=1 SV=2
Q96JS3	41.5	93.8	1	PiggyBac transposable element-derived protein 1 OS=Homo sapiens GN=PGBD1 PE=1 SV=1
Q8TDY4-3	27.1	98.9	1	Isoform 3 of Arf-GAP with SH3 domain, OS=Homo sapiens GN=ASAP3
Q8TDY4-3	27.1	98.9	1	Isoform 3 of Arf-GAP with SH3 domain, OS=Homo sapiens GN=ASAP3
Q8TC76	46.9	41	2	Protein FAM110B OS=Homo sapiens GN=FAM110B PE=1 SV=1
Q8N8E2-2	40.2	52.9	1	Isoform 2 of Zinc finger protein 513 OS=Homo sapiens GN=ZNF513
Q8N4C8-2	35.8	146.4	1	Isoform MiNK-1 of Misshapen-like kinase 1 OS=Homo sapiens GN=MINK1
Q8IZF6-2	32.2	303	1	Isoform 2 of Probable G-protein coupled receptor 112 OS=Homo sapiens GN=GPR112
Q86W50-2	35	25.8	1	Isoform 2 of Putative methyltransferase METT10D OS=Homo sapiens GN=METT10D
Q86VP1-2	41.9	87.2	2	Isoform 2 of Tax1-binding protein 1 OS=Homo sapiens GN=TAX1BP1
Q86UX2-3	33.5	81	1	Isoform 3 of Inter-alpha-trypsin inhibitor heavy chain H5 OS=Homo sapiens GN=ITIH5
Q7LBC6	61.3	193.2	4	Lysine-specific demethylase 3B OS=Homo sapiens GN=KDM3B PE=1 SV=2
Q76KP1	57.9	117	3	N-acetyl-beta-glucosaminyl-glycoprotein 4-beta-N-acetylgalactosaminyltransferase
Q5VT06	131.5	352.3	9	Centrosome-associated protein 350 OS=Homo sapiens GN=CEP350 PE=1 SV=1
Q4KWH8	109	191.2	6	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase eta-1
Q32MQ0	35.1	78.2	1	Protein ZNF750 OS=Homo sapiens GN=ZNF750 PE=2 SV=1
Q15878-2	35.6	256.5	1	Isoform Alpha-1E-1 of Voltage-dependent R-type calcium channel subunit alpha-1E

1H6 IP (with and without TMPyP4) minus IgG control IP (with and without TMPyP4) continued

Accession	IonScore	Mass	Unique Peptides	Description
Q15018	48.3	47.1	2	BRISC complex subunit Abro1 OS=Homo sapiens GN=FAM175B PE=1 SV=2
Q14596-2	16.6	104.9	1	Isoform 2 of Next to BRCA1 gene 1 protein OS=Homo sapiens GN=NBR1
Q13404-6	57	11.9	2	Isoform UEV-1As of Ubiquitin-conjugating enzyme E2 variant 1 OS=Homo sapiens GN=UBE2V1
Q12996	26.9	83.3	1	Cleavage stimulation factor subunit 3 OS=Homo sapiens GN=CSTF3 PE=1 SV=1
Q0VDD7	30.8	70.1	2	Uncharacterized protein C19orf57 OS=Homo sapiens GN=C19orf57 PE=1 SV=2
Q07020	49	21.7	1	60S ribosomal protein L18 OS=Homo sapiens GN=RPL18 PE=1 SV=2
Q05682-2	64	65.7	1	Isoform WI-38 L-CAD I of Caldesmon OS=Homo sapiens GN=CALD1
Q05639	253.5	50.8	5	Elongation factor 1-alpha 2 OS=Homo sapiens GN=EEF1A2 PE=1 SV=1
Q02878	43.5	32.8	1	60S ribosomal protein L6 OS=Homo sapiens GN=RPL6 PE=1 SV=3
P68032	200.9	42.3	4	Actin, alpha cardiac muscle 1 OS=Homo sapiens GN=ACTC1 PE=1 SV=1
P63241-2	124.8	20.4	5	Isoform A of Eukaryotic translation initiation factor 5A-1 OS=Homo sapiens GN=EIF5A
P62979	55.1	10.7	1	HA-Ubiquitin-KO ubiquitin mutant
P61513	34.3	10.5	1	60S ribosomal protein L37a OS=Homo sapiens GN=RPL37A PE=1 SV=2
P60866	49.4	13.5	1	40S ribosomal protein S20 OS=Homo sapiens GN=RPS20 PE=1 SV=1
P60709	287.2	42.1	5	Actin, cytoplasmic 1 OS=Homo sapiens GN=ACTB PE=1 SV=1
P56730	28.4	99.4	1	Neurotrypsin OS=Homo sapiens GN=PRSS12 PE=2 SV=2
P49458	59.4	10.2	1	Signal recognition particle 9 kDa protein OS=Homo sapiens GN=SRP9 PE=1 SV=2
P46779	40.4	15.8	1	60S ribosomal protein L28 OS=Homo sapiens GN=RPL28 PE=1 SV=3
P46779	149.6	15.8	4	60S ribosomal protein L28 OS=Homo sapiens GN=RPL28 PE=1 SV=3
P39019	75.4	16.1	2	40S ribosomal protein S19 OS=Homo sapiens GN=RPS19 PE=1 SV=2
P36957	24.1	49	1	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial OS=Homo sapiens GN=DLST PE=1 SV=3
P35030-2	96.3	28.8	3	Isoform B of Trypsin-3 OS=Homo sapiens GN=PRSS3
P26232-2	56.5	101.1	3	Isoform AlphaN-catenin I of Catenin alpha-2 OS=Homo sapiens GN=CTNNA2

			Unique	
Accession	IonScore	Mass	Peptides	Description
P0C0S5	84.6	13.5	2	Histone H2A.Z OS=Homo sapiens GN=H2AFZ PE=1 SV=2
P09012	62.8	31.3	2	U1 small nuclear ribonucleoprotein A OS=Homo sapiens GN=SNRPA PE=1 SV=3
P07355-2	53.5	40.7	2	Isoform 2 of Annexin A2 OS=Homo sapiens GN=ANXA2
P06899	148.2	13.9	4	Histone H2B type 1-J OS=Homo sapiens GN=HIST1H2BJ PE=1 SV=3
P04406	79.9	36.2	2	Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens GN=GAPDH PE=1 SV=3
P04406	106.3	36.2	2	Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens GN=GAPDH PE=1 SV=3
P04075	66.1	39.9	1	Fructose-bisphosphate aldolase A OS=Homo sapiens GN=ALDOA PE=1 SV=2
P02774-2	49.2	40.7	2	Isoform 2 of Vitamin D-binding protein OS=Homo sapiens GN=GC
O95714	55.9	533.5	2	E3 ubiquitin-protein ligase HERC2 OS=Homo sapiens GN=HERC2 PE=1 SV=2
O95197-2	39.6	111.2	2	Isoform A3b of Reticulon-3 OS=Homo sapiens GN=RTN3
O14639	24.9	89.5	1	Actin-binding LIM protein 1 OS=Homo sapiens GN=ABLIM1 PE=1 SV=3

1H6 IP (with and without TMPyP4) minus IgG control IP (with and without TMPyP4) continued

Table 8. G4-ChIP-mass spectrometry of proteins associated with 1H6 antibody indicate some abundance of nucleic acid binding proteins.
Under modified conditions for protein isolation anti-G-quadruplex antibody 1H6 was used to immunoprecipitate chromatin from fixed human HeLa cells. Mass spectrometry was used to examine the protein components of G4-ChIP with HeLa cells treated and untreated with TMPyP4. Proteins listed above comprise all proteins found in 1H6 antibody fractions from the combination of TMPyP4 and untreated immunoprecipitations minus non-specific bound proteins from IgG controls. Green highlighting illustrates ion scores above 50. Proteins with low ion scores (<50) and only 1 unique peptide may be contamination events.

4.4 Discussion

The ultimate goal of this project was to look for G-quadruplex nucleic acids in human cells. No scientific group to date has shown that G-quadruplex nucleic acid structures naturally exist in mammalian cells. In this regard we endeavored to use our anti-G-quadruplex antibody 1H6 in various experiments to explore the possibility of discovering the existence of naturally occurring G-quadruplex structures in human cells. If naturally occurring G4 structures are present in human cells and the 1H6 antibody can bind those structures, we would expect that these structures can be isolated using modified immunoprecipitation (IP) experiments. We verified that 1H6 binds DNA in excess over an isotype control (Figure 35). When the same experiment was performed with chromatin prepared from cells treated with TMPyP4 we isolated twice as much DNA. We conclude that cells treated with TMPyP4 stabilized the G4 structures in cells and stabilization of these structures allowed for more 1H6 antibody binding and therefore greater yields of DNA in our IPs.

DNA quantification after IP does not describe the types of sequences pulled-down in the reaction. We do not know if TMPyP4 is having a non-specific effect and subsequently increasing non-specific DNA binding during IP or if the conditions during IP result in high specificity binding of 1H6 to G4 structures.

We, thus, investigated the sequences eluted during IP using conventional and quantitative PCR to probe the eluted DNA for putative quadruplex forming regions thought to form *in vivo*. Many other groups have described sequences throughout the human genome that are proposed to form G4 structures *in vivo* (48, 50, 53, 54, 56, 194, 198-200), principally found in the promoters of oncogenes. Using the literature as a guide we designed primers for

157

areas of the genome thought to form G4 structures *in vivo* and another set of primers for

areas (at least 1000 bp away from any G4 forming sequences) incapable of forming G4

structures. Interestingly we found by conventional PCR that some G4 forming sequences

could be detected in the eluted DNA from 1H6 IP experiments while none were detected in

the matched isotype control (Figure 36). This finding suggested to us that our initial

conditions for IP warranted further exploration.

The differences observed between the isotype control and the 1H6 antibody could be

explained by an induction effect caused by the binding of the 1H6 antibody to guanine-rich

sites or as previously stated, random variation observed in IP. In this case, the lack of bands

observed in the isotype control would only indicate that the IgG_{2b} antibody cannot similarly

induce binding at these sites. However, previous CD experiments in Chapter 2 exploring the

induction effects of the 1H6 antibody on guanine-rich DNA sequences indicated no induction

effects. However CD experiments were done *in vitro* and utilized pure DNA as opposed to

the chromatin used during IP reactions which certainly change the physiological environment

for 1H6 binding. Nonetheless, the *myc* band observed in the 1H6 antibody IPs that was

absent in the input control lane could indicate enrichment of guanine-rich *myc* sequences

facilitated by the binding of 1H6.

However using quantitative PCR, we did not detect G4 enrichment between IgG_{2b} IPs

and 1H6 antibody IPs over NON-G4 forming sequences (Figure 37). We did detect

enrichment of G4 forming sequences found in *c-Kit* and *p53* but also found enrichment of the

NON-G4 forming sequence in Rb. There are many possibilities as to why we did not detect

enrichment in all of our G4 forming sequences over our NON-G4 forming sequences. First

and foremost guanine-rich DNA forms unwanted secondary structures which inadvertently

affect DNA sequencing and polymerase chain reaction (PCR) amplification and guanine-rich

sequencing is often challenging because of extensive self-association (43). Quantitative PCR

is challenging through guanine-rich templates due inherently to this issue. However, we

have taken steps to alleviate this issue by including PCR additives known to inhibit the

formation of these structures during PCR. Betaine and DMSO are the most common PCR

additives for this purpose but two other organic additives have also been employed ethylene

glycol and 1,2-propanediol (206). In an attempt to alleviate some of the unwanted secondary

structures we used DMSO in all of our PCR reactions. During our initial optimization of

PCR conditions, we observed inefficient amplification of selected genes without the use of

PCR additives (indicated by very high CT values, skewed amplification plots and missing

bands) (data not shown). We assumed that the 1H6 antibody was pulling down G4 DNA

structures but the sequences were not efficiently amplified during PCR and qPCR. Our

current data indicate that the use of PCR additives has helped this issue but there may be

remaining differences in amplification efficiency associated with secondary structure

formation. Future testing of other PCR additives may help to elucidate this issue.

Secondly, some of the targeted promoter sequences have not been experimentally

proven to fold into G4 structures *in vitro* and thus could have created an experimental design

flaw that resulted in varying experimental results. Although sequences in the literature have

been proposed to form G4 structures *in vivo*, none have been directly detected

experimentally. Possibly examining only the promoters of genes that have been well

characterized to form G4 structures may well clarify these experimental results.

Finally, as indicated by our previous specificity work in Chapter 2 the 1H6 antibody

has varying affinity across different G4 structures. Different G4 topologies would be

differentially bound by the 1H6 antibody during an IP experiment. The 1H6 antibody could be enriching for particular structures and sequences over other structures. It is plausible that the large enrichment of the *kit* gene in both conventional PCR and qPCR indicate that the kit G4 structure is preferentially bound by the 1H6 antibody in comparison to other G4 structures. In the future, we could examine this possibility by recreating the *kit* G4 structure synthetically and applying it in an ELISA specificity assay.

Although our experiments by ELISA and gel-shift indicate that the 1H6 antibody is specifically binding G4 DNA we cannot eliminate the possibility that the 1H6 antibody is also binding DNA in a non-specific manner. By qPCR we observed a 20 fold enrichment of Rb NON in the 1H6 antibody IP experiments over IgG, that could be explained by random variations observed in IP, non-specific interactions with DNA or an unknown artifact. Regardless, this experimental result suggests that the 1H6 antibody is not performing as we had hypothesized as we cannot explain this qPCR result.

While G4-ChIP-qPCR proved to be variable, the real test for providing *in vivo* evidence of the formation of G4 structures in humans would be to isolate guanine-rich sequences from 1H6 antibody IP experiments. Furthermore aligning those guanine-rich sequences to the human genome would give us insight into the sequences that naturally form G4 structures *in vivo*. Using isolated DNA from 1H6 antibody IP experiments and the corresponding input chromatin as a control we built Illumina ready libraries for next-generation sequencing. Initially quality control experiments examining the size distribution of adaptor ligated libraries suggested that proper library construction had been successfully completed (Figure 38 and 39). While proper library construction was suggested, Figure 38 and 39 do not indicate the type of DNA sequences that were incorporated into the libraries.

Unfortunately, in our first attempt at sequencing G4-ChIP material all sequenced libraries failed quality control (QC) done by the GSC. There are many possible reasons however, in general the issue was a low signal-to-noise relationship between the experimental libraries (1H6) and the control (Input) libraries. Control libraries did not show the expected random distribution of reads across the genome and when compared to the experimental libraries, masked all signal in the experimental libraries. Given that *in vitro* data in chapters 2 and 3 suggest that the 1H6 antibody is binding G4 nucleic acids our first thoughts are that the conditions for IP have not been fully optimized for sequencing. Additionally, we are unaware of the type of DNA that was incorporated into the libraries during construction. After IP with 1H6 there might be a bias for double-stranded DNA incorporation into the libraries that overshadows any enrichment of single-stranded G4 DNA structures pulled-down by the 1H6 antibody. The chemistry required for the successful preparation of Illumina libraries requires the addition of adaptors to double stranded DNA prior to sequencing. After sonication end-repair, A-tailing and adaptor ligation are required for the successful generation of Illumina ready sequences. One or more of these steps may bias for double-stranded DNA and influence the sequencing results with 1H6 antibody IP material. Moreover, there are very few reports of performing ChIP with antibodies to nucleic acids other than the highly published method of 5-methylcytosine ChIP, MeDIP (201). Our conditions during IP were different than those performed for MeDIP but future experiments could compare the two methods for further optimization.

Nevertheless, optimization of the IP conditions for sequencing will require alternative read-outs outside of next-generation sequencing, due to the increased wait times (upwards of 5-6 months) between an experiment and obtaining sequencing results (Dr. Martin Hirst

personal communication). In this regard we hope to optimize the IP conditions by qPCR and mass spectrometry.

In addition to examining the DNA sequences pulled-out of IP reactions with the 1H6 antibody we were curious to explore the proteins bound to these sequences in similar IP experiments. Proteins isolated from IP reactions with 1H6 may help to elucidate the type of nucleic acids that the 1H6 antibody preferentially binds. If in fact the 1H6 antibody is binding naturally occurring G4 structures we could also gain insight into the types of proteins important in the biology of G4 structures. In this regard we attempted to combine G4-ChIP with mass spectrometry to decipher the unknown proteins bound by nucleic acid sequences in 1H6 IP experiments. Only a few groups have attempted this type of mass spectrometry analysis and to our knowledge we are the first group to combine anti-nucleic acid antibodies in ChIP with mass spectrometry. In our first attempt at this procedure we did not modify the G4-ChIP protocol too drastically and only changing the lysis detergent to 0.5% NP-40 instead of 1% SDS. While SDS does have some protein denaturing capabilities further modification of the G4-ChIP procedure may help to isolate even great number of proteins associated with G4 structures *in vivo*. Even so, we did isolate and detect proteins after G4-ChIP (Table 8). By subtracting proteins also found in IgG lanes (that are assumed to be non-specifically interacting proteins) we were able to focus on the proteins found only in 1H6 antibody IP reactions. Under similar experimental conditions to G4-ChIP-sequencing, experimental results from G4-ChIP-mass spectrometry were limited in part by quality control. The overall ion score associated with each protein is relatively low and many proteins contained only 1 uniquely mapped peptide.

Only 23 proteins passed QC, of which over 50% were located in the nucleus. If the 1H6 antibody is binding G4 DNA structures then we expect most of the proteins associated with these structures should be located in the nucleus. However, we also identified 4 proteins that have been associated with RNA. Proteins found to be specific to RNA suggests at least some affinity for the 1H6 antibody to RNA structures, quite possibly G4 RNA structures. RNA sequences do naturally occur in the cell as single-stranded nucleic acid entities and may therefore have a higher propensity to form unimolecular G4 structures in comparison to DNA that tends to be in double-stranded form. In spite of this possibility, our specificity ELISA indicated no binding to at least one G4 RNA structure (Chapter 2) and treatment of cells with RNAse enzymes had little to no effect on the staining patterns observed during fluorescent microscopy with the 1H6 antibody (Chapter 3). Overall, the ChIP-mass spectrometry experiments do not conclusively indicate if the 1H6 antibody preferentially binds DNA, RNA or protein components.

In the future we hope to continue exploring the specific sequences with which the 1H6 antibody binds. We would like to decipher if there are guanine-rich sequences to which the antibody has preference and whether or not the 1H6 antibody truly binds G4 RNA structures in addition to G4 DNA structures. Further validation of proteins bound to G4 structures will help elucidate the complex biological interactions that have been proposed for G4 structures.

Chapter 5: General discussion and conclusion

5.1 Summary of findings

In chapter 2 we describe the successful creation of novel antibodies to G-quadruplex nucleic acids. The specificity of the antibodies was examined by ELISA and it was found that the purified antibodies can discriminate between double-stranded, single-stranded (unfolded) and quadruplex structures. The 1H6 antibody was tested against other quadruplex folds and found to bind to multiple conformations of quadruplex folds. This finding was confirmed by a gel-shift assay in a collaborating laboratory.

In chapter 3, we tested the subcellular localization of structures recognized by the 1H6 antibody which includes potential G4 DNA structures. The antibody was found to give a spotted nuclear staining in most cells *in vitro* and *in vivo*. In Chapter 3 the nature of the 1H6 foci is examined using various staining conditions. Addition of synthetic G-quadruplex structures to the antibody abolished staining, while treatment with different G-quadruplex stabilizing drugs increased the intensity and number of foci detected. By combining immunostaining and FISH, 1H6 foci were detected at a subset of telomeres. However, without further quantification the detection of foci at telomeres could be considered coincidental.

In chapter 4 attempts were made to confirm the *in vivo* existence of G-quadruplex structures using the 1H6 antibody. A new technique aiming to enrich for putative quadruplex forming regions in the genome (G4-ChIP) was used to isolate and purify both the DNA sequences bound by the 1H6 antibody and the proteins attached to those sequences. Illumina sequencing was performed to validate the enrichment. Unfortunately sequencing results

lacked the required signal-to-noise ratio to derive conclusive results. In addition, endeavors to combine G4-ChIP with mass spectrometry were performed. Evidence for enrichment of ribosomal proteins in 1H6 immunoprecipitates was obtained. These results need to be validated further.

5.2   Overall analysis

This project has been built on knowledge from a diverse group of disciplines. Many of the techniques described in this thesis are practiced in diverse fields of study including fundamental sciences like chemistry, biology and physics, but also more specialized sciences such as genomics and proteomics. Furthermore, little to no past research in many aspects of this project has required me to find my way around existing and develop new techniques. These efforts required innovation and perseverance.

The existence of G-quadruplex nucleic acids has yet to be established in mammalian cells, but the work described in this thesis adds to our current understanding of G-quadruplex nucleic acids. A significant portion of the work has been focused on the 1H6 antibody. Our data support that 1H6 may be the first monoclonal antibody to detect G-quadruplex structures *in vitro* and possibly in living cells. We cannot conclusively say that the foci detected in fluorescence microscopy by 1H6 are in fact G-quadruplex nucleic acid structures but we have no data to the contrary.

Although other scientific groups have established the development of antibodies and antibody fragments to different quadruplex structures (108, 171, 172, 176) we are not aware of any groups providing visual evidence of quadruplex structures in mammalian cells.

Continued research with the 1H6 antibody may ultimately provide the support required to conclusively validate the presence of G-quadruplex nucleic acids in mammalian cells.

5.3   Comments on strengths and limitations

One of the main strengths of this project is that we have endeavored to repeat experimental findings across a wide range of techniques. All techniques have limitations. By gathering data from different techniques to address similar questions we have tried to mitigate the inherent limitations in each individual technique. Throughout this project we have attempted to maintain this cross-examination to help strengthen our results and conclusions.

For instance, the initial development of G-quadruplex structures for immunization purposes was straightforward. The classical native gel mobility retardation of tetramolecular structures and their characteristic CD spectrum made it relatively easy to produce these structures in relatively large quantities and establish their topology. However, when antibodies were to be interrogated for their specificity it became fundamental to produce distinct quadruplex topologies and identify their characteristics. Relying on native gel mobility patterns only to classify G-quadruplex folding has resulted in the publication of erroneous or misleading results most often because mixture of "purified" G-quadruplex structures were analyzed (42, 177-179). Supplemental to all of our mobility assays we confirmed the nature of our purified structures by CD. CD is not without its own limitations, however, in conjunction with gel mobility assays, it has allowed us to increase our level of certainty regarding the purification of single quadruplex species and assigning topology. CD

is limited by its ability to assign topology to new sequences and structures that have not already been well established using more refined techniques such as NMR or crystallography. As more and more structures are characterized their CD spectrum recorded, we can ultimately create a library of CD spectrum that can be used to assign topologies to many different G-quadruplex structures.

Similarly, during our specificity assays we have strengthened our results by examining specificity across different techniques (ELISA and gel-shift) and across different conditions (direct ELISA vs. competition ELISA). Again, using this approach we have limited inaccuracies in our findings by asking a similar question using different techniques. One of the major limitations of the ELISA system is that we cannot visualize the structures. This is a fundamental issue when we are trying to establish that the antibody in question is binding a particular folded structure. G-quadruplexes are flexible and may fold and unfold in different types of experiments, even during an ELISA test. The benefits of gel-shift assays are that we can directly visualize the binding of antibodies to different structures. In the gel-shift assay we can confirm that a tetramolecular structure remains and is bound by antibodies by the mobility in the gel, while in ELISA we could not tell if the tetramolecular structure had unfolded or changed conformation. More importantly, in gel-shift we can confirm the unfolded nature of a single-strand and visualize that no antibody binding occurs.

Our in-cell localization data was also strengthened in a similar way. We used fluorescence microscopy in conjunction with flow cytometry to visualize changes in antibody binding. Some of the limitations of microscopy stem from the limitations of the technology itself including limitations on resolution due to diffraction, but also include weaknesses with results due to the subjective nature of imaging. We have tried to alleviate some of this

167

natural bias by combining flow cytometry analysis with our results. Again, flow cytometry has its limitations but in addition to the results gathered by microscopy they provide supplementary assurance by corroborating our results.

In view of our observations using different techniques we are confident that our results will be confirmed by collaborative laboratories in the field and individuals performing similar experiments using our antibodies. We have had several collaborators and individuals use the 1H6 antibody for the purpose of their own experiments (See all results in the Appendix). Many of our results have been reproduced and we look forward to additional results from our collaborators.

One of the major critiques of this work is the variability of results observed using the 1H6 antibody. Although we continue to suggest that the 1H6 antibody is binding G4 DNA we have not yet proven that the 1H6 antibody does not also bind an unknown protein or RNA component. We have tested the 1H6 antibody under different experimental conditions thought to stabilize G4 DNA (TMPyP4 and telomestatin) and observed increased mean fluorescent signal. But other than the incubation of the 1H6 antibody in the presence of synthetic G4 DNA we have not been able to decrease the mean fluorescent signal. These findings in conjunction with the lack of proof observed by next-generation sequencing, require further experimentation to provide evidence that the 1H6 antibody is binding G4 DNA.

Proving that G-quadruplex nucleic acids naturally exist in mammalian cells without a doubt has been outside the reach of this dissertation, however we believe we have made good strides in the right direction and laid the foundations for such studies.

5.4   Discussion of applications and possible future research

The goal of this project has always been to establish the existence of G-quadruplex nucleic acids in human cells using novel monoclonal antibodies. Given that our data suggests that our antibodies are indeed binding different nucleic acid structures it will be of future interest to establish whether or not these antibodies can isolate G-quadruplex structures in different organisms. The optimization of our initial experiments using the 1H6 antibody immunoprecipitation in combination with next-generation sequencing and mass spectrometry will be of high priority in the future. To understand if, given the right conditions, the 1H6 antibody can enrich for guanine-rich nucleic acids it has been suggested that we should first prepare genomic libraries, denature those libraries and then subsequently perform immunoprecipitation experiments under conditions in which guanine-rich sequences fold into quadruplex structures with the 1H6 antibody. We could combine this type of experiment with quantitative-PCR to establish that the 1H6 antibody can enrich for putative quadruplex forming sequences.

Additionally, to further explore 1H6 binding to RNA G4 structures we could extract total RNA from cells and treat them under conditions in which they would fold into quadruplex structures. Again, performing immunoprecipitation experiments, examine the amount of RNA captured by this antibodies. Examining native RNA during follow-up experiments may help provide evidence for the establishment of RNA G4 structures in mammalian cells.

5.5   Concluding remarks

Many groups have been trying to elucidate the natural existence of G-quadruplex nucleic acid structures in human cells for many decades. Our novel antibodies have started to provide support that these interesting structures indeed exist in all type of cells. Future work establishing the full capabilities of these antibodies may further our understanding of G-quadruplex structures and one day decipher their biological authenticity.

References

1. Nordhoff E, Kirpekar F, Roepstorff P. Mass spectrometry of nucleic acids. Mass Spectrom Rev. 1996 MAR-APR;15(2):67-138.

2. Watson JD, Crick FHC. Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. Nature. 1953;171(4356):737-8.

3. Rich A, Nordheim A, Wang AHJ. The chemistry and biology of left-handed Z-dna. Annu Rev Biochem. 1984;53:791-846.

4. SantaLucia J, Hicks D. The thermodynamics of DNA structural motifs. Annu Rev Biophys Biomol Struct. 2004;33:415-40.

5. Huppert JL. Hunting G-quadruplexes. Biochimie. 2008 AUG;90(8):1140-8.

6. Frankkamenetskii MD, Mirkin SM. Triplex dna structures. Annu Rev Biochem. 1995;64:65-95.

7. Chan PP, Glazer PM. Triplex DNA: Fundamentals, advances, and potential applications for gene therapy. J Mol Med. 1997 APR;75(4):267-82.

8. Moffat AS. Triplex dna finally comes of age. Science. 1991 JUN 7;252(5011):1374-5.

9. Bang I. Examination or the guanyle acid. Biochem Z. 1910;26:293-311.

10. Gellert M, Lipsett MN, Davies DR. Helix formation by guanylic acid. Proc Natl Acad Sci U S A. 1962;48(12):2013,&.

11. Huppert JL. Structure, location and interactions of G-quadruplexes. Febs J. 2010 SEP;277(17):3452-8.

12. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: Sequence, topology and structure. Nucleic Acids Res. 2006 NOV;34(19):5402-15.

13. Huppert JL. Four-stranded nucleic acids: Structure, function and targeting of G-quadruplexes. Chem Soc Rev. 2008;37(7):1375-84.

14. Pinnavaia TJ, Marshall CL, Mettler CM, Fisk CI, Miles HT, Becker ED. Alkali-metal ion specificity in solution ordering of a nucleotide, 5'-guanosine monophosphate. J Am Chem Soc. 1978;100(11):3625-7.

15. Sen D, Gilbert W. A sodium-potassium switch in the formation of 4-stranded G4-dna. Nature. 1990 MAR 29;344(6265):410-4.

16. Sessler JL, Sathiosatham M, Doerr K, Lynch V, Abboud KA. A G-quartet formed in the absence of a templating metal cation: A new 8-(N,N-dimethylaniline)guanosine derivative. Angew Chem -Int Edit. 2000;39(7):1300,+.

17. Williamson JR. G-quartet structures in telomeric dna. Annu Rev Biophys Biomol Struct. 1994;23:703-30.

18. Henderson E, Hardin CC, Walk SK, Tinoco I, Blackburn EH. Telomeric dna oligonucleotides form novel intramolecular structures containing guanine guanine base-pairs. Cell. 1987 DEC 24;51(6):899-908.

19. Sen D, Gilbert W. Formation of parallel 4-stranded complexes by guanine-rich motifs in dna and its implications for meiosis. Nature. 1988 JUL 28;334(6180):364-6.

20. Woody RW. Circular-dichroism. Methods Enzymol. 1995;246:34-71.

21. Kypr J, Kejnovská I, Renčiuk D, Vorlíčková M. Circular dichroism and conformational polymorphism of DNA. Nucleic Acids Research. 2009 April 01;37(6):1713-25.

22. Giraldo R, Rhodes D. The yeast telomere-binding protein Rap1 binds to and promotes the formation of dna quadruplexes in telomeric dna. EMBO J. 1994 MAY 15;13(10):2411-20.

23. Hardin CC, Watson T, Corregan M, Bailey C. Cation-dependent transition between the quadruplex and watson-crick hairpin forms of d(CGCG3GCG). Biochemistry (N Y). 1992 01/01;31(3):833-41.

24. Fojtík P, Kejnovská I, Vorlíčková M. The guanine-rich fragile X chromosome repeats are reluctant to form tetraplexes. Nucleic Acids Research. 2004 January 16;32(1):298-306.

25. Green JJ, Ying L, Klenerman D, Balasubramanian S. Kinetics of unfolding the human telomeric DNA quadruplex using a PNA trap. J Am Chem Soc. 2003 04/01;125(13):3763-7.

26. Uddin MK, Kato Y, Takagi Y, Mikuma T, Taira K. Phosphorylation at 5′ end of guanosine stretches inhibits dimerization of G-quadruplexes and formation of a G-quadruplex interferes with the enzymatic activities of DNA enzymes. Nucleic Acids Research. 2004 January 01;32(15):4618-29.

27. Virgilio A, Esposito V, Randazzo A, Mayol L, Galeone A. 8-methyl-2′-deoxyguanosine incorporation into parallel DNA quadruplex structures. Nucleic Acids Research;33(19):6188-95.

28. Ito H, Tanaka S, Miyasaka M. Circular dichroism spectra of DNA quadruplexes [d(G(5)T(5))](4) as formed with G(4) and T-4 tetrads and [d(G(5)T(5)) center dot d(A(5)C(5))](2) as formed with watson-crick-like (G-C)(2) and (T-A)(2) tetrads. Biopolymers. 2002 OCT 15;65(2):61-80.

29. Giraldo R, Suzuki M, Chapman L, Rhodes D. Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: A circular dichroism study. Proceedings of the National Academy of Sciences. 1994 August 02;91(16):7658-62.

30. Hardin CC, Henderson E, Watson T, Prosser JK. Monovalent cation induced structural transitions in telomeric DNAs: G-DNA folding intermediates. Biochemistry (N Y). 1991 05/01;30(18):4460-72.

31. Chang C, Kuo I-, Ling I-, Chen C, Chen H, Lou P, et al. Detection of quadruplex DNA structures in human telomeres by a fluorescent carbazole derivative. Anal Chem. 2004 08/01;76(15):4490-4.

32. Keniry MA, Owen EA, Shafer RH. The contribution of thymine—thymine interactions to the stability of folded dimeric quadruplexes. Nucleic Acids Research. 1997 November 01;25(21):4389-92.

33. Qi J, Shafer RH. Covalent ligation studies on the human telomere quadruplex. Nucleic Acids Research;33(10):3185-92.

34. Ren J, Chaires JB. Sequence and structural selectivity of nucleic acid binding ligandsâ€ . Biochemistry (N Y). 1999 12/01;38(49):16067-75.

35. Rezler EM, Bearss DJ, Hurley LH. Telomeres and telomerases as drug targets. Current Opinion in Pharmacology. 2002 AUG;2(4):415-23.

36. Rujan IN, Meleney JC, Bolton PH. Vertebrate telomere repeat DNAs favor external loop propeller quadruplex structures in the presence of high concentrations of potassium. Nucleic Acids Research;33(6):2022-31.

37. Smirnov I, Shafer RH. Lead is unusually effective in sequence-specific folding of DNA. J Mol Biol. 2000 2/11;296(1):1-5.

38. Vorlickova M, Chladkova J, Kejnovska I, Fialova M, Kypr J. Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. Nucleic Acids Res. 2005;33(18):5851-60.

39. Hardin CC, Perry AG, White K. Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids. Biopolymers. 2000;56(3):147-94.

40. Baumann P. G-quadruplex DNA: Methods and protocols. G-Quadruplex DNA: Methods and Protocols. 2010.

41. Sundquist WI, Klug A. Telomeric dna dimerizes by formation of guanine tetrads between hairpin loops. Nature. 1989 DEC 14;342(6251):825-9.

42. Williamson JR, Raghuraman MK, Cech TR. Mono-valent cation induced structure of telomeric dna - the G-quartet model. Cell. 1989 DEC 1;59(5):871-80.

43. Davis JT. G-quartets 40 years later: From 5 '-GMP to molecular biology and supramolecular chemistry. Angewandte Chemie-International Edition. 2004;43(6):668-98.

44. Todd AK, Johnston M, Neidle S. Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res. 2005;33(9):2901-7.

45. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. Nucleic Acids Res. 2005;33(9):2908-16.

46. Eddy J, Maizels N. Gene function correlates with potential for G4 DNA formation in the human genome. Nucleic Acids Res. 2006;34(14):3887-96.

47. Rawal P, Kummarasetti VBR, Ravindran J, Kumar N, Halder K, Sharma R, et al. Genome-wide prediction of G4 DNA as regulatory motifs: Role in escherichia coli global regulation. Genome Res. 2006 MAY;16(5):644-55.

48. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proc Natl Acad Sci U S A. 2002 SEP 3;99(18):11593-8.

49. Simonsson T, Pecinka P, Kubista M. DNA tetraplex formation in the control region of c-myc. Nucleic Acids Res. 1998 MAR 1;26(5):1167-72.

50. Phan AT, Modi YS, Patel DJ. Propeller-type parallel-stranded g-quadruplexes in the human c-myc promoter. J Am Chem Soc. 2004 JUL 21;126(28):8710-6.

51. Rankin S, Reszka AP, Huppert J, Zloh M, Parkinson GN, Todd AK, et al. Putative DNA quadruplex formation within the human c-kit oncogene. J Am Chem Soc. 2005 AUG 3;127(30):10584-9.

52. Fernando H, Reszka AP, Huppert J, Ladame S, Rankin S, Venkitaraman AR, et al. A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. Biochemistry (N Y). 2006 JUN 27;45(25):7854-60.

53. Phan AT, Kuryavyi V, Burge S, Neidle S, Patel DJ. Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. J Am Chem Soc. 2007 APR 11;129(14):4386-92.

54. De Armond R, Wood S, Sun DY, Hurley LH, Ebbinghaus SW. Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1 alpha promoter. Biochemistry (N Y). 2005 DEC 13;44(49):16341-50.

55. Martinybaron G, Marme D. Vegf-mediated tumor angiogenesis - a new target for cancer-therapy. Curr Opin Biotechnol. 1995 DEC;6(6):675-80.

56. Sun DY, Guo KX, Rusche JJ, Hurley LH. 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 Res. 2005;33(18):6070-80.

57. Catasti P, Chen X, Moyzis RK, Bradbury EM, Gupta G. Structure-function correlations of the insulin-linked polymorphic region. J Mol Biol. 1996 DEC 6;264(3):534-45.

58. Johnson JE, Cao K, Ryvkin P, Wang L, Johnson FB. Altered gene expression in the werner and bloom syndromes is associated with sequences having G-quadruplex forming potential. Nucleic Acids Res. 2010 MAR;38(4):1114-22.

59. Arthanari H, Bolton PH. Functional and dysfunctional roles of quadruplex DNA in cells. Chem Biol. 2001 MAR;8(3):221-30.

60. Sun H, Karow JK, Hickson ID, Maizels N. The bloom's syndrome helicase unwinds G4 DNA. J Biol Chem. 1998 OCT 16;273(42):27587-92.

61. Sun H, Bennett RJ, Maizels N. The saccharomyces cerevisiae Sgs1 helicase efficiently unwinds G-G paired DNAs. Nucleic Acids Res. 1999 MAY 1;27(9):1978-84.

62. Wu Y, Brosh RM, Jr. G-quadruplex nucleic acids and human disease. Febs J. 2010 SEP;277(17):3470-88.

63. Fry M, Loeb LA. Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)(n). J Biol Chem. 1999 APR 30;274(18):12797-802.

64. Neff NF, Ellis NA, Ye TZ, Noonan J, Huang K, Sanz M, et al. The DNA helicase activity of BLM is necessary for the correction of the genomic instability of bloom syndrome cells. Mol Biol Cell. 1999 MAR;10(3):665-76.

65. Chakraverty RK, Hickson ID. Defending genome integrity during DNA replication: A proposed role for RecQ family helicases. Bioessays. 1999 APR;21(4):286-94.

66. Auerbach AD, Verlander PC. Disorders of DNA replication and repair. Curr Opin Pediatr. 1997 Dec., 1997;9(6):600-16.

67. Lee SK, Johnson RE, Yu SL, Prakash L, Prakash S. Requirement of yeast SGS1 and SRS2 genes for replication and transcription. Science. 1999 DEC 17;286(5448):2339-42.

68. Heo SJ, Tatebayashi K, Ohsugi I, Shimamoto A, Furuichi Y, Ikeda H. Bloom's syndrome gene suppresses premature ageing caused by Sgs1 deficiency in yeast. Genes Cells. 1999 NOV;4(11):619-25.

69. Auerbach AD. Fanconi anemia and its diagnosis. Mutat Res -Fundam Mol Mech Mutag. 2009 JUL 31;668(1-2):4-10.

70. Thompson LH, Hinz JM. Cellular and molecular consequences of defective fanconi anemia proteins in replication-coupled DNA repair: Mechanistic insights. Mutat Res - Fundam Mol Mech Mutag. 2009 JUL 31;668(1-2):54-72.

71. Cheung I, Schertzer M, Rose A, Lansdorp PM. Disruption of dog-1 in caenorhabditis elegans triggers deletions upstream of guanine-rich DNA. Nat Genet. 2002 AUG;31(4):405-9.

72. Ding H, Schertzer M, Wu XL, Gertsenstein M, Selig S, Kammori M, et al. Regulation of murine telomere length by rtel: An essential gene encoding a helicase-like protein. Cell. 2004 JUN 25;117(7):873-86.

73. Barber LJ, Youds JL, Ward JD, McIlwraith MJ, O'Neil NJ, Petalcorin MIR, et al. RTEL1 maintains genomic stability by suppressing homologous recombination. Cell. 2008 OCT 17;135(2):261-71.

74. Masuda-Sasa T, Polaczek P, Peng XP, Chen L, Campbell JL. Processing of G4 DNA by Dna2 helicase/nuclease and replication protein a (RPA) provides insights into the mechanism of Dna2/RPA substrate recognition. J Biol Chem. 2008 SEP 5;283(36):24359-73.

75. Masuda-Sasa T, Imamura O, Campbell JL. Biochemical analysis of human Dna2. Nucleic Acids Res. 2006;34(6):1865-75.

76. Kao HI, Campbell JL, Bambara RA. Dna2p helicase/nuclease is a tracking protein, like FEN1, for flap cleavage during okazaki fragment maturation. J Biol Chem. 2004 DEC 3;279(49):50840-9.

77. Kang HY, Choi E, Bae SH, Lee KH, Gim BS, Kim HD, et al. Genetic analyses of schizosaccharomyces pombe dna2(+) reveal that Dna2 plays an essential role in okazaki fragment metabolism. Genetics. 2000 JUL;155(3):1055-67.

78. Weitao T, Budd M, Hoopes LLM, Campbell JL. Dna2 helicase/nuclease causes replicative fork stalling and double-strand breaks in the ribosomal DNA of saccharomyces cerevisiae. J Biol Chem. 2003 JUN 20;278(25):22513-22.

79. Choe W, Budd M, Imamura O, Hoopes L, Campbell JL. Dynamic localization of an okazaki fragment processing protein suggests a novel role in telomere replication. Mol Cell Biol. 2002 JUN;22(12):4202-17.

80. Tomita K, Kibe T, Kang HY, Seo YS, Uritani M, Ushimaru T, et al. Fission yeast Dna2 is required for generation of the telomeric single-strand overhang. Mol Cell Biol. 2004 NOV;24(21):9557-67.

81. Paeschke K, McDonald KR, Zakian VA. Telomeres: Structures in need of unwinding. FEBS Lett. 2010 SEP 10;584(17):3760-72.

82. Fry M. Tetraplex DNA and its interacting proteins. Front Biosci. 2007 MAY 1;12:4336-51.

83. Raghavan SC, Swanson PC, Wu XT, Hsieh CL, Lieber MR. A non-B-DNA structure at the bcl-2 major breakpoint region is cleaved by the RAG complex. Nature. 2004 MAR 4;428(6978):88-93.

84. Bonaglia MC, Giorda R, Mani E, Aceti G, Anderlid B, Baroncini A, et al. Identification of a recurrent breakpoint within the SHANK3 gene in the 22q13.3 deletion syndrome. J Med Genet. 2006 OCT;43(10):822-8.

85. Wells RD. Molecular basis of genetic instability of triplet repeats. J Biol Chem. 1996 FEB 9;271(6):2875-8.

86. Sinden RR. Biological implications of the DNA structures associated with disease causing triplet repeats. Am J Hum Genet. 1999 FEB;64(2):346-53.

87. Darnell JC, Jensen KB, Jin P, Brown V, Warren ST, Darnell RB. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. Cell. 2001 NOV 16;107(4):489-99.

88. Moine H, Mandel JL. Biomedicine - do G quartets orchestrate fragile X pathology? Science. 2001 DEC 21;294(5551):2487-8.

89. Brown V, Jin P, Ceman S, Darnell JC, O'Donnell WT, Tenenbaum SA, et al. Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. Cell. 2001 NOV 16;107(4):477-87.

90. Blackburn EH. Structure and function of telomeres. Nature. 1991 APR 18;350(6319):569-73.

91. McEachern MJ, Krauskopf A, Blackburn EH. Telomeres and their control. Annu Rev Genet. 2000;34:331-58.

92. Palm W, de Lange T. How shelterin protects mammalian telomeres. Annu Rev Genet. 2008;42:301-34.

93. Makarov VL, Hirose Y, Langmore JP. Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. Cell. 1997 MAR 7;88(5):657-66.

94. McElligott R, Wellinger RJ. The terminal DNA structure of mammalian chromosomes. EMBO J. 1997 JUN 16;16(12):3705-14.

95. Wright WE, Tesmer VM, Liao ML, Shay JW. Normal human telomeres are not late replicating. Exp Cell Res. 1999 SEP 15;251(2):492-9.

96. Neidle S, Parkinson G. Telomere maintenance as a target for anticancer drug discovery. Nature Reviews Drug Discovery. 2002 MAY;1(5):383-93.

97. Klobutcher LA, Swanton MT, Donini P, Prescott DM. All gene-sized dna-molecules in 4 species of hypotrichs have the same terminal sequence and an unusual 3' terminus. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences. 1981;78(5):3015-9.

98. Blackburn EH, Szostak JW. The molecular-structure of centromeres and telomeres. Annu Rev Biochem. 1984;53:163-94.

99. Zakian VA. Structure and function of telomeres. Annu Rev Genet. 1989;23:579-604.

100. Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H, et al. Mammalian telomeres end in a large duplex loop. Cell. 1999 MAY 14;97(4):503-14.

101. Lacroix L, Lienard H, Labourier E, Djavaheri-Mergny M, Lacoste J, Leffers H, et al. Identification of two human nuclear proteins that recognise the cytosine-rich strand of human telomeres in vitro. Nucleic Acids Res. 2000 APR 1;28(7):1564-75.

102. Wang Y, Patel DJ. Solution structure of the human telomeric repeat D[ag(3)(t(2)ag(3))3] G-tetraplex. Structure. 1993;1(4):263-82.

103. Parkinson GN, Lee MPH, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. Nature. 2002;417(6891):876-80.

104. Phan AT, Kuryavyi V, Luu KN, Patel DJ. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K+ solution. Nucleic Acids Res. 2007 OCT;35(19):6517-25.

105. Dai J, Carver M, Punchihewa C, Jones RA, Yang D. 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 Res. 2007 AUG;35(15):4927-40.

106. Wang Y, Patel DJ. Solution structure of a parallel-stranded G-quadruplex dna. J Mol Biol. 1993 DEC 20;234(4):1171-83.

107. Prescott DM. The dna of ciliated protozoa. Microbiol Rev. 1994 JUN;58(2):233-67.

108. Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps HJ, Pluckthun A. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with stylonychia lemnae macronuclei. Proc Natl Acad Sci U S A. 2001 JUL 17;98(15):8572-7.

109. Paeschke K, Simonsson T, Postberg J, Rhodes D, Lipps HJ. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. Nature Structural & Molecular Biology. 2005 OCT;12(10):847-54.

110. Smith JS, Chen Q, Yatsunyk LA, Nicoludis JM, Garcia MS, Kranaster R, et al. Rudimentary G-quadruplex-based telomere capping in saccharomyces cerevisiae. Nat Struct Mol Biol. 2011 APR;18(4):478-U119.

111. Lustig AJ. Cdc13 subcomplexes regulate multiple telomere functions. Nat Struct Biol. 2001 APR;8(4):297-9.

112. Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, et al. A highly conserved repetitive dna-sequence, (ttaggg)n, present at the telomeres of human-chromosomes. Proc Natl Acad Sci U S A. 1988 SEP;85(18):6622-6.

113. Delange T, Shiue L, Myers RM, Cox DR, Naylor SL, Killery AM, et al. Structure and variability of human-chromosome ends. Mol Cell Biol. 1990 FEB;10(2):518-27.

114. Kipling D, Cooke HJ. Hypervariable ultra-long telomeres in mice. Nature. 1990 SEP 27;347(6291):400-2.

115. Lejnine S, Makarov VL, Langmore JP. Conserved nucleoprotein structure at the ends of vertebrate and invertebrate chromosomes. Proc Natl Acad Sci U S A. 1995 MAR 14;92(6):2393-7.

116. Watson JD. Origin of concatemeric T7 dna. Nature-New Biology. 1972;239(94):197,&.

117. Olovniko.Am. Theory of marginotomy - incomplete copying of template margin in enzymic-synthesis of polynucleotides and biological significance of phenomenon. J Theor Biol. 1973;41(1):181-90.

118. Harley CB, Futcher AB, Greider CW. Telomeres shorten during aging of human fibroblasts. Nature. 1990 MAY 31;345(6274):458-60.

119. Wu A, Ichihashi M, Ueda M. Correlation of the expression of human telomerase subunits with telomerase activity in normal skin and skin tumors. Cancer. 1999 NOV 15;86(10):2038-44.

120. Fujioka T, Hasegawa M, Suzuki Y, Suzuki T, Sugimura J, Tanji S, et al. Telomerase activity in human renal cell carcinoma. Int J Urol. 2000 JAN;7(1):16-21.

121. Shimada M, Hasegawa H, Gion T, Utsunomiya T, Shirabe K, Takenaka K, et al. The role of telomerase activity in hepatocellular carcinoma. Am J Gastroenterol. 2000 MAR;95(3):748-52.

122. Bryan TM, Reddel RR. Telomere dynamics and telomerase activity in in vitro immortalised human cells. Eur J Cancer. 1997 APR;33(5):767-73.

123. Hurley LH, Wheelhouse RT, Sun D, Kerwin SM, Salazar M, Fedoroff OY, et al. G-quadruplexes as targets for drug design. Pharmacol Ther. 2000 MAR;85(3):141-58.

124. Jenkins TC. Targeting multi-stranded DNA structures. Curr Med Chem. 2000 JAN;7(1):99-115.

125. de Lange T, Jacks T. For better or worse? telomerase inhibition and cancer. Cell. 1999 AUG 6;98(3):273-5.

126. Mergny JL, Mailliet P, Lavelle F, Riou JF, Laoui A, Helene C. The development of telomerase inhibitors: The G-quartet approach. Anticancer Drug Des. 1999 AUG;14(4):327-39.

127. Neidle S, Kelland LR. Commentary: Telomerase as an anti-cancer target: Current status and future prospects. Anticancer Drug Des. 1999 AUG;14(4):341-7.

128. Wang J, Hannon GJ, Beach DH. Cell biology - risky immortalization by telomerase. Nature. 2000 JUN 15;405(6788):755-6.

129. Zahler AM, Williamson JR, Cech TR, Prescott DM. Inhibition of telomerase by G-quartet dna structures. Nature. 1991 APR 25;350(6320):718-20.

130. Neidle S, Parkinson G. Telomere maintenance as a target for anticancer drug discovery. Nat Rev Drug Discov. 2002 MAY;1(5):383-93.

131. Han HY, Hurley LH. G-quadruplex DNA: A potential target for anti-cancer drug design. Trends Pharmacol Sci. 2000 APR;21(4):136-42.

132. Perry PJ, Arnold JRP, Jenkins TC. Telomerase inhibitors for the treatment of cancer: The current perspective. Expert Opin Investig Drugs. 2001 DEC;10(12):2141-56.

133. Kerwin SM. G-quadruplex DNA as a target for drug design. Curr Pharm Des. 2000 MAR;6(4):441-71.

134. Chen Z, Corey DR. Telomerase inhibitors: A new option for chemotherapy. Adv Cancer Res. 2003;87:31-58.

135. Sun DY, Thompson B, Cathers BE, Salazar M, Kerwin SM, Trent JO, et al. Inhibition of human telomerase by a G-quadruplex-interactive compound. J Med Chem. 1997 JUL 4;40(14):2113-6.

136. Fedoroff OY, Salazar M, Han HY, Chemeris VV, Kerwin SM, Hurley LH. NMR-based model of a telomerase-inhibiting compound bound to G-quadruplex DNA. Biochemistry (N Y). 1998 SEP 8;37(36):12367-74.

137. Han HY, Cliff CL, Hurley LH. Accelerated assembly of G-quadruplex structures by a small molecule. Biochemistry (N Y). 1999 JUN 1;38(22):6981-6.

138. Arthanari H, Basu S, Kawano TL, Bolton PH. Fluorescent dyes specific for quadruplex DNA. Nucleic Acids Res. 1998 AUG 15;26(16):3724-8.

139. Anantha NV, Azam M, Sheardy RD. Porphyrin binding to quadruplexed T(4)G(4). Biochemistry (N Y). 1998 MAR 3;37(9):2709-14.

140. Wheelhouse RT, Sun DK, Han HY, Han FXG, Hurley LH. Cationic porphyrins as telomerase inhibitors: The interaction of tetra-(N-methyl-4-pyridyl)porphine with quadruplex DNA. J Am Chem Soc. 1998 APR 8;120(13):3261-2.

141. Izbicka E, Wheelhouse RT, Raymond E, Davidson KK, Lawrence RA, Sun DY, et al. Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. Cancer Res. 1999 FEB 1;59(3):639-44.

142. Shin-ya K, Wierzba K, Matsuo K, Ohtani T, Yamada Y, Furihata K, et al. Telomestatin, a novel telomerase inhibitor from streptomyces anulatus. J Am Chem Soc. 2001 FEB 14;123(6):1262-3.

143. Kim MY, Gleason-Guzman M, Izbicka E, Nishioka D, Hurley LH. The different biological effects of telomestatin and TMPyP4 can be attributed to their selectivity for interaction with intramolecular or intermolecular G-quadruplex structures. Cancer Res. 2003 JUN 15;63(12):3247-56.

144. Kang C, Zhang XH, Ratliff R, Moyzis R, Rich A. Crystal-structure of 4-stranded oxytricha telomeric dna. Nature. 1992 MAR 12;356(6365):126-31.

145. Kettani A, Kumar RA, Patel DJ. Solution structure of a dna quadruplex containing the fragile-X syndrome triplet repeat. J Mol Biol. 1995 DEC 8;254(4):638-56.

146. Simonsson T. G-quadruplex DNA structures - variations on a theme. Biol Chem. 2001 APR;382(4):621-8.

147. Smith F,Flint W.Juli. Quadruplex structure of oxytricha telomeric DNA oligonucleotides. Nature. 1992 03/12;356(6365):164-8.

148. Carrasco C, Rosu F, Gabelica V, Houssier C, De Pauw E, Garbay-Jaureguiberry C, et al. Tight binding of the antitumor drug ditercalinium to quadruplex DNA. Chembiochem. 2002 DEC 2;3(12):1235-41.

149. Wu Y, Rawtani N, Thazhathveetil AK, Kenny MK, Seidman MM, Brosh RM. Human replication protein A melts a DNA triple helix structure in a potent and specific manner. Biochemistry (N Y). 2008;47(18):5068-77.

150. Pantophlet R, Saphire EO, Poignard P, Parren PWHI, Wilson IA, Burton DR. Fine mapping of the interaction of neutralizing and nonneutralizing monoclonal antibodies with the CD4 binding site of human immunodeficiency virus type 1 gp120. J Virol. 2003 JAN;77(1):642-58.

151. Lin CW, Wu SC. A functional epitope determinant on domain III of the japanese encephalitis virus envelope protein interacted with neutralizing-antibody combining sites. J Virol. 2003 FEB;77(4):2600-6.

152. Wang ZD, Raifu M, Howard M, Smith L, Hansen D, Goldsby R, et al. Universal PCR amplification of mouse immunoglobulin gene variable regions: The design of degenerate primers and an assessment of the effect of DNA polymerase 3 ' to 5 ' exonuclease activity. J Immunol Methods. 2000 JAN 13;233(1-2):167-77.

153. Balagurumoorthy P, Brahmachari SK. Structure and stability of human telomeric sequence. Journal of Biological Chemistry. 1994 August 26;269(34):21858-69.

154. Macaya RF, Schultze P, Smith FW, Roe JA, Feigon J. Thrombin-binding dna aptamer forms a unimolecular quadruplex structure in solution. Proc Natl Acad Sci U S A. 1993 APR 15;90(8):3745-9.

155. Smith FW, Feigon J. Quadruplex structure of oxytricha telomeric dna oligonucleotides. Nature. 1992;356(6365):164-8.

156. Cheong CJ, Moore PB. Solution structure of an unusually stable rna tetraplex containing G-quartet and U-quartet structures. Biochemistry (N Y). 1992;31(36):8406-14.

157. Dapic V, Abdomerovic V, Marrington R, Peberdy J, Rodger A, Trent JO, et al. Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. Nucleic Acids Res. 2003 APR 15;31(8):2097-107.

158. Balagurumoorthy P, Brahmachari SK, Mohanty D, Bansal M, Sasisekharan V. Hairpin and parallel quartet structures for telomeric sequences. Nucleic Acids Research. 1992 August 11;20(15):4061-7.

159. Giraldo R, Suzuki M, Chapman L, Rhodes D. Promotion of parallel dna quadruplexes by a yeast telomere binding-protein - a circular-dichroism study. Proc Natl Acad Sci U S A. 1994 AUG 2;91(16):7658-62.

160. Olsen CM, Gmeiner WH, Marky LA. Unfolding of G-quadruplexes: Energetic, and ion and water contributions of G-quartet stacking. J Phys Chem B. 2006 APR 6;110(13):6962-9.

161. Marathias VM, Bolton PH. Determinants of DNA quadruplex structural type: Sequence and potassium binding. Biochemistry (N Y). 1999 APR 6;38(14):4355-64.

162. Ying LM, Green JJ, Li HT, Klenerman D, Balasubramanian S. Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. Proc Natl Acad Sci U S A. 2003 DEC 9;100(25):14629-34.

163. Risitano A, Fox KR. Inosine substitutions demonstrate that intramolecular DNA quadruplexes adopt different conformations in the presence of sodium and potassium. Bioorg Med Chem Lett. 2005 APR 15;15(8):2047-50.

164. Keniry MA. Quadruplex structures in nucleic acids. Biopolymers. 2000;56(3):123-46.

165. Patel PK, Hosur RV. NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by saccharomyces cerevisiae telomere repeats. Nucleic Acids Res. 1999 JUN 15;27(12):2457-64.

166. Laughlan G, Murchie AIH, Norman DG, Moore MH, Moody PCE, Lilley DMJ, et al. The high-resolution crystal-structure of a parallel-stranded guanine tetraplex. Science. 1994 JUL 22;265(5171):520-4.

167. Jin RZ, Gaffney BL, Wang C, Jones RA, Breslauer KJ. Thermodynamics and structure of a dna tetraplex - a spectroscopic and calorimetric study of the tetramolecular complexes of D(tg3t) and D(tg3t2g3t). Proc Natl Acad Sci U S A. 1992 SEP 15;89(18):8832-6.

168. Dapic V, Abdomerovic V, Marrington R, Peberdy J, Rodger A, Trent JO, et al. Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. Nucleic Acids Res. 2003 APR 15;31(8):2097-107.

169. Kaushik M, Bansal A, Saxena S, Kukreti S. Possibility of an antiparallel (tetramer) quadruplex exhibited by the double repeat of the human telomere. Biochemistry (N Y). 2007 JUN 19;46(24):7119-31.

170. Zhang N, Phan AT, Patel DJ. (3+1) assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. J Am Chem Soc. 2005 DEC 14;127(49):17277-85.

171. Brown JC, Brown BA, Li YQ, Hardin CC. Construction and characterization of a quadruplex DNA selective single-chain autoantibody from a viable motheaten mouse hybridoma with homology to telomeric DNA binding proteins. Biochemistry (N Y). 1998 NOV 17;37(46):16338-48.

172. Brown BA, Li YQ, Brown JC, Hardin CC, Roberts JF, Pelsue SC, et al. Isolation and characterization of a monoclonal anti-quadruplex DNA antibody from autoimmune "viable motheaten" mice. Biochemistry (N Y). 1998 NOV 17;37(46):16325-37.

173. Lafer EM, Moller A, Nordheim A, Stollar BD, Rich A. Antibodies specific for left-handed Z-dna. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences. 1981;78(6):3546-50.

174. Nordheim A, Pardue ML, Lafer EM, Moller A, Stollar BD, Rich A. Antibodies to left-handed Z-dna bind to interband regions of drosophila polytene chromosomes. Nature. 1981;294(5840):417-22.

175. Lafer EM, Sousa R, Ali R, Rich A, Stollar BD. The effect of anti-Z-dna antibodies on the B-dna-Z-dna equilibrium. J Biol Chem. 1986 MAY 15;261(14):6438-43.

176. Fernando H, Rodriguez R, Balasubramanian S. Selective recognition of a DNA G-quadruplex by an engineered antibody. Biochemistry (N Y). 2008 SEP 9;47(36):9365-71.

177. Chen FM. Sr2+ facilitates intermolecular G-quadruplex formation of telomeric sequences. Biochemistry (N Y). 1992 APR 21;31(15):3769-76.

178. Petracek ME, Berman J. Chlamydomonas-reinhardtii telomere repeats form unstable structures involving guanine-guanine base-pairs. Nucleic Acids Res. 1992 JAN 11;20(1):89-95.

179. Sen D, Gilbert W. Novel dna superstructures formed by telomere-like oligomers. Biochemistry (N Y). 1992 JAN 14;31(1):65-70.

180. Isalan M, Patel SD, Balasubramanian S, Choo Y. Selection of zinc fingers that bind single-stranded telomeric DNA in the G-quadruplex conformation. Biochemistry (N Y). 2001 JAN 23;40(3):830-6.

181. Guth AM, Zhang XH, Smith D, Detanico T, Wysocki LJ. Chromatin specificity of anti-double-stranded DNA antibodies and a role for arg residues in the third complementarity-determining region of the heavy chain. Journal of Immunology. 2003 DEC 1;171(11):6260-6.

182. Krishnan MR, Jou NT, Marion TN. Correlation between the amino acid position of arginine in VH-CDR3 and specificity for native DNA among autoimmune antibodies. Journal of Immunology. 1996 SEP 15;157(6):2430-9.

183. Oganesian L, Moon IK, Bryan TM, Jarstfer MB. Extension of G-quadruplex DNA by ciliate telomerase. EMBO J. 2006 MAR 8;25(5):1148-59.

184. Zaug AJ, Podell ER, Cech TR. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. Proc Natl Acad Sci U S A. 2005 AUG 2;102(31):10864-9.

185. Fang GW, Cech TR. The beta-subunit of oxytricha telomere-binding protein promotes G-quartet formation by telomeric dna. Cell. 1993 SEP 10;74(5):875-85.

186. Li J, Correia JJ, Wang L, Trent JO, Chaires JB. Not so crystal clear: The structure of the human telomere G-quadruplex in solution differs from that present in a crystal. Nucleic Acids Res. 2005;33(14):4649-59.

187. Bashkirov VI, Scherthan H, Solinger JA, Buerstedde JM, Heyer WD. A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. J Cell Biol. 1997 FEB 24;136(4):761-73.

188. Sundquist WI, Heaphy S. Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus-1 genomic rna. Proc Natl Acad Sci U S A. 1993 APR 15;90(8):3393-7.

189. Christiansen J, Kofod M, Nielsen FC. A guanosine quadruplex and 2 stable hairpins flank a major cleavage site in insulin-like growth-factor-ii messenger-rna. Nucleic Acids Res. 1994 DEC 25;22(25):5709-16.

190. Horsburgh BC, Kollmus H, Hauser H, Coen DM. Translational recoding induced by G-rich mRNA sequences that form unusual structures. Cell. 1996 SEP 20;86(6):949-59.

191. Sharma S, Raymond E, Soda H, Sun D, Hilsenbeck SG, Sharma A, et al. Preclinical and clinical strategies for development of telomerase and telomere inhibitors. Annals of Oncology. 1997;8(11):1063-74.

192. Perry PJ, Kelland LR. Telomeres and telomerase: Targets for cancer chemotherapy? Expert Opinion on Therapeutic Patents. 1998;8(12):1567-86.

193. Paeschke K, Capra JA, Zakian VA. DNA replication through G-quadruplex motifs is promoted by the saccharomyces cerevisiae Pif1 DNA helicase. Cell. 2011 MAY 27;145(5):678-91.

194. Uribe DJ, Guo K, Shin Y, Sun D. Heterogeneous nuclear ribonucleoprotein K and nucleolin as transcriptional activators of the vascular endothelial growth factor promoter through interaction with secondary DNA structures. Biochemistry (N Y). 2011 MAY 10;50(18):3796-806.

195. Law MJ, Lower KM, Voon HPJ, Hughes JR, Garrick D, Viprakasit V, et al. ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. Cell. 2010 OCT 29;143(3):367-78.

196. Lambert J, Baetz K, Figeys D. Of proteins and DNA-proteomic role in the field of chromatin research. Mol Biosyst. 2010;6(1):30-7.

197. Lambert J, Fillingham J, Siahbazi M, Greenblatt J, Baetz K, Figeys D. Defining the budding yeast chromatin-associated interactome. Mol Syst Biol. 2010 DEC;6:448.

198. Cogoi S, Paramasivam M, Membrino A, Yokoyama KK, Xodo LE. The KRAS promoter responds to myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. J Biol Chem. 2010 JUL 16;285(29):22003-16.

199. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. 2007 JAN;35(2):406-13.

200. Patel DJ, Phan AT, Kuryavyi V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: Diverse higher order DNA and RNA targets for cancer therapeutics. Nucleic Acids Res. 2007 DEC;35(22):7429-55.

201. Weber M, Davies J, Wittig D, Oakeley E, Haase M, Lam W, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet. 2005 AUG;37(8):853-62.

202. Smith JS, Chen Q, Yatsunyk LA, Nicoludis JM, Garcia MS, Kranaster R, et al. Rudimentary G-quadruplex-based telomere capping in saccharomyces cerevisiae. Nat Struct Mol Biol. 2011 APR;18(4):478-U119.

203. Maizels N. Dynamic roles for G4 DNA in the biology of eukaryotic cells. Nature Structural & Molecular Biology. 2006 DEC;13(12):1055-9.

204. Zou L, Elledge S. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science. 2003 JUN 6;300(5625):1542-8.

205. Schulman M, Wilde CD, Kohler G. A better cell line for making hybridomas secreting specific antibodies. Nature (London). 1978;276(5685):269-70.

206. Zhang Z, Yang X, Meng L, Liu F, Shen C, Yang W. Enhanced amplification of GC-rich DNA with two organic reagents. BioTechniques. 2009 SEP;47(3):775-8.

207. Lane AN. The stability of intramolecular DNA G-quadruplexes compared with other macromolecules. Biochimie. 2011, doi: 10.1016/j.biochi.2011.08.004

208. Cahoon, L. A. & Seifert, H. S. An Alternative DNA Structure Is Necessary for Pilin Antigenic Variation in Neisseria gonorrhoeae. Science. 2009;764-767.

Appendices

Appendix A

A.1   Specificity of 1H6 antibody is confirmed by gel-shift assays
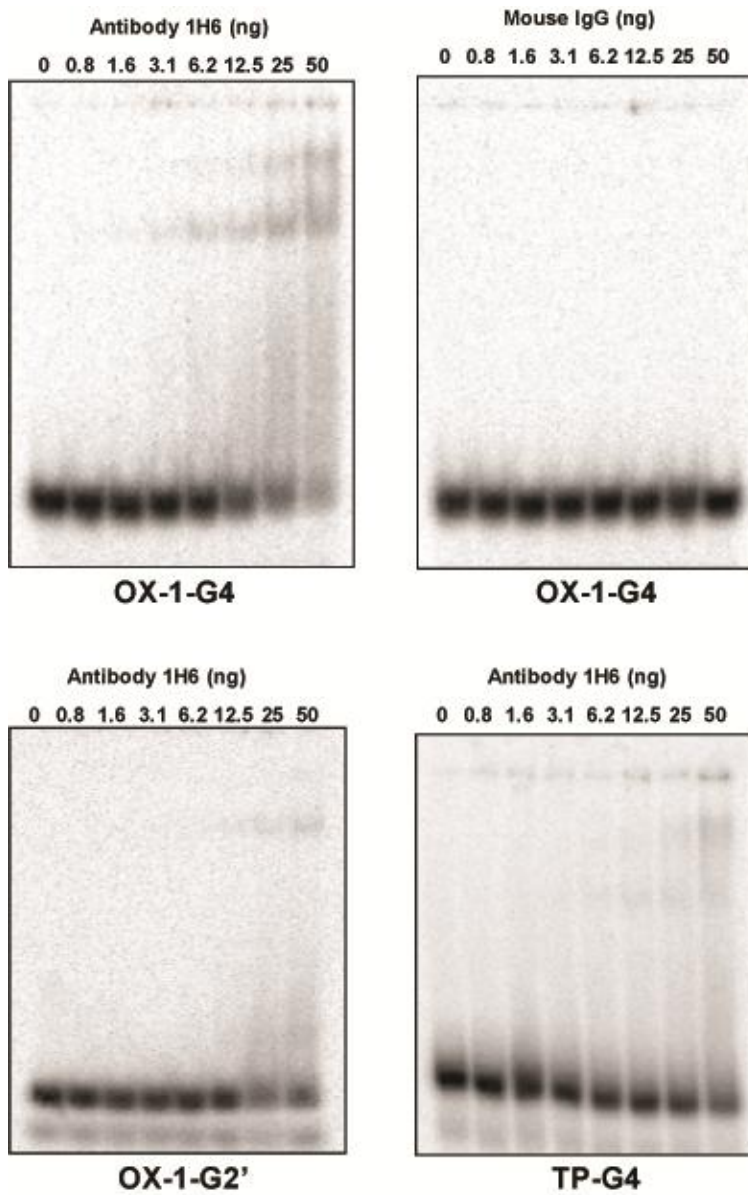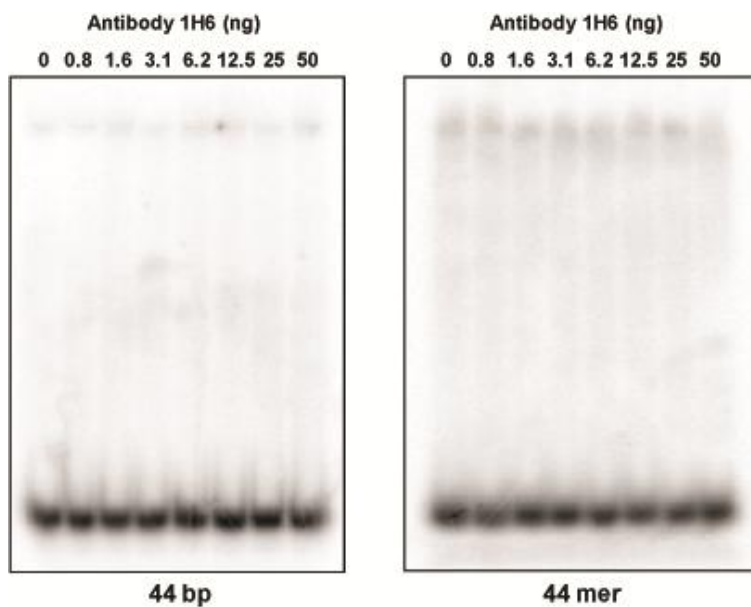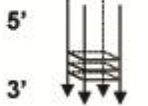
Figure 41. Gel-shift assays indicate 1H6 antibody specificity to various G4 structures
Collaborative results from the Brosh laboratory at NIH. The 1H6 antibody appears to be binding parallel (OX-1-G4 high molecular weight bands) and antiparallel (OX-1-G2' high molecular weight bands) G4 structures. No detectable shift was found with single stranded or double stranded DNA. See table below for substrate details.

Substrate name	Substrate structure	Oligomer name / Sequence (5'-3')
OX-1-G2'	5' / 3' 5'	OX-1: ACTGTCGTACTTGATATTTTT<u>GGGG</u>TTTT<u>GGGG</u>
OX-1-G4	5' / 3'	OX-1: ACTGTCGTACTTGATATTTTT<u>GGGG</u>TTTT<u>GGGG</u>
TP-G4	5' / 3'	TP: TGGACCAGACCTAGCAGCTA<u>GGGGG</u>AGCT<u>GGGG</u>AA<u>GG</u>T<u>GGG</u>AATGTGA
Single-stranded DNA	_____	T_{STEM}: GCACTGGCCGTCGTTTTACGGTCGTGACTGGGAAAACCCTGGCG
Double-stranded DNA	_____ _____	T_{STEM}: GCACTGGCCGTCGTTTTACGGTCGTGACTGGGAAAACCCTGGCG T_{STEM}COMP: CGCCAGGGTTTTCCCAGTCACGACCGTAAAACGACGGCCAGTGC

The guanine residues that compose the G-quartet structures in oligonucleotides are underlined.

Table 9. Oligonucleotides used in gel-shift assays.

Oligonucleotides used in the gel-shift assays and the folded structures that were tested for binding to the 1H6 antibody. Performed in collaboration with Dr. Brosh (NIH).

Appendix B

B.1   Flow cytometry analysis of different fixation techniques using 1H6 antibody
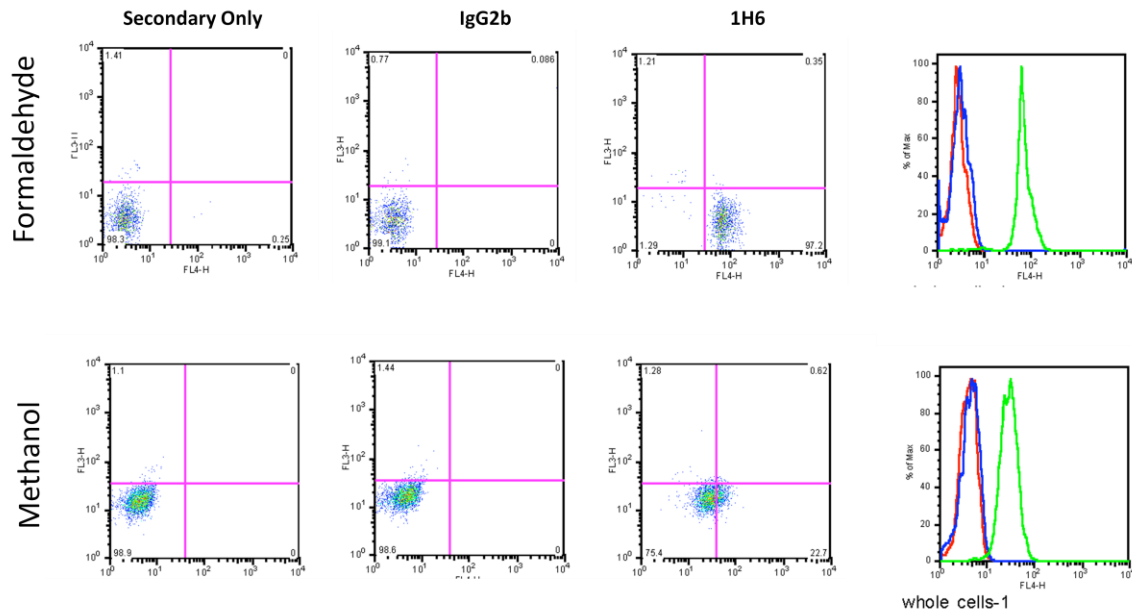
Figure 42. Different fixatives change the absolute mean fluorescent signal detected using 1H6 antibody.
HeLa cells were fixed in either formaldehyde or methanol and prepared for flow cytometry analysis (See Experimental Methodologies). Absolute mean fluorescence of detected using formaldehyde is higher than cells fixed in methanol. Research results provided by Liz Chavez.

B.2   Unfixed cell staining appears the same as fixed cell staining using the 1H6 antibody
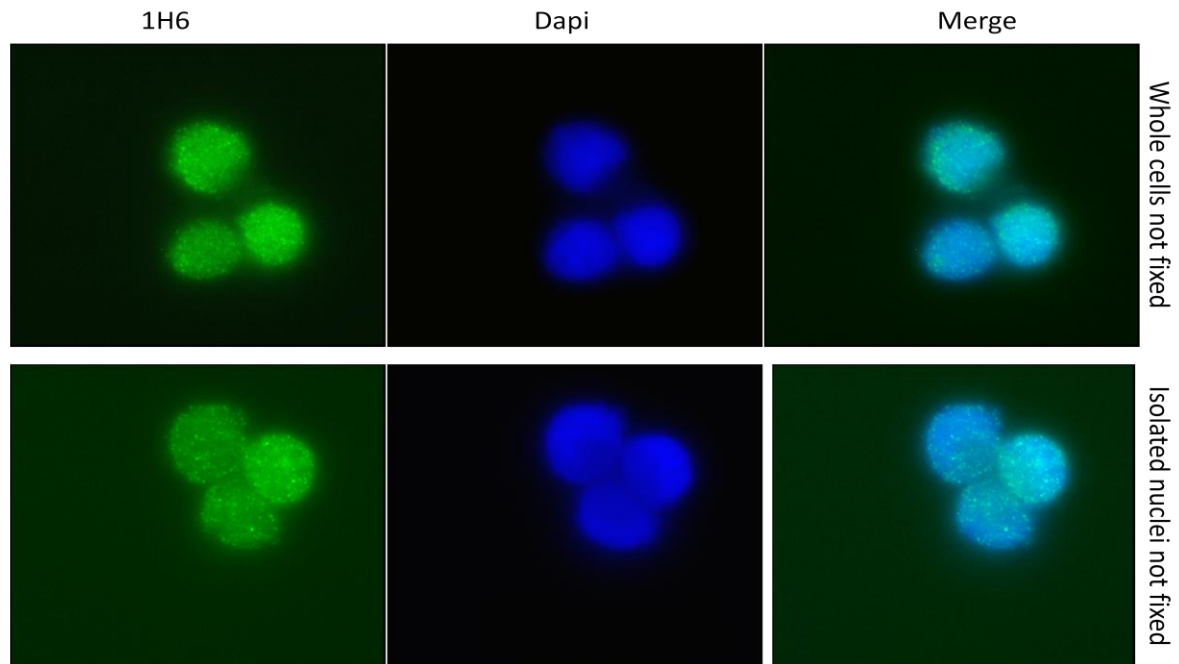
Figure 43. Anti-G-quadruplex antibody 1H6 has clear nuclear staining without fixation of cells.
In collaboration with Liz Chavez HeLa cells were stained and visualized without the use of fixatives. Research results provided by Liz Chavez.

B.3   Treatment of cells with RNAse has no effect on signal detected from 1H6 antibody
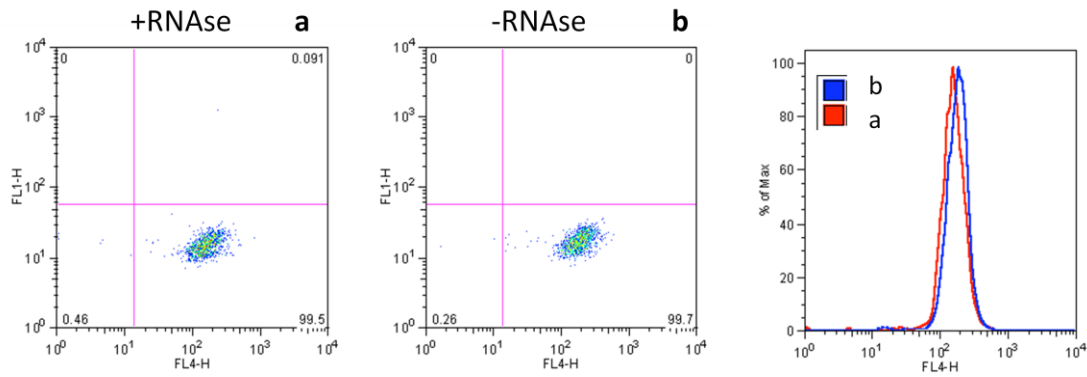
Figure 44. Treatment with RNAse does not affect the overall mean fluorescence signal from 1H6 antibody.

In collaboration with Liz Chavez HeLa cells were fixed and treated with RNAse A prior to staining with 1H6 antibody.

B.4 Different G4 stabilizing drugs increase the mean fluorescence detected using the 1H6 antibody
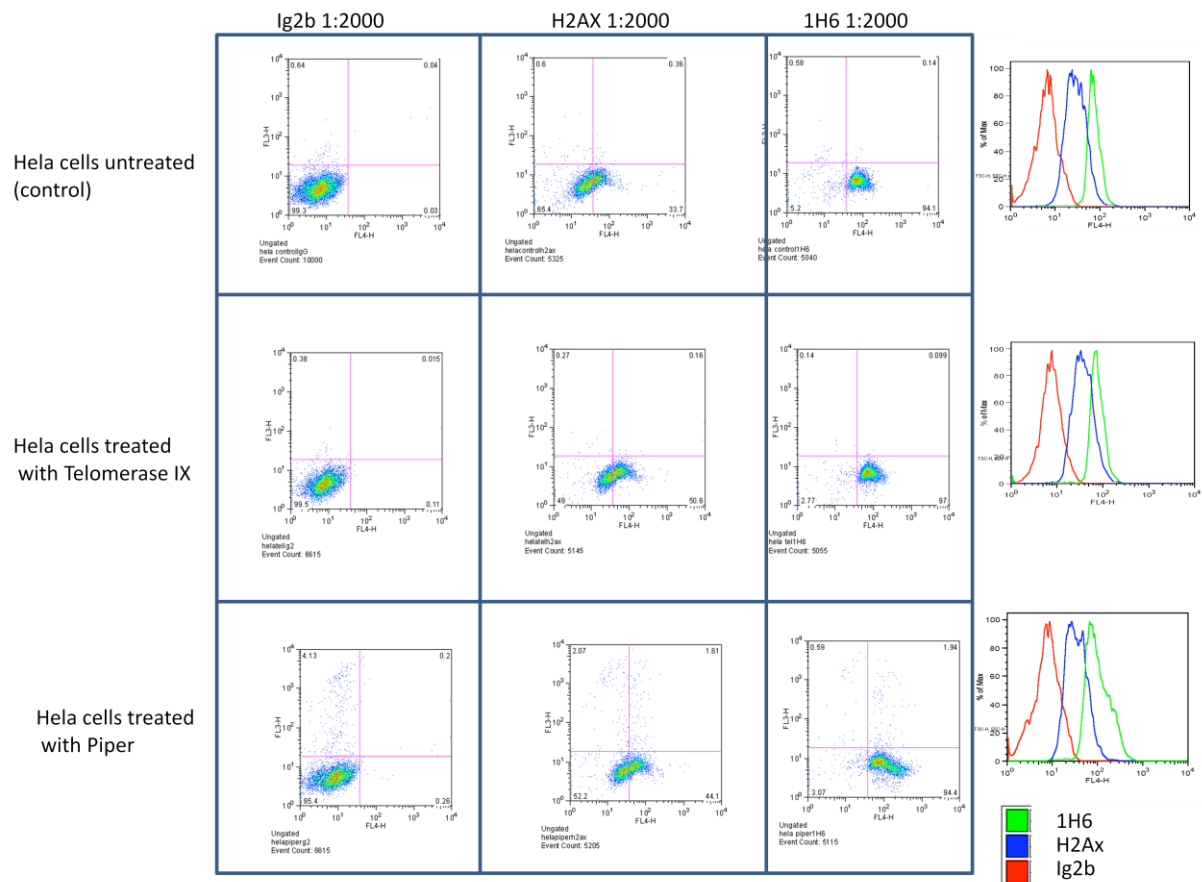
Figure 45. G4 stabilizing drug piper show increased mean fluorescence signal with 1H6 antibody. Treatment of cells with piper increases the mean fluorescence in a subset of cells using the 1H6 antibody. Experiments were in collaboration with Liz Chavez.

B.5   TMPyP4 increases the mean fluorescent signal detected when staining with the 1H6 antibody
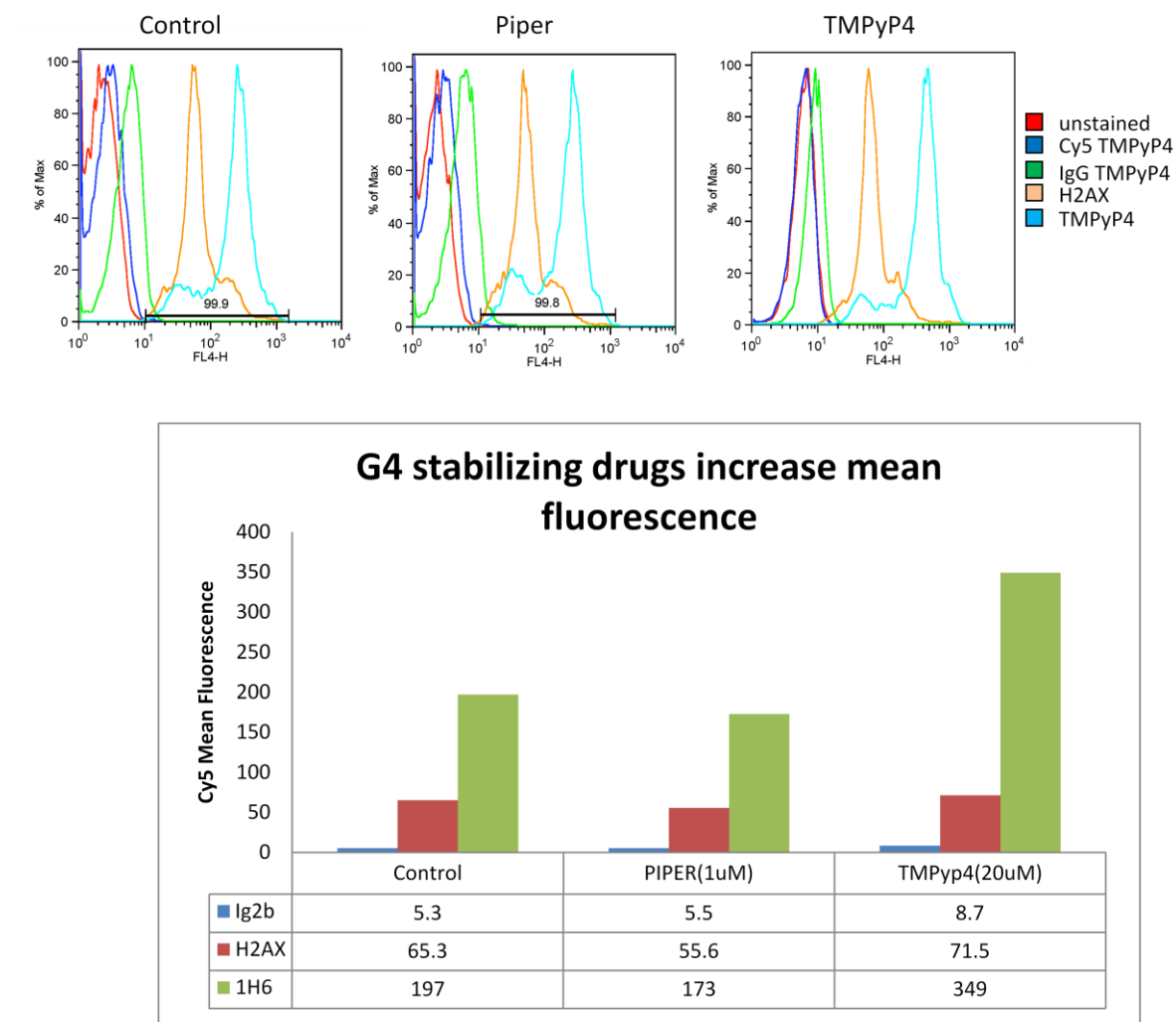
Figure 46. G4 stabilizing drugs increase the mean fluorescence of cells stained with anti-G-quadruplex antibody 1H6.

Fixed HeLa cells treated with TMPyP4 were prepared for flow cytometry analysis and stained with anti-G-quadruplex antibody 1H6. There is an increased mean fluorescence detected from treated cells compared to untreated cells. Experiments were in collaboration with Liz Chavez.

B.6   Time course experiment of U2OS cells treated with telomestatin
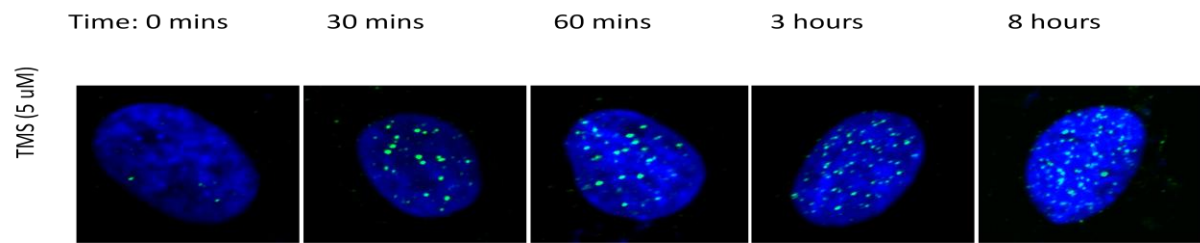
Figure 47. Fluorescence microscopy analysis of fixed cells after time course experiment of telomestatin shows increased number of antibody foci when stained with 1H6 antibody.
U2OS cells were treated with telomestatin for the indicated amount of time prior to fixation and staining with antibody 1H6. Experiments were done in collaboration with Dr. Brosh.
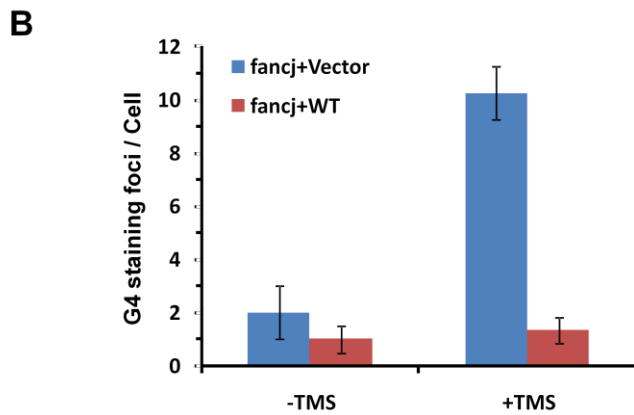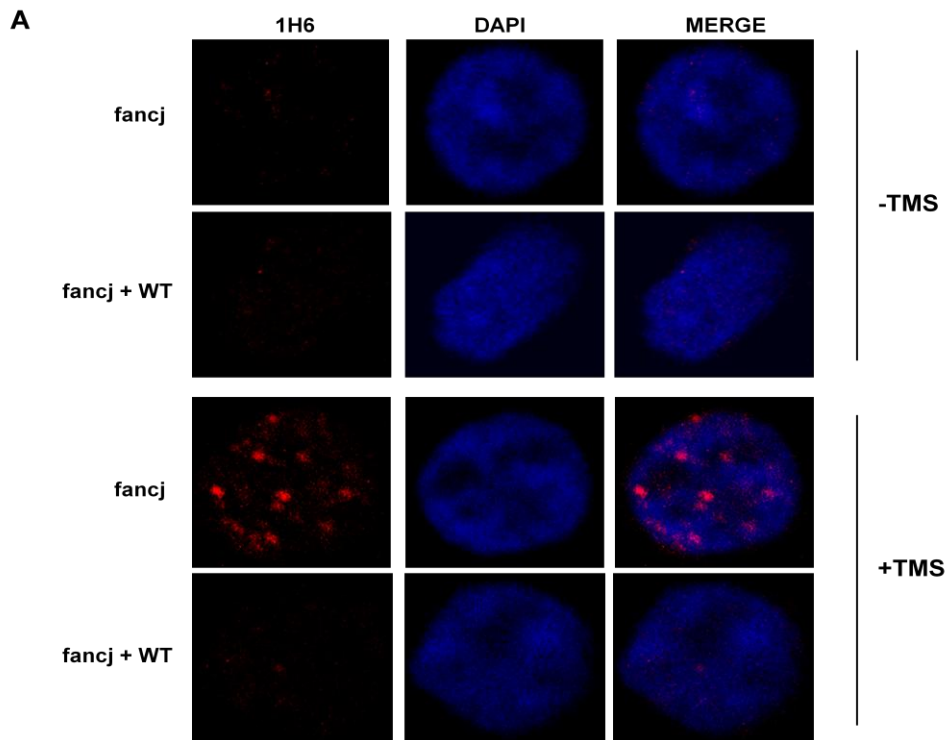
B.7   Staining of Fancj mutant and corrected cells with 1H6

Figure 48. TMS induces elevated numbers of G4 (1H6) foci in fancj DT40 cells.
Fancj DT40 cells or cells expressing human FANCJ-WT were either exposed to 5 μM TMS (telomestatin) for 2 h or left unexposed, fixed with formaldehyde, and stained with anti-G4 antibody (1H6) and DAPI. (B) The graph shows the mean and s.e.m. of the G4 foci in each cell. More than 100 cells were scored for each sample. Experiments were done in collaboration with Dr. Brosh.

B.8   HeLa cells treated with telomestatin create co-localization between 53bp1 and 1H6
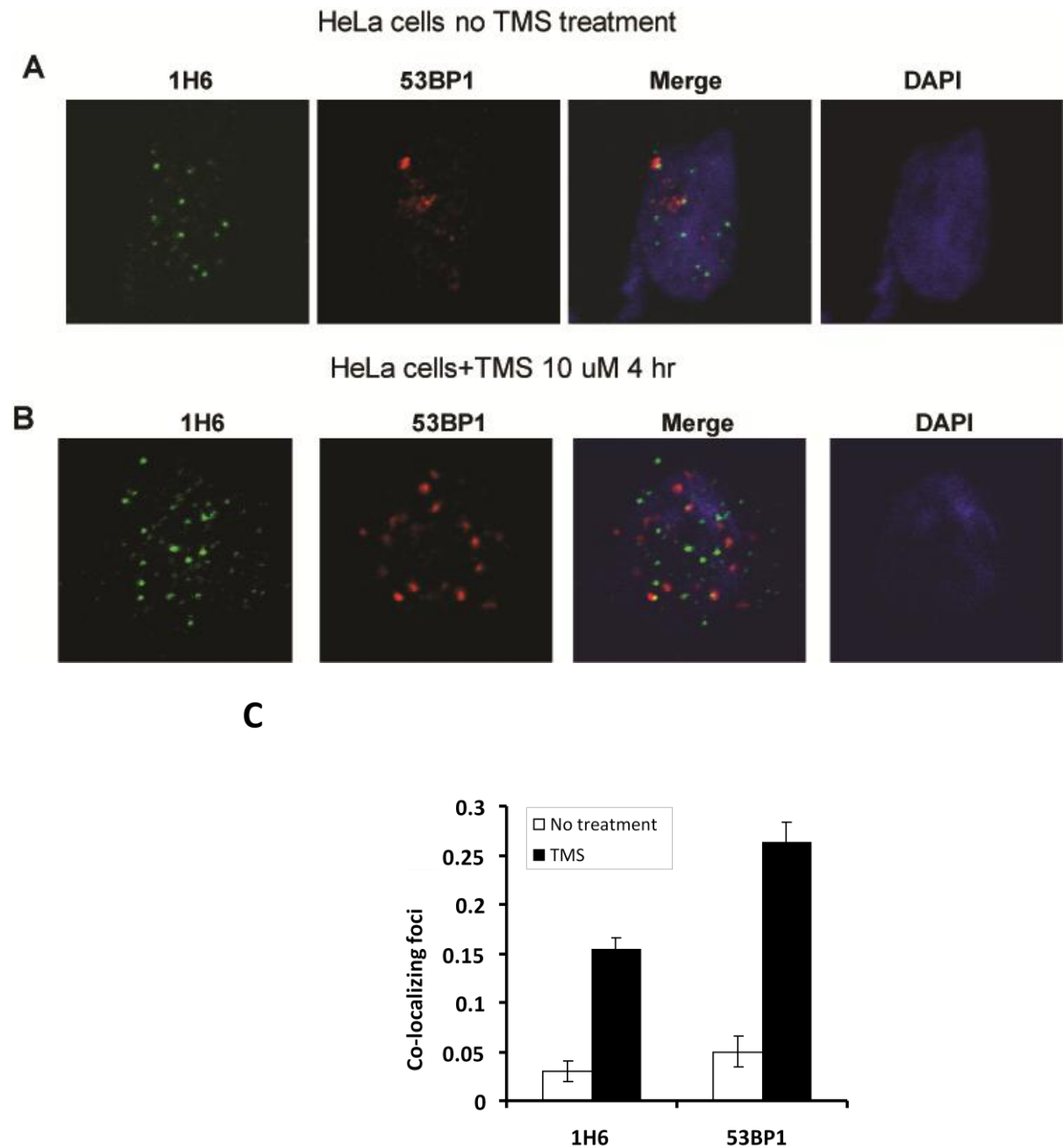

Figure 49. Treatment of HeLa cells with telomestatin increases the co-localization of 1H6 and 53bp1. Experiments were done in collaboration with Dr. Brosh. Increased DNA damage foci colocalize with 1H6 antibody upon treatment with telomestatin.

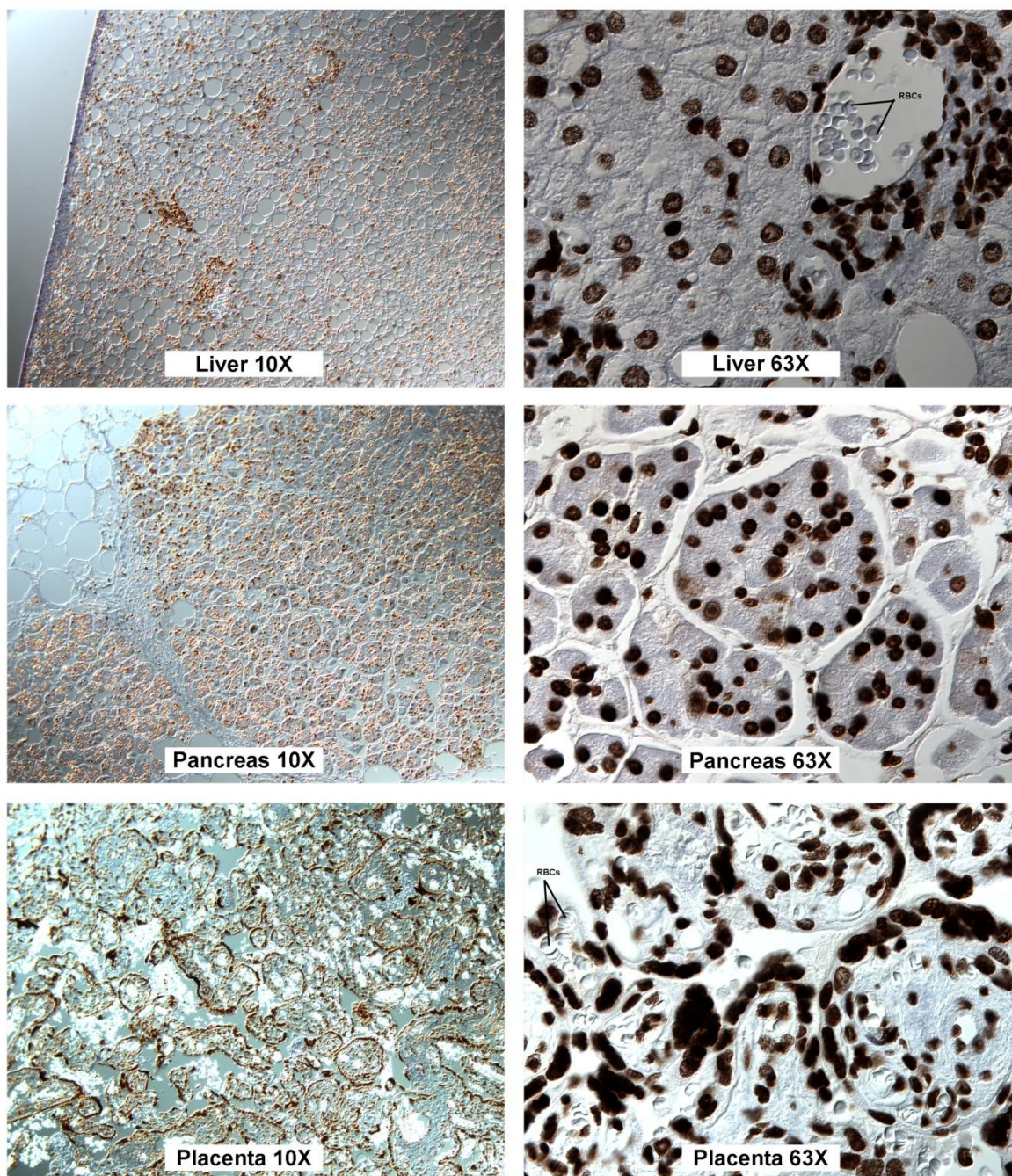## B.9   Normal human tissues stained with anti-G-quadruplex antibody 1H6

Figure 50. Histology of normal human tissue stained with the 1H6 antibody

In collaboration with CTAG normal human tissues were fixed, sectioned and stained with the
1H6 antibody. All nucleated cells stain positive except a subset of cells in the testis.
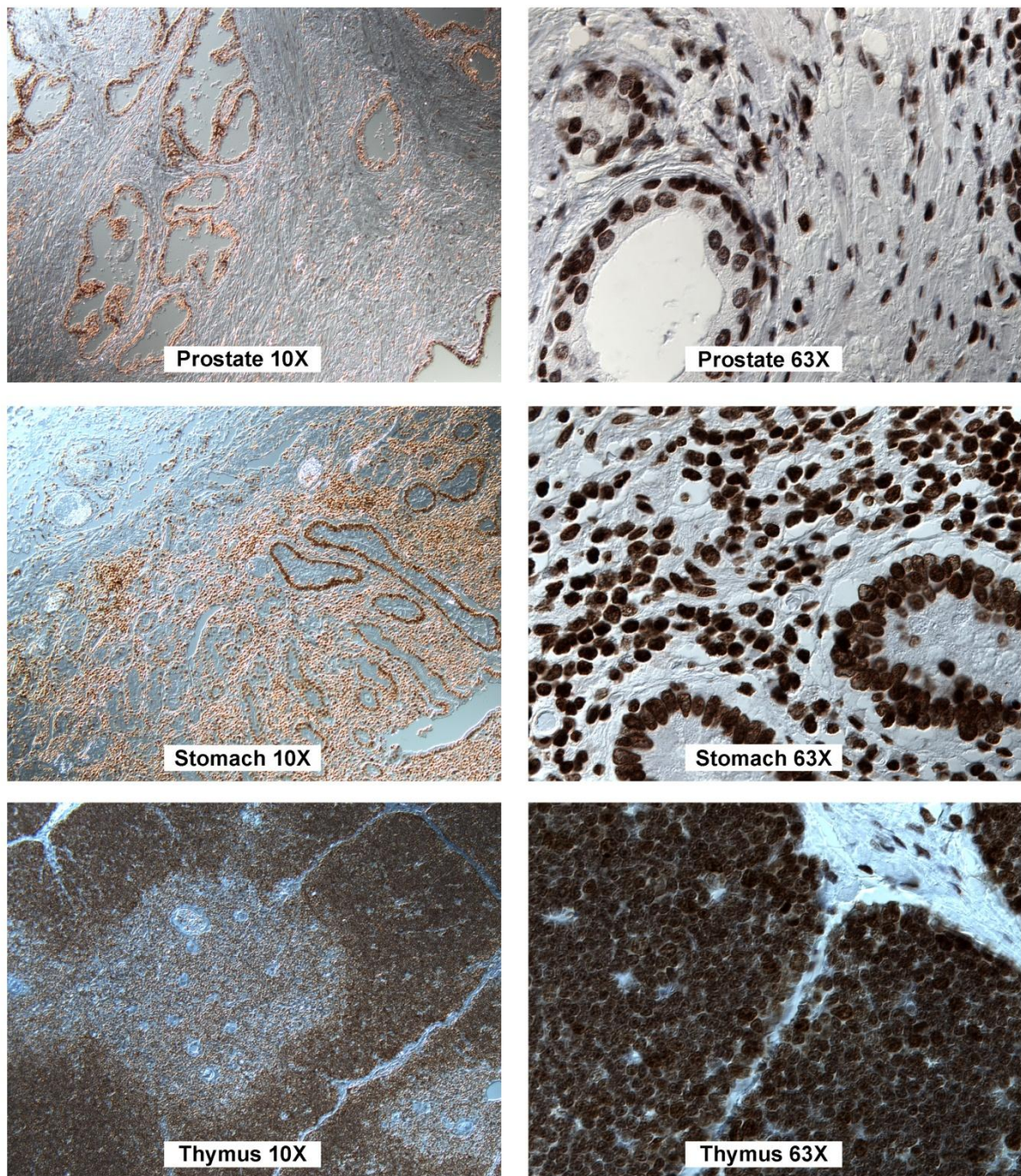
Figure 51. Histology of normal human tissue stained with the 1H6 antibody

In collaboration with CTAG normal human tissues were fixed, sectioned and stained with the 1H6 antibody. All nucleated cells stain positive except a subset of cells in the testis.
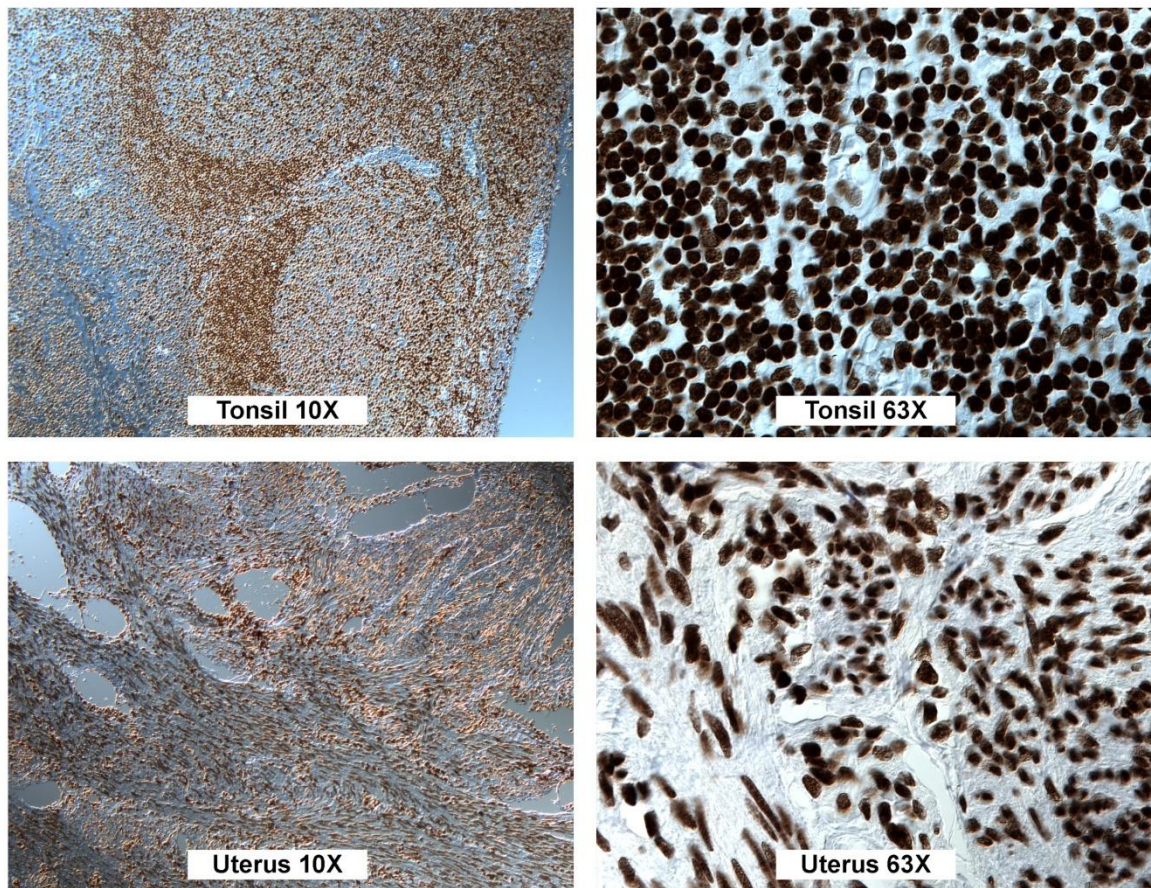
Figure 52. Histology of normal human tissue stained with the 1H6 antibody

In collaboration with CTAG normal human tissues were fixed, sectioned and stained with the 1H6 antibody. All nucleated cells stain positive except a subset of cells in the testis.

B.10   Possible cytoplasmic staining with 1H6 antibody
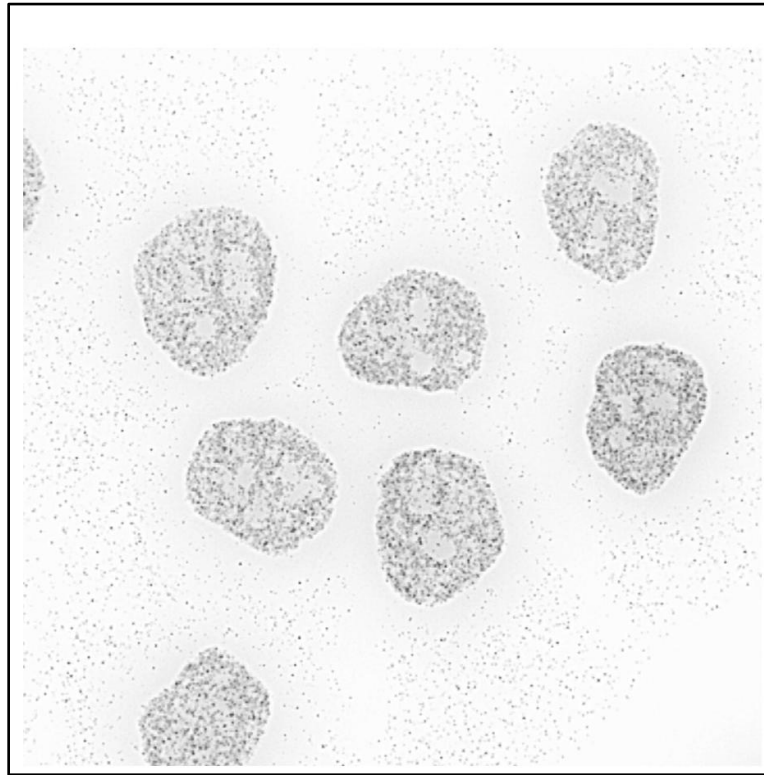

Figure 53. Detection of 1H6 foci outside the nucleus in some preparations.

The majority of 1H6 foci appear in the nucleus of cells (strong black staining). However, a proportion of paraformaldehyde fixed HeLa cells show 1H6 antibody foci outside of the nucleus (faint black spots). The foci outside of the nucleus were variable from experiment to experiment. Images have been inverted for clarity.

B.11   Detection of 1H6 foci after proteinase K digestion

Figure 54. Detection of 1H6 foci after proteinase K digestion of fixed HeLa cells.
Paraformaldehyde fixed HeLa cells were over digested with proteinase K (10 mg/ml) for one hour at 37 °C. 1H6 antibody (black foci) appear to be located on the DNA fibers (grey fibers stained with DAPI) of exploding cells. Images have been inverted for clarity.