Structure and dynamics of G-quadruplex DNA studied by ensemble and single-molecule techniques

Sengar, Anjali

2017

Sengar, A. (2017). Structure and dynamics of G-quadruplex DNA studied by ensemble and single-molecule techniques. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/69535>

<https://doi.org/10.32657/10356/69535>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**STRUCTURE AND DYNAMICS OF G-QUADRUPLEX DNA
STUDIED BY ENSEMBLE AND SINGLE-MOLECULE
TECHNIQUES**

SENGAR ANJALI

**SCHOOL OF PHYSICAL AND MATHEMATICAL
SCIENCES**

2017

**STRUCTURE AND DYNAMICS OF G-QUADRUPLEX DNA
STUDIED BY ENSEMBLE AND SINGLE-MOLECULE
TECHNIQUES**

SENGAR ANJALI

School of Physical and Mathematical Sciences

**A thesis submitted to the Nanyang Technological
University in partial fulfillment of the requirement for
the degree of Doctor of Philosophy**

2017

Acknowledgements

Firstly, I would like to express my sincere appreciation and thanks to my supervisor Prof. Phan Anh Tuấn, for giving me the opportunity to work in his lab. I am grateful for his constant guidance, motivation and patience throughout the course of my PhD research and allowing me to grow as a researcher.

I would like to express my gratitude towards Brahim, Nerea and Thuan, for the help and collaborations I have received. I am incredibly thankful to the undergraduate student Jeya for her contributions in the PARP-1 project. My appreciation extends to the past and present members of biophysics lab Kah Wai, Vineeth, Chris, Michael, Ranjani, Dung, Gitali, Vee Vee, Wan Jun, Ngoc, Krishnashish, Kim, Hong Anh, Fernaldo, Blaz, Derrick, Arijit, Tuan Anh, Huy, Zijian, Glenn and Viet for their advice, discussions and all the fun time that we shared.

I want to convey my heartfelt thanks to all friends for their pleasant company and the memories. Lastly, I would like to avail this opportunity to express my heartiest sense of reverence and immense love to all my family members for their everlasting encouragement, blessings and love.

Table of contents

Acknowledgements	i
Table of contents.....	ii
List of figures.....	v
Abbreviations.....	vii
Abstract	ix
Chapter 1	1
Introduction to G-quadruplex DNA: Overview	1
1.1. Structural basis of G-quadruplex DNA	2
1.2. Stability of G-quadruplexes	4
1.3. Biological relevance of G-quadruplex structures	5
1.3.1. Telomeric G-quadruplexes	7
1.3.2. Promoter G-quadruplexes	8
1.4. G/C tracts and genomic instability	10
1.5. G-quadruplex aptamers and nanodevices	10
1.6. Dynamics of G-quadruplex structures.....	11
Research objectives.....	13
Chapter 2	14
Overview of characterization methods	14
2.1. Nuclear Magnetic Resonance (NMR)	15
2.1.1. NMR basics.....	15
2.1.2. Two-Dimensional NMR.....	17
2.1.3. G-quadruplex structure determination by NMR	18
2.2. Ultraviolet Absorption	19
2.3. Circular dichroism spectroscopy	19
2.4. Gel electrophoresis	20

2.5.	Single-molecule methods.....	20
2.5.1.	Förster resonance energy transfer (FRET).....	22
2.5.2.	Single-molecule FRET	24
Chapter 3		26
Formation of G-quadruplexes in poly-G DNA sequences		26
3.1.	Introduction	27
3.2.	Results and Discussion.....	29
3.2.1.	G-quadruplexes formed by poly-G oligonucleotides of different lengths	29
3.2.2.	NMR structure of a <i>G15</i> sequence	31
3.2.3.	Single-residue propeller loops	35
3.2.4.	Effect of loop residue on the stability of parallel G-quadruplexes	37
3.2.5.	NMR structure of a <i>G22</i> sequence	39
3.2.6.	Effect of mutations on the <i>G22</i> sequence.....	41
3.2.7.	G-quadruplex folding topology of <i>G22</i> and <i>T7</i> sequence	45
3.2.8.	Examination of polymorphism of <i>G22</i> sequence; equilibrium of higher order structure and monomer	50
3.3.	Conclusions	54
Chapter 4		55
G-quadruplex structure in PARP-1 promoter		55
4.1.	Introduction	56
4.2.	Results and Discussion.....	57
4.2.1.	G-quadruplex of <i>P3</i> sequence and effect of flanking bases	57
4.2.2.	(3+1) hybrid intramolecular G-quadruplex formation by <i>TP3</i> sequence	59
4.2.3.	Spectral Assignments of <i>TP3</i> sequence	60
4.2.4.	Effect of mutations on <i>TP3</i> sequence	62
4.2.5.	Spectral assignments of <i>TP3-T6</i> sequence	63
4.2.6.	(3+1) G-quadruplex fold in <i>TP3-T6</i> sequence	65
4.3.	Conclusions	69

Chapter 5	70
Single molecule FRET study of structural dynamics in G-quadruplex DNA	
.....	70
5.1. Introduction	71
5.2. Instrumentation	73
5.2.1. Excitation: Total internal reflection (TIR) microscopy:	73
5.2.2. Collection and detection part	74
5.2.3. Data processing	75
5.2.4. Fluorescent dyes selection and labelling	77
5.2.5. DNA construct for sm-FRET	77
5.3. Results and discussions	78
5.3.1. Control experiments on hairpin	78
5.3.2. Effect of salt concentration on hairpin dynamics	80
5.3.3. Effect of stem size on hairpin dynamics	80
5.3.4. Structural stability of parallel G-quadruplexes	81
5.3.5. G-quadruplex conformation and dynamics of <i>T95-1T-12</i>	85
5.3.6. G-quadruplex conformation and dynamics of <i>T95-1T-9L</i>	87
5.4. Conclusions	89
Chapter 6	90
Summary and future work.....	90
Appendices	110
Experimental protocols	110

List of figures

Figure 1.1. Schematic representation of a G-tetrad	2
Figure 1.2. Schematic of polymorphism in G-quadruplexes.....	3
Figure 1.3. Glycosidic conformations of guanine.	4
Figure 1.4. Possible G-quadruplex structure locations in cells.....	6
Figure 1.5. Biology of telomeric DNA and G-quadruplex structure.	8
Figure 1.6. Model of G-quadruplex regulated gene transcription.	9
Figure 2.1. The precession of nuclear spin magnetic moment in the external magnetic field	15
Figure 2.2. NMR chemical shift pattern.	16
Figure 2.3. Schematic representation of a 2D NMR spectrum.....	17
Figure 2.4. The principle of Forster resonance energy transfer (FRET)	22
Figure 2.5. FRET efficiency as a function of the distance	23
Figure 2.6. Single molecule FRET	25
Figure 3.1. Imino proton spectra of <i>Gn</i> sequences	29
Figure 3.2.1. CD spectra and native PAGE analysis of <i>G15</i>	31
Figure 3.2.2. Proton NMR spectra of <i>T95-2T</i> and <i>G15</i>	32
Figure 3.2.3. Solution structure of <i>G15</i> determined by NMR	33
Figure 3.2.4. Comparison between the structures of <i>G15</i> and <i>T95- 2T</i>	35
Figure 3.2.5. ¹ H NMR spectra of the <i>G15</i> , <i>T95-2T</i> , <i>C95-2T</i> , <i>A95-2T</i> and <i>S95- 2T</i> sequences	36
Figure 3.2.6. Effect of loop residue on CD spectrum.	37
Figure 3.2.7. Effect of loop residue on the stability of G-quadruplex.....	38
Figure 3.3.1. Guanine imino and aromatic proton assignments of <i>G22</i>	40
Figure 3.3.2. Imino proton spectra of <i>G22</i> and D ₂ O solvent exchange	42
Figure 3.3.4. Guanine imino and aromatic proton assignments of <i>T7</i> mutated sequence.....	43
Figure 3.3.5. NMR imino and aromatic proton spectra of single G-to-T mutations on the <i>G22</i> sequence	44

Figure 3.3.6. Imino proton spectrum of <i>T7</i> mutated sequences and D ₂ O solvent exchange.....	45
Figure 3.3.7. Determination of G-quadruplex folding topology of <i>T7</i> & <i>G22</i>	46
Figure 3.3.8. NOESY spectra of <i>T7</i> and <i>G22</i>	47
Figure 3.3.9. Structural model of <i>G22</i> determined by NMR	49
Figure 3.3.10. Native PAGE, CD and imino proton spectra of <i>G22</i> gel purified fractions.....	51
Figure 3.3.11. Thermal denaturation CD spectra of <i>G22</i> purified fragments...	53
Figure 4.1. Imino proton NMR spectra of <i>P3</i> with different flanking bases	58
Figure 4.2. CD and native PAGE of <i>TP3</i>	60
Figure 4.3. Imino proton resonance assignments of <i>TP3</i>	61
Figure 4.4. Time evolution of <i>TP3</i> monomers and higher order aggregates...	62
Figure 4.5. Imino proton spectra region of <i>TP3</i> and mutated <i>TP3</i> sequences	63
Figure 4.6. Imino (H1) and aromatic (H8) proton assignments for the <i>TP3-T6</i> sequence.....	64
Figure 4.7. Folding topology of <i>TP3-T6</i>	65
Figure 4.8. Imino proton spectrum of <i>TP3-T6</i>	66
Figure 4.9. NOESY spectrum of <i>TP3-T6</i>	67
Figure 4.10. NOESY spectrum of <i>TP3-T6</i> showing H1-H1	68
Figure 5.1. Schematics of sm-FRET set up showing beam path.	74
Figure 5.2. Excitation and emission spectra of Cy3 and Cy5.....	75
Figure 5.3. Sm-FRET data analysis	76
Figure 5.4. Sm-FRET DNA construct.	78
Figure 5.5. Hairpin dynamics.....	79
Figure 5.7. Effect of stem size on hairpin dynamics	81
Figure 5.8. Imino-proton NMR spectra and proposed schematic of sm-FRET DNA construct.	82
Figure 5.9. Sm-FRET analysis of <i>T95-1T</i> G-quadruplex.....	83
Figure 5.10. Imino proton NMR spectra with schematic of parallel G-quadruplexes formed by modified sequences.....	84
Figure 5.11. FRET efficiency time traces for <i>T95-1T-12</i> G-quadruplex.	87

Abbreviations

1D	One-dimensional
2D	Two-dimensional
A	Adenine
CD	Circular dichroism
C	Cytosine
Cy3, Cy5	Cyanine dyes
D ₂ O	Deuterium oxide
DNA	Deoxyribonucleic acid
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
Duplex	Double helix
EMCCD	Electron multiplying charge coupled device
FRET	Förster resonance energy transfer
FANCI	Fanconi Anemia
G	Guanine
G-tetrad	Guanine tetrad
G-wire	Long molecular wire arranged of stacked G-tetrads
HIV	Human immunodeficiency virus
JR	Jump-and-return
KPi	Potassium phosphate
NMR	Nuclear magnetic resonance
NOE	Nuclear overhauser effect
NOESY	Nuclear overhauser effect spectroscopy
nt	Nucleotide
PARP-1	Poly(ADP-ribose) polymerase-1
PBP	Pellin-Broca prism
PDB	Protein data bank
PPM	parts per million
RMSD	Root-mean-square deviation

RNA	Ribonucleic acid
mRNA	Messenger RNA
sm-FRET	single molecule Förster resonance energy transfer
TIRF	Total internal reflection fluorescence
Tris	Tris(hydroxymethyl)aminomethane
Trolox	(6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid)
T	Thymine
T _m	Melting temperature
TBA	Thrombin-binding aptamer
TOCSY	Total correlation spectroscopy
UTR	Untranslated region
UV	Ultraviolet

Abstract

G-rich DNA sequences have a propensity to fold into four-stranded G-quadruplex structures. As G-quadruplexes are highly polymorphic structures understanding and manipulation of structural features offer great possibilities in the field of structural biology, medicinal chemistry and nanotechnology. This thesis explores the G-quadruplex formation in different sequence contexts. Two novel intramolecular G-quadruplex structures formed by biologically relevant poly-G tracts of 15 and 22 continuous guanines are presented. Another structural study has been carried out to investigate the G-quadruplex structure in the G-rich promoter sequence of the DNA repair protein PARP-1, showing three-layered, (3+1) hybrid G-quadruplex scaffold which could serve as a potential target for anti-cancer therapy. Structural dynamics of parallel G-quadruplexes and effect of loop length and inosine substitution on it have been studied using both ensemble and single molecule FRET techniques. Insights obtained from both ensemble as well as single molecule techniques contribute towards the understanding of G-quadruplex structures and dynamics.

Chapter 1

Introduction to G-quadruplex DNA: Overview

DNA is a flexible biological polymer which is known to carry genetic information. DNA structure is mostly represented as a double helix, but it can also adopt many other conformations. G-quadruplex is one of such non-canonical four-stranded DNA structure formed by guanine-rich sequences. Guanines can self-associate to form G-tetrads and stacking of G-tetrads leads to the formation of G-quadruplex structure. These structures are very diverse in terms of topology, molecularity and strand directionality. G-quadruplexes have been found in a number of biologically relevant sites in the human genome such as in gene promoters and telomeres; the ends of mammalian chromosomes. These sites are considered as attractive drug targets for gene regulation and anticancer activity. As G-quadruplexes are highly polymorphic, understanding and manipulation of conformational diversity offer great possibilities in structure biology, medicinal chemistry and promising candidate for biosensors and nanotechnology.

1.1. Structural basis of G-quadruplex DNA

G-rich DNA sequences have a unique property to self-assemble and to form a square planar structural unit termed as G-tetrad or quartet. Each guanine is donor and acceptor of two hydrogen bonds in a G-tetrad connected by cyclic Hoogsteen hydrogen bonds (1,2). The G-quadruplex structure is built from the stacking of these G-tetrads (3,4). A unique feature of G-tetrad structural motif is a central cavity for cation, in which coordination of cations to electronegative carbonyl oxygen stabilize the formation of G-tetrads by reducing the repulsion among oxygen of guanines and stabilizing stacking between adjacent tetrads (Figure 1.1)

Figure 1.1. Schematic representation of a G-tetrad. Hoogsteen hydrogen-bonded guanines and cation coordination with central oxygen. G-tetrads stack on one another to form four-stranded structures.

G-rich DNA sequences display an extraordinary polymorphism controlled by various factors such as DNA sequence, crowding/hydration, cation type and concentration. Minor changes in these specified factors could result either in disruption of the structure or transition from one form to another (5). Structural versatility of G-quadruplexes can be classified in terms of number of strands involved in formation of the structure (i.e., intramolecular for one and inter-molecular for two or more strands), loops; connecting nucleotide linkers in between the G stacks (either edgewise or diagonal or double chain reversal (propeller) and strand directionality (i.e., all strands pointing in one direction: parallel, three in one and one in the other direction: (3+1) hybrid, two stands

pointing in the one direction rest of the two in another direction: antiparallel)
(Figure 1.2).

Figure 1.2. Schematic of polymorphism in G-quadruplexes. A. G-quadruplex monomer, dimer and tetramer based on the number of strands involved in G-quadruplex formation. B. Three main types of loops edgewise, diagonal and propeller. C. Strand orientations; all strands in one direction parallel, three strand in one direction hybrid and two strands pointing in same direction antiparallel.

Guanine bases that contribute to G-tetrad formation adopt two different glycosidic conformations usually (i.e., anti or syn) with respect to sugar (Figure 1.3) (6).

Figure 1.3. Glycosidic conformations of guanine. anti and syn guanines.

The orientations adopted by guanine bases have significant consequences for stacking of tetrads and the overall shape of grooves. Grooves are voids between the sugar-phosphate backbones with various dimensions. In G-quadruplex DNA, groove dimensions are classified as narrow (12Å), medium (16Å) and wide (19Å). Four G-quadruplex building strands can give rise to four different types of grooves depending upon the strand directionality. In the case of a parallel G-quadruplex structure (4+0), the same glycosidic conformation of all guanines will generate four medium grooves. In the case of (3+1) hybrid topology and antiparallel (2+2) G-quadruplexes, medium, wide and narrow grooves will be generated according to the cyclic connectivities of syn/anti guanines in a tetrad going in the direction of G-tetrad polarity. The direction of tetrad polarity is defined by the N1-H→O6 or N2-H→N7 direction (5). These various grooves add an additional important factor to the polymorphism of G-quadruplexes and have been demonstrated as an important recognition and binding site by proteins and small molecule ligands (7,8).

1.2. Stability of G-quadruplexes

G-quadruplex structures are generally stable under physiologically relevant conditions. Stability of the G-quadruplex structures are due to contribution of various factors, such as hydrogen bonding of guanines, hydration structure in the grooves, hydrophobic and van der Waals stacking of G-tetrads (9-11). Apart from these factors, a major contribution to G-quadruplex stability comes

from the coordination of metal cations with negatively charged carbonyl oxygen in central cavity neutralizing the electrostatic repulsion of oxygen atoms. Many monovalent and divalent cations are shown to stabilize G-quadruplex structures, where K^+ and Na^+ being the most studied owing to their physiological relevance. The difference in the size of metal cation affects the position of their coordination site, for example, K^+ ions, are accommodated in the cavity between two tetrads while Na^+ ions are located at the center of planar G-tetrads due to their smaller size (12-15).

Among various G-quadruplexes in same solution, stability may vary widely depending on the DNA sequence, folding topology, number of tetrads, loops sequence and strand stoichiometry. Loop sequence and loop length play very critical role in guiding the folding topology of G-quadruplex structures, short loop length may offer a constraint in folding and long loops may connect in different fashions. Mostly short loops are connected edgewise or propeller type, long loops can make a diagonal loop in addition to edgewise and propeller. One nucleotide propeller type linkers are found to be the most stable for three stacked tetrads G-quadruplex (16). General trend of inverse correlation between G-quadruplex stability and loop size have been shown (17,18).

1.3. Biological relevance of G-quadruplex structures

Computational studies for investigating the putative quadruplex sequences in the human genome have found 376,000 sequences distributed throughout the human genome (19,20). Highest possibility of finding G-rich regions are in telomeres (21), Genes coding for ribosomal RNA (rDNA) (22), immunoglobulin heavy chain switch regions (23,24), Gene promoters (25), simple sequence repeats (26-29) and some of the single copy Genes (30-32). G-rich RNA sequences were found at RNA telomeric repeats, non-coding RNA (31-34) and enriched at the regulatory regions like 5'UTR and 3'UTR of mRNA (35-37).

Figure 1.4. Possible G-quadruplex structure locations in cells. G-rich sequences are mostly found in telomeres and promoters. Transient formation of the G-quadruplex structure during **A. Transcription** **C. Replication** **B. Single strand G-rich telomeric overhang.** **D. G-rich mRNA can form G-quadruplex structure outside the nucleus, to name an example during Translation.** Figure adapted from ref (32).

Most used sequence motif for putative quadruplex sequences prediction is $G_3+N_{1-7}G_3+N_{1-7}G_3+N_{1-7}G_3+$, where three or more runs of continuous guanines involve in forming G-quadruplex column and non-guanine residues comprise in the connecting loops (19,20). However, some unusual G-quadruplex sequence motifs are also reported including bulges (38), abasic site (39), long loops (40) and stem loops (41). Bioinformatic survey on long central loop in mRNA G-quadruplexes had predicted 1453 G-quadruplex forming sequences and model based on stem loops had identified 80307 stem-loop DNA quadruplex forming sequences (42,43). Sequencing of human genome had motifs identified 716310 G-quadruplex sequence motifs including long loops and bulges, increasing the number of predicted G-quadruplex sequence motifs in earlier studies.

Genomic duplex DNA needs to be separated into single strands in order to fold into G-quadruplex structures *in vivo*, it could happen transiently during replication, transcription and recombination, while G-rich mRNA does not need to compete with duplex formation making G-quadruplexes more likely (Figure 1.4) (32). Various studies have shown that G-quadruplex structures are

preferentially located in the functional region of genomes such as in telomeres, double-strand break sites and transcriptional start sites (promoters) (34). Available increasing evidence for the *in vivo* existence and more recently direct visualization in telomeres supports the function and biological relevance of G-quadruplexes (44-47)

1.3.1. Telomeric G-quadruplexes

Most extensive research on G-quadruplex structures was triggered by their presence in telomeric repeats with potential applications as an attractive drug target for cancer and anti-aging, discussed and summarized in articles (31,32,48-50). Telomere is the nucleoprotein complex at ends of a linear chromosome to form a protective cap that prevents chromosome from degradation and end to end fusion. Telomeric DNA (for vertebrates) is highly G-rich; double-stranded DNA consisting tandem repeats (TTAGGG) of short sequences (3-15 kb) and a 3' G-rich (100-300 nt) overhang (21). Telomere length shorten during each cell division and telomere shortening below a critical length leads to cell death. An enzyme telomerase maintains the telomere length by adding new DNA (51). Telomerase is found to be over-expressed in more than >85% of cancer cells (52). Adoption of a G-quadruplex structure by telomere DNA sequence has been shown to suppress the activity of telomerase because it does not recognize G-quadruplex DNA as a substrate (53). The anticancer potential of quadruplexes arises from the resulting loss of telomerase activity by the formation of G-quadruplex structure in cancer cells leading to cell death. Small molecules stabilizing quadruplex structures are an attractive tool to inhibit telomerase activity. G-quadruplex structure in human telomeric DNA has shown to be highly polymorphic can adopt a variety of G-quadruplex structures in the presence of Na⁺ and K⁺ cations (5).

Figure 1.5. Biology of telomeric DNA and G-quadruplex structure. The G-quadruplex structure can form at the 3' G-rich overhang and stabilization of G-quadruplex structure can inhibit telomerase activity. Figure adapted from ref (54).

1.3.2. Promoter G-quadruplexes

Outside telomeres, G-quadruplexes are prevalent in gene promoters (upstream of transcription start sites). Where they are hypothesized to have a regulatory role during transcription and hence in controlling gene expression (55). First discovery of *in vivo* transcriptional control by forming G-quadruplex structure was on the c-myc proto-oncogene, that encodes for the transcription factor involved in cell proliferation and found to be overexpressed in various human tumors (25). The G-rich sequence located in the nuclease hypersensitivity element (NHE) III₁, upstream of the c-myc promoter (P1) controls up to 90% of its transcription. 27 nt long G-rich region located from -142 to -115 upstream of P1 promoter has been shown to form intramolecular G-quadruplex structure (56). A single G to A mutation to disrupt the G-quadruplex structure has been found to increase the transcriptional level three times, further by using G-quadruplex structure stabilizing ligand downregulated transcription significantly. This discovery provided the proof of concept for the G-quadruplex mediated

transcriptional control and attracted discovery of new G-quadruplex specific ligands for therapeutics (25). G-rich motifs were also identified in c-kit (57) (58), bcl-2 (59) (60), VEGF (61), and HIF-1 α (62) oncogene-promoter region (63). Since these quadruplexes in promoter region may affect gene transcription and therefore they could influence a broad range of processes. G-quadruplex motifs are also found near promoter sites of yeast (64), mouse, rat and chimpanzee genes (65).

Figure 1.6. Model of G-quadruplex regulated gene transcription. G-quadruplex structure formation during transcription can down regulate the gene expression. Figure adapted from Wikipedia, modelled from ref (66).

Recent reports highlighted the interaction of PARP-1 (Poly ADP-ribose polymerase-1) protein with *c-kit*, *c-myc*, *KRAS* and telomeric promoter G-quadruplexes and have been found to influence their transcriptional regulation (67-70). Sequencing results have identified the potential G-quadruplex forming G-rich regions in the promoter of the *PARP-1* which could affect its transcription (personal communication S. Balasubramanian). Stabilization of G-quadruplex in *PARP-1* promoter could act as *PARP-1* inhibitor; an attractive target for cancer therapy. Understanding of G-quadruplex structure in the *PARP-1* gene promoter could facilitate the rationale ligand design for its inhibition. The G-quadruplex structure adopted by *PARP-1* promoter is explored in the fourth chapter of this thesis.

1.4. G/C tracts and genomic instability

Genomic instability seems to be related to the capability to DNA fragments to form unusual secondary structures such as hairpins, triplexes or quadruplexes (71). These structures can obstruct the progression of replication fork and induce DNA mutations. In some cases, genomic instability is associated with diseases like Bloom syndrome (BLM), Werner syndrome (WRN), Fanconi anemia (FANCF) and identified as cancer hallmark (72).

Genomic instability was first studied in *Caenorhabditis elegans*. In *dog-1* deficient *C.elegans*, these G-tracts impose a structural barrier to DNA replication fork and G tract deletions were observed, suggesting G-quadruplex formation in G-tracts and its implications in *FANCF* mutation-associated cancers (73,74). Particularly most studied G-tracts are in *C. elegans*, where approximately 400 poly-G regions of various lengths (~10-30 nucleotides, nt) are identified in different chromosomes (74,75). These G-tracts might have some biological function as suggested by the existence of G-quadruplex-specific helicase and their role in the maintenance of G-tracts in *C.elegans*. However, structural details of G-quadruplex formation in such poly G-tracts had remained to be understood. Hence, G-quadruplex structure formation in poly-G tracts have been studied and presented in the third chapter of this thesis.

1.5. G-quadruplex aptamers and nanodevices

A unique structural feature of G-quadruplex DNA is useful for various applications. G-quadruplex structures can be used for molecular recognition e.g. aptamers. Aptamers are an interesting class of oligonucleotides having a unique structure to bind with the specific molecular targets (76). G-quadruplex-based aptamers have several advantages over other unstructured aptamers: they are relatively stable against thermal, chemical denaturing and have enhanced cellular uptake (77,78). These advantages make G-quadruplex extremely viable for using as therapeutic agents for cancer therapy and many other diseases. For example, Thrombin binding aptamer, TBA (anti-coagulant) (79), *AGRO100* (80), *T30695* (81) (anti-cancer and anti-HIV) and *93del* (82) (ant-HIV) are significant G-quadruplex aptamers.

G-rich sequences can self-associate and form extended wires on a micrometer scale. G-wires have been found as a potential candidate for use in nanoelectronics due to their high stiffness, versatile and controllable structures (83,84). G-wires have been shown to form in different G-rich sequences such as poly-G and $G_4T_2G_4$ (85-87). The length and stability of these G-wires could be manipulated by different engineering schemes for G-quadruplex structure. G-quadruplexes also have great potential in the further development of nanodevices; e.g. Ion channels, switches and electrochemical biosensors (88-91).

1.6. Dynamics of G-quadruplex structures

Despite a large amount of research focused on G-quadruplex structure and stability, dynamics in G-quadruplex structures are much less understood. Understanding of the dynamics in G-quadruplex structures could be helpful in two ways. First, it can directly provide the information about the intermediate states that can help in understanding reaction pathways. Second, it can give information about the timescale of structural changes in biologically relevant events such as ligand binding and protein interactions.

G-quadruplex folding/unfolding has been investigated by various spectroscopic techniques that monitor different aspects of the dynamics, given that different techniques can observe different time scales (92,93). Single-molecule methods have also been used to study kinetics with their exceptional capability to examine individual subpopulations from a heterogeneous mixture and reveal their dynamics (94-97). Particularly, single molecule fluorescence resonance energy transfer (sm-FRET) has been used and already demonstrated the conformational diversity in G-quadruplexes.

First sm-FRET study on human telomeric DNA has demonstrated the coexistence of parallel and antiparallel G-quadruplex conformations with unfolding time ($\sim 700 - 800$ s) by hybridization with complementary strands (98). Another sm-FRET study on human telomeric G-quadruplex also supported the conformational diversity. These parallel and antiparallel structures have been shown to be divided into long-lived (minutes) and short-lived (seconds) states.

These short-lived states have been proposed to be the transient intermediates of stable G-quadruplex formation (99). Subsequent sm-FRET studies on human telomeric G-quadruplexes observed the various intermediates such as hairpin, triplet state, hybrid and chair type G-quadruplex structure from a mixture of conformations (100,101). Most recent study on human telomeric G-quadruplexes in Na⁺ solution has identified four distinct states showing fast (<~10 s) and slow dynamics (>10 s), basket type antiparallel being the most stable state via chair-type, hybrid or triplet G-quadruplex intermediate states (102). The presence of various intermediates has also been proposed by a number of other single molecule and ensemble studies on human telomeric DNA (92,93,103-106).

Understanding of the dynamics of G-quadruplex structures are largely based on the human telomeric sequence. Besides, few studies have examined G-quadruplex dynamics in c-myc promoter sequence and detected highly stable (longer unfolding time) parallel G-quadruplex structure (95,96,107). Structural dynamics of parallel G-quadruplex structures and effect of sequence modification on it has been explored in the fifth chapter of this thesis using the sm-FRET technique.

Research objectives

The motivation underlying this work is the better understanding of G-quadruplex structures and dynamics adopted by various G-rich sequences using both ensemble techniques (NMR, UV, CD spectroscopy and gel electrophoresis) and single molecule FRET. Insights gained could be used to understand the interaction mechanism of ligand and protein binding to the G-quadruplex structure. Single molecule methods could provide the unique information on the structural dynamics of G-quadruplexes extending the present understanding from ensemble studies leading towards the advancement in understanding the G-quadruplex structures and their biological functions, making them druggable for therapeutics as well as a building block for many nanostructure devices. These goals can be divided into two main categories.

- Study of G-quadruplex structure and dynamics in poly-G tracts and understanding of the G-quadruplex folding topology in PARP-1 promoter sequence using ensemble techniques.
- Optimization and validation of custom built single molecule FRET set up and investigation of the structural dynamics of parallel G-quadruplexes by using ensemble techniques and single molecule FRET.

Chapter 2

Overview of characterization methods

Ensemble methods used in this work for the structural studies on G-quadruplexes are Nuclear Magnetic Resonance (NMR), Ultraviolet spectroscopy (UV), Circular dichroism (CD) and gel electrophoresis. Nuclear Magnetic Resonance (NMR) is a powerful technique and capable of giving a wealth of detailed structural information at atomic-resolution. NMR was used in conjunction with relatively simple yet direct methods of CD and gel electrophoresis for determining the folding topology of G-quadruplexes. Apart from these ensemble methods, single molecule FRET was used to study conformation dynamics of G-quadruplexes and their kinetics. This chapter focuses on the aspects of different methods used and discusses the basic principles. Detailed experimental protocols are provided in the appendix A1& A2.

2.1. Nuclear Magnetic Resonance (NMR)

2.1.1. NMR basics

Spin $\frac{1}{2}$ nuclei such as ^1H , ^{13}C , ^{15}N , ^{31}P , ^{19}F are mostly used for NMR detection, they possess nuclear spin magnetic momentum (M_0). In the presence of an external magnetic field (B_0), nuclear spin magnetic momentum aligns in the direction of the external magnetic field. When a radiofrequency pulse is applied perpendicular to external magnetic field (along the x-axis) to generate an oscillating magnetic field which tilts the nuclear spin magnetic momentum. As a result, spin magnetic momentum (M_0) starts to precess around external magnetic field (B_0), depends upon the strength of external magnetic field and the property of nucleus called gyromagnetic ratio (γ) as follows,

$$\omega_o = \gamma B_o$$

Figure 2.1. The precession of nuclear spin magnetic moment in the external magnetic field.

ω_o is the precession frequency and known as Larmor frequency. Among all spin $\frac{1}{2}$ nuclei ^1H ($\gamma = 42.576 \text{ MHz/T}$) is the most sensitive and useful for studying biomolecules.

After irradiation, the magnetization vector the system will revert to equilibrium position (“relax”) over time. The precession of the magnetization vector will induce a current (free induction signal), recorded as a function of time in the xy-

plane. Free induction decay gives the amplitude of NMR signal with time and Fourier transformation converts it into the frequency domain.

Different protons in the molecule precess at different frequencies depends on their environment. The external magnetic field around the proton is modified by surrounding chemical-/physical-environment (electron cloud) can decrease (“shielding”) or increase (“de-shielding”) the effect of external magnetic field on a nuclear spin magnetic moment and influence its precession. Hence, nuclei in different chemical-/physical-environment have different resonance frequencies. The change of frequency in comparison to reference is expressed as chemical shift value (δ), which is a fundamental concept for structural study by NMR spectroscopy. The references used in this work is 4,4- dimethyl-4-silapentane-1-sulfonic acid (DSS), commonly used for NMR. Chemical shifts of a proton (^1H) involved in nucleic acids are shown in Figure 2.2.

Figure 2.2. NMR chemical shift pattern. A. Chemical shift pattern for nucleic acid protons. B. schematic guanine nucleotide with phosphate. C. Adenine D. Thymine E. cytosine bases. The numbering of the protons in bases is shown in pink.

In G-quadruplex studies, Imino proton of guanines (N1-H—O) in G-quadruplex structure shows characteristic resonances at the chemical shift from 10-12 ppm which is well separated from other DNA conformations, making it an excellent marker for G-quadruplex structure. Number of imino peaks correspond with the number of guanines involves in tetrads.

2.1.2. Two-Dimensional NMR

Two-dimensional (2D) NMR is frequently used for elucidating the molecular structure of macromolecules, such as the G-quadruplex. The signal of 2D NMR is recorded as a function of two time variables, t_1 and t_2 and the resulting data are Fourier transformed twice to yield a two-dimensional spectrum with chemical shifts F_1 and F_2 . Two dimensional spectrum usually consists of self-peaks and symmetrical cross peaks, shown in the black and white circle, respectively. Since it is symmetrical, the analysis is only interpreted on one side of the spectrum (Figure 2.3.).

Figure 2.3. Schematic representation of a 2D NMR spectrum. The system consists of two hypothetical protons on the diagonal. Off diagonal peaks represent the inter-proton cross-peaks between the two protons.

Two types of 2D correlation experiment are available, in which the correlation is either by through-bond or through space. Through-space correlations, such as NOESY, are dependent on the dipole-dipole interactions between two

neighboring magnetic nuclei that give rise to the NOE (Nuclear Overhauser Effect). The magnitude (contour intensity) of the NOE is dependent on the inter-nuclear distance (R), which is inversely proportional to R^6 . NOE is only detectable when the inter-nuclear distance is less than or equal to 5 Å.

On the other hand, through-bond correlations are reliant on the J -coupling constant, in which the strength of the J -coupling is determined by the electronic interactions through the chemical bonds of a molecule. There are two main types, specifically homo-nuclear and heteronuclear correlation. In homonuclear experiments both frequency domains belong to protons, giving information on the H-H J -coupling, whereas in the heteronuclear provides information between heteroatoms for instance ^1H - ^{15}N . Considering the low natural abundance of ^{15}N (0.37%), heteronuclear bond correlation experiments require higher sensitivity for signal detection.

In this work, nuclear Overhauser effect spectroscopy (NOESY) and total Correlation Spectroscopy (TOCSY) is used to determine the base and sugar proton assignments of the G-quadruplexes.

2.1.3. G-quadruplex structure determination by NMR

In this study, jump and return (JR) water suppression pulse sequence is mostly used to suppress the huge (55 M) water signal (108). In 1D ^1H spectrum, the guanine imino protons being hydrogen-bonded to an oxygen in the G-tetrad, appear at characteristic chemical shift ~ 10.5 – 12.5 ppm (Figures 2.2). Four imino proton peaks should be observed in this region corresponding to a G-tetrad. In general, the number of peaks in 10.5–12.5 region provide the initial information about the number of guanines involved in the G-quadruplex formation. Upon forming a G-tetrad, each guanine imino proton is sufficiently close to the H8 proton of one of the neighboring guanines (only on one side) to exhibit an NOE (H1-H8) cross-peak. This leads to a cyclic imino–H8 connectivity within a G-tetrad and the directionality of the hydrogen bonds which all together help to establish the overall topology of the quadruplex (109).

Identification of spin system of all the nucleotides in a sequence and sequential connectivity between them is required for the prediction of 3D structure.

Individual guanine H1 & H8 assignments were unambiguously assigned by using site-specific low-level ^{15}N labeling (110). Spectra assignments were completed by TOCSY and NOESY experiments. Interproton distances were measured by the intensity of the NOE cross peak at different mixing times. The collection of all interatomic distances and topological information is used to for the structure computation. Structure computations in this study were carried out with the program XPLOR-NIH (111). Two general steps of structure calculation are: (i) distance geometry-simulated annealing, (ii) distance-restrained molecular dynamics refinement. During structure calculations hydrogen bond restraints, interproton distance restraints, dihedral restraints as well as planarity restraints were provided. Ten lowest-energy structures that best satisfy the set of NMR restraints are eventually selected and displayed using PYMOL (112).

2.2. Ultraviolet Absorption

DNA bases absorb ultraviolet (UV) radiations maximum at close to 260 nm. To measure DNA concentration, DNA sample is exposed to UV radiation and absorption is recorded, the higher the concentration more the absorption. UV absorption at a particular wavelength can be recorded with temperature to at different temperatures to measure the thermodynamic stability. The nature of the melting and cooling spectra can also provide insights into the folding/unfolding kinetics. Change in absorbance at 295 nm, allows to precisely monitoring guanine G-quadruplex formation and dissociation. Hypochromic shift at 295 nm is specific for G-quadruplex structure (113).

2.3. Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy is wavelength-dependent differential absorption of left- and right-handed light by a asymmetric (chiral) molecule. Hence a plot of the ellipticity (difference in the amount of left- and right- circularly polarized light) across different wavelengths can be used to characterize the folding topology of a DNA structure (114,115). Various G-quadruplexes topologies (parallel, antiparallel and hybrid) are correlated with different syn/anti guanines conformations in G-tetrads, which in turn affect the G-tetrad stacking pattern and hence a unique CD signature pattern for each topology. Simple and

direct approach of CD measurements have been used to empirically determine the folding topology of different G-quadruplexes and contributed significantly in understanding the structural diversity of G-quadruplexes together with high resolution techniques (115,116). Parallel G-quadruplexes give typical CD signal with a positive peak around 265 nm and negative peak close to 240 nm. Antiparallel G-quadruplexes give a positive peak at ~ 295 nm and negative peak at 265 nm, while the hybrid form gives two positive peaks at 265, 295 nm and a negative peak at 240 nm (117). CD can also be used to study the conformational transitions with the change in environment. Thermal stability of G-quadruplexes can be studied by monitoring changes in the CD signal with temperature. CD spectrum alone is not sufficient to determine the structure; and hence the CD data should be supported by other biophysical techniques.

2.4. Gel electrophoresis

Gel electrophoresis is a useful method for separating G-quadruplex based on their mobility in a gel matrix under the electric field. Electrophoretic separation of molecules relies on their difference in molecular size, net charge, and conformation. Under an electric field, negatively charged DNA would travel through the matrix towards the positive electrode. Gel Electrophoresis acts as a molecular sieve where small, folded molecules migrate faster farther than large, unfolded molecules. G-quadruplex forming sequences of similar sizes are compared with references of known structure to understand their folding topology. G-quadruplex monomers migrate faster than dimers and among monomers, parallel G-quadruplex with propeller loops migrates slower than the antiparallel or hybrid G-quadruplexes. After the electrophoresis is done, the gel with DNA can either be visualized under UV light or directly by using fluorescent dye/staining the sample.

2.5. Single-molecule methods

A new class of methods to probe single molecules is being developed in the past two decades, which allowed us to monitor and manipulate the individual molecules. Single molecule methods give the opportunity to unravel the molecular and temporal heterogeneity, which otherwise obscured by the

population averaging of exceedingly large number ($\sim 10^{14}$ - 10^{23}) molecules (118). Single molecule methods can be broadly divided into two classes: one to visualize individual molecule using fluorescence-based detection (119) and other using mechanical force to detect and manipulate each molecule e.g. optical tweezers, magnetic tweezers and atomic force microscope (120).

Fluorescence microscopy is a powerful detection technique in which fluorophore is attached to host molecule and used as a reporter. Localization of emitted photons from fluorophore contains information about the location of the molecule, fluctuations in fluorescence intensity gives information about the dynamics of the molecule and by studying polarization response; it is possible to study the orientation of the molecule. Additionally, all the above-stated information of a molecule can be studied as the response to the changes in local environment i.e. pH, ion concentration, viscosity. Single molecule FRET (Förster Resonance Energy Transfer) can be used to detect distance dependent structural changes between two fluorophores; FRET often termed as molecular ruler in nanometer scale (1-10 nm) (121).

In force based single molecule methods, the mechanical force has been applied to the molecular system to characterize the strength of different chemical bonds; intermolecular, intramolecular and macromolecular. Structural changes in molecules are often accompanied by the chemical bond rearrangements; hence different structures have different mechanical stabilities. Forces can be applied by different ways; physical (atomic force microscopy (AFM), microneedles), optical (optical tweezers) and magnetic (magnetic tweezers) and they have different ranges of manipulation and spatial sensitivity. Many important biological processes generate forces in picoNewtons such as the activity of DNA/RNA enzymes and motor proteins (120). To summarize, different optical and mechanical manipulation single molecule techniques or a combination of two can be used to study the structure, mechanical stability, equilibrium dynamics and kinetics and ligand binding interactions in detail without the population and temporal averaging.

Among the various methods used to study single molecule, we have used fluorescence-based single-molecule FRET (Forster Resonance Energy

Transfer). FRET has widely used method due to its high sensitivity of distance measurement in a range of 1-10 nm, in this regime most of the dynamical changes in the structure of biomolecules such as DNA, protein and cellular processes takes place. Availability of suitable fluorophore, imaging instrument and analysis tools in the past two decades made it a popular choice. FRET is very powerful to probe the conformational heterogeneity and dynamics of DNA. The remaining chapter is devoted to describing the study of DNA G-quadruplex dynamics by single-molecule FRET technique.

2.5.1. Förster resonance energy transfer (FRET)

Theodor Förster in 1946, first established the FRET principle with correct distance (122,123). Briefly, FRET is a non-radiative energy transfer via dipole-dipole interaction from an excited fluorophore (donor dye) to another nearby fluorophore (acceptor dye). According to the classical theory of energy transfer, an excited donor will generate an oscillating dipole electromagnetic field, which induces an oscillating dipole of nearby acceptor; and hence the energy is transferred from excited donor to the acceptor and produces donor fluorescence (depicted in Figure 2.4).

Figure 2.4. The principle of Förster resonance energy transfer (FRET). The excited donor may relax to ground state by emission of a photon. If an acceptor is in close proximity, the donor energy can be transferred to the acceptor by a dipole-dipole interaction. For this to happen, there should be a spectral overlap between donor emission and acceptor excitation.

Energy transfer efficiency (E_F) is strongly distance dependent being inversely proportional to the sixth power of the distance (R) between the two dyes. Hence

any small changes in distance R result in sharp observable changes in energy transfer efficiency.

$$E_F = \left(1 + \frac{R^6}{R_0^6} \right)^{-1}$$

Figure 2.5. FRET efficiency as a function of the distance between donor and acceptor. The FRET efficiency, E , decreases as 6th power of distance R . At $R = R_0$, FRET efficiency is equal to 50%. Figure adopted from ref (124).

Forster radius (R_0) is different for different dye pairs and depends upon the donor and acceptor dye quantum yield, extinction coefficient. For most popular classical FRET dye pair of Cy3 and Cy5, R_0 is calculated as ~ 6 nm (125). Changes in R between two molecules about 3 nm (near R_0 , from 4 to 7 nm) will change FRET efficiency (E_F) from 0.8 to 0.2. Forster radius (R_0) for any donor-acceptor dye pair can be calculated as

$$R_0 = \left(\frac{8.79 \cdot 10^{-25} \cdot \phi_D \cdot \kappa^2 \cdot J^2}{n^4} \right)$$

Where κ^2 factor depicts the relative dipolar orientation of the dyes, for freely rotating dyes approximated as $2/3$. ϕ_D is the donor quantum yield, n is refractive index of medium, J is the normalized spectral overlap. Higher the value of

spectral overlap integral (J) more energy transfer will take place, which can be measured as

$$J = \frac{\int_0^{\infty} \varepsilon_A(\lambda) \lambda^4 I_D(\lambda) d\lambda}{\int_0^{\infty} I_D(\lambda) d\lambda}$$

Where λ is wavelength, ε_A is absorption extinction coefficient of acceptor; I_D is emission spectrum of the donor.

Experimentally energy transfer efficiency (E_F) can be measured in terms of emitted fluorescence intensity from a donor (I_D) and acceptor (I_A) as:

$$E_F = \frac{I_A}{I_A + I_D}$$

At Forster radius (R_0), spontaneous fluorescence decay and energy transfer to acceptor dye are equally probable and hence E_F would be 0.5. The relative fluorescence intensity of the two emissions (I_D) and (I_A) changes the FRET efficiency E_F with a strong dependence on R .

2.5.2. Single-molecule FRET

Single molecule FRET can provide the wealth of information on the conformational dynamics of each molecule, even the rarely visited conformations can be observed. smFRET can detect conformational heterogeneities of molecules in equilibrium and allows us to follow their pathways in real time during their functional, folding or binding events.

Figure 2.6. Single molecule FRET. A. mixture of different conformations of molecules. B. Single-molecule FRET reveals the dynamics of individual molecules in real time. Figure adopted from ref (126).

The first experiment on single molecule FRET is demonstrated by Taekjip Ha et al. in 1996 (127,128). It was a proof of FRET concept, shown the different FRET energy states ($E_F = 0.65$ and 0.32) using the DNA of different length 10 and 20 nts. Afterwards many researchers have used this technique to study the enzymatic reactions, protein folding and DNA-protein interactions. Significant advancement to the method is introduced by using polarized excitation, three color dyes and multiple color LASER excitations (129).

Chapter 3

Formation of G-quadruplexes in poly-G DNA sequences

Poly-G sequences are found in different genomes including human and have the potential to form higher-order structures with various applications. Long poly-G sequences were thought to lead to multiple possible ways of G-quadruplex folding, rendering their structural characterization challenging. Here we investigate the structure of G-quadruplexes formed by poly-G sequences d(TTG_nT), where $n = 12$ to 27. Our data show the presence of multiple and higher-order G-quadruplex structures in most sequences. Strikingly, NMR spectra of the TTG₁₅T and TTG₂₂T sequence containing a stretch of 15 and 22 continuous guanines are exceptionally well-resolved and indicate the formation of a well-defined G-quadruplex structure. NMR solution structure of G15 sequence revealed a propeller-type parallel-stranded G-quadruplex containing three G-tetrad layers and three single-guanine propeller loops while structure model of G22 displayed four-layer antiparallel G-quadruplex structure showing unusual exchange dynamics among guanines connecting edgewise loops and in an adjacent tetrad. In principle, these two structures can form anywhere along a long G_n stretch, making them unique for molecular recognition by other cellular molecules.¹

Part of this chapter was adapted from the authors work: Anjali Sengar, Brahim Heddi, and Anh Tuấn Phan. (2014) Formation of G-Quadruplexes in Poly-G Sequences: Structure of a Propeller-Type Parallel-Stranded G-Quadruplex Formed by a G₁₅ Stretch, *Biochemistry*, 53, 7718–7723.

Part of this chapter is being written for publication: “Structure of a Four-Layer Chair Type Antiparallel G-Quadruplex in Poly-G Sequence: Unprecedented Dynamics among Guanine in Loops and G-tetrad Core”, Anjali Sengar, Brahim Heddi, Nerea Martin-Pintado and Anh Tuấn Phan.

3.1. Introduction

Poly-G tracts are runs of continuous guanines and increasingly identified in various genomes, for instance, *Fungi*, *Propionibacterium*, *Helicobacter pylori*, *Caenorhabditis elegans* including human (130-136). Particularly most studied G-tracts are in *C. elegans*, where approximately 400 poly-G regions of various lengths (~10-30 nts) are identified in different chromosomes (74,75). In *dog-1* helicase deficient *C.elegans*, these G-tracts impose a structural barrier to DNA replication fork and G tract deletions were observed, suggesting G-quadruplex formation by G-tracts *in vivo* and its implications in *FANCI* mutation-associated cancers (73). G-repeats have been shown to play an important role in host adaptation of pathogenic bacteria particularly in *H. pylori* where tract length controls the gene expression via posttranscriptional regulation (136,137). Poly-G regions are found in the human genome too. For instance, in chromosome 2 there is a 427-nt region, where only 24 out of 427 nts are non-guanines (138). The frequency, distribution and length of G-tract vary extensively among different genomes and short length tracts are more frequent than the longer ones (75,134). Such sequences are inherently unstable and prone to mutations and length instability (130,139). Therefore, in the majority of genomes, G-tracts have been shown in the non-coding regions and suggested to have a regulatory role except some they were found in exons. (e.g. *H. pylori*, *C. elegans*, *Pirelulla* sp) (140,141). Due to these distinctive features of poly-G repeats serve as excellent genetic markers for lineage tracing, As such, they were used to construct fate maps and characterize intratumor heterogeneity (142,143). Interestingly, contiguous runs of twenty or more guanines in the human genome have shown to have natural anti G20-anibodies (144).

Evidently the unique characteristics of poly-G repeats are associated with its tract length. So we set out to study the poly-G tract length dependent propensity to form the G-quadruplex structure. Most used sequence motif for G-quadruplex prediction is $G_3+N_{1-7}G_3+N_{1-7}G_3+N_{1-7}G_3+$, where three or more runs of continuous guanines will form G-quadruplex column and non-guanine residues will comprise the connecting loops (19,20). However, some new unusual G-quadruplex sequence motifs are also reported including bulges (38), abasic site

(39), long dynamic loops (40) and structured loops (41). G-quadruplex structures in poly-G sequences were thought to be fundamentally intriguing as such continuous guanine runs could simultaneously form multiple possible conformations, since non-guanine residues pose limitation on the number of tetrads, molecularity and loop length. Despite the general belief that long poly-G sequences would form tetrameric G-quadruplexes, recent mass spectrometry experiments showed that these sequences folded into G-quadruplexes of lower stoichiometry (145), in accordance with an early model of a monomeric fold-back G-quadruplex (146). However, high-resolution structures of G-quadruplexes formed by long *Gn* tracts have not been studied.

In this work, using Nuclear Magnetic Resonance (NMR) and Circular Dichroism (CD) spectroscopy, we investigate the structure of G-quadruplexes formed in DNA poly-G sequences d(TTG_{*n*}T) referred as *Gn* (*italicized*) henceforth, where *n* ranges from 12 to 27. Two thymines at the 5'-end and one at the 3'-end were added to prevent the stacking of G-quadruplex monomeric units through planer surfaces which could possibly occur at higher DNA concentrations required for NMR structural study. Previous study on stacking interactions of parallel G-quadruplex had revealed the effect of flanking sequences, temperature, DNA and salt concentrations on stacking equilibrium, keeping the structure of monomeric G-quadruplex intact (147). Terminal bases have been shown to stack on the outer G-tetrads and hence hinder the possible stacking interaction (38,40,148).

Unexpectedly, we discovered that the *G15* (TTG₁₅T) and *G22* (TTG₂₂T) sequences displayed exceptionally well-resolved NMR spectra. The structure of *G15* revealed a three-layer parallel-stranded G-quadruplex containing three single-guanine propeller loops while structure model of *G22* displayed four-layer antiparallel G-quadruplex containing three two-guanine edgewise loops. The model of *G22* structure showed unusual dynamics among guanines in edgewise loops and in an adjacent tetrad (flipping bases). Structural plasticity of guanines in loops could be related to the response of changes *in vivo* environments and might have unique recognition properties. Therefore, flexible G loops could be attractive targets as aptamers.

3.2. Results and Discussion

3.2.1. G-quadruplexes formed by poly-G oligonucleotides of different lengths

Figure 3.1. Imino proton spectra of *G_n* sequences.

Imino proton NMR spectra of DNA sequences d(TTGnT) (termed *Gn*, *italicized*) containing *n* continuous guanines (*n* = 12–27) in a solution of Tris-HCl, pH 7.0, supplemented with 2 mM KCl, are shown in Figure 3.1.

For *n* < 15, imino proton spectra show a broad hump, suggesting the formation of high-order structures and/or a mixture of multiple conformations. Remarkably, for *n* = 15 the *G15* sequence displays an extremely well-resolved spectrum with distinct imino proton peaks in the 10–12 ppm range, characteristic of the formation of a well-defined G-quadruplex conformation. For *n* > 15, more peaks appeared in addition to the set of peaks observed in *G15*, suggesting that a similar G-quadruplex as *G15* coexist with additional G-quadruplex conformation(s). The number of peaks increases with guanine tract length and several peaks start to merge giving rise to broadened peaks, consistent with the formation of multiple conformations.

Noticeably, the spectrum of *G22* shows well-resolved imino peaks corresponding to one predominant form. Increasing tract length from 22 (*G23* to *G27*) promotes increasingly minor conformations leading to a hump like spectrum. Although from 22 to 27 the patterns of major peaks remain conserved suggest that major form (*G22*) remains there and coexist with other minor forms in longer tracts. Thus, we decided to study the *G22* for having unusual G-quadruplex spectral features and its structural prevalence with other minor forms in longer tracts.

We set out to determine the solution structure of the G-quadruplex formed by the *G15* & *G22* sequence containing 15 & 22 continuous guanines, which showed distinctive NMR spectra compared to other poly-G sequences.

3.2.2. NMR structure of a *G15* sequence

The imino proton spectrum of *G15*, which displays 12 sharp peaks, could be three layer quadruplex. The CD spectra of *G15* shows a maximum around 260 nm and a minimum around 240 nm (Figure 3.2.1.A), characteristic of parallel-stranded G-quadruplexes (117). The monomeric nature of *G15* was indicated in a gel electrophoresis experiment (Figure 3.2.1 B).

Figure 3.2.1. CD spectra and native PAGE analysis of *G15*. A. CD spectra of the *G15*. B. Native PAGE. lane 1, the *93del* (GGGGTGGGAGGAGGGT), reference of an interlocked dimeric G-quadruplex (82); lane 2, the *G15* sequence; lane 3, the *T95-2T* (TTGGGTGGGTGGGTGGGT) sequence (147), reference of a monomeric propeller-type parallel-stranded three-layered G-quadruplex.

NMR spectra of *G15* is almost identical to that of the reference sequence *T95-2T* studied previously; the aromatic proton spectrum of *G15* is also very similar to that of *T95-2T* except for the positions of peaks of residues 6, 10, and 14 (guanine residue in loops in positions 6, 10 and 14 *i.e.* G6, G10, and G14 in *G15* vs. T6, T10, and T14 in *T95-2T*, respectively) (Figure 3.2.2 A and B). The *T95-2T* sequence was previously shown to adopt a propeller-type parallel-stranded three-layered G-quadruplex with three single-residue propeller

Figure 3.2.2. Proton NMR spectra of *T95-2T* and *G15*. (A and B) Imino and aromatic proton spectra of A. *T95-2T* and B. *G15*. NMR spectra of the two sequences are almost identical, except for the aromatic protons of the loop residues 6, 10, and 14 (labeled in red). C. NOESY spectra of *G15* (mixing time, 300 ms): (Left) Imino-H8 proton region is showing characteristic guanine imino-H8 cross-peaks for G-tetrads, which are framed and labeled with the assignment of the imino proton in the first position and that of the H8 proton in the second position. (Right) H8/H6-H1' proton region is showing NOE sequential connectivities. Intraresidue cross-peaks are labeled with the corresponding residue numbers. Moderate intensities of the cross-peaks between H8 and H1' protons indicate anti glycosidic conformations of all guanines. D. Folding topology by following tetrad alignment in three tetrads.

or (double-chain-reversal) loops T6, T10, and T14. The striking similarity between the 1D and 2D NMR spectra of *G15* and those of *T95-2T* (Figure 3.2.2. and data not shown) indicates that *G15* adopts the same fold as that of *T95-2T*, i.e. a propeller-type parallel-stranded G-quadruplex involving three G-tetrad layers, where three guanines G6, G10, and G14, form three single-residue propeller loops (Figure 3.2.2. D). All guanines adopt an anti glycosidic conformation. This fold was confirmed by the NOESY spectra of *G15* (Figure 3.2.2. C).

High-resolution structure of *G15* (Figure 3.2.3) was determined by NMR restraints (Table 3.1). This structure is a three-layer propeller-type parallel-stranded G-quadruplex.

Figure 3.2.3. Solution structure of *G15* determined by NMR. (A) Side view and (B) top view of 10 superimposed lowest-energy structures and ribbon view of the representative structure on the right. The bases of thymines are colored orange; guanines, cyan; backbone and sugar, grey; O4' atoms, yellow; phosphorus atoms, red.

with three single- residue propeller loops. The first two thymines T1 and T2 are positioned on top of the upper G-tetrad with T2 being well stacked on the G-tetrad and T1 slightly tilted. The thymine at the 3' end T18 is also stacked at the bottom G-tetrad.

Table 3.1. Statistics of the computed structures of the *G15* G-quadruplex ^a

A. NMR restraints		
distance restraints	D ₂ O	H ₂ O
intraresidue	155	0
sequential (i, i+1)	91	11
long-range (i, ≥i+2)	5	43
other restraints		
hydrogen bond	48	
dihedral angle	12	
repulsive	16	
B. Structure statistics		
NOE violations		
number (>0.2 Å)		0.1±0.0
maximum violation (Å)		0.158±0.025
RMSD of violations (Å)		0.018±0.002
deviations from the ideal covalent geometry		
bond lengths (Å)		0.002±0.000
bond angles (deg)		0.665±0.004
impropers (deg)		0.328±0.004
pairwise all heavy atom rmsd values (Å)		
all heavy atoms except G6, G10, G14		0.73±0.13
all heavy atoms		1.77±0.39

^aPDB ID 2MB2.

Indeed, the structure of *G15* is very similar to that of *T95-2T* (pairwise r.m.s.d. < 1 Å for all heavy atoms, excluding residues in the propeller loops), except that single-residue thymine loops are replaced by single-residue guanine loops (Figure 3.2.4).

Figure 3.2.4. Comparison between the structures of *G15* (blue) and *T95- 2T* (red). (A) Side view and (B) top view of a representative structure.

3.2.3. Single-residue propeller loops

We hypothesized that any nucleotide (G, T, C, A, or even an abasic residue) could form a single-residue propeller loop bridging three G-tetrad layers of a parallel-stranded G-quadruplex. The loop residues G6, G10, G14 of *G15* were substituted by T, C, A, or S (a dSpacer–deoxyribose with no base) (Table 3.2).

The imino proton NMR spectra of these sequences were almost identical to that of *G15* (Figure 3.2.5), confirming that these sequences adopt the same propeller-type G-quadruplex fold irrespective of the loop residue.

Table 3.2. List of DNA sequences analyzed in this study ^a

Name	Sequence (5'-3')
<i>G15</i>	TTGGG <u>G</u> GGG <u>G</u> GGG <u>G</u> GGGT
<i>T95-2T</i>	TTGGG <u>T</u> GGG <u>T</u> GGG <u>T</u> GGGT
<i>C95-2T</i>	TTGGG <u>C</u> GGG <u>C</u> GGG <u>C</u> GGGT
<i>A95-2T</i>	TTGGG <u>A</u> GGG <u>A</u> GGG <u>A</u> GGGT
<i>S95-2T^b</i>	TTGGG <u>S</u> GGG <u>S</u> GGG <u>S</u> GGGT

^aLoop residues are underlined. ^bS represents d(spacer) – deoxyribose with no base

Figure 3.2.5. ¹H NMR spectra of the *G15*, *T95-2T*, *C95-2T*, *A95-2T* and *S95-2T* sequences. (Left) Imino proton spectra of these sequences, each displaying 12 sharp peaks, are almost identical. (Right) Aromatic proton spectra of these sequences are also very similar, except for the peaks from the loop residues 6, 10, and 14 (labeled in red).

This conclusion was supported by the CD spectra of these sequences showing a positive peak at 260 nm, characteristic of a parallel G-quadruplex (Figure 3.2.6) (117).

Figure 3.2.6. Effect of loop residue on CD spectrum. CD spectra of *G15*, *T95-2T*, *C95-2T*, *A95-2T* and *S95-2T* sequences.

3.2.4. Effect of loop residue on the stability of parallel G-quadruplexes

The effect of loop residue on the stability of this propeller- type G-quadruplex was investigated using UV melting experiments (Figure 3.2.7). The UV- melting data (Table 3.3) showed the highest melting temperature (T_m) for the sequence with loops containing no bases (dSpacer), very similar T_m values for sequences with thymine or cytosine loops, and a lower T_m value for the sequence with guanine loops, followed by that with adenine loops. This behavior can be attributed to the size differences between purines and pyrimidines and the ability of these bases to be hydrated or involved in secondary interactions. Our results agree well with the results previously reported for G- quadruplexes containing single-residue loops(149,150).

Figure 3.2.7. Effect of loop residue on the stability of G-quadruplex: UV melting curves of the *G15*, *T95-2T*, *C95-2T*, *A95-2T* and *S95-2T*. For clarity, data for *S95-2T* are not shown at temperatures below 50 °C, where an additional transition (~30% absorbance change) was observed.

Table 3.3. Thermodynamic Parameters for Propeller-Type G- Quadruplexes Containing Different Single-Residue Loops^a

Sequence name	T_m (°C) ^b	ΔH (kcal/ mol)	ΔS (kcal/ mol/ K)	$\Delta G(37^\circ\text{C})$ (kcal/ mol)
<i>A95-2T</i>	60.3	52.0	0.156	3.6
<i>G15</i>	66.0	58.4	0.172	4.9
<i>C95-2T</i>	75.6	79.1	0.227	8.7
<i>T95-2T</i>	75.6	86.8	0.249	9.5
<i>S95-2T</i>	>79.5	94.5	0.268	11.4

^aUV melting experiments were carried out in a buffer solution containing 10 mM Tris (pH 7) and 5 mM KCl. The oligonucleotide concentration was 4-6 μM . All parameters listed were obtained using the transition direction from the G-quadruplex to single-strand (unfolding event). ^b T_m values for the heating (unfolding) and cooling (folding) curves differ by less than 1 °C.

3.2.5. NMR structure of a G22 sequence

Noticeably, the spectrum of G22 showed major imino peaks corresponding to one predominant form in contrast with other poly-G sequences (other than G15) forming a mixture of G-quadruplexes (Figure 3.1). Most of the imino peaks displayed broad overlapping resonances. Twelve imino peaks and seventeen aromatic peaks could only be observed in the G22 sequence (Figure 3.3.3), whereas there should be twenty-five aromatics peaks corresponding to twenty-five residues in the sequence (22 G (H8) and 3 T (H6)). There are two possible explanations for undetectable aromatic peaks; (i) extreme peak broadening by internal motions among guanine residues, and (ii) overlapping of multiple peaks due to quasi-symmetry of the structure.

The direct and decisive approach of unambiguous resonance assignments was first used to probe the unusual structural characteristics of G22 G-quadruplex. Site-specific 4% ^{15}N - enrichment of all twenty-two guanines in the sequence (one guanine at a time), assisted in the identification of ten imino and aromatic guanine peaks labeled at their position in the sequence (Figure 3.3.1 A) (110). Two extra imino and aromatic peaks appear at higher temperature (35 °C), corresponding to guanine at 5th and 17th positions (Figure 3.3.1 B). This behavior could be attributed to the effects of internal motions on linewidths.

Assignments of 13 and 14th guanine was achieved by site-specific deuterium substitution, (151) (Figure 3.3.1 C) suggesting their involvement in the loops. Aromatic (H6) proton peak of three thymine at 1st, 2nd and 25th positions (Figure 3.3.4) were assigned by TOCSY and NOESY assignments (data not shown). Surprisingly, no NMR signal (imino and/or aromatic peaks) could be observed in site-specific ^{15}N -labeling for eight guanines belonging to two tracts of four consecutive guanines (G6,G7,G8,G9) and (G18,G19,G20,G21).

D₂O solvent exchange of G22 (Figure 3.3.2) showed slower exchange (~2 days) for six imino proton peaks indicating the presence of additional protecting layer structure. Usually, four guanines peaks are protected corresponding to central tetrad in a three-layer G-quadruplex.

Figure 3.3.1. Guanine imino and aromatic proton assignments of G22. A. Imino and aromatic protons were assigned by ^{15}N -filtered spectra of samples. The reference spectrum is shown on the top with assignments indicated over the spectrum. Below are the individual spectra of 4% ^{15}N -labeled samples with indicated positions.

Figure 3.3.1. B. Individual spectra of 4% ^{15}N -labeled, G22 samples with indicated positions at 35 °C. C. Assignment of G13 and G14 aromatic (H8) proton of G22 sequence by deuterium substitution.

3.2.6. Effect of mutations on the G22 sequence

Several variant sequences (G-to-T mutation) have been synthesized to understand the role of missing guanines tracts (G6,G7,G8,G9) and (G18,G19,G20,G21) on NMR spectrum of G22. The rationale behind these mutations was to devise the discontinuity in the guanine tract such that apparent symmetry could be broken, and internal dynamics could be inhibited in the

Figure 3.3.2. Imino proton spectra of G22 sequences in solution (top) and D₂O solvent exchange after ~ 2 days.

structure to facilitate the NMR study. Thymine base substitution at 7th guanine (termed as *T7*) and 19th guanine (*T19*) exhibited the similar imino proton spectrum highlighting the symmetrical environment in the structure at these loop positions (Figure 3.3.5).

Due to the great spectral correspondence of G22 and *T7*, we focused our study on *T7*. NMR spectrum of *T7* is well resolved to display sixteen imino and twenty-five aromatic peaks (Figure 3.3.3). The spectral similarity of G22 and *T7* suggests the adoption of similar G-quadruplex structure and most importantly in *T7* mutation enabled the appearance of the undetectable (missing/broaden) NMR peaks in native G22 sequence. Unambiguous assignments of *T7* guanine imino protons were achieved based on site-specific 4% ¹⁵N-enrichment of samples, one residue at a time (Figure 3.3.4). The similarity in the NMR spectra

Figure 3.3.3. Imino and aromatic proton spectra of G22 and *T7* mutated sequence. Peak assignments are written on top. Assignments indicated in red color correspond to peaks appear in *T7* only.

Figure 3.3.4. Guanine imino and aromatic proton assignments of *T7* mutated sequence. Imino and aromatic protons were assigned by ^{15}N -filtered spectra of samples. The reference spectrum is shown on the top with assignments indicated over the spectrum. Below are the individual spectra of 4% ^{15}N -labeled samples with indicated positions.

of *T7* and *G22* facilitated the assignments for both cases; resonances of *T7* could be deduced based on the spectral resemblance between the two. Structural similarity of *G22* and *T7* G-quadruplexes is also evident from the similar CD pattern as well as similar native PAGE mobility (data not shown).

Mutations at *T8* (or *T20*, structurally symmetric positions) showed the similar spectrum to *G22* establishing the formation of same parent fold and suggesting their role in forming loops. Conversely, mutations at *T6* and *T9* (or *T18* and *T21*) positions are unfavorable, suggesting their involvement in G-tetrad core formation in *G22* G-quadruplex structure (Figure 3.3.5).

Figure 3.3.5. NMR imino and aromatic proton spectra of single G-to-T mutations on the *G22* sequence.

Solvent exchange experiment with D₂O (Figure 3.3.6, ~ 2days) shows the presence of eight protected peaks, corresponding to the formation of four-layer G-quadruplex structure.

Figure 3.3.6. Imino proton spectrum of *T7* mutated sequences and D₂O solvent exchange after ~ 2 days. Eight imino peaks corresponding to two outer tetrads disappears and 8 peaks from two middle tetrads are protected from an exchange with D₂O and stays there.

3.2.7. G-quadruplex folding topology of *G22* and *T7* sequence

With the help of spectral assignments of *G22* and *T7*, G-tetrad alignments were determined by using imino (H1)-aromatic (H8) NOESY connectivity pattern (Figure 3.3.7). G-tetrads alignment established for *T7* are G3•G12•G15•G24 (red), G4•G11•G16•G23 (blue), G5•G9•G17•G22 (green) and G6•G10•G18•G21 (black) by following specific NOE cross peaks (Figure 3.3.7 A). Similarly, G-tetrads alignment established for *G22* are G3•G12•G15•G24 (red), G4•G11•G16•G23 (blue) and G5•G9•G17•G22 (green) by following specific NOE cross peaks (Figure 3.3.7 B). *T7* G-quadruplex topology built from alignments of four tetrads shows chair type antiparallel structure (Figure 3.3.7 C) and *G22* G-quadruplex topology built from alignments of three tetrads shows chair type antiparallel structure (Figure 3.3.7 D).

Figure 3.3.7. Determination of G-quadruplex folding topology of *T7* and *G22*. **A.** and **B.** NOESY spectra of (mixing time, 250 ms) *T7* (Left) and *G22* (Right): Imino-H8 proton region is showing characteristic guanine imino-H8 cross-peaks for G-tetrads. G-tetrad alignments based on NOE pattern. **C.** *T7* topology confirmed by NOE pattern. **D.** *G22* topology, bottom region of *G22* topology with three tetrads is confirmed by NMR assignments and proposed dynamic structure of top tetrad based on the similarity of *T7* topology.

Solvent exchange experiments independently confirms the proposed folding topology for *T7* and *G22*, supporting central positioning of G4•G11•G16•G23 and G5•G9•G17•G22 protected tetrads (Figure 3.3.2 and 3.3.6) except for the fact that in *G22* imino proton peaks corresponding to G9 and G17 were not observable at room temperature (Figure 3.3.2).

Figure 3.3.8. NOESY spectra of *T7* and *G22* (mixing time, 300 ms): H8/H6-H1' region is showing NOE sequential connectivities. Intra-residue cross-peaks are labeled with the corresponding residue numbers. Syn guanines are framed (magenta). Rectangular NOE connectivities show syn- to -anti connection.

Intraresidue peaks H1'-H8/H6 cross peaks are assigned for *T7* and *G22* (Figure 3.3.8). High-intensity intra-residue H1' -H8 cross peaks (framed in magenta) for syn guanines were observed for G3,G5,G9,G11,G15,G17,G21 and G23 in *T7* and G3,G5,G11,G15,G17 and G21 for *G22*. Classical NOE sequential connectivity between residues (H8(*n*)-H1'(*n*)-H8(*n*+1),5'-3' direction) strongly depends on upon the glycosidic conformation. H8/H6-H1' NOE sequential connectivity from the syn-to-anti connection is confirmed by rectangular NOE pattern (Figure 3.3.8) correspondingly eight syn-to-anti rectangular NOE patterns were observed for *T7* and six for *G22*. Other guanines in the tetrad adopt anti glycosidic conformation, assignments indicated at intra-residue peak.

T7 and *G22* G-quadruplex adopts antiparallel strand orientations where guanine residues along the G-column were oriented in syn-anti and anti-syn conformations. Consequently, hydrogen-bond directionalities of four tetrads were anti-clockwise, clockwise, anti-clockwise and clockwise respectively. Hence two narrow and two wide grooves were created, two nts edgewise loop (three GG loops in *G22* and two GG and one GT in *T7*) spans over the narrow and wide groove.

In summary, it is demonstrated that *T7* and *G22* G-quadruplexes are conformers, and *T7* mutation enabled us to characterize the chair-type, four-layer, antiparallel G-quadruplex structure adopted by *G22*.

Structural model of *G22* G-quadruplex (Figure 3.3.9) was determined by available NMR restraints. The structure corroborates to the *G22* topology as chair-type antiparallel G-quadruplex with three stable G-tetrads and one apparent dynamic tetrad on top formed by swapping of guanines in the loops and G-tetrad. Top tetrad and two edgewise loops spanning over the wide groove did not converge since no NMR restraints were available for equivalent pairs of (G6, G7, G8, G9 and G18, G19, G20, G21) residues. Molecular dynamic simulations could be run in future to gain structural insights on G-tetrad and loops formation by these guanines. Central edgewise loop (G13, G14) connecting across narrow groove is converged, and these residues are partially stacked on the G-tetrad. Thymines T1, T2 and T25 are positioned on top of the G-tetrad with T1 and T25 being partially stacked on the G-tetrad and T2 tilted

outward. These stacking interactions with outer tetrads compact the structure, which is evident from native PAGE experiments (Figure 3.3.9 A) where mobility of four layer G22 is similar to three-layer, hybrid G-quadruplex reference.

Figure 3.3.9. Structural model of G22 determined by NMR. A. Side-view of 10 lowest-energy superimposed solution structures. B. Side-view of a representative structure. The bases of thymine are colored green while guanines in tetrad core are colored cyan and guanines in loops are colored orange. Backbone and sugar are colored grey; O4' atoms are in red.

Four-layer, chair-type, antiparallel G-quadruplex structure adopted by G22 is unique and different from singular reported a member of four layer, chair-type antiparallel G-quadruplexes family adopted by $(G_4C_2)_4$ repeats (152,153). $(G_4C_2)_4$ G-quadruplex has three CC edgewise loops in contrast to GG loops in G22. An interesting observation is that in the structure of $(G_4C_2)_4$, two edgewise loops connect narrow groove and a central loop connects wide groove while in the structure of G22, two edgewise loops connect wide groove and one central loop connects narrow groove.

3.2.8. Examination of polymorphism of G22 sequence; equilibrium of higher order structure and monomer

G-quadruplex structures are known for their polymorphism, the presence of higher order structures along with predominant monomer form has been demonstrated in G22 sequence by native PAGE and CD spectroscopy. Native PAGE separation of G22 (Figure 3.3.10 A. lane (c)) revealed two species; one lower band (higher mobility) corresponding to a major G-quadruplex monomer ($\geq 50\%$) and another diffused band on top (lower mobility) corresponding to the formation of higher order G-wires.

These two different bands of G22 were separated and reexamined; represented by lane (d) and lane (b) (Figure 3.3.10 A). G-quadruplex monomer fraction appeared as single predominant conformation (lane (d)), some re-equilibrium has been observed for higher order fraction (lane (b)). Comparing the migration rate with other G-quadruplex references; lane (a) (93 *del*: interlocked dimer (82)) and lane (e) (Lo2: (3+1) monomer (154)), It is evident that migration of major G22 monomer is similar to G-quadruplex monomer Lo2 while the higher order structures migrate slower than G-quadruplex stacked dimer reference 93~~del~~. These results suggest that in G22 sequence in solution forms G-quadruplex monomer and other higher order structures in equilibrium.

G22 and its gel separated fragments (lane (b), (c) and (d)) were further investigated by CD spectroscopy (Figure 3.3.10 B). Higher order fragments (in blue) showed the characteristic feature of parallel structure (positive peak at 260 nm and negative at 240nm) while the major G-quadruplex fraction (in red) showed the feature of antiparallel (2+2) structure (positive peak at 295 nm and 240 nm and negative at 260 nm). Native G22; a mixture of both G-quadruplex species (in black) showed two positive peaks at 260 and 295 nm and a negative peak ~240 nm, which would be similar to the arithmetic sum of CD spectrum of individual species.

Figure 3.3.10. Native PAGE, CD and imino proton spectra of G22 gel purified fractions. A. Native 20% PAGE of (lane a) 93 *del* d(GGGGTGGGAGGAGGGT), reference of an interlocked dimeric G-quadruplex (82) (lane b) gel-purified higher order fragment of G22, (lane c) G22, (lane d) gel-purified lower band fragment of G22 (lane e) Lo2 DNA (d(TTGGGTAGGGTTAGGGTTAGGGA), three-layered, monomeric, (3+1) hybrid) (154). B. CD spectra of G22 (c) and its gel –purified fragments (b) and (d). C. Imino proton NMR spectra of G22 (c) and gel purified fragment of G22 (d).

Finally, imino proton NMR spectrum of native G22 and its monomer fragment showed similar imino peak distribution, suggesting that higher order G-quadruplex does not show any signatures in imino proton peaks could be contributing to the background (large molecules to study by NMR). Hence, the solution structure of G22 represents the predominant monomeric, antiparallel (2+2) G-quadruplex. Slight peak shift pattern in (Figure 3.3.10 C.-(d)) is due to higher salt concentration (data not shown).

The equilibrium between monomeric G-quadruplex and higher order structures can be altered by sample preparation conditions, particularly with annealing and with time. It has been observed that the monomeric antiparallel G-quadruplex form evolves with time (data not shown). Upon heating (20 °C to 90 °C) antiparallel G-quadruplex structure unfolds and converts into parallel higher order structures (Figure 3.3.11 A) as observed by the changes in CD spectra. Same sample upon cooling (90 °C to 20 °C) does not convert backs to the antiparallel G-quadruplex but a mixture of parallel higher order structures and antiparallel G-quadruplex (Figure 3.3.11 B).

Thus, CD thermal denaturing experiment showed annealing process disturbs the conformational equilibrium. Due to the irreversibility of CD melting, the thermal stability of four-layer, antiparallel G22 and T7 G-quadruplex is calculated by following the decrease in CD signal at 295 nm on heating (Figure 3.3.11 C). The G22 structure is quite stable with $T_m \sim 68.5$ °C and T7 mutation increased the stability by $\sim 3^\circ\text{C}$ ($T_m \sim 71.5$ °C) in 5 mM K⁺ solution.

Figure 3.3.11. Thermal denaturation CD spectra of G22 gel purified fragment. A. Heating, original antiparallel form converts to parallel. B. Cooling, original parallel form converts to a mixture of parallel and antiparallel. C. Thermal stability of antiparallel G22 and T7 G-quadruplex (T_m at ~ 68.5 and ~ 71.5 °C) calculated by following the CD signal at 295 nm for the unfolding (heating) event.

Summarizing the above arguments, we report that the G22 sequence exists as a mixture of monomeric, antiparallel G-quadruplex and higher-order parallel G-quadruplexes in solution, both forms thermally quite stable. The equilibrium between the two forms depends on sample conditions particularly on evolution time and annealing.

3.3. Conclusions

In this study, poly-G DNA G-quadruplex forming sequence motifs have been explored and two novel intramolecular G-quadruplex structures were presented. G-quadruplex structures formed by a continuous tract of 15 and 22 guanines represent unique, well-defined robust scaffold, observed as a single major conformation. The structure of *G15* is the propeller-type G-quadruplex involving three G-tetrad layers and three single-residue propeller loops, while the structure of *G22* is four layer, chair-type antiparallel quadruplex with three edgewise GG loops displaying unique feature of dynamic swapping between guanines in the G-tetrad core and loops.

Poly-G tract length dependent study of G-quadruplex formation also provides insights about the natural conformational selection for the three-layer G-quadruplex with three single residue propeller loops for *G15* and four-layer, antiparallel G-quadruplex with three double residues edgewise loops over myriad of accessible G-quadruplex conformations for different poly-G tract lengths. In principle, multiple combination of G-quadruplex conformations involving different structure features such as multiple or single DNA strand, number of stacked G-tetrads, varying loop length and geometry, flanking bases, abasic site and bulges could possibly coexist in solution.

Such unique G-quadruplex motifs can be formed by a fragment of 15/22 continuous guanines anywhere along a poly-G sequence. Our findings broaden the description of G-quadruplex forming sequences especially in context with poly-G sequences.

Coincidentally, among the poly-G tracts present in the genome of DOG-1-deficient *C. elegans*, DNA deletions were found exclusively for those with $n \geq 15$ and $n \geq 22$ (73,74). Given the role of DOG-1 in protecting *G15* and *G22* tracts from deletion, it is possible that unique G-quadruplex structures in these G-tracts can be biologically relevant, being specifically recognized by proteins or other cellular molecules.

Chapter 4

G-quadruplex structure in PARP-1 promoter

Poly(ADP-ribose) polymerase-1 (PARP-1) is a DNA repair protein. PARP-1 is found to be upregulated in cancer cells, suggesting that cancerous cells rely more on the PARP-1 for DNA repair likely due to defects in other pathways and hence PARP-1 inhibitors has emerged as an attractive target for cancer therapy. Recently, G-rich sequences have been identified in the PARP-1 promoter and hence targeting of G-quadruplex structure in *PARP-1* promoter could serve as an alternative approach for PARP inhibitors. Thus, in this work using NMR, UV, CD spectroscopy and Gel electrophoresis, we investigate the structure of G-quadruplexes formation in 23-nt G-rich DNA sequence in *PARP-1* promoter. Our study has revealed the (3+1) hybrid, intramolecular, three-layered G-quadruplex topology. Unique structural features of *PARP-1* G-quadruplex are adenine residue bulge and capping structure on top formed by the central edgewise loop and 5' flanking terminal which could be exploited as a recognition element by small molecule ligands, providing a platform for the development of alternative drugs for PARP-1 inhibition.²

² The work in this chapter is being written for publication: "G-quadruplex structure with a bulge in *PARP-1* promoter", Anjali Sengar, J.Jeya Vandana, Vicki S Chambers, Shankar Balasubramanian and Anh Tuân Phan.

Part of this work has been presented as a final year project report "Structural Characterization of a DNA G-quadruplex in PARP-1" by J.Jeya Vandana.

4.1. Introduction

Poly (ADP-ribose) polymerase-1 (PARP-1) is an abundant, ubiquitously expressed multifunctional nuclear enzyme, encoded by the highly conserved *PARP-1* gene. PARP-1 catalyzes post-transcription modification of several nuclear proteins by Poly(ADP-ribosyl)ation reaction. ADP-ribosylation is the addition of ADP-ribose moieties to the various acceptor proteins using cellular nicotinamide adenine dinucleotide (NAD^+) substrate (155-157). Numerous studies have suggested the important role of PARP1 in diverse cellular processes, such as DNA repair, replication, transcriptional regulation of various proteins (158), maintenance of genomic integrity, regulation of telomerase activity (159), cell signaling (157) and in chromatin remodeling (160-162).

The main contribution of PARP1 is in DNA repair pathways (163). The involvement of PARP-1 in single strand break repair (SSB) is well known. However, some recent studies have also underlined its contribution in double strand break repair (DSB) (164). PARP-1 activates the process of SSB repair by recognition and binding at the site of single strand DNA breaks (SSBs) via zinc finger motifs within PARP1(165,166), initiating ADP-ribosylation reaction which is followed by recruitment of multiprotein complex for the DNA repair and cell signaling (167). Furthermore, PARP1 also modifies the chromatin structure by poly ADP-ribosylation of the histones and enabling the access of repair proteins to damaged DNA (168). In cancer cells, PARP1 is found to be upregulated (169,170), suggesting that cancerous cells rely more on the PARP1 for DNA repair likely due to defects in other pathways and hence making PARP1 inhibition an attractive target for cancer therapy(171). Some PARP-1 inhibitors are already in clinical trials and are showing promise in the treatment of breast, ovarian and prostate cancers. Most notably PARP-1 inhibition is targeted in the Homologous-recombination-deficient tumors such as in tumors containing BRCA mutations(172,173).

Few studies have investigated the PARP-1-G-quadruplex binding interaction with *c-kit*, *c-myc* and human telomere G-quadruplex (70). It has been found to influence the transcriptional regulation of the human *c-myc* gene via interaction with the G-quadruplex structure present in the *c-myc* promoter. PARP-1

converts the G-quadruplex structure into the more transcriptionally active B-DNA form, thus potentially upregulating the expression of the *c-myc* gene (174,175).

G-quadruplex DNA sequencing analysis of human genome results have identified the potential G-quadruplex forming sequence in the promoter region of the *PARP-1* gene, where G-quadruplex structure were found to be present and stable in chromosomal DNA context. (personal communication with S. Balasubramanian). In this work, we investigated the G-quadruplex structure formation in 23-nucleotide G-rich sequence in *PARP-1* promoter. Our study has revealed the (3+1) hybrid, intramolecular, three-layered G-quadruplex topology with unique structural features of adenine bulge and capping structure on top formed by the central edgewise loop and 5' flanking end residues. G-quadruplex structure formation in this sequence G-quadruplex structures in this gene may possess a regulatory role in the production of the *PARP-1* protein. Stabilization of G-quadruplex could help in downregulation (or silencing) of the *PARP-1* expression which could help in inducing synthetic lethality in cancerous cells. Thus, G-quadruplexes in the *PARP-1* gene may be promising as anti-cancer therapeutic targets. Also, such G-quadruplexes may also serve as therapeutic targets for the prevention of other diseases that occur due to overexpression of *PARP-1* induced by cellular stress and subsequently influence the expression of other genes.

4.2. Results and Discussion

4.2.1. G-quadruplex of *P3* sequence and effect of flanking bases

G-quadruplex structure formation in *PARP-1* promoter DNA sequence (termed as *P3*) located 126 nt upstream from the transcription start site (TSS) is examined by NMR with naturally linked flanking sequences. NMR spectra of the *P3* with various flanking ends are studied in a solution containing 70 mM KCl, 20 mM KPi, pH 7 at 25 °C (Figure 4.1, Table 4.1). The *P3* sequence shows twelve sharp imino peaks from 10.6 – 12.0 ppm, indicative of the formation of the three-layer G-quadruplex fold as a major structure supplemented with the

fraction of other higher order structures characterized by the broad hump in the background.

Figure 4.1. Imino proton NMR spectra of *P3* with different flanking bases.

Likewise, other *P3* sequences with 5' flanking bases (*TP3*, *CTP3* and *CCTP3*) show similar chemical shift distribution pattern of imino proton peaks suggesting similar G-quadruplex structure formation in all sequences with 5' flanking ends. Furthermore, the presence of flanking bases at 5' end reduces the background hump in the spectrum significantly compared to *P3*, most likely disfavoring aggregation of structures to form higher order structures. Additional peak(s) at > 12.0 ppm, mark the presence of other base pairs can be seen in all the sequences and the effect of flanking bases on them. However, the effect flanking sequences at 3' is different and shows a mixture of G-quadruplex conformations. In this chapter, we focus on the *TP3* G-quadruplex; owing to its superior spectral quality as it represents the major G-quadruplex structure presented in other *P3* sequences with different 5' flanking ends.

Table 4.1. List of sequences found in PARP-1 promoter with flanking bases.

Name	Sequence
<i>P3</i>	GGGGGCCGAGGCGGGGCTTGGG
<i>TP3</i>	T <u>GGGGGCCGAGGCGGGGCTTGGG</u>
<i>CTP3</i>	CT GGGGGCCGAGGCGGGGCTTGGG
<i>CCTP3</i>	CCT GGGGGCCGAGGCGGGGCTTGGG
<i>P3C</i>	GGGGGCCGAGGCGGGGCTTGGG C
<i>TP3C</i>	T GGGGGCCGAGGCGGGGCTTGGG C

4.2.2. (3+1) hybrid intramolecular G-quadruplex formation by *TP3* sequence

CD and native PAGE experiments suggest the (3+1) hybrid G-quadruplex structure in *TP3* (Figure 4.2). The CD spectrum of *TP3* G-quadruplex displays two positive peaks at 265 and 290 nm along with a negative peak at 240 nm, characteristic of a (3+1) hybrid G-quadruplex (Figure 4.2 A.) (117). A noticeable difference here is the enhanced peak intensity at 265 nm, which could be due to the presence of other parallel stranded higher order structures in the sample. Native PAGE analysis compares the migration of *TP3* G-quadruplex structure with respect to other references G-quadruplex structures (Figure 4.2 B.) Migration of *TP3* sample is comparable to that of *Lo2* (three-layered, monomeric and (3+1) hybrid) (154) and faster than other reference samples like *T95-2T* (three-layered, monomeric and propeller-type parallel) and *T95* (stacked dimer) (147), which suggests that *TP3* forms a three-layer monomeric G-quadruplex. The presence of faint smeared upper band in *TP3* suggests that a fraction of other higher order structures co-exist in the sample. This idea is also supported by CD (enhanced peak at 260 nm) and NMR (background hump) results.

Figure 4.2. CD and native PAGE of *TP3*. (A) CD spectrum of *TP3* in 70 mM KCl, 20 mM KPi, pH 7 at 25 °C. (B) Native PAGE mobility shift analysis of *TP3* with known G-quadruplex structure references; *T95* (d(GGGTGGGTGGGTGGG), stacked dimer) (147), *Lo2* (d(TTGGGTAGGGTTAGGGTTAGGGA), three-layered, monomeric, (3+1) hybrid) (154) and *T95-2T* (d(TTGGGTGGGTGGGTGGGT), three-layered, monomeric, propeller-type parallel) (147).

4.2.3. Spectral Assignments of *TP3* sequence

Unambiguous assignments of guanine imino (H1) (Figure 4.3) and aromatic (H8) (Data not shown) protons of *TP3* were carried out via site-specific 4% ^{15}N -labeling (110), and site-specific ^2H (151) substitution approaches.

TGGGGCCGAGGCGGGCTTGGG

Highlighted in blue are the guanines participating in G-quadruplex core formation observed by imino peak identification in site 4% ^{15}N -labeling experiments.

Figure 4.3. Imino proton resonance assignments of *TP3* indicated on the reference spectrum (top). Below are the individual spectra of 4% ^{15}N labeled guanines at indicated positions.

Imino protons assignments revealed that from the first G-tract of five guanines; G2 and G6 are not involved in G-quadruplex core formation and from the third G-tract of four guanines; G14 is not involved. We can speculate that adenine interruption in the second G-tract could be incorporated as a bulge. Additional two-dimensional NMR experiments (NOESY) need to be done to understand the imino-H8 proton cyclic connectives around G-tetrads and elucidate the topology of the *TP3* G-quadruplex. Due to the poor long-term stability of G-quadruplex structure in *TP3* with time at room temperature (Figure 4.4), we could not execute the long (> 15 h) two-dimensional NMR experiments. Imino

proton NMR and native PAGE studies have indicated the conversion of the *TP3* G-quadruplex form into various other higher order structures at room temperature over time (Figure 4.4). Imino proton peaks disappeared gradually with the appearance of hump in imino proton NMR (Figure 4.4. A) and the corresponding band representing (3+1) fold in native PAGE also disappeared (Figure 4.4. B).

Figure 4.4. Time evolution of *TP3* monomers and higher order aggregates at room temperature A. Imino proton resonances of *TP3* at varying time intervals B. Time-dependent native PAGE analysis of *TP3*.

4.2.4. Effect of mutations on *TP3* sequence

Variant of *TP3* sequence were examined where non-tetrad forming guanine residues (2,6 and 14) were replaced by thymine (G to T substitution) (Figure 4.5). Imino proton spectra of mutants show that single thymine mutations at G2 and G14 is not tolerated and results in the co-existence of a mixture of G-quadruplexes. Markedly, guanine to thymine substitution at position 6 (*TP3-T6*) shows the extreme spectral similarity to that of *TP3* indicating the formation of same G-quadruplex folding in both the sequences. Furthermore, *TP3-T6* G-quadruplex structure displays improved stability of the G-quadruplex structure with time upon exposure at room temperature (> 2 days, data not shown), enabling a detailed NMR structural study viable. Thus, the *TP3-T6* sequence is studied to elucidate the G-quadruplex structure topology.

Figure 4.5. Imino proton spectra region of *TP3* and mutated *TP3* sequences (G to T substitution at indicated positions).

Increased temporal stability of *TP3-T6* compared to *TP3*, could be due to the limitation posed by (G to T mutation) at G6 to the dynamic equilibrium of co-existence of multiple possible G-quadruplexes formed by inclusion of other guanines from the tract and subsequently to the aggregation. However, mutations of G2 and G14 shows completely different effects suggesting that these guanines residues might involve in other interactions. Two-point and three-point thymine mutations at 2,6 and 14th guanines show that mutation of G6 with others (*TP3-T2,6* and *TP3-T2,6,14*) results in the twelve imino peaks, forming single three-layered G-quadruplex structure while *TP3-T6,14* shows mixture of two G-quadruplex conformations.

4.2.5. Spectral assignments of *TP3-T6* sequence

Spectral assignment of *TP3-T6* was carried out by site-specific ¹⁵N-enrichment

Figure 4.6. Imino (H1) and aromatic (H8) proton assignments for the *TP3-T6* sequence. A. Imino (H1) and aromatic (H8) proton assignments indicated on the reference spectrum for *TP3*. B. Imino (H1) and aromatic (H8) proton assignments indicated on the reference spectrum on top for *TP3-T6*. Below are the individual spectra of 4% ^{15}N labeled guanine bases at indicated positions.

(4%) approach. Guanine imino (H1) and aromatic (H8) protons were unambiguously assigned for *TP3-T6* (Figure 4.6. B) and some of the assignments were also extrapolated from *TP3* (Figure 4.6. A) as proton's chemical shifts in both the sequences are extremely similar.

4.2.6. (3+1) G-quadruplex fold in *TP3-T6* sequence

Folding topology was determined using cyclic NOE connectivities between the imino (H1) proton of a guanine and the aromatic (H8) proton of neighboring guanine in the same tetrad. G-tetrads alignment established for *TP3-T6* are G3•G12•G15•G21 (red), G4•G11•G16•G22 (blue) and G5•G9•G17•G23 (green) by following specific NOE cross peaks (Figure 4.7 (A) & (B)). Adenine interruption in the 3rd guanine tract (A10) is incorporated as bulge between two tetrads.

Figure 4.7. Folding topology of *TP3-T6*. A. NOESY spectrum (mixing time, 200 ms) showing the H1/H8 connectivity of G-tetrads. Guanine H1-H8 cross-peaks for G-tetrads are framed and labeled. The residues in different G-tetrads are indicated by different colors. B. Specific H1-H8 connectivity pattern. Schematic of the folding topology of *TP3-T6*.

Solvent exchange experiment with D₂O independently confirms the proposed folding topology, where centrally positioned G4, G11, G16 and G22 are among the protected residues (Figure 4.8) most protected from the solvent exchange.

Figure 4.8. Imino proton spectrum of *TP3-T6* recorded after 20 min exposure in D₂O, showing protected peaks corresponding to middle tetrad (in blue). Peaks marked with an asterisk are discussed later.

High-intensity intra-residue NOEs ($H8(n) - H1'(n)$) were observed for G3, G9, G11, G15 and G21 (Figure 4.9, framed in black), indicating the *syn* glycosidic conformation adopted by them. NOE sequential connectivity between residues ($H8(n) - H1'(n) - H8(n+1)$, 5'-3' direction) strongly depends on the glycosidic conformation. Syn-to-anti inter-residue connection shows both $H1'(n) - H8(n+1)$ and $H1'(n+1) - H8(n)$ cross peaks marked by rectangular NOE pattern (Figure 4.9, color coded). Other guanines in the tetrad adopt anti glycosidic conformation, assignments indicated at intra-residue peak.

Figure 4.9. NOESY spectrum of *TP3-T6* (mixing time, 300 ms). Intra-residue NOE correlations are labeled according to its respective nucleotide number. Cross-peaks of strong intensity are framed in black and implicates residue 3, 9, 11, 15, and 21 adopt the syn glycosidic orientation. Sequential syn-to-anti (3-4, 11-12, 15-16 and 21-22) walk traced by blue, magenta, red and green rectangles respectively.

Imino-imino (H1-H1) NOE peaks of nearby guanines corresponding to the stacked tetrads can also be seen observed. Distances of (H1) -(H1) depend on the tetrad polarity (directions of imino (H1) to O6 bond), tetrads of the same polarity have same glycosidic conformation of adjacent guanines. (H1) -(H1) NOEs are of different intensities in different tetrad stacking modes has been discussed in the review paper (109). Medium intensity cross peaks for the guanine pairs (G4/G5, G9/G11, G16/G17 and G22/G23) in adjacent tetrads is observed (Figure 4.10), confirming the same polarity for central and bottom tetrad (clockwise) and opposite polarity of the top tetrad (anticlockwise).

Figure 4.10. NOESY spectrum of *TP3-T6* showing H1-H1 through space correlations labeled as indicated on the spectrum.

Linkers in *TP3-T6* joins the G-columns by edgewise, edgewise and propeller loops respectively forming (3+1) hybrid scaffold. The groove widths are wide, narrow, medium and medium correspondingly. First edgewise loop (TCC) connects across the wide groove at the bottom, second edgewise loop (CG) connects across the narrow groove at the top, while third propeller loop (CTT) traverses across the medium groove connecting two parallel stands.

Observation of additional broad imino proton peaks at ~12.2–12.4 ppm could be due to the hydrogen bonding interaction between terminal G2 with G14 of the edgewise loop and capping the top tetrad. Protection of imino protons peaks of G12 and G15 in the D₂O solvent exchange experiment (Figure 4.8) supports this argument where G2 and G14 pair could stack on the top tetrad such that imino protons of G12 and G15 remains protected while imino proton of G3 and G21 are exposed for deuterium exchange. Structural details of such interactions

between edgewise loop residues and 5' terminal residues in *TP3-T6* quadruplex structure remain to be fully characterized. This unique base pairing can form a capping structure on top and provide the additional stability.

4.3. Conclusions

We have revealed the (3+1) hybrid G-quadruplex folding topology adopted by promoter sequence of the PARP-1 gene. Unique structural features of the *TP3-T6* G-quadruplex are adenine residue bulge and unique capping structure formed by the central edgewise loop and 5' flanking terminal, could be exploited as a recognition element for small molecule ligands.

In particular, it may be possible to target specifically the G-quadruplex in the PARP-1 gene with rationally designed ligand to serve as a specific PARP-1 inhibitor. Hence, the investigation of the structural properties of the *TP3* G-quadruplex provides a platform for the development of alternative drugs for PARP-1 inhibition with enhanced potencies for anti-cancer therapeutics and other related diseases.

Chapter 5

Single molecule FRET study of structural dynamics in G-quadruplex DNA

G-quadruplex formation inside the cell is a dynamic process, hence better understanding of the G-quadruplex structural dynamics is important for understanding their biological functions. Single molecule FRET can provide the wealth of information on the conformational dynamics of G-quadruplex. This chapter describes the single molecule FRET instrumentation and demonstrates the functionality of our in-house built sm-FRET set up by observing two-state dynamics of a model hairpin system. Using sm-FRET technique, we have studied the structural dynamics in parallel stranded G-quadruplexes. We have demonstrated that the folding-unfolding dynamics of a stable parallel G-quadruplex structure could be manipulated by sequence modifications such as guanine to inosine substitution and loop length increment, preserving the parallel G-quadruplex topology. We have observed single molecule FRET transitions between folded and unfolded structures involving two other states, which could correspond to intermediate structures. These intermediate structures could potentially affect the dynamic interactions of G-quadruplexes with other small molecule ligands and proteins *in vivo*.

5.1. Introduction

As discussed in the introduction, G-quadruplex structures have been found to exist in multiple conformations. Numerous studies have proposed the regulatory roles of G-quadruplex structure in cellular processes and as potential drug targets. Understanding of the conformational heterogeneity and dynamics of G-quadruplex structures is important since G-quadruplex formation inside the cell is a dynamic process interacting with other molecules.

Different G-quadruplex topologies (parallel, antiparallel and hybrid) can be adopted depending on the sequence composition and experimental conditions. These G-quadruplex topologies give rise to very different shape features such as groove width and projection of linker residues. Interestingly, out of all G-quadruplex topologies; intermolecular, parallel stranded G-quadruplex with three propeller loops is most distinctive due to its compact globular shape. In intramolecular, parallel G-quadruplexes propeller loops are projected outside in such a way that outer tetrads are exposed for molecular recognition and binding (176). Proteins and small molecules have been shown to use of these planar tetrads for recognition and stacking interactions (63,177,178). These structures have been identified in many biologically relevant G-rich regions such as human telomere (preferred in crowded solution) (179), the dominant structure in most of promoters available so far (60,62,180-183), CEB1 and CEB 25 minisatellites (40,184) making them an interesting class of G-quadruplex structures to study.

Most of the sm-FRET studies on structure and dynamics of G-quadruplexes have been focused on the human telomeric sequence which forms a mixture of conformations (98,99,101,102,185,186). Except couple of studies, focused on the G-quadruplex dynamics in *c-myc* promoter sequence and detected highly stable (longer unfolding time) parallel G-quadruplex structure (95,96,107). Thus, we have focused our study on parallel, propeller G-quadruplex using a simple model of a highly stable, robust, parallel G-quadruplex formed by the *T95-2T* sequence, in which four tracts of three consecutive guanines are connected by three single loops (147). In this study, we have investigated the influence of sequence modifications specifically guanine to inosine substitution and loop length modification on folding-unfolding dynamics on parallel G-quadruplexes

using the sm-FRET technique. Previous ensemble studies have shown that the thermal stability in G-quadruplexes drops with the increase in loop size (18,187). Similar negative effects on the thermal stability of G-quadruplexes have been observed by the inosine substitution (188).

Our approach is unique in a manner of using NMR spectroscopy and sm-FRET jointly to validate the G-quadruplex conformations. The result of our investigation showed parallel G-quadruplex structure formed in *T95-1T* [T(G₃T)₃] sequence is highly stable; no transitions could be seen within observation time, even at low salt concentration. Sequence modifications in *T95-2T* sequence such as guanine to inosine base substitution and the inclusion of long central loop (9T), make the structure more dynamic by destabilizing the folded conformation. Particularly, we observed structural transitions involving two intermediate states in the folding-unfolding dynamic of parallel G-quadruplexes.

5.2. Instrumentation

Single molecule FRET setup requires three main parts (i) excitation of fluorophores (ii) collection of emitted fluorescence (iii) signal detection.

5.2.1. Excitation: Total internal reflection (TIR) microscopy:

Prism type-TIRF (total internal reflection) excitation is used to gain high imaging sensitivity by reducing the excitation volume and background auto fluorescence from the solvent. The incident light is reflected at the glass-water interface by total internal reflection creating local evanescent field illumination. Evanescent field decays exponentially from the interface, the distance of field inside the sample is given by term penetration depth, calculated as

$$d_p = \frac{\lambda}{4\pi\sqrt{(NA^2 - 1)}}$$

In our case λ wavelength of excitation laser (Nd:YAG DPSS laser, 50 mW) is 532 nm and numerical aperture (NA) of the objective lens (Olympus, 60x) is 1.2, which leads to penetration depth (d_p) ~ 64 nm.

Short penetration depth in TIRF excites only the dye molecules attached to the surface. TIRF excitation schematic is shown in figure 5.1 (boxed). The incident light is focused onto the prism (PBP, the refractive index of 1.46) surface at the critical angle (66° from the vertical plane) using mirror and lens. Index matching oil with the refractive index of 1.51 was used to match the interface of prism and quartz slide (same refractive index, i.e. 1.46) on which sample was surface immobilized. TIR occurs at the quartz - solvent (at the discontinuity of refractive index from 1.46 to 1.3) interface. The sample is excited in a wide field offering a possibility to measure hundreds of molecules (~50 μm x 100 μm) simultaneously.

Figure 5.1. Schematics of sm-FRET set up showing beam path. Modified figure adapted from ref (124).

5.2.2. Collection and detection part

The emitted fluorescence signal from the molecules is collected using objective. Through the objective, the part of excitation light (unabsorbed, scattered) passes with the emission signal, which will pass through a customized band filter (Chroma 532/632m-PH special, filter profile is shown in Figure 5.2). Filter profile is such that it will block the direct excitation and allows only the emitted fluorescence signal from dyes (Cy3 and Cy5). A dichroic mirror (Chroma, 630DCXR) was used to separate the donor (Cy3) and acceptor (Cy5) emission and focused onto the sensor of Andor Ixon 897 EMCCD (Electron Multiplying Charge-Coupled Device) camera for detection. To minimize the noise low noise optics is chosen and camera with high photon sensitivity with low dark current.

Figure 5.2. Excitation and emission spectra of Cy3 and Cy5. 532 nm LASER is used to excite Cy3 dye. Donor and Acceptor band of detection filter is also shown.

5.2.3. Data processing

Donor and acceptor fluorescence intensity (I_D and I_A) is used to extract the FRET efficiency from single-molecule fluorescence measurements. Donor and acceptor spots in both channels are mapped by IDL software using an image with fluorescent microbeads (~ 200 nm, F8806, Invitrogen). Intense fluorescence signal from beads can be seen in both channels; manually three bead molecules are mapped in both channels by generating a mapping table between the two channels for correction for the offset displacement. An automated algorithm (IDL software) is used for mapping.

Time variation in the intensity of each peak is determined by tracing it through the video recorded and time-dependent FRET efficiencies are calculated for each molecule. Output data from a single molecule experiment is shown in figure 5.3. From time traces of donor and acceptor intensities, FRET efficiency trace is generated to extract information. The amount of time a molecule spends in one state before transiting into other state is called dwell time and dwell time of various states can be calculated from the time trace analysis to understand kinetics of molecule.

Figure 5.3. Sm-FRET data analysis. A. EMCCD image showing dual channel for different emission wavelengths. Green region showing signal from Cy3 and red for Cy5 from the same molecule. B. Donor (green) and acceptor (red) time trace showing anti-correlation. C. Calculated FRET Efficiency (E_F) as a function of time.

FRET efficiency histogram can be generated from FRET efficiency exhibited by ~ 1000 of molecules showing the distribution of different conformational states. Data acquisition and analysis software package is obtained from Taekjip Ha's Lab (<https://cplc.illinois.edu/software/>).

5.2.4. Fluorescent dyes selection and labelling

Many fluorescent dye pairs have been developed and used for sm-FRET study. For sm-FRET experiments a dye pair should have following qualities; high extinction coefficient, high quantum yield and sufficient spectral overlap in donor emission and acceptor absorption (121,189). Cy3-Cy5 is one of most popular dye pair for single molecule FRET experiments due to its higher quantum yield and considerable overlap between donor emission and acceptor absorption spectra for FRET.

Dye molecules can permanently or temporarily stay in the non-fluorescent states, these effects are called photobleaching and photoblinking. Interaction between molecular oxygen and dye in triplet state is one of the major cause of photobleaching and hence it can be controlled by the adding the oxygen scavenging system which removes reactive oxygen from solution, allowing the dye to live longer before photobleaching. Enzymatic oxygen scavenger system consists of glucose oxidase and catalase with glucose (190). Photoblinking is primarily due to triplet state induced processes, it has been observed that reducing agents such as Trolox can quench the triplet state to restore the dye ground state (191).

5.2.5. DNA construct for sm-FRET

Dye labelled DNA was purchased from IDT DNA. Two separate DNA strands were purchased, one strand with Cy 5 on 5' terminal of the molecule of interest with a 29 nt stem and 3' biotin. Another strand is complementary to the 29 nt stem with Cy3 attached to the thymine residue. Two DNA strands were hybridized by mixing in the 1:1.2 ratio in the presence of 100 mM salt (NaCl for hairpin, LiCl for G-quadruplex) and heating the mixture to ~ 85 °C for 5 minutes in order to anneal out any self-pairing within the individual strands. Afterwards the DNA is allowed to cool down to room temperature using 2 °C per minute cooling rate. Annealed DNA is then aliquoted and transferred to a -30 °C freezer for long-term storage. Typical DNA sequence construct for sm-FRET are shown in schematic (Figure 5.4) is same as used in previous studies (99). For all the

measurements we used 10 mM Tris-HCL pH-8 buffer with specified salt concentration.

Figure 5.4. sm-FRET DNA construct. G-quadruplex DNA with 29 nt duplex stem. Cy5 is attached at the 5' end of G-quadruplex forming sequence and Cy3 is linked to thymine in the complementary stem sequence. 3' Biotin-neutravidin binding is used for surface tethering. Distance between the Cy3-Cy5 dye pair changes with the structural changes.

5.3. Results and discussions

Using the constructed sm-FRET system, we first performed an analysis on the Hairpin DNA sequence and compared our data with the previously published data from Nir E. group using sm-FRET (192) The consistency of obtained results (see below) validates our sm-FRET system.

5.3.1. Control experiments on hairpin

The dynamics of the hairpin with A31 loop and 6 base pairs (bp) stem was

studied and opening, and closing events were distinctly observed (Figure 5.5). The closed state characterized by high FRET ($E_F \sim 0.8$) and the open state by low FRET ($E_F \sim 0.2$). Multiple transition events between open and closed states are observed in more than $\sim 40\%$ of the hairpin time traces studied. Opening and closing rates are calculated by fitting the dwell time histogram with exponential distributions. For hairpins with a loop size of A31 and a stem of 6 bp, open and closed-state lifetimes are observed as 156 ± 1.5 ms and 33 ± 6.3 ms in 70 mM NaCl. Our results are in good agreement with the pioneering study by sm-FRET on hairpin with A40 loop and 7 bp stem in 100 mM NaCl with open and closed state lifetimes as 45 ± 2.4 and 133 ± 5.5 (193). However, observed closing rates are approximately 10 times and opening rates are 3 times smaller than the reported on same hairpin sequence (192). This discrepancy in rates could arise due to time resolution limit (100 ms) of our system.

Figure 5.5. Hairpin dynamics. A. Schematic of the open and close state transition of a DNA hairpin. B. The sequence of hairpin and duplex stem with the position of the Cy3 and Cy5 fluorophores. C. Folding-unfolding dynamics of hairpin is shown in FRET efficiency time trace, where a closed state is

characterized by high FRET and an open state by low FRET. Time traces are fitted by HAMMY for kinetic measurements (194).

5.3.2. Effect of salt concentration on hairpin dynamics

Hairpin dynamics under varying NaCl concentration has been studied (Figure 5.6). Higher salt concentration favors the compact folded state by stabilizing base pair interaction and reducing columbic repulsions. As observed by the decrease in closing rates and increase in opening rate with increasing salt. This behavior of salt concentration on hairpin dynamics is in accordance with the previously reported study (192).

Figure 5.6. Calculated closing and opening rates of the A31-stem6 hairpin with varying concentration of NaCl.

5.3.3. Effect of stem size on hairpin dynamics

We also monitored the hairpin dynamics as a function of the number of base pairs in hairpin stem sequence. Two hairpin sequences studied are 6 bp stem with A31 loop (*A31-stem 6*) and 8 bp stem with A31 loop (*A31-stem8*) (Figure 5.7. A). FRET Histograms obtained for both sequences in no salt condition is shown in Figure 5.7. D and E. Two peaks are observed low-FRET and high-FRET centered at 0.25 and 0.68 respectively. Peak centered at 0.12 is an

artifact represent the donor only molecules, which could be due to partial labeling of DNA or inactive acceptors and hereafter is ignored. *A31-stem8* displayed significant stabilizing effect on the folded state as observed by presence prominent folded Peak while folded peak cannot be observed for *A31-stem6* in the similar conditions. Time trajectory of the representative molecule of *A31-stem8* and *A31-stem6* are shown in Figure 5.7. B & C. Closing and opening rates were calculated from dwell time distributions are 0.12 ± 0.04 and 0.23 ± 0.06 for *A31-stem8*, whereas 0.15 ± 0.07 and 1.15 ± 0.6 for *A31-stem6*. This observation is consistent with the predicted, opening rate (or closing time) dependence on the stem length.

Figure 5.7. Effect of stem size on hairpin dynamics. A. Schematic of A31-stem8 and A31-stem6 hairpin in the closed state. D. and E. Histogram of FRET efficiency. B. and C. Time traces showing folding-unfolding dynamics. Peak with star in the histograms is an artifact representing donor only population in the sample.

5.3.4. Structural stability of parallel G-quadruplexes

After successful validation of our single molecule FRET system on hairpins, we

went a step further to test structural dynamics of G-quadruplex structures. We focused our study on *T95-1T* ($T(G_3T)_4$) sequence reported forming a stable, robust, three-layer parallel G-quadruplex structure with three single residues (147). Parallel G-quadruplex structure in *T95-1T* is established by 1D proton NMR spectrum showing 12 imino peaks from 10.0 to 12.0 ppm (Figure 5.8 A). In single molecule FRET DNA construct a 35 bp duplex stem has to be added for anchoring the molecule of interest on the surface. Here we tested 1D NMR spectra to verify that duplex stem formation does not interfere with the G-quadruplex structure. We used a 7 bp long foldback hairpin as a model sequence for the duplex stem, which shows 7 imino proton peaks from 12.0 to 14.0 ppm region (Figure 5.8 B). In a sequence containing both hairpin and G-quadruplex forming motifs (Figure 5.8 C), peaks corresponding to hairpin and G-quadruplex remains the same demonstrating both structures are compatible.

Figure 5.8. Imino-proton NMR spectra and proposed schematic of sm-FRET DNA construct. A. duplex hairpin B. G-quadruplex in *T95-1T* sequence. C. G-quadruplex and hairpin structure in the presence of 2 mM KCl. Stars and circles denote the duplex and G-quadruplex peaks, respectively.

Single molecule FRET efficiency histogram for *T95-1T* parallel G-quadruplex composed of three peak distributions in the presence of 2 mM KCl. The distribution at ~ 0.15 is an artifact representing the donor only population, which could be due to partial labeling of DNA or inactive acceptors and therefore ignored. A broad peak centered at ~ 0.25 represents the unfolded population in the sample, denoted as **U**. Another FRET peak at ~ 0.64 termed as **F**, assuming that it represents the folded parallel G-quadruplex conformation as shown by NMR studies (Figure 5.9. A and B)(147).

Figure 5.9. Sm-FRET analysis of *T95-1T* G-quadruplex. A. Schematic of *T95-1T* parallel G-quadruplex. B. Histogram of FRET efficiency. C. Time trace analysis of FRET efficiency. Peak with star in the histograms is an artifact representing donor only population in the sample.

Single molecule–FRET time trace analysis showed that most of the molecules remain constantly in the folded state within the observation time (~3 min) (Figure 5.9. C). In some traces, a single transition between folded (~0.64) and unfolded state (~ 0.25) states have been observed. Forming a stable G-quadruplex structure, dynamic behavior of parallel G-quadruplex in *T95-1T* sequence could not be studied.

In order to observe structural dynamics in parallel stranded G-quadruplex structures, the *T95-1T* sequence has been modified. Two different sequence modification approaches have been followed to increase the structural dynamic yet maintaining the parallel topology of G-quadruplexes (Figure 5.10).

Figure 5.10. Imino proton NMR spectra with schematic of parallel G-quadruplexes formed by modified sequences. A. Parallel propeller G-quadruplex with single G to I substitution, indicated by the blue rectangle in the middle tetrad (NMR spectrum adopted from ref (147)) B. G-quadruplex with long central loop (9 thymine) (long loop highlighted in red). C. Long loop G-quadruplex with hairpin structure. Stars and circles denote the duplex and G-quadruplex peaks, respectively.

In the first approach the guanine in the G-column has been substituted with inosine (*T95-1T-12*); structural analogue of guanine which forms one less hydrogen bond in the G-tetrad. In the second approach, we increased the central loop length to nine thymines (*T95-1T-9L*). Imino proton peak pattern of both modified sequences is similar to that of *T95-1T* (Figure 5.8.B and 5.10), establishing the parallel folding topology. In a sequence containing both hairpin and G-quadruplex with long central loop (Figure 5.10 C), peaks corresponding to hairpin and G-quadruplex remains the same demonstrating both structures does not interfere with each other formation in single molecule construct. Both the above sequence modification approaches have been shown to decrease the thermal stability of G-quadruplexes (147,187,188).

5.3.5. G-quadruplex conformation and dynamics of *T95-1T-12*

Some of the representative time traces for *T95-1T-12* sequence in 2 mM KCl are shown in Figure 5.11. Four different states have been observed in the time traces of *T95-1T-12*. Folded (**F**, $E_F \sim 0.64$) and unfolded (**U**, $E_F \sim 0.2$) states, peak positions for folded (**F**) and unfolded (**U**) state are in accordance with those observed for the *T95-1T* sequence (Figure 5.9), which establishes the same parallel topology adopted by *T95-1T-12* sequence. Interestingly, structural transitions between two new states (**F**₁, $E_F \sim 0.4$) and (**F**₂, $E_F > 0.8$) have been observed (Figure 5.11). These states were undetectable in *T95-1T* sequence (with 100 ms time resolution), suggesting that inosine substitution stabilizes the intermediate states.

Figure 5.11. FRET efficiency time traces for *T95-1T-12* G-quadruplex.

5.3.6. G-quadruplex conformation and dynamics of *T95-1T-9L*

Some of the representative time traces observed for *T95-1T-9L* sequence in 2 mM KCl are shown in Figure 5.12. Structural transitions were observed between four different states attributed to unfolded state (**U**, $E_F \sim 0.25$), intermediate states (**F**₁, $E_F \sim 0.48$), (**F**₂, $E_F \sim 0.77$) and folded state (**F**, $E_F \sim 0.65$) (Figure 5.12). FRET peak positions for folded (**F**) state is in accordance with that observed for the *T95-1T* and *T95-1T-12* sequence, which establishes the same parallel G-quadruplex topology adopted by *T95-1T-9L* sequence. Transitions involving intermediate states (**F**₁) and (**F**₂) were detected for both modified sequences *T95-1T-12* and *T95-1T-9L*. (**F**₂) state is significantly stabilized with increased lifetime by a long central loop as visible in time traces (Figure 5.12. B and F).

Observed intermediate states could possibly correspond to G-G hairpin and triplex structures as proposed for the other G-quadruplex structures (93,106). There is no sequential pathway is observed in these transitions suggesting it could arise from local rearrangements. Identification of corresponding intermediate structures could be done in future by introducing strategic mutations in the sequence to favor one of the intermediate structure. More single molecule traces would be acquired in future for quantitative analysis showing histograms of FRET state distribution.

Figure 5.12. FRET efficiency time traces for *T95-1T-9L* G-quadruplex.

5.4. Conclusions

We have successfully validated our custom built prism-type TIRF system by demonstrating two-state folding-unfolding dynamics of the hairpin system, in agreement with previously reported results.

Our studies on parallel G-quadruplex structure have revealed that the folding-unfolding dynamics of stable parallel G-quadruplex structure could be manipulated by either inosine substitution or by increasing loop length preserving the same parallel folding topology. Structural transitions with two intermediate states have been observed. The sequence with nine thymine loop stabilizes one of the intermediate states significantly. Such sequence modifications have been reported previously to decrease the thermal stability. However, their influence on folding dynamics was unexplored. Influence on the temporal stability of intermediate states could potentially affect the dynamic interactions of G-quadruplexes with other small molecule ligands and proteins *in vivo*.

Chapter 6

Summary and future work

G-rich sequences have the tendency to fold into non-canonical, four-stranded G-quadruplex nucleic acid structure via multiple stacking of Hoogsteen hydrogen bonded planar G-tetrads. Various G-quadruplex structures have been demonstrated with varying structural features such as strand orientations, glycosidic conformations, groove widths, connecting loops, the number of tetrads and molecularity. Structural diversity in G-quadruplex structures arise mainly due to the sequence composition and length. Herein, this thesis was set out to examine the formation of G-quadruplex structure in different sequences.

For the first time, we have systematically studied the G-quadruplex formation in poly-G sequences and presented two novel intramolecular G-quadruplex structures using NMR, UV and CD spectroscopy. The structure in a *G15* tract, which is the propeller-type G-quadruplex involving three G-tetrad layers and three single-residue propeller loops and the structure in a *G22* tract, which is four layer, chair-type antiparallel G-quadruplex with three edgewise loops displaying unique dynamic swapping of guanines between the G- tetrad core and loops.

Such poly-G tract length dependent study of G-quadruplex formation also provide insights about the natural conformational selection for the *G15* and *G22* G-quadruplex structure over myriad of accessible G-quadruplex conformations as they appear as single major conformation in solution. In principle, multiple combination of G-quadruplex conformations involving different structure features such as multiple or single DNA strand, number of stacked G-tetrads, varying loop length and geometry, flanking bases, abasic site and bulges could possibly coexist in solution. Our results suggest that for the structure of *G15*, guanine residue in the propeller can be substituted with other residues, while maintaining the same G-quadruplex fold. For the structure of *G22*, the substitution of two guanines in the edgewise loops can alter the loop geometry connecting narrow or wide grooves as evident from particular G-quadruplex structure of $(G_4C_2)_2$ sequence (152).

FANCI is G-quadruplex structure specific human helicase, recognized to resolve G-quadruplex structures during replication and hence maintains genomic integrity, failure to which can lead to many diseases including Fanconi Anemia. In the absence of Dog-1 (functional analogue of FANCI in *C.elegans*) deletions were observed particularly, at the G-tract of 15 and 22 in *C.elegans*. Understanding of the G-quadruplex structure of G₁₅ and G₂₂ tracts could be used in future for uncovering the mechanism of G-quadruplex recognition and unfolding by helicases (Dog-1 or FANCI) at both ensemble and single molecule level. Although a recent sm-FRET study had shed light on the FANCI interaction with telomeric and parallel G-quadruplexes but the unfolding mechanism remains ambiguous due to structural polymorphism of telomeric G-quadruplex (195).

Inhibition of DNA repair protein PARP-1 has been actively pursued as an attractive target for anticancer therapy. Identification of G-quadruplex forming region in the *PARP-1* promoter by sequencing of G-quadruplex structure in human genome led us to study the G-quadruplex formation in this sequence revealing three layered, intramolecular, (3+1) hybrid G-quadruplex structure. This structure exhibits an adenine bulge and a unique capping structure formed by the central edgewise loop and 5' flanking terminal residues. Such unique structural feature of *PARP-1* G-quadruplex could be targeted by small molecule ligands to regulate the PARP-1 expression. Thus, Understanding of TP3 G-quadruplex in *PARP-1* promoter could be helpful for its regulatory role as an alternative approach for PARP-1 silencing or downregulation of PARP-1 expression.

Aside from the structural ensemble study on G-quadruplexes, we have also studied structural dynamics of G-quadruplexes using single molecule FRET observations. We have successfully validated the in-house prism-type TIRF sm-FRET system by demonstrating two-state folding-unfolding dynamics of the hairpin system, in agreement with the previously reported results. After that, we went a step further to study structural dynamics of parallel, intramolecular G-quadruplexes. Our single molecule studies revealed that the folding-unfolding dynamics of stable parallel G-quadruplex structure could be manipulated by

sequence modifications. Structural transitions have been observed for parallel G-quadruplexes involving two intermediate states, which might have implications on the dynamic interactions of G-quadruplexes with other small molecule ligands and proteins *in vivo*. Further investigations will be needed for the identification of intermediate states by introducing strategic sequence mutations such that to stabilize individual intermediate state by using ensemble methods as well as sm-FRET.

Sm-FRET technique provided the opportunity to unravel the rarely visited states in the G-quadruplex structure, which otherwise obscured by the population averaging of exceedingly large number molecules used for ensemble studies.

Our current understanding on G-quadruplexes from both ensemble and single-molecule level could be exploited in combination to understand the structural and dynamic aspect of G-quadruplex interactions with proteins and small molecule ligands.

References

1. Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U. S. A.*, **48**, 2013.
2. Sen, D. and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364-366.
3. Davis, J.T. (2004) G-quartets 40 years later: From 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed. Engl.*, **43**, 668-698.
4. Simonsson, T. (2001) G-quadruplex DNA structures - variations on a theme. *Biol. Chem.*, **382**, 621-628.
5. Phan, A.T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.*, **277**, 1107-1117.
6. Karsisiotis, A.I., O'Kane, C. and Webba da Silva, M. (2013) DNA quadruplex folding formalism-a tutorial on quadruplex topologies. *Methods*, **64**, 28-35.
7. Monchaud, D. and Teulade-Fichou, M.P. (2008) A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.*, **6**, 627-636.
8. Roeland Boer, D., Canals, A. and Coll, M. (2009) DNA-binding drugs caught in action: the latest 3D pictures of drug-DNA complexes. *Dalton Trans.*, 399-414.
9. Williamson, J.R. (1994) G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 703-730.
10. Phillips, K., Dauter, Z., Murchie, A.I., Lilley, D.M. and Luisi, B. (1997) The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution. *J. Mol. Biol.*, **273**, 171-182.
11. Haider, S., Parkinson, G.N. and Neidle, S. (2002) Crystal structure of the potassium form of an *Oxytricha nova* G-quadruplex. *J. Mol. Biol.*, **320**, 189-200.
12. Williamson, J.R. (1993) Guanine quartets. *Curr. Opin. Struct. Biol.*, **3**, 357-362.
13. Hud, N.V., Smith, F.W., Anet, F.A.L. and Feigon, J. (1996) The selectivity for K⁺ versus Na⁺ in DNA quadruplexes is dominated by relative free energies of hydration: A thermodynamic analysis by H-1 NMR. *Biochemistry*, **35**, 15383-15390.

14. Gonzalez-Rodriguez, D., van Dongen, J.L.J., Lutz, M., Spek, A.L., Schenning, A. and Meijer, E.W. (2009) G-quadruplex self-assembly regulated by Coulombic interactions. *Nat. Chem.*, **1**, 151-155.
15. Pinnavaia, T.J., Marshall, C.L., Mettler, C.M., Fisk, C.I., Miles, H.T. and Becker, E.D. (1978) ALKALI-METAL ION SPECIFICITY IN SOLUTION ORDERING OF A NUCLEOTIDE, 5'-GUANOSINE MONOPHOSPHATE. *J. Am. Chem. Soc.*, **100**, 3625-3627.
16. Piazza, A., Adrian, M., Samazan, F., Heddi, B., Hamon, F., Serero, A., Lopes, J., Teulade-Fichou, M.-P., Anh Tuan, P. and Nicolas, A. (2015) Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J.*, **34**, 1718-1734.
17. Hazel, P., Huppert, J., Balasubramanian, S. and Neidle, S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.*, **126**, 16405-16415.
18. Guedin, A., Gros, J., Alberti, P. and Mergny, J.L. (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.*, **38**, 7858-7868.
19. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908-2916.
20. Todd, A.K., Johnston, M. and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901-2907.
21. Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L. and Wu, J.R. (1988) A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 6622-6626.
22. Hanakahi, L.A., Sun, H. and Maizels, N. (1999) High affinity interactions of nucleolin with G-G-paired rDNA. *J. Biol. Chem.*, **274**, 15908-15912.
23. Dunnick, W., Hertz, G.Z., Scappino, L. and Gritzmacher, C. (1993) DNA sequences at immunoglobulin switch region recombination sites. *Nucleic Acids Res.*, **21**, 365-372.
24. Maizels, N. (2005) Immunoglobulin gene diversification. *Annu. Rev. Genet.*, **39**, 23-46.
25. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 11593-11598.

26. Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A.J. (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.*, **51**, 269-288.
27. Kozlowski, P., Sobczak, K. and Krzyzosiak, W.J. (2010) Trinucleotide repeats: triggers for genomic disorders? *Genome Med.*, **2**, 29-29.
28. Di Prospero, N.A. and Fischbeck, K.H. (2005) Therapeutics development for triplet repeat expansion diseases. *Nat. Rev. Genet.*, **6**, 756-765.
29. Bacolla, A., Larson, J.E., Collins, J.R., Li, J., Milosavljevic, A., Stenson, P.D., Cooper, D.N. and Wells, R.D. (2008) Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties. *Genome Res.*, **18**, 1545-1553.
30. Maizels, N. (2012) G4 motifs in human genes. *Ann. N. Y. Acad. Sci.*, **1267**, 53-60.
31. Maizels, N. and Gray, L.T. (2013) The G4 Genome. *PLoS Genet.*, **9**, e1003468.
32. Rhodes, D. and Lipps, H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627-8637.
33. Xu, Y., Suzuki, Y., Ito, K. and Komiyama, M. (2010) Telomeric repeat-containing RNA structure in living cells. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 14579-14584.
34. Bochman, M.L., Paeschke, K. and Zakian, V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.*, **13**, 770-780.
35. Huppert, J.L., Bugaut, A., Kumari, S. and Balasubramanian, S. (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.*, **36**, 6260-6268.
36. Patel, D.J., Phan, A.T. and Kuryavyi, V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: Diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic acids research*, **35**, 7429-7455.
37. Beaudoin, J.D. and Perreault, J.P. (2010) 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.*, **38**, 7022-7036.
38. Mukundan, V.T. and Phan, A.T. (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.*, **135**, 5017-5028.
39. Heddi, B., Martin-Pintado, N., Serimbetov, Z., Kari, T.M. and Phan, A.T. (2016) G-quadruplexes with (4n - 1) guanines in the G-tetrad core:

formation of a G-triad.water complex and implication for small-molecule binding. *Nucleic Acids Res.*, **44**, 910-916.

40. Adrian, M., Ang, D.J., Lech, C.J., Heddi, B., Nicolas, A. and Phan, A.T. (2014) Structure and conformational dynamics of a stacked dimeric G-quadruplex formed by the human CEB1 minisatellite. *J.Am. Chem. Soc.*, **136**, 6297-6305.
41. Lim, K.W., Nguyen, T.Q. and Phan, A.T. (2014) Joining of multiple duplex stems at a single quadruplex loop. *J. Am. Chem. Soc.*, **136**, 17969-17973.
42. Jodoin, R., Bauer, L., Garant, J.M., Mahdi Laaref, A., Phaneuf, F. and Perreault, J.P. (2014) The folding of 5'-UTR human G-quadruplexes possessing a long central loop. *RNA (New York, N.Y.)*, **20**, 1129-1141.
43. Lim, K.W., Jenjaroenpun, P., Low, Z.J., Khong, Z.J., Ng, Y.S., Kuznetsov, V.A. and Phan, A.T. (2015) Duplex stem-loop-containing quadruplex motifs in the human genome: a combined genomic and structural study. *Nucleic Acids Res.*, **43**, 5630-5646.
44. Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F. and Maizels, N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev*, **18**, 1618-1629.
45. Muller, S., Kumari, S., Rodriguez, R. and Balasubramanian, S. (2010) Small-molecule-mediated G-quadruplex isolation from human cells. *Nat. Chem.*, **2**, 1095-1098.
46. Lam, E.Y., Beraldi, D., Tannahill, D. and Balasubramanian, S. (2013) G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.*, **4**, 1796.
47. Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182-186.
48. Zakian, V.A. (1989) Structure and function of telomeres. *Annu. Rev.Genet.*, **23**, 579-604.
49. Neidle, S. and Parkinson, G. (2002) Telomere maintenance as a target for anticancer drug discovery. *Nat. Rev. Drug Discov.*, **1**, 383-393.
50. Neidle, S. (2010) Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.*, **277**, 1118-1125.
51. de Lange, T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes. Dev.*, **19**, 2100-2110.
52. Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L. and Shay, J.W. (1994)

Specific association of human telomerase activity with immortal cells and cancer. *Science*, **266**, 2011-2015.

53. Wang, Y. and Patel, D.J. (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure*, **1**, 263-282.
54. Yaku, H., Fujimoto, T., Murashima, T., Miyoshi, D. and Sugimoto, N. (2012) Phthalocyanines: a new class of G-quadruplex-ligands with many potential applications. *Chem. Commun.*, **48**, 6203-6216.
55. Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406-413.
56. Gonzalez, V. and Hurley, L.H. (2010) The c-MYC NHE III(1): function and regulation. *Annu. Rev. Pharmacol. Toxicol.*, **50**, 111-129.
57. Rankin, S., Reszka, A.P., Huppert, J., Zloh, M., Parkinson, G.N., Todd, A.K., Ladame, S., Balasubramanian, S. and Neidle, S. (2005) Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.*, **127**, 10584-10589.
58. Fernando, H., Reszka, A.P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A.R., Neidle, S. and Balasubramanian, S. (2006) A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. *Biochemistry*, **45**, 7854-7860.
59. Dexheimer, T.S., Sun, D. and Hurley, L.H. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.*, **128**, 5404-5415.
60. Agrawal, P., Lin, C., Mathad, R.I., Carver, M. and Yang, D. (2014) The major G-quadruplex formed in the human BCL-2 proximal promoter adopts a parallel structure with a 13-nt loop in K⁺ solution. *J. Am. Chem. Soc.*, **136**, 1750-1753.
61. Sun, D., Guo, K., Rusche, J.J. and Hurley, L.H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.*, **33**, 6070-6080.
62. De Armond, R., Wood, S., Sun, D., Hurley, L.H. and Ebbinghaus, S.W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry*, **44**, 16341-16350.
63. Balasubramanian, S., Hurley, L.H. and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.*, **10**, 261-275.

64. Hershman, S.G., Chen, Q., Lee, J.Y., Kozak, M.L., Yue, P., Wang, L.-S. and Johnson, F.B. (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **36**, 144-156.
65. Verma, A., Halder, K., Halder, R., Yadav, V.K., Rawal, P., Thakur, R.K., Mohd, F., Sharma, A. and Chowdhury, S. (2008) Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *J. Med. Chem.*, **51**, 5641-5649.
66. Bugaut, A. and Balasubramanian, S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*, **40**, 4727-4741.
67. Salvati, E., Scarsella, M., Porru, M., Rizzo, A., Iachettini, S., Tentori, L., Graziani, G., D'Incalci, M., Stevens, M.F.G., Orlandi, A. *et al.* (2010) PARP1 is activated at telomeres upon G4 stabilization: possible target for telomere-based therapy. *Oncogene*, **29**, 6280-6293.
68. Cogoi, S., Paramasivam, M., Membrino, A., Yokoyama, K.K. and Xodo, L.E. (2010) 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.*, **285**, 22003-22016.
69. Brazda, V., Haronikova, L., Liao, J.C.C. and Fojta, M. (2014) DNA and RNA Quadruplex-Binding Proteins. *Int. J. Mol. Sci.*, **15**, 17493-17517.
70. Soldatenkov, V.A., Vetcher, A.A., Duka, T. and Ladame, S. (2008) First evidence of a functional interaction between DNA quadruplexes and poly(ADP-ribose) polymerase-1. *ACS Chem. Biol.*, **3**, 214-219.
71. Moore, H., Greenwell, P.W., Liu, C.-P., Arnheim, N. and Petes, T.D. (1999) Triplet repeats form secondary structures that escape DNA repair in yeast. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 1504-1509.
72. De, S. and Michor, F. (2011) DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nat. Struct. Mol. Biol.*, **18**, 950-955.
73. Kruisselbrink, E., Guryev, V., Brouwer, K., Pontier, D.B., Cuppen, E. and Tijsterman, M. (2008) Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCDJ-defective *C. elegans*. *Curr. Biol.*, **18**, 900-905.
74. Cheung, I., Schertzer, M., Rose, A. and Lansdorp, P.M. (2002) Disruption of dog-1 in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA. *Nat. Genet.*, **31**, 405-409.
75. Zhao, Y., O'Neil, N.J. and Rose, A.M. (2007) Poly-G/poly-C tracts in the genomes of *Caenorhabditis*. *BMC Genom.*, **8**, 403.

76. Collie, G.W. and Parkinson, G.N. (2011) The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem. Soc. Rev.*, **40**, 5867-5892.
77. Li, T., Shi, L., Wang, E. and Dong, S. (2009) Multifunctional G-quadruplex aptamers and their application to protein detection. *Chemistry*, **15**, 1036-1042.
78. Tucker, W.O., Shum, K.T. and Tanner, J.A. (2012) G-quadruplex DNA aptamers and their ligands: structure, function and application. *Curr. Pharm. Des.*, **18**, 2014-2026.
79. Macaya, R.F., Schultze, P., Smith, F.W., Roe, J.A. and Feigon, J. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl. Acad. Sci. U. S. A.*, **90**, 3745-3749.
80. Girvan, A.C., Teng, Y., Casson, L.K., Thomas, S.D., Juliger, S., Ball, M.W., Klein, J.B., Pierce, W.M., Jr., Barve, S.S. and Bates, P.J. (2006) AGRO100 inhibits activation of nuclear factor-kappaB (NF-kappaB) by forming a complex with NF-kappaB essential modulator (NEMO) and nucleolin. *Mol. Cancer Ther.*, **5**, 1790-1799.
81. Jing, N., Rando, R.F., Pommier, Y. and Hogan, M.E. (1997) Ion selective folding of loop domains in a potent anti-HIV oligonucleotide. *Biochemistry*, **36**, 12498-12505.
82. Phan, A.T., Kuryavyi, V., Ma, J.B., Faure, A., Andreola, M.L. and Patel, D.J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 634-639.
83. Livshits, G.I., Stern, A., Rotem, D., Borovok, N., Eidelstein, G., Migliore, A., Penzo, E., Wind, S.J., Di Felice, R., Skourtis, S.S. *et al.* (2014) Long-range charge transport in single G-quadruplex DNA molecules. *Nat. Nanotechnol.*, **9**, 1040-1046.
84. Troha, T., Drevensek-Olenik, I., Webba da Silva, M. and Spindler, L. (2016) Surface-Adsorbed Long G-Quadruplex Nanowires Formed by G:C Linkages. *Langmuir : the ACS journal of surfaces and colloids*.
85. Marsh, T.C. and Henderson, E. (1994) G-wires: self-assembly of a telomeric oligonucleotide, d(GGGGTTGGGG), into large superstructures. *Biochemistry*, **33**, 10718-10724.
86. Chiorcea-Paquim, A.M., Santos, P.V., Eritja, R. and Oliveira-Brett, A.M. (2013) Self-assembled G-quadruplex nanostructures: AFM and voltammetric characterization. *Phys. Chem. Chem. Phys.*, **15**, 9117-9124.
87. Chiorcea-Paquim, A.M., Pontinha, A.D., Eritja, R., Lucarelli, G., Sparapani, S., Neidle, S. and Oliveira-Brett, A.M. (2015) Atomic Force Microscopy and Voltammetric Investigation of Quadruplex Formation

between a Triazole-Acridine Conjugate and Guanine-Containing Repeat DNA Sequences. *Anal. Chem.*, **87**, 6141-6149.

88. Ruttkay-Nedecky, B., Kudr, J., Nejdl, L., Maskova, D., Kizek, R. and Adam, V. (2013) G-quadruplexes as sensing probes. *Molecules*, **18**, 14760-14779.
89. Ge, B., Huang, Y.C., Sen, D. and Yu, H.Z. (2010) A robust electronic switch made of immobilized duplex/quadruplex DNA. *Angew. Chem. Int. Ed. Engl.*, **49**, 9965-9967.
90. Simmel, F.C. and Dittmer, W.U. (2005) DNA nanodevices. *Small*, **1**, 284-299.
91. Cavalieri, F., Ng, S.L., Mazzuca, C., Jia, Z., Bulmus, V., Davis, T.P. and Caruso, F. (2011) Thin multilayer films and microcapsules containing DNA quadruplex motifs. *Small*, **7**, 101-111.
92. Gray, R.D., Trent, J.O. and Chaires, J.B. (2014) Folding and unfolding pathways of the human telomeric G-quadruplex. *J. Mol. Biol.*, **426**, 1629-1650.
93. Zhang, A.Y. and Balasubramanian, S. (2012) The kinetics and folding pathways of intramolecular G-quadruplex nucleic acids. *J. Am. Chem. Soc.*, **134**, 19297-19308.
94. Koirala, D., Dhakal, S., Ashbridge, B., Sannohe, Y., Rodriguez, R., Sugiyama, H., Balasubramanian, S. and Mao, H. (2011) A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands. *Nat. Chem.*, **3**, 782-787.
95. Shirude, P.S., Ying, L. and Balasubramanian, S. (2008) Single molecule conformational analysis of the biologically relevant DNA G-quadruplex in the promoter of the proto-oncogene c-MYC. *Chem. Commun.*, 2007-2009.
96. You, H., Wu, J., Shao, F. and Yan, J. (2015) Stability and kinetics of c-MYC promoter G-quadruplexes studied by single-molecule manipulation. *J. Am. Chem. Soc.*, **137**, 2424-2427.
97. Deniz, A.A., Mukhopadhyay, S. and Lemke, E.A. (2008) Single-molecule biophysics: at the interface of biology, physics and chemistry. *J. R. Soc. Interface*, **5**, 15-45.
98. Ying, L., Green, J.J., Li, H., Klenerman, D. and Balasubramanian, S. (2003) 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.*, **100**, 14629-14634.
99. Lee, J.Y., Okumus, B., Kim, D.S. and Ha, T. (2005) Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 18938-18943.

100. Long, X., Parks, J.W., Bagshaw, C.R. and Stone, M.D. (2013) Mechanical unfolding of human telomere G-quadruplex DNA probed by integrated fluorescence and magnetic tweezers spectroscopy. *Nucleic Acids Res.*, **41**, 2746-2755.
101. Okamoto, K., Sannohe, Y., Mashimo, T., Sugiyama, H. and Terazima, M. (2008) G-quadruplex structures of human telomere DNA examined by single molecule FRET and BrG-substitution. *Bioorg. Med. Chem.*, **16**, 6873-6879.
102. Noer, S.L., Preus, S., Gudnason, D., Aznauryan, M., Mergny, J.L. and Birkedal, V. (2016) Folding dynamics and conformational heterogeneity of human telomeric G-quadruplex structures in Na⁺ solutions by single molecule FRET microscopy. *Nucleic Acids Res.*, **44**, 464-471.
103. Lane, A.N., Chaires, J.B., Gray, R.D. and Trent, J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, **36**, 5482-5515.
104. Xue, Y., Liu, J.Q., Zheng, K.W., Kan, Z.Y., Hao, Y.H. and Tan, Z. (2011) Kinetic and thermodynamic control of G-quadruplex folding. *Angew. Chem. Int. Ed. Engl.*, **50**, 8046-8050.
105. Koirala, D., Mashimo, T., Sannohe, Y., Yu, Z., Mao, H. and Sugiyama, H. (2012) Intramolecular folding in three tandem guanine repeats of human telomeric DNA. *Chem. Commun.*, **48**, 2006-2008.
106. Mashimo, T., Yagi, H., Sannohe, Y., Rajendran, A. and Sugiyama, H. (2010) Folding pathways of human telomeric type-1 and type-2 G-quadruplex structures. *J. Am. Chem. Soc.*, **132**, 14910-14918.
107. Tippiana, R., Xiao, W. and Myong, S. (2014) G-quadruplex conformation and dynamics are determined by loop length and sequence. *Nucleic Acids Res.*, **42**, 8106-8114.
108. Plateau, P. and Gueron, M. (1982) Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences. *J. Am. Chem. Soc.*, **104**, 7310-7311.
109. Adrian, M., Heddi, B. and Phan, A.T. (2012) NMR spectroscopy of G-quadruplexes. *Methods*, **57**, 11-24.
110. Phan, A.T. and Patel, D.J. (2002) A site-specific low-enrichment (15)N,(13)C isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids. *J. Am. Chem. Soc.*, **124**, 1160-1161.
111. Schwieters, C.D., Kuszewski Jj Fau - Tjandra, N., Tjandra N Fau - Clore, G.M. and Clore, G.M. The Xplor-NIH NMR molecular structure determination package.
112. Delano, W.L. (2002) The PyMOL Molecular Graphics System.

113. Mergny, J.L. and Lacroix, L. (2009) UV Melting of G-Quadruplexes. *Curr. Protoc. Nucleic. Acid. Chem.*, **Chapter 17**, Unit 17.11.
114. Beychok, S. (1966) Circular dichroism of biological macromolecules. *Science*, **154**, 1288-1299.
115. Kypr, J., Kejnovská, I., Renčiuk, D. and Vorlíčková, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713-1725.
116. Vorlickova, M., Kejnovska, I., Sagi, J., Renciuik, D., Bednarova, K., Motlova, J. and Kypr, J. (2012) Circular dichroism and guanine quadruplexes. *Methods*, **57**, 64-75.
117. Gray, D.M., Wen, J.D., Gray, C.W., Repges, R., Repges, C., Raabe, G. and Fleischhauer, J. (2008) Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality*, **20**, 431-440.
118. Cornish, P.V. and Ha, T. (2007) A survey of single-molecule techniques in chemical biology. *ACS Chem. Biol.*, **2**, 53-61.
119. Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. and Ha, T. (2008) Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.*, **77**, 51-76.
120. Neuman, K.C. and Nagy, A. (2008) Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods.*, **5**, 491-505.
121. Roy, R., Hohng, S. and Ha, T. (2008) A Practical Guide to Single Molecule FRET. *Nature methods*, **5**, 507-516.
122. Forster, T. (1946) Energiewanderung und Fluoreszenz. *Naturwissenschaften*, **33**, 166-175.
123. Clegg, R.M. (2006) In Geddes, C. D. and Lakowicz, J. R. (eds.), *Reviews in Fluorescence 2006*. Springer US, Boston, MA, pp. 1-45.
124. Roy, R., Hohng, S. and Ha, T. (2008) A practical guide to single-molecule FRET. *Nat. Methods.*, **5**, 507-516.
125. Murphy, M.C., Rasnik, I., Cheng, W., Lohman, T.M. and Ha, T. (2004) Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophysical journal*, **86**, 2530-2537.
126. Selvin, P.R. and Ha, T. (2008) *Single-molecule techniques*. Cold Spring Harbor Laboratory Press.
127. Ha, T., Enderle, T., Ogletree, D.F., Chemla, D.S., Selvin, P.R. and Weiss, S. (1996) Probing the interaction between two single molecules:

- fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 6264-6268.
128. Moerner, W.E. and Kador, L. (1989) Optical detection and spectroscopy of single molecules in a solid. *Phys. Rev. Lett.*, **62**, 2535-2538.
 129. Hohng, S., Lee, S., Lee, J. and Jo, M.H. (2014) Maximizing information content of single-molecule FRET experiments: multi-color FRET and FRET combined with force or torque. *Chemical Society reviews*, **43**, 1007-1013.
 130. Bacolla, A., Zhu, X., Chen, H., Howells, K., Cooper, D.N. and Vasquez, K.M. (2015) Local DNA dynamics shape mutational patterns of mononucleotide repeats in human genomes. *Nucleic Acids Res.*, **43**, 5065-5080.
 131. Siddle, K.J., Goodship, J.A., Keavney, B. and Santibanez-Koref, M.F. (2011) Bases adjacent to mononucleotide repeats show an increased single nucleotide polymorphism frequency in the human genome. *Bioinformatics*, **27**, 895-898.
 132. van Passel, M.W. and de Graaff, L.H. (2008) Mononucleotide repeats are asymmetrically distributed in fungal genes. *BMC Genom.*, **9**, 596.
 133. Denver, D.R., Morris, K., Kewalramani, A., Harris, K.E., Chow, A., Estes, S., Lynch, M. and Thomas, W.K. (2004) Abundance, distribution, and mutation rates of homopolymeric nucleotide runs in the genome of *Caenorhabditis elegans*. *J. Mol. Evol.*, **58**, 584-595.
 134. Murray, V. (2015) The frequency of poly(G) tracts in the human genome and their use as a sensor of DNA damage. *Comput. Biol.Chem.*, **54**, 13-17.
 135. Scholz, C.F., Bruggemann, H., Lomholt, H.B., Tettelin, H. and Kilian, M. (2016) Genome stability of *Propionibacterium acnes*: a comprehensive study of indels and homopolymeric tracts. *Sci. Rep.*, **6**, 20662.
 136. Pernitzsch, S.R., Tirier, S.M., Beier, D. and Sharma, C.M. (2014) A variable homopolymeric G-repeat defines small RNA-mediated posttranscriptional regulation of a chemotaxis receptor in *Helicobacter pylori*. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E501-E510.
 137. Snyder, L.A., Loman, N.J., Linton, J.D., Langdon, R.R., Weinstock, G.M., Wren, B.W. and Pallen, M.J. (2010) Simple sequence repeats in *Helicobacter canadensis* and their role in phase variable expression and C-terminal sequence switching. *BMC genomics*, **11**, 67.
 138. Huppert, J. (2008) Hunting G-quadruplexes. *Biochimie*, **90**, 1140-1148.
 139. Sia, E.A., Kokoska, R.J., Dominska, M., Greenwell, P. and Petes, T.D. (1997) Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol. Cell. Biol.*, **17**, 2851-2858.

140. Coenye, T. and Vandamme, P. (2005) Characterization of mononucleotide repeats in sequenced prokaryotic genomes. *DNA Res.*, **12**, 221-233.
141. Gu, T., Tan, S., Gou, X., Araki, H. and Tian, D. (2010) Avoidance of long mononucleotide repeats in codon pair usage. *Genetics*, **186**, 1077-1084.
142. Salipante, S.J. and Horwitz, M.S. (2006) Phylogenetic fate mapping. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 5448-5453.
143. Naxerova, K., Brachtel, E., Salk, J.J., Seese, A.M., Power, K., Abbasi, B., Snuderl, M., Chiang, S., Kasif, S. and Jain, R.K. (2014) Hypermutable DNA chronicles the evolution of human colon cancer. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, E1889-1898.
144. Fattal, I., Shental, N., Ben-Dor, S., Molad, Y., Gabrielli, A., Pokroy-Shapira, E., Oren, S., Livneh, A., Langevitz, P., Zandman-Goddard, G. *et al.* (2015) Guanine polynucleotides are self-antigens for human natural autoantibodies and are significantly reduced in the human genome. *Immunology*, **146**, 401-410.
145. Joly, L., Rosu, F. and Gabelica, V. (2012) d(TG(n)T) DNA sequences do not necessarily form tetramolecular G-quadruplexes. *Chem. Commun.*, **48**, 8386-8388.
146. Panyutin, I.G., Kovalsky, O.I., Budowsky, E.I., Dickerson, R.E., Rikhirev, M.E. and Lipanov, A.A. (1990) G-DNA: a twice-folded DNA structure adopted by single-stranded oligo(dG) and its implications for telomeres. *Proc. Natl. Acad. Sci. U. S. A.*, **87**, 867-870.
147. Do, N.Q. and Phan, A.T. (2012) Monomer-dimer equilibrium for the 5'-5' stacking of propeller-type parallel-stranded G-quadruplexes: NMR structural study. *Chemistry*, **18**, 14752-14759.
148. Do, N.Q., Lim, K.W., Teo, M.H., Heddi, B. and Phan, A.T. (2011) Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res*, **39**, 9448-9457.
149. Rachwal, P.A., Brown, T. and Fox, K.R. (2007) Sequence effects of single base loops in intramolecular quadruplex DNA. *FEBS Lett.*, **581**, 1657-1660.
150. Guedin, A., De Cian, A., Gros, J., Lacroix, L. and Mergny, J.L. (2008) Sequence effects in single-base loops for quadruplexes. *Biochimie*, **90**, 686-696.
151. Huang, X., Yu, P., LeProust, E. and Gao, X. (1997) An efficient and economic site-specific deuteration strategy for NMR studies of homologous oligonucleotide repeat sequences. *Nucleic Acids Res.*, **25**, 4758-4763.

152. Brcic, J. and Plavec, J. (2015) Solution structure of a DNA quadruplex containing ALS and FTD related GGGGCC repeat stabilized by 8-bromodeoxyguanosine substitution. *Nucleic Acids Res.*, **43**, 8590-8600.
153. Zhou, B., Liu, C., Geng, Y. and Zhu, G. (2015) Topology of a G-quadruplex DNA formed by C9orf72 hexanucleotide repeats associated with ALS and FTD. *Sci. Rep.*, **5**, 16673.
154. Luu, K.N., Phan, A.T., Kuryavyi, V., Lacroix, L. and Patel, D.J. (2006) Structure of the human telomere in K⁺ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.*, **128**, 9963-9970.
155. D'Amours, D., Desnoyers, S., D'Silva, I. and Poirier, G.G. (1999) Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. *Biochem. J.*, **342** (Pt 2), 249-268.
156. Chambon, P., Weill, J.D. and Mandel, P. (1963) Nicotinamide mononucleotide activation of new DNA-dependent polyadenylic acid synthesizing nuclear enzyme. *Biochem. Biophys. Res. Commun.*, **11**, 39-43.
157. Luo, X. and Kraus, W.L. (2012) On PAR with PARP: cellular stress signaling through poly(ADP-ribose) and PARP-1. *Genes Dev*, **26**, 417-432.
158. Kraus, W.L. and Lis, J.T. (2003) PARP goes transcription. *Cell*, **113**, 677-683.
159. Salvati, E., Scarsella, M., Porru, M., Rizzo, A., Iachettini, S., Tentori, L., Graziani, G., D'Incalci, M., Stevens, M.F., Orlandi, A. *et al.* (2010) PARP1 is activated at telomeres upon G4 stabilization: possible target for telomere-based therapy. *Oncogene*, **29**, 6280-6293.
160. Schreiber, V., Dantzer, F., Ame, J.C. and de Murcia, G. (2006) Poly(ADP-ribose): novel functions for an old molecule. *Nat. Rev. Mol. Cell. Biol.*, **7**, 517-528.
161. Kraus, W.L. (2008) Transcriptional control by PARP-1: chromatin modulation, enhancer-binding, coregulation, and insulation. *Curr. Opin. Cell. Biol.*, **20**, 294-302.
162. Hassa, P.O., Haenni, S.S., Elser, M. and Hottiger, M.O. (2006) Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? *Microbiol. Mol. Biol. Rev.*, **70**, 789-829.
163. Satoh, M.S. and Lindahl, T. (1992) Role of poly(ADP-ribose) formation in DNA repair. *Nature*, **356**, 356-358.
164. De Vos, M., Schreiber, V. and Dantzer, F. (2012) The diverse roles and clinical relevance of PARPs in DNA damage repair: current state of the art. *Biochem. Pharmacol.*, **84**, 137-146.

165. Eustermann, S., Brockmann, C., Mehrotra, P.V., Yang, J.C., Loakes, D., West, S.C., Ahel, I. and Neuhaus, D. (2010) Solution structures of the two PBZ domains from human APLF and their interaction with poly(ADP-ribose). *Nat. Struct. Mol. Biol.*, **17**, 241-243.
166. Gradwohl, G., Menissier de Murcia, J.M., Molinete, M., Simonin, F., Koken, M., Hoeijmakers, J.H. and de Murcia, G. (1990) The second zinc-finger domain of poly(ADP-ribose) polymerase determines specificity for single-stranded breaks in DNA. *Proc. Natl. Acad. Sci. U. S. A.*, **87**, 2990-2994.
167. Haince, J.F., McDonald, D., Rodrigue, A., Dery, U., Masson, J.Y., Hendzel, M.J. and Poirier, G.G. (2008) PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites. *J. Biol. Chem.*, **283**, 1197-1208.
168. Rouleau, M., Aubin, R.A. and Poirier, G.G. (2004) Poly(ADP-ribosyl)ated chromatin domains: access granted. *J. Cell. Sci.*, **117**, 815-825.
169. Jagtap, P. and Szabo, C. (2005) Poly(ADP-ribose) polymerase and the therapeutic effects of its inhibitors. *Nat Rev Drug Discov*, **4**, 421-440.
170. Rouleau, M., Patel, A., Hendzel, M.J., Kaufmann, S.H. and Poirier, G.G. (2010) PARP inhibition: PARP1 and beyond. *Nat. Rev. Cancer.*, **10**, 293-301.
171. Miwa, M. and Masutani, M. (2007) PolyADP-ribosylation and cancer. *Cancer Sci.*, **98**, 1528-1535.
172. Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C. *et al.* (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, **434**, 917-921.
173. Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M.J. *et al.* (2009) Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.*, **361**, 123-134.
174. Pyndiah, S., Tanida, S., Ahmed, K.M., Cassimere, E.K., Choe, C. and Sakamuro, D. (2011) c-MYC suppresses BIN1 to release poly(ADP-ribose) polymerase 1: a mechanism by which cancer cells acquire cisplatin resistance. *Sci. Signal.*, **4**, ra19.
175. Fekete, A., Kenesi, E., Hunyadi-Gulyas, E., Durgo, H., Berko, B., Dunai, Z.A. and Bauer, P.I. (2012) The Guanine-Quadruplex Structure in the Human c-myc Gene's Promoter Is Converted into B-DNA Form by the Human Poly(ADP-Ribose)Polymerase-1. *PLoS ONE*, **7**, e42690.
176. Sengar, A., Heddi, B. and Phan, A.T. (2014) Formation of G-quadruplexes in poly-G sequences: structure of a propeller-type parallel-

stranded G-quadruplex formed by a G₁₅ stretch. *Biochemistry*, **53**, 7718-7723.

177. Heddi, B., Cheong, V.V., Martadinata, H. and Phan, A.T. (2015) Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex. *Proc. Natl. Acad. Sci. U. S. A.*, **112**, 9608-9613.
178. Chung, W.J., Heddi, B., Hamon, F., Teulade-Fichou, M.P. and Phan, A.T. (2014) Solution structure of a G-quadruplex bound to the bisquinolinium compound Phen-DC(3). *Angew. Chem. Int. Ed. Engl.*, **53**, 999-1002.
179. Heddi, B. and Phan, A.T. (2011) Structure of human telomeric DNA in crowded solution. *J. Am. Chem. Soc.*, **133**, 9824-9833.
180. Agrawal, P., Hatzakis, E., Guo, K., Carver, M. and Yang, D. (2013) Solution structure of the major G-quadruplex formed in the human VEGF promoter in K⁺: insights into loop interactions of the parallel G-quadruplexes. *Nucleic Acids Res.*, **41**, 10584-10592.
181. Phan, A.T., Modi, Y.S. and Patel, D.J. (2004) Propeller-Type Parallel-Stranded G-Quadruplexes in the Human c-myc Promoter. *J. Am. Chem. Soc.*, **126**, 8710-8716.
182. Lim, K.W., Lacroix, L., Yue, D.J., Lim, J.K., Lim, J.M. and Phan, A.T. (2010) Coexistence of two distinct G-quadruplex conformations in the hTERT promoter. *J. Am. Chem. Soc.*, **132**, 12331-12342.
183. Guo, K., Pourpak, A., Beetz-Rogers, K., Gokhale, V., Sun, D. and Hurley, L.H. (2007) Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. *J. Am. Chem. Soc.*, **129**, 10220-10228.
184. Amrane, S., Adrian, M., Heddi, B., Serero, A., Nicolas, A., Mergny, J.L. and Phan, A.T. (2012) Formation of pearl-necklace monomorphic G-quadruplexes in the human CEB25 minisatellite. *J. Am. Chem. Soc.*, **134**, 5807-5816.
185. Lee, J.Y. and Kim, D.S. (2009) Dramatic effect of single-base mutation on the conformational dynamics of human telomeric G-quadruplex. *Nucleic Acids Res.*, **37**, 3625-3634.
186. Ray, S., Bandaria, J.N., Qureshi, M.H., Yildiz, A. and Balci, H. (2014) G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 2990-2995.
187. Lim, K.W., Khong, Z.J. and Phan, A.T. (2014) Thermal stability of DNA quadruplex-duplex hybrids. *Biochemistry*, **53**, 247-257.
188. Risitano, A. and Fox, K.R. (2005) Inosine substitutions demonstrate that intramolecular DNA quadruplexes adopt different conformations in the

presence of sodium and potassium. *Bioorg. Med. Chem. Lett.*, **15**, 2047-2050.

189. Ha, T. and Tinnefeld, P. (2012) Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annu. Rev. Phys. Chem.*, **63**, 595-617.
190. Shi, X., Lim, J. and Ha, T. (2010) Acidification of the oxygen scavenging system in single-molecule fluorescence studies: in situ sensing with a ratiometric dual-emission probe. *Anal. Chem.*, **82**, 6132-6138.
191. Cordes, T., Vogelsang, J. and Tinnefeld, P. (2009) On the mechanism of Trolox as antiblinking and antibleaching reagent. *J. Am. Chem. Soc.*, **131**, 5018-5019.
192. Tsukanov, R., Tomov, T.E., Masoud, R., Drory, H., Plavner, N., Liber, M. and Nir, E. (2013) Detailed study of DNA hairpin dynamics using single-molecule fluorescence assisted by DNA origami. *J. Phys Chem. B.*, **117**, 11932-11942.
193. Grunwell, J.R., Glass, J.L., Lacoste, T.D., Deniz, A.A., Chemla, D.S. and Schultz, P.G. (2001) Monitoring the conformational fluctuations of DNA hairpins using single-pair fluorescence resonance energy transfer. *J. Am. Chem. Soc.*, **123**, 4295-4303.
194. McKinney, S.A., Joo, C. and Ha, T. (2006) Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling. *Biophys. J.*, **91**, 1941-1951.
195. Wu, C.G. and Spies, M. (2016) G-quadruplex recognition and remodeling by the FANCDJ helicase. *Nucleic Acids Res.*, **18**, 8742-8753.

Appendices

Experimental protocols

Appendix A1

Sample preparation

Sample Preparation. Unlabeled and site-specific labelled DNA oligonucleotides were chemically synthesized on an ABI 394 DNA synthesizer using products from Glen Research and Cambridge Isotope Laboratories and then purified following the protocol from Glen Research. The samples (concentration, 0.02–0.20 mM) were dialyzed successively against water, 2 mM KCl, and water again. DNA oligonucleotides were frozen, lyophilized, and dissolved in buffer containing 2 mM KCl and 10 mM Tris-HCl buffer (pH 7). The DNA concentration was expressed in strand molarity using the nearest neighbor approximation for the 260-nm molar extinction coefficient of the unfolded species.

UV Spectroscopy

The stability of G-quadruplexes was measured in UV melting experiments conducted on a JASCO V-650 spectrophotometer. Experiments were performed with 1-cm path length quartz cuvettes. The absorbance at 295 nm was recorded as a function of temperature ranging from 20 to 90 °C. Heating and cooling were done at a rate of 0.2 °C/min. The DNA concentration ranged from 4 to 6 µM in different buffer conditions. To calculate the DNA folded fraction, all UV melting curves are manually fitted for two-state folding where two baselines are drawn at low and high temperatures corresponding to completely folded and unfolded states. For each sequence, the melting temperature (T_m) for unfolding is presented. The difference in the T_m values from folding and unfolding was less than 1 °C.

Circular Dichroism

Circular dichroism (CD) spectra were recorded on a JASCO- 815 spectropolarimeter using a 1-cm path length quartz cuvette in a reaction volume of 500 µl at 25 °C. Scans from 220 to 320 nm were performed with a rate of 200 nm/min, 1-nm pitch, and 1-nm bandwidth. The DNA concentration ranged from 4 to 6 µM dissolved in different buffer conditions mentioned in the chapters. For each measurement, an average of three scans was taken, the spectral

contribution of the buffer was subtracted and data were zero-corrected at 320 nm.

Gel Electrophoresis

The molecular size of oligonucleotides was visualized by non-denaturing gel electrophoresis. Oligonucleotides were prepared in the mentioned buffer conditions. The samples were loaded on a 20% polyacrylamide gel, supplemented with 10 mM KCl, 40 % sucrose was added just before loading. The gel was viewed by UV shadowing.

Additionally, Gel purification was used to separate different bands of G22 sample. Native PAGE analysis of the G22 sequence was carried out showing to bands corresponding to higher order and single G-quadruplex conformation. Two major bands of G22 sequence were excised, crushed and dissolved in 10 mM Tris-HCl, pH 7 buffer with 2 mM KCl and subsequently dialyzed against water and lyophilized. The lyophilized samples were dissolved in a buffer containing 10 mM Tris-HCl, pH 7 buffer with 2 mM KCl and CD and NMR spectroscopy was performed on the purified samples corresponding to different bands.

NMR Spectroscopy

NMR experiments were performed on Bruker Advance spectrometers operating at 600 or 700 MHz at 25 °C unless specified. The DNA concentration for NMR experiments was typically 0.2–2.0 mM in respective buffer conditions. Spectra analyses were performed using SpinWorks (<http://home.cc.umanitoba.ca/~wolowiec/spinworks/>), FELIX (Felix NMR, Inc.) and SPARKY (<https://www.cgl.ucsf.edu/home/sparky/>) programs.

Appendix A2

Slide preparation protocol

Slides cleaning

Figure 1-A shows a picture of a template slide for 5 channels. A template slide was used to mark the holes on the new slides that needed to be drill. Holes were drilled on the quartz slide with a drilling speed between 6000-9000rpm using 0.75 mm drill bits. Drilled slides were rinsed with Deionized water and placed into the slide cleaning containers. Coverslips were placed into another container. Both slide and coverslip containing jars have been sonicated with Deionized water for 10 min, with Alconox for 20 mins, with Deionized water for 5 min, with acetone for 15 mins, followed by 3 times rinse with Deionized water, sonicated with KOH 1N solution for 40 – 60 mins, rinsed again with Deionized water 3 times and sonicate with Deionized water for 5 mins. After cleaning, slides were blow dried with nitrogen gas flow and burn for ~30 seconds with the propane torch (~5 seconds a glass coverslip) to remove any organic molecules on their surface. Cooled slides were placed in a methanol-filled jar.

Used quartz slides were also reused. Tape and coverslips was removed from used quartz slide chamber by boiling in water for 5 mins. While it is hot, the coverslip and glue can be easily removed. These relatively cleaned slides were rinsed and wiped with acetone and methanol. Rest of the procedure is same as new slides.

PEG coating protocol

1. Jars for amino silanization were rinsed with methanol and filled with methanol.
2. Amino silanization mixture was prepared as 100 ml methanol, 5 ml Acetic acid, 1 ml aminosilane, mixed and replaced methanol with this mixture, incubated for 20 mins.
3. After incubation amino silanization mixture in the jars was replaced with methanol, each slide and coverslip were rinsed with (i) deionized water; (ii) methanol; then (iii) dried by blowing with nitrogen gas flow.

PEG coating

1. PEG coating boxes were prepared by filling with de-ionized water to provide a humid environment. Excess water on the surface was wiped.
2. PEG buffer was prepared as 84 mg NaHCO_3 in 10 ml de-ionized water.
3. Biotin-PEG and m-PEG were weighed out in a 1.5 ml tube, mix thoroughly by tapping. Added the PEG buffer and mixed gently by flipping the tube. The amount of biotin-PEG and mPEG used per slide was 0.33 mg bPEG, 17 mg mPEG with 70 μl PEG buffer.
4. PEG mixture was centrifuged at 10,000 rpm for 1 minute. 75 μl of PEG mixture was spread on the slide for coating.

Disassembly and storage of PEG slides

1. Non-PEG coated side of the slide was labelled as number "2" to distinguish PEG-coated surface.
2. Disassemble and rinse each slides thoroughly with de-ionized water.
3. Dried with nitrogen gas flow and stored in separate 50 mL tubes at $-30\text{ }^{\circ}\text{C}$

Chamber fabrication protocol

1. Double sided tape was used to stick the coverslip and quartz slide in order to make a chamber. Double sided tape of desired thickness was used to make 5 channels. A coverslip was placed on the tape. 5-minutes-epoxy was used to seal the ends of the chamber.

Figure 1.A. Picture of double side tapes sticks on the quartz slide to make 5 channels.

Chemical preparations

T0: 10 mM Tris-HCl, pH 8 and 0.1 mM EDTA.

T50: 10 mM Tris-HCl, pH 8 and 0.1 mM EDTA and 50 mM NaCl/LiCl/KCl.

T100: 10 mM Tris-HCl, pH 8 and 0.1 mM EDTA and 100 mM NaCl/LiCl/KCl.

Trolox: (Sigma-Aldrich, # 38813)

30 mg Trolox is mixed with 10 mL of deionized water vortexed for ~ 15 h. Filter twice by 0.22 μ m, pH 8 adjusted by 1M LiOH solution. Stored at 4 °C and used within a month.

Gloxy: Glucose Oxidase (Sigma-Aldrich, # G2133) and catalase (Roche, # 10106810001)

10mg of the glucose oxidase powder is weighed in an empty 600uL Eppendorf. Added 100uL T50 solution and 20uL of catalase. Gently mixed by pipetting. Spin down for 2min at 10,000g stored at 4 °C and used within a month.

Neutravidin (N): (Thermo Fisher, # A2666)

Mixed 2ml of T50 to the stock bottle of 10 mg (the original bottle from the company). Gently mixing by tapping. Stored at 4 °C.

Dilution factor: N -> N/25 -> N/2500 in T50 buffer.

Imaging buffer:

823µl Trolox solution

80 µl glucose solution

67 µl T₀ solution

10 µl gloxy solution

Mixed with the desired amount of salt concentration. Gloxy solution was added just before use.

The following steps are usually followed in order to achieve optimal condition of single-molecule immobilization on the quartz surface for sm-FRET study:

1. Flushed 100 µl of the T50 buffer to each channel to clean the channel.
2. Flushed in 50 µl of (1/2500) of Neutravidin stock solution and incubated for 5 minutes.
3. Flushed in 100 µl T50 to clean the residual Neutravidin molecules, leaving only the ones attached to Biotin on the surface.
4. Flushed in 50 µl of 20 – 50 pM DNA sample (diluted in the T0 buffer) and incubated for 5 minutes

5. Flushed in 100 μ l T50 in order to clean the rest of the DNA molecules that didn't attach to the surface.
6. Checked sm-FRET to see whether there are enough spots (~ 300 – 400 spots) on the surface. If there are not enough spots, step 4 and 5 were repeated.
7. Flushed in 100 μ l T0-Tris to clear the residual ions on the channels.
8. Flushed 50 μ l imaging buffer consisting of the oxygen scavenger system and Trolox in appropriate buffer and salt conditions into the channel.
9. Acquired at least ~ 15 sets of movie data with 0.1 seconds integration time and ~100 – 200 seconds duration for histogram and time trace analysis.
10. Imaging buffer becomes acidic over the time, so it is recommended to clean the imaging buffer in the channel by T0 after data acquisition.