Investigating the solution structure of G-rich DNA and RNA using AFM

Bose, Krishnashish

2018

Bose, K. (2018). Investigating the solution structure of G-rich DNA and RNA using AFM.
Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/73074>

<https://doi.org/10.32657/10356/73074>

**INVESTIGATING THE SOLUTION STRUCTURE
OF G-RICH DNA AND RNA USING AFM**

KRISHNASHISH BOSE

School of Physical and Mathematical Sciences

2017

**INVESTIGATING THE SOLUTION STRUCTURE
OF G-RICH DNA AND RNA USING AFM**

**By
KRISHNASHISH BOSE**

A thesis submitted to

**Nanyang Technological University
Division of Physics and Applied Physics
School of Physical and Mathematical Sciences**

In partial fulfillment for the degree of Doctor of Philosophy (PhD)

November 2017

Acknowledgements

I shall be eternally grateful to Prof. Anh Tuan Phan for offering me a PhD position in his lab. I consider myself fortunate to have gotten so much of his attention, guidance, support and dedication; which shall help me become a great scientist. I am especially grateful to Prof. John van Noort (Leiden University) and Dr. Iwan A.T. Schaap (formerly Göttingen University) for their guidance and help with Atomic Force Microscopy. I am incredibly thankful to our collaborators Prof. Neil A. Schneider (Columbia University) and Prof. Jack D. Griffith (UNC Chapel Hill) for their choice to work with us and for sending their lab's precious DNA samples from USA.

I am grateful to Dr. Christopher J. Lech, Dr. Heddi Brahim, Dr. Cheong Vee Vee and Dr. Sengar Anjali for their constant support both as a friend and as a fellow senior researcher. I am thankful to all my lab members for being so nice and friendly with me, and for being eager to offer any kind of help I needed.

All that I have achieved till now and will achieve in future is the grace of the creators of this universe. I shall remain eternally grateful to them for offering me a healthy human body with a magnificent neuro-genetic architecture. Until the end of my existence, I shall make the best possible use of the greatest creation in the multiverse: "*consciousness*".

Last but not the least, I have no words to express my gratitude to my wife Karnika for her unconditional love and dedication for me. Without my family's love and emotional support, it would have been impossible for me to be emotionally stable. The biggest reason for my success is "LOVE", which to me is the only thing that gives life a true meaning in this seemingly meaningless universe built with unfathomable perfection.

Abstract

G-rich DNA/RNA sequences are highly dynamic and diverse in nature as they can adopt a myriad of secondary structures in physiologically relevant conditions. Several biologically relevant G-rich repeat motif DNA sequences are known to form higher-order G-quadruplexes, but little is known about their atomic level structure. Atomic Force Microscopy (AFM) has been widely used to investigate higher-order G-quadruplexes for the past 3 decades. In the current work, AFM has been pushed to the limits to develop a new methodology for probing biomolecular structure with sub-nanometer resolution and thus bridge the gap between single-molecule and high-resolution ensemble techniques.

The protocol for visualization of G-quadruplex DNA and RNA in aqueous solution was developed to resolve their intricate structural features. Higher-order structures formed by three different biologically relevant sequences containing G-rich repeat motif were investigated, viz., GGGGTT, GGGGCC and TTAGGG. It was found that short and long G-wires are formed by d[G₄T₂G₄], d[G₄C₂G₄] and r[G₄C₂G₄]. All these G-wires revealed similar morphology (uniform height, variable length) and periodic surface pattern. We built several molecular models and found one that can explain the experimentally observed features of G-wires. Our work hint towards a unified architecture adopted by G-wires formed by different DNA and RNA oligonucleotides. These findings are useful for the design of G-wire based nanodevices and may find relevance in biology.

Transcription of PCR-amplified DNA containing several repeats of TTAGGG was used to generate telomeric repeat containing RNA (TERRA), which was investigated using high-resolution AFM in aqueous solution. The higher-order structures formed by TERRA were grouped into two broad classes based on their heights and intricate structural periodicities. One class of observed structures were like the pearl necklace as proposed earlier for telomeric DNA and RNA.

In summary, we conclude that G-rich DNA and RNA sequences are capable of self-assembling into higher-order intermolecular G-quadruplexes which are consistent with a model of stacked G-tetrads having molecular periodicities which are multiples of the distance between individual G-tetrad layers. However, individual strands of G-rich DNA and RNA could exist as a pearl necklace, where the beads are likely G-quadruplexes, which are connected by a single-stranded linker. The higher-order structures reported here could be broadly classified into just two groups: four-stranded parallel G-wire and single-stranded pearl necklace.

Commonly used abbreviations

1D	One-dimensional
A	Adenine
AFM	Atomic Force Microscopy
ALS	Amyotrophic Lateral Sclerosis
bp	base pair
BPL	Biotinylated Poly-L-Lysine
CD	Circular Dichroism
C	Cytosine
DNA	Deoxyribonucleic acid
dsDNA	double stranded DNA
Duplex	Double helix
FTD	Frontotemporal dementia
G	Guanine
GA	Glutaraldehyde
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
kbp	kilo base pair
KPi	$K_2HPO_4 + KH_2PO_4$
MD	Molecular Dynamics
NMR	Nuclear Magnetic Resonance
nt	nucleotide
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
ppm	parts per million

RNA	Ribonucleic acid
ssDNA	single stranded DNA
T	Thymine
TERRA	Telomeric Repeat containing RNA
UV	Ultraviolet

Table of Contents

CHAPTER 1: INTRODUCTION	1
1.1. NUCLEIC ACIDS	1
1.2. DNA/RNA G-QUADRUPLX	4
1.3. TELOMERE	6
1.4. DNA G-WIRES	8
1.5. GGGGCC REPEATS	10
1.6. TERRA	12
CHAPTER 2: EXPERIMENTAL TECHNIQUES	15
2.1. G-WIRE SYNTHESIS	15
2.2. SYNTHESIS OF LONG DNA AND RNA	16
2.3. GEL ELECTROPHORESIS	20
2.4. ATOMIC FORCE MICROSCOPY	22
2.5. CIRCULAR DICHROISM	26
2.6. NMR SPECTROSCOPY	27
CHAPTER 3: DEVELOPMENT OF SOLUTION AFM IMAGING	31
3.1. INTRODUCTION	31
3.2. TIP-SAMPLE INTERACTION	32
3.3. OVERCOMING THE “FOREST OF PEAKS”	33
3.4. USING THE CYPHER ES	34
3.5. ULTRA-SHORT CANTILEVERS	38
3.6. DNA IMMOBILIZATION	40
3.7. IMMOBILIZATION OF LONG G-WIRES	47
3.8. HIGH-RESOLUTION IMAGING	49
3.9. CONCLUSIONS AND FUTURE WORK	53
CHAPTER 4: DNA NANOMECHANICAL PROPERTIES	57
4.1. INTRODUCTION	57
4.2. VARIATION OF ATTRACTIVE FORCE	61
4.3. VARIABLE FORCE IMAGING OF G-WIRES IN AIR	62
4.4. DEFORMATION OF G-WIRES IN SOLUTION	67
4.5. HIGH-RESOLUTION NANOMECHANICAL MAPPING IN SOLUTION	69
4.6. DISCUSSIONS AND FUTURE WORK	73

CHAPTER 5: SOLUTION AFM STRUCTURE OF DNA G-WIRES	75
5.1. INTRODUCTION	75
5.2. ENSEMBLE CHARACTERIZATION OF G-WIRES	75
5.3. ROLE OF ANNEALING	78
5.4. HEIGHT OF G-WIRES IN SOLUTION	80
5.5. HIGH-RESOLUTION IMAGING OF G-WIRES	86
5.6. PERIODICAL FEATURES ON G-WIRES	88
5.7. 1 NM SUB-FEATURES	91
5.8. HANDEDNESS OF PERIODIC FEATURES	92
5.9. STATISTICAL ANALYSIS OF PERIODICITIES	95
5.10. BUILDING STRUCTURAL MODELS	101
5.11. SIMULATIONS AND COMPARISON WITH EXPERIMENTAL DATA	110
5.12. DISCUSSIONS AND FUTURE WORK	115
CHAPTER 6: HIGHER-ORDER STRUCTURES OF GGGGCC	117
6.1. INTRODUCTION	117
6.2. NMR SPECTROSCOPY	117
6.3. CD SPECTROSCOPY	120
6.4. RNASE TREATMENT OF RNA G-WIRES	123
6.5. FORMATION OF G-WIRES	125
6.6. PURIFICATION OF G-WIRES	129
6.7. DUPLEX DNA VS. G ₄ CC WIRES	132
6.8. PERIODICAL FEATURES ON G ₄ CC-WIRES	134
6.9. MODELLING AND SIMULATION	136
6.10. DISCUSSIONS AND FUTURE WORK	140
CHAPTER 7: SOLUTION AFM IMAGING OF TERRA	142
7.1. INTRODUCTION	142
7.2. THE PRST5 PLASMID	142
7.3. TRANSCRIPTION WITH T7 RNA POLYMERASE	148
7.4. AFM IMAGING OF TERRA IN SOLUTION	152
7.5. INTRICATE STRUCTURAL DETAILS OF TERRA	156
7.6. DISCUSSIONS AND FUTURE WORK	158
CHAPTER 8: SCOPE OF AFM IN G-QUADRUPLEX RESEARCH	161
REFERENCES	165
APPENDIX	179

List of Figures

Figure 1.1 Nucleic acids	2
Figure 1.2 Base pairing via hydrogen bonding	3
Figure 1.3 G-tetrad primary structure.....	4
Figure 1.4 G-quadruplex with different number of layers	5
Figure 1.5 The structure of telomere	7
Figure 1.6 DNA duplex and G-wires imaged in air and propanol	9
Figure 1.7 BG4 stained foci and area in fibroblasts and astrocytes	11
Figure 1.8 EM visualization of C-rich and G-rich TERRA	13
Figure 2.1 Polymerase Chain Reaction	18
Figure 2.2 Transcription by T7 RNA polymerase.....	19
Figure 2.3 Gel electrophoresis setup	20
Figure 2.4 AFM imaging modes	23
Figure 2.5 Principle of Tapping mode.....	24
Figure 2.6 Principle of ScanAsyst mode	25
Figure 2.7 CD spectra of different types of G-quadruplexes	27
Figure 2.8 Proton chemical shifts in nucleic acids	29
Figure 3.1 Force-time curve in air vs. water	32
Figure 3.2 Cantilever tuning.....	33
Figure 3.3 AFM schematic	35
Figure 3.4 AFM setup.....	36
Figure 3.5 Ultra-short cantilevers.....	39
Figure 3.6 Thermal noise spectra in air vs. liquid for FastScan-D cantilever	40
Figure 3.7 Floating G-wires on non-prewashed Mica.....	42
Figure 3.8 Disappearance of weakly immobilized G-wires	43
Figure 3.9 Reproducible imaging of G-wires.....	44
Figure 3.10 Reproducible imaging of duplex DNA with just 0.2 mM NiCl ₂	45
Figure 3.11 Duplex DNA and G-wires imaged on BPL mica.....	46

Figure 3.12 Disappearance of weakly immobilized DNA on BPL mica	46
Figure 3.13 DNA/RNA do not attach with just glutaraldehyde	47
Figure 3.14 Duplex DNA and G-wires immobilized with Spermidine.....	48
Figure 3.15 Clean imaging of G-wires with spermidine	49
Figure 3.16 Programmed setup for unattended imaging	52
Figure 3.17 Resolving the minor and major grooves of double helix	53
Figure 4.1 Cantilever amplitude over Mica vs. Purple membrane.....	59
Figure 4.2 Attractive forces in air vs. solution	62
Figure 4.3 Height variation with variable imaging force	64
Figure 4.4 Selective plastic deformation of G-wires in air	65
Figure 4.5 Collapsing G-wires in air	66
Figure 4.6 Variable force imaging in solution.....	68
Figure 4.7 Mechanical mapping of minor and major grooves of duplex DNA	70
Figure 4.8 Height vs. Loss tangent for G-wires	71
Figure 4.9 4 nm periodicity in topography and mechanical properties	72
Figure 5.1 Proton NMR spectra of d[GGGGTTGGGG].....	76
Figure 5.2 CD spectra of d[G ₄ T ₂ G ₄].....	77
Figure 5.3 Decreased ellipticity at high temperature.....	77
Figure 5.4 d[G ₄ T ₂ G ₄] imaged before and after annealing	78
Figure 5.5 Height of d[G ₄ T ₂ G ₄] before and after annealing.....	79
Figure 5.6 Length histogram of d[G ₄ T ₂ G ₄] before and after annealing	79
Figure 5.7 Height of duplex DNA vs. G-wires	81
Figure 5.8 Molecules and their splines from which height was analyzed.....	82
Figure 5.9 Height histogram of images with duplex and G-quadruplex DNA	83
Figure 5.10 Height histogram from splines of duplex and G-quadruplex DNA ...	84
Figure 5.11 Reproducible high-resolution imaging of G-wires in solution	86
Figure 5.12 Grooves of duplex DNA resolved throughout its entire length	88
Figure 5.13 Type I and Type II G-wires.....	89
Figure 5.14 3 types of periodical features in the same G-wire.....	90
Figure 5.15 Sub-periodic features in Type I G-wires.....	91

Figure 5.16	Height profiles obtained from selected sections of Type I G-wires ..	92
Figure 5.17	Handedness of type I G-wires	93
Figure 5.18	Handedness of type II G-wires.....	93
Figure 5.19	Correcting tip dilation for G-wire	94
Figure 5.20	Correcting tip dilation for duplex DNA	95
Figure 5.21	Flowchart for periodicity analysis.....	96
Figure 5.22	Collage of a subset of G-wires cropped from AFM images	98
Figure 5.23	Splines of 100 G-wires with least standard deviation in periodicity ..	98
Figure 5.24	Obtaining periodicity histogram for type I G-wires.....	99
Figure 5.25	Obtaining periodicity histogram for type II G-wires	100
Figure 5.26	Guanine arrangement for (2,2) adjacent slip model.....	102
Figure 5.27	Guanine arrangement for (2,2) diagonal slip model	103
Figure 5.28	Guanine arrangement for (3,1) single slip model	104
Figure 5.29	Guanine positions for less than half-strand slipping arrangement...	105
Figure 5.30	Anti-glycosidic dihedral constraints for Guanines.....	106
Figure 5.31	Planarity constraints for Guanines	106
Figure 5.32	Distance constraints for Guanines in a G-tetrad	107
Figure 5.33	Distance constraints for carbon atoms in consecutive G-tetrads	107
Figure 5.34	Half-slipping structures computed using XPLOR	108
Figure 5.35	Partial slipping structures computed with XPLOR.....	109
Figure 5.36	Guanine arrangement for an anti-parallel G-wire	110
Figure 5.37	The starting structures for MD simulation.....	112
Figure 5.38	Simulated AFM images	113
Figure 5.39	Agreement between experimental data with predicted structure	114
Figure 6.1	1-D proton NMR spectra for r[G ₄ CCG ₄]	118
Figure 6.2	Integral area ratio before and after annealing	119
Figure 6.3	Imino proton spectra for d[G ₄ CCG ₄]	120
Figure 6.4	UV-CD heating spectra for r[G ₄ CCG ₄]	121
Figure 6.5	UV-CD heating spectra for d[G ₄ CCG ₄].....	122
Figure 6.6	White residue formation by d[G ₄ CCG ₄] after CD experiment	122

Figure 6.7 Sharp decline in DNA concentration at high temperature	123
Figure 6.8 Incubation of RNA G-wires with RNase	124
Figure 6.9 Native PAGE of G ₄ CC DNA and RNA G-wires	125
Figure 6.10 d[G ₄ CCG ₄] immobilized with Spermidine in solution	127
Figure 6.11 r[G ₄ CCG ₄] immobilized with Spermidine in solution	128
Figure 6.12 d[G ₄ TTG ₄] immobilized with Spermidine in solution	129
Figure 6.13 Native PAGE purification of DNA G-wires	130
Figure 6.14 AFM images of PAGE purified DNA G-wires	131
Figure 6.15 G-wires of d[G ₄ T ₂ G ₄] before and after purification	131
Figure 6.16 Solution images of G ₄ CC structures with duplex DNA	132
Figure 6.17 Height histograms for d[G ₄ CCG ₄] with duplex B-DNA	133
Figure 6.18 Height histograms for r[G ₄ CCG ₄] with duplex B-DNA	134
Figure 6.19 High-resolution images of DNA G ₄ CC-wires in solution	135
Figure 6.20 High-resolution images of RNA G ₄ CC-wires in solution	135
Figure 6.21 Sub-nm resolution showing 1 nm periodicity	135
Figure 6.22 Computed molecular structure of d[G ₄ CCG ₄]	136
Figure 6.23 Few snapshots showing the process of AFM simulation	138
Figure 6.24 Simulated AFM images for G-wire formed by d[G ₄ CCG ₄]	139
Figure 7.1 Plasmid map of pRST5 containing d[TTAGGG] ₉₆	143
Figure 7.2 Cell culture for pRST5 plasmid extraction	144
Figure 7.3 Linear digestion of pRST5 plasmid	145
Figure 7.4 Length variation in linearized pRST5	146
Figure 7.5 Multiple sequence alignment	147
Figure 7.6 Reproducible sequencing chromatogram	148
Figure 7.7 Gel visualization of telomeric-repeat containing DNA/RNA	150
Figure 7.8 Non-telomeric PCR amplification of pRST5 plasmid	152
Figure 7.9 Solution AFM images of TERRA	153
Figure 7.10 Height histograms for TERRA	154
Figure 7.11 Mechanical property contrast between TERRA I and TERRA II ...	155
Figure 7.12 Periodical features of TERRA	157

This page has been intentionally left blank

Chapter 1: Introduction

1.1. Nucleic acids

Deoxyribonucleic acid (DNA) is an organic macromolecule that carries the molecular identity (genetic code) of every known living organism. DNA carries the genetic code for thousands of years while allowing sufficient changes required for evolution. In nature, DNA is used as a template to generate Ribonucleic acid (RNA), by the process called transcription. Nucleic acids are composed of nitrogenous bases namely Cytosine (C), Thymine (T), Uracil (U), Adenine (A) and Guanine (G). The latter two (A, G) are broadly classified as Purines, while the former three (C, T, U) are classified as Pyrimidines. The nucleobases are attached to a phosphate backbone with oxyribose or deoxyribose for RNA and DNA respectively (Figure 1.1). RNA differs from DNA by presence of a hydroxyl group (-OH) at the 2'-Carbon atom of the sugar moiety and absence of a C5 methyl group in Uracil. The absence of a C5 methyl group almost certainly has a destabilizing effect. However, the presence of 2'-OH group could be either stabilizing or destabilizing based on the complex formed by the RNA. For instance, it has been observed that RNA duplexes are more stable than DNA duplexes, for the same set of purines and pyrimidines¹. Hydroxyl groups in the purine strand gives additional stability (0.54 ± 0.10 kcal/mol per base pair) in pairing with a pyrimidine RNA complement.² On the other hand, the same 2'-OH group makes single-stranded RNA highly susceptible to degradation due to alkaline hydrolysis.

The nucleobases have the unique ability to pair with itself or other nucleobases via Watson-Crick and Hoogsteen hydrogen bonding (Figure 1.2).³ Watson-crick hydrogen bonding in A-T and G-C are one of the most stable in aqueous solution, which are thus referred to as complimentary bases. Their pairing allows the formation of stable double-stranded DNA (dsDNA), RNA (dsRNA) and DNA-RNA hybrids. However, Hoogsteen hydrogen bonding can be more stable than Watson-crick interactions in crowded molecular conditions that attempt to mimic the cellular environment.⁴ The higher order structure formation of nucleic acids is mediated by the nucleobases via specific classifications of H-bond complementarity, base-stacking interactions as well as glycosidic bond angles and ribose sugar geometry.

Figure 1.1 Nucleic acids

The molecular structure of five principal nitrogenous bases which are the building blocks of nucleic acids: Cytosine (C), Thymine (T), Uracil (U), Guanine (G) and Adenine (A). These nucleobases are attached to oxyribose (for RNA) and deoxyribose (for DNA), which is further attached to a phosphate group that acts as the backbone for formation of a long polymer chain. RNA has an extra oxygen atom in the sugar which has been highlighted in red color.

Figure 1.2 *Base pairing via hydrogen bonding*

(A) Watson-crick and (B) Hoogsteen hydrogen bonding between nucleobases.

In living cells, dsDNA exists as double stranded polymer called B-DNA, which is stabilized by Watson-Crick hydrogen bonding between complementary nucleobases. It is a right-handed double helical structure with average base-pair separation of 3.34 ± 0.1 Å and has a diameter of 2.4 ± 0.2 nm in aqueous solution.⁵ It exhibits long-range periodic features in the form of minor and major grooves of width 1.4 nm and 2.0 nm respectively.⁶ For certain cellular processes like transcription and replication, parts of dsDNA are intermittently unzipped into two single stranded DNA (ssDNA) by certain enzymes like helicases and polymerases.⁷ The nucleobases of a single strand of RNA/DNA can form stable inter/intra-molecular hydrogen bonded pairs to keep its hydrophobic bases away from water. The structure of single stranded nucleic acids has thus remained elusive owing to the myriad of possibilities of hydrogen bonding. For instance, hydrogen bonding between nucleobases can lead to formation of triads⁸, tetrads⁹ and hexads¹⁰.

This PhD thesis addresses the structure(s) formed by repeats of d[GGGGTT], r[GGGGCC] and r[UUAGGG]. These 3 sequences are of extreme biological significance, and the structures formed by them are barely known/understood.

1.2. DNA/RNA G-quadruplex

G-tetrads are composed of four planar Guanine bases bonded via 8 Hoogsteen hydrogen bonds (Figure 1.3). G-tetrads can stack upon each other by π - π interactions to form multi-layered structures called G-quadruplexes (Figure 1.4), which is further stabilized by a cationic core of appropriate size and ionic strength, ex: K^+ ,¹¹ Na^+ ,¹² NH_4^+ ,¹³ Sr^{++} .¹⁴ Since potassium concentration inside cells is much higher compared to other ions¹⁵, G-quadruplexes stabilized by potassium (K^+) core are of biological relevance and has thus been the main focus of our study.

The discovery of G-quadruplexes could be traced a century back to 1910 when it was observed that guanine mono-phosphate can form gels, which could only be explained in 1962 with the help of X-ray crystallography.¹⁶ Over 375,000 potential G-quadruplex (G-4) forming sequences^{17, 18} have been computationally estimated to exist in the human genome. These are found in introns/promoters of several genes¹⁹ and at chromosome-ends²⁰.

Figure 1.3 G-tetrad primary structure

Four guanines connected by 8 Hoogsteen hydrogen bonds form a G-tetrad.

Figure 1.4 *G-quadruplex with different number of layers*

Solution NMR structure of (A) 2-layer and (B) 3-layer dimeric²¹ DNA G-quadruplexes. (C) X-ray crystal structure of a 4-layer dimeric²² G-quadruplex DNA. The residues are colored as: aquamarine (Guanine), purple (Inosine), green (Adenine) and orange (Thymine). The tetrads are highlighted with thicker bonds. The respective PDB IDs are shown in parentheses.

High-resolution ensemble techniques like NMR spectroscopy, X-ray crystallography and small-angle X-ray scattering (SAXS)²³ have been used to determine the molecular structure of DNA and RNA G-quadruplexes. However, for large molecules it is exceedingly difficult due to polymorphism and/or formation of higher-order structures. For instance, peak broadening is observed in solution NMR for polymeric and supramolecular systems. Single molecule studies using Atomic Force Microscopy (AFM) could offer solution to these challenges, which we show for the first time.

G-quadruplex forming (G-4) sequence(s) in the promoter of a gene could act as a transcription regulator.²⁴⁻²⁷ The same could be applicable for G-4 sequences found in replication origins, and thus act as a regulator for DNA replication.²⁸⁻³⁰ Since G-quadruplexes are highly polymorphic, using it as a regulator gives high degree of specificity and control. Albeit, the specific helicase-binding and G-quadruplex-unfolding mechanisms in vivo, are barely known.³¹⁻³⁴

1.3. Telomere

DNA of eukaryotic cells are packed tightly as a set of linear chromosomes into a small compartment referred to as the nucleus. The chromosomes terminate with short tandem repeats of duplex DNA followed by a small single-stranded DNA overhang. A simple termination like that could easily be recognized as double-strand breaks resulting in detrimental chromosome re-arrangements and thus genomic instability. To prevent the chromosome ends being recognized as double-strand breaks, it is capped by proteins collectively referred to as the shelterin complex. This whole DNA-protein complex with which linear eukaryotic chromosomes are terminated, is called Telomere. In many instances, telomeric DNA alone is also referred to as telomere for simplicity. The telomere of all mammals (which includes humans) is composed of varying number of TTAGGG repeats. A list of telomeric repeats for different classes of animals, plants, fungi, etc. could be found here: http://telomerase.asu.edu/sequences_telomere.html.

In humans, shelterin complex is composed of six main proteins: telomeric repeat-binding factors 1 and 2 (TRF1 and TRF2), protection of telomeres 1 (POT1), TRF1-interacting nuclear factor 2 (TIN2), TIN2- and POT1-interacting nuclear protein 1 (TPP1), repressor and activator protein 1 (RAP1). TRF1 and TRF2 can directly recognize and bind³⁵ duplex TTAGGG repeats, while POT1 binds to the single-stranded TTAGGG overhang³⁶. They are interconnected by TIN2, TPP1 and RAP1.³⁷ The shelterin complex enables the cellular DNA damage and repair (DDR) machinery to distinguish telomeric DNA from sites of chromosome breaks, and thus prevent inappropriate DNA damage repair pathways at telomeres which could otherwise lead to genomic instability. In presence of TRF2, it has been observed that telomeres form large duplex loops, called t-loops,^{38, 39} in which the single-stranded terminus is embedded within the double-stranded part of the telomeric tract (Figure 1.5).

With each cell division, telomere length is reduced by ~50 to 200 bp⁴⁰ primarily because the lagging strand of DNA synthesis is unable to replicate the

extreme 3' end of the chromosome. Eventually, telomeres become so short that cells lose the ability to replicate further. In most eukaryotic cells, this problem is tackled by the action of a reverse transcriptase called Telomerase^{41, 42}. In cells with active telomerase, such as cancer cells, the telomere length is continually being built up and shortened in a regulated way that maintains telomere length homeostasis and retains telomere functionality.⁴³ Another similar mechanism is called ALT^{44, 45} which stands for “alternative lengthening of telomeres” which have been found in some cancer cells where telomerase expression was specifically suppressed.

Figure 1.5 The structure of telomere

Chromosomes terminate with a single-stranded 3'-overhang which is folded into a T-loop by TRF1 and TRF2. This end is further protected by 4 other shelterin proteins. Telomerase can increase the length of telomere by reverse transcribing the C-rich telomeric strand.⁴⁶

1.4. DNA G-wires

Formation of higher-order structures by Guanine-rich DNA sequences is very well-known. *Tetrahymena* telomeric sequence d[G₄T₂G₄] was one of the first reported to form wire-like structures, called G-wires⁴⁷. The *Oxytricha* telomeric sequence d[G₄T₄G₄] and the mammalian telomeric sequence d[(TTAGGG)_{2,4}] have also been shown to form DNA G-wires.^{48,49} Telomeres are long repetitive sequences found at both ends of a chromosome.

G-wires have previously been visualized in dehydrated or partially hydrated environments using scanning probe microscopy.⁵⁰ AFM imaging of duplex DNA along with G-wires formed by d[GGGGTTGGGG] revealed that G-wires are stiffer and higher (Figure 1.6).⁵⁰ In air, the height of G-wires were found to be double that of supercoiled duplex DNA and 4 times the height of duplex DNA. G-wires maintained similar average height of ~2.2 nm in propanol, but the height of duplex DNA was further reduced, suggesting that G-wires were less affected by dehydration in a non-polar solvent. However, the height of fully hydrated G-wires in aqueous solution has not been reported earlier. The observed height of G-wires is comparable to expected diameter of a typical G-quadruplex from crystal structure, and it has been independently shown that d[TGGGGT] can form tetramolecular DNA G-quadruplexes^{51, 52}. Thus, the underlying assumption for the self-assembly of d[G₄T₂G₄] has been the formation of interlocked DNA G-quadruplexes.⁴⁷

DNA G-wires have received significant interest from the nanotechnology community for applications in nano-electronics,⁵³⁻⁵⁶ nanosensors⁵⁷⁻⁵⁹ and nanodevices^{60, 61}. G-wires are highly stable at room temperature and upon dehydration, making them ideally suited for harsh environments in nanotechnological applications. Long-range charge transport in G-wires by hopping may allow their use as conductive nanowires.⁶² Recent experimental studies have revealed conductive behavior in single G-wires over lengths of over 100 nm.⁵⁴ Despite the popularity of G-wires in nanotechnology, the knowledge and understanding about its structure is very limited. Understanding the structure of G-

wires is crucial for the controlled assembly of systems⁶³ with optimal nano-mechanical and electrical properties.

Figure 1.6 DNA duplex and G-wires imaged in air and propanol

Topography images of duplex DNA and d[G₄T₂G₄] on Mica acquired using AFM in (a) air and (c) propanol. From the height profile across the marked line, it can be seen that height of G-wires (dark arrows) is double of supercoiled duplex DNA (gray arrows) and 4 times the height of duplex DNA (white arrows) in (c) air. G-wires maintain the same average height of 2.2 nm in (d) propanol, but the height of duplex DNA is further reduced. Vertical height scale is 10 nm. Figure adapted with permission from James Vesenska.⁵⁰

1.5. GGGGCC repeats

In 2011, a major genetic abnormality was identified in patients suffering from Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal dementia (FTD).⁶⁴
⁶⁵ The promoter and (or) the first intron of C9orf72 gene of some ALS/FTD patients and their family members contained unusually long d[GGGGCC] repeats. For ALS, this accounted for 23–47% of familial cases and 4–5% of sporadic cases.^{64–66} However, only about 10% of ALS cases are familial, and the remaining 90% cases are sporadic⁶⁷, which means that GGGGCC repeats in the C9orf72 gene accounts for just 6–9 % of all ALS cases. More than 20 genes have been implicated in the development of ALS⁶⁸ (<http://alsod.iop.kcl.ac.uk>), among which, C9orf72 and superoxide dismutase 1 (SOD1) have been diagnosed to be the most commonly mutated genes⁶⁸. Very little is known about the molecular basis of FTD.

Four principal disease mechanisms have been proposed to be associated with G₄C₂ mutation in the C9orf72 gene: (1) loss of function due to reduced levels of C9orf72 mRNA^{69, 70}; cell toxicity induced by (2) RNA G-quadruplexes^{71, 72}; (3) poly-dipeptide repeats^{73, 74}; and (4) sequestration of RNA binding proteins^{75–79}. Four repeats of G₄C₂ DNA^{72, 80} and RNA^{81, 82} have been shown to form G-quadruplex using solution NMR. It has been observed that transcription of long G₄CC repeats lead to abortive transcripts of variable length⁷¹. RNases in the cell could chop down these abortive transcripts into smaller fragments, including 1.5 repeats of r[G₄C₂]. By immunostaining with BG4, it has been shown that C9 ALS patient fibroblasts and astrocytes have more abundant RNA G-quadruplexes compared to those without the expansion (Figure 1.7)⁸³. BG4 antibody is known to have very high affinity for intramolecular and intermolecular DNA G-quadruplexes (K_d = 0.5–1.6 nM and 2.0 nM, respectively) with no detectable binding to RNA hairpin, single-stranded DNA or double-stranded DNA.²⁰

Figure 1.7 BG4 stained foci and area in fibroblasts and astrocytes

Immuno-staining C9ALS (a) fibroblasts and (b) astrocytes with BG4 showed increased fluorescence compared to control. All stained area in pink are above 10% threshold in red, while the areas in white correspond to particularly dense staining. Inset depicts source image showing only the red (BG4-FLAG) channel. (c) Stained foci in C9 fibroblasts is 1.8 times more while in C9 astrocytes it is 2 times more compared to non-C9-ALS controls. (d) Stained area in C9 fibroblasts is 2.3 times more while in C9 astrocytes it is 2.8 times more compared to non-C9-ALS controls. Figure adapted from eLife 2016;5:e17820.⁸³

1.6. TERRA

The ends of mammalian chromosomes, called telomeres, are composed of varying number of TTAGGG repeats. Initially, it was considered transcriptionally silent, but in the past decade it was found otherwise. RNA polymerase II (RNAPII) transcribes it into telomeric repeat-containing RNA (TERRA) in a variety of eukaryotes including mammals, zebra fish, budding yeast, etc.⁸⁴ The length of TERRA varies from 100 – 9000 bases depending on the length of telomere and the promoter (start site) for transcription which is usually located in the sub-telomeric region. Transcription of TERRA is regulated by several factors including cell phase, telomere length, cellular stress, cancer stages and chromatin structure. Low TERRA level has been observed in undifferentiated cells, carcinogenic cells and in closed chromatin structure; whereas high TERRA levels are observed in cells with long telomeres and under stress.⁸⁵

RNA transcripts from both C-rich and G-rich strands of mammalian telomere have been found in the cell nucleus. We investigated the transcripts of r[(UUAGGG)_n], generated from the C-rich DNA template. Based on gel electrophoresis and UV/CD spectroscopy, it has been proposed that d[(TTAGGG)₁₂₋₇₂] forms “beads on a string” structure with each “bead” consisting of three G-quartet bundles stacked in an antiparallel arrangement, joined by flexible TTA linkers.⁸⁶ The structure of short G-rich telomeric DNA can provide some insight into the structure of short TERRA, but the biologically relevant transcripts are 10 to 100 times longer.

TERRA has been associated with several important biological functions like telomere protection^{85, 87-89}, maintenance of telomere length^{85, 90}, telomerase inhibition^{85, 91} and telomere hetero-chromatin formation^{85, 88, 89, 92-95}. Mammalian TERRA primarily localize to telomeres in both interphase cells and transcriptionally inactive metaphase cells.^{92, 94}

The structure of long TERRA has been previously investigated using transmission electron microscopy⁹⁶. The G-rich transcripts appeared to be mostly

blobs and short rods, while the C-rich transcripts seemed to be longer even though both were supposed to be ~0.6 kb long (Figure 1.8). We received the plasmid containing telomeric repeats from Jack D. Griffith, and after transcribing TERRA, its structure was investigated using AFM. The solution AFM images showed interesting structural features which were never observed before.

Figure 1.8 *EM visualization of C-rich and G-rich TERRA*

C-rich (A) and G-rich (B) RNA molecules in 100 mM KCl were prepared for EM by mounting on thin carbon supports, dehydrating, and rotary shadow casting with tungsten. The C-rich RNA appears as an extended thread with kinks and compact regions. The G-rich RNA appears as a mixture of balls and thick rods (arrows). The thickness of the rods is significantly greater than that of the C-rich RNA or duplex DNA. C-rich (C) and G-rich (D) RNA molecules were mounted in 10 mM KCl as in A and B. The C-rich RNA again appears extended with kinks, whereas the G-rich RNA appears as mostly balls as opposed to a mixture of rods and balls. The bar is equivalent to 100 nm. Figure adapted with permission from Jack D. Griffith.⁹⁶

Chapter 2: Experimental techniques

2.1. G-wire synthesis

DNA/RNA oligonucleotides were usually purchased from IDT, which were purified by standard desalting columns. The lyophilized samples were diluted to a concentration of 500 μM in 10 mM HEPES (pH 7) as per the quantity (nmoles) of oligonucleotide mentioned by IDT. HEPES was purchased from GE Healthcare as 1 M solution (free acid, pH 7). The DNA/RNA concentrations were also cross-verified using UV spectroscopy. For G-wire synthesis, the very first oligonucleotide tried, was the *Tetrahymena* telomeric sequence d[GGGGTTGGGG], which is well-known to form DNA G-wires. The 500 μM sample was diluted to required concentrations (10 μM , 50 μM , 100 μM and 250 μM) in buffer containing either 50 mM or 100 mM KCl, 10 mM MgCl_2 and 10/20 mM HEPES (pH 7). In the living cell, the concentration of K^+ and Mg^{++} has been estimated to be ~ 130 mM and ~ 10 mM (1 mM free) respectively.¹⁵ The diluted sample was then heated rapidly in boiling water and cooled very slowly so as to have a temperature gradient of $<0.6^\circ\text{C}/\text{min}$. The temperature gradient was controlled by adjusting the temperature of the hot plate on which the beaker containing the water bath was kept. The beaker was covered with aluminum foil to minimize rate of cooling while a thermometer was poked through a small hole. The same method of G-wire synthesis was used for several other DNA/RNA oligonucleotides most of which were tested using AFM (air) and PAGE (native) to check for G-wire formation (see Appendix).

Length (bases)	Sequence
7	d[GGG T GGG]
8	d[GGG TT GGG]
9	d[GGG TTA GGG]
9	d[GGGG T GGGG]
9	d[GGGG A GGGG]
10	d[GGGG TC GGGG]
10	d[GGGG CC GGGG]
10	r[GGGG CC GGGG]
10	r[GGGG UU GGGG]
11	d[GGGG TTA GGGG]
12	d[GGGG TTTT GGGG]
12	d[TT GGGG TT GGGG]
16	d[GGGG TT GGGG TT GGGG]
16	d[GGGG CC GGGG CC GGGG]
17	d[C GGGG CC GGGG CC GGGG]

2.2. Synthesis of long DNA and RNA

Polymerase Chain Reaction (PCR) is one of the most commonly used techniques for DNA amplification/modification. It was invented in 1983 by Kary Mullis and later developed by Michael Smith, for which both won the Nobel Prize in Chemistry in 1993. PCR can be used to amplify trace amounts of DNA (<100 ng) using suitable primers and DNA polymerase. The PCR reaction mixture is passed through several heating/cooling cycles as shown in Figure 2.1.

Primers are short single stranded DNA that can bind to the ends of the DNA sequence to be amplified. A primer which is not sufficiently unique and have high

binding affinity, could bind to random sites and lead to undesired amplification of the template DNA. The primer should also not have repeats, as that could lead to slipping. Formation of primer-dimers and other secondary structures can decrease the amount of PCR product. Several programs and online websites are freely available to design suitable primers. However, for more control, a program in Mathematica was written that can search for primers in the desired location within certain range of melting temperatures. The primer sequences are then broken into smaller pieces to check for match with both strands of duplex DNA template. The program lists all matches and their melting temperatures. The most unique primers are thereby chosen as forward (left) and reverse (right) primer and subsequently checked using PerlPrimer or Primer3 for the formation of primer-dimers or hairpin. SnapGene program was used to simulate virtual agarose gel with the chosen primers, to check for the success of PCR and length of PCR product.

Pfu polymerase was used for DNA amplification using PCR. It is highly thermostable and have a high extension temperature of 72–75°C which is enough for most secondary structures to unfold into single stranded DNA and thus clear obstacles in the path of DNA polymerase while synthesizing complimentary strand of DNA from 5' to 3' direction. Pfu also has endonuclease activity from 3' to 5' direction that automatically corrects misincorporated dNTPs. For long amplicons (>2 kb), Q5[®] polymerase could be used that is much faster, has similar extension temperature and better proof-reading ability.

Figure 2.1 Polymerase Chain Reaction

Schematic showing the basic mechanism of a single PCR cycle.

RNA was obtained by in-vitro transcription using T7 RNA polymerase, which uses the sequence (d[TAATACGACTCACTATAGG]) for recognition. The T7 promoter could be inserted into any duplex DNA either by ligation or by designing suitable PCR primers with T7 promoter that is an overhang or partial mismatch. During transcription, the RNA polymerase synthesizes RNA from the end of the T7 promoter, starting with 'GG' (Figure 2.2).

Figure 2.2 Transcription by T7 RNA polymerase

Schematic showing transcription in progress as the T7 polymerase binds to the opposite strand of the T7 promoter (red) and synthesizes RNA (blue) by creating a ripple as it progresses from 3' to 5' direction.

Following both DNA and RNA synthesis, purification was done to get rid of proteins, salts, etc. For DNA, this could be easily done using filters containing silica, called spin columns. Buffer PB (from Qiagen) containing ~5M guanidine hydrochloride and ~30% isopropanol binds DNA/RNA to silica. Buffer PE (from Qiagen) containing 80% ethanol and 10 mM Tris-HCl pH 7.5 precipitates long DNA/RNA (>50 nt) in the column, while washing away everything else. After the cleanup, DNA/RNA was eluted in 10 mM HEPES pH 7 (DNase/RNase free). RNA was purified using RNA certified spin columns (from Epoch Life Science). The exact composition of buffer PB and PE are confidential, but some compositional reference was obtained from http://www.openwetware.org/wiki/Qiagen_Buffers, which helped understand the mechanism of purification and judge their suitability for RNA purification even though they were originally meant only for DNA purification.

2.3. Gel electrophoresis

Both DNA and RNA are negatively charged and thus get attracted towards a positive electrode (Figure 2.3). This phenomenon can be used to separate DNA/RNA molecules based on their charge density, size and shape. The resolution of separation depends on a variety of factors, but mainly the media of propagation which is typically composed of polyacrylamide or agarose. Polyacrylamide gel is formed by copolymerization of acrylamide and bis-acrylamide (typically with ratio of 37.5:1 or 29:1). The polymerization is initiated by ammonium persulfate (APS) and catalyzed by Tetramethylethylenediamine (TEMED).⁹⁷ Agarose is a complex polysaccharide obtained from Seaweed which is dissolved in an aqueous buffer by heating, and is subsequently cooled to form a gel. The composition and concentration of polyacrylamide/agarose can be tuned to control the porosity of the gel matrix, and thus the resistance felt by molecules moving through it. Usually, the pore size in polyacrylamide gel is much smaller and more uniform than agarose, and thus offer better resolving power for small molecules (up to 300 nucleotides). Large/long molecules have a hard time crawling through small pores, and so agarose gel is used for large molecules ($10^2 - 10^6$ nucleotides).

Figure 2.3 *Gel electrophoresis setup*

A high voltage DC power supply is used to create an electric field across the gel. Negatively charged molecules interact with the electric field and migrate towards the positive electrode.

The DNA/RNA molecules trapped in the gel can be visualized either from absorption of UV light by DNA/RNA or fluorescence emitted by a dye bound to DNA/RNA. In most of my experiments, a fluorescent dye named SYBR gold was directly added to the samples prior to loading in the polyacrylamide/agarose gel. It was observed that 2 μ L of 10x SYBR gold was sufficient to show bands containing as little as 5 ng DNA/RNA. When SYBR gold is bound to nucleic acids, it has very high fluorescence quantum yield (~ 0.7) with two fluorescence excitation maxima at ~ 300 nm and ~ 495 nm, leading to emission at ~ 537 nm.⁹⁸

Gel electrophoresis is not only useful for visualization, but also for purification. Polyacrylamide gel electrophoresis (PAGE) was used to purify DNA G-wires. The gel was visualized by UV shadowing and a portion of the gel containing DNA molecules of desired size was excised. 1 mL of buffer containing 50 mM KCl, 10 mM MgCl₂ and 10 mM HEPES was added to about 0.2 g of crushed gel and gently vortexed overnight (12 hours) at 50°C. The samples were centrifuged at 16,000g and the supernatant was passed through a desalting column to get rid of dissolved polyacrylamide and excess salts. The purified samples were too dilute, but was still sufficient for AFM imaging. Instead of passing the sample through a desalting column, it could be passed through a simple 0.2 μ m filter and dialyzed in appropriate buffer. The purified samples could be lyophilized and re-suspended in desired buffer and volume.

Purification from agarose gel was much quicker and straightforward using the gel extraction kit from Qiagen/ NEB. The excised gel was easily dissolved in buffer QG (from Qiagen) by vortexing for 15 to 30 minutes at 50°C. Buffer QG contained 5.5 M guanidine thiocyanate (GuSCN) and 20 mM Tris HCl pH 6.6. The dissolved gel solution was then passed through a spin column, which traps DNA molecules. The spin column was rinsed with buffer PE to wash away salts and other chemicals. The trapped DNA was then eluted using 40 μ L of 10 mM HEPES buffer.

2.4. Atomic Force Microscopy

An Atomic Force Microscope (AFM) is a high-resolution microscope that detects force instead of electromagnetic radiation. An AFM can be operated in gaseous or liquid environment, and is the only technique that can be used for atomic-level visualization of dynamics in liquid. For AFM, there is no need to stain the samples with anything for visualization or enhancing contrast.

A very sharp tip with end-radius of <10 nm interacts with the sample surface, while a laser beam deflects depending on movement of the tip. This deflection signal is used as feedback to control (and record) the height of the sample while maintaining certain parameter(s) constant. Based on the feedback parameter(s), several modes of AFM imaging can be defined which are used to measure different properties of a sample (Figure 2.4). AFM was developed by Binnig, Quate and Gerber in 1986,⁹⁹ for which (along with STM) they won the Nobel Prize in Physics in the same year.

One of the simplest mode of AFM operation is called the “Contact mode”, in which the AFM tip hovers at a constant height (< 1 nm) over the sample surface such that the interaction force between the tip and sample is constant. One of the drawbacks of such a technique is lateral forces generated on encountering an obstacle. When the tip encounters a fragile or weakly immobilized sample it can drag and damage it. An AFM mode frequently used for imaging soft and fragile samples is amplitude modulated AC mode (AM mode); also known as dynamic mode, tapping mode or intermittent-contact mode.

Feedback parameter	AFM mode	Measured Parameter
Deflection	Contact mode PeakForce™ mode Jumping™ mode QI™ mode	Height Force Current Contact Resonance
Amplitude	AM-AC mode Tapping™ mode Dynamic mode	Height Electric field Magnetic field Surface potential
Drive-amplitude	DAM mode	Height
Frequency	FM-AC mode	Height Force Electric field Surface potential
Phase	Phase-locked mode	Height

Figure 2.4 AFM imaging modes

Various modes of imaging using AFM based on feedback parameter that is kept constant.

In tapping mode, the AFM cantilever is driven continuously at its fundamental resonant frequency while its amplitude is used as a feedback parameter to record the sample height (Figure 2.5). In tapping mode, the AFM tip may come in contact with the sample occasionally or periodically owing to stickiness of the surface. Stiffer cantilevers have lesser tendency to be affected by this, and are thus the preferred choice for imaging in air. However, it cannot be ensured that the tip never comes in contact with the sample; and when that happens the sample/tip are prone to damage or deformation due to the very high contact forces of stiff cantilevers. Occasionally, the height obtained in tapping mode can be less accurate (usually underestimate) compared to that obtained using contact mode, as the damping of the AFM cantilever oscillation is not necessarily always proportional to

the tip-sample distance. Even for the same tip-sample distance, the tip-sample interactions may vary (e.g.: Mica vs. DNA). The “reduced height artifact” is frequently observed for DNA samples imaged in air^{100, 101} and thus we refrained from using Tapping mode in air.

Figure 2.5 Principle of Tapping mode

AFM tip trajectory in AM/Tapping mode. As the tip approaches the sample and is sufficiently close, the cantilever oscillation is damped owing to attractive tip-sample interaction forces. When the amplitude is damped sufficiently so as to reach a user-defined amplitude difference, the tip retracts until the cantilever attains its free amplitude.

ScanAsyst™ is an imaging mode based on PeakForce™ tapping which combines the best of contact mode and non-contact mode. Instead of hovering over the sample at a constant height, the AFM cantilever is subject to forced oscillations at a constant non-resonant frequency. In the Multimode 8 AFM (from Bruker), a single up/down tap of AFM cantilever takes about 0.5 ms, during which the tip gets close to the sample till a user-defined value of cantilever deflection is reached and then moves away from the sample. This way, the height of the sample is recorded at every point the AFM tip taps. Figure 2.6 shows the cantilever trajectory and the resultant deflection (V) as a function of time.

Figure 2.6 Principle of ScanAsyst mode

AFM tip trajectory and cantilever deflection (in volts) as a function of time in ScanAsyst™ mode. When the tip is far away from the sample, it experiences negligible force (**A**). As it approaches towards the sample, it suddenly jumps on-to-contact with the sample if it is sufficiently close to the sample (**B**). The tip then feels a repulsive force as it is trying to deform the sample until a user-defined value is reached (**C**). The tip then tries to move away from the sample, but it must overcome significant attractive force (**D**). Finally, the tip moves far away from the sample where it experiences negligible force once again (**E**). The entire cycle takes about 0.5 ms in the AFM used.

The voltage when multiplied with the deflection sensitivity¹⁰² (nm/V) of the cantilever, gives the actual deflection of the cantilever, which in turn can be multiplied with the stiffness (nN/nm) of the cantilever to obtain the interaction force between the AFM tip and sample. The deflection sensitivity, also known as the inverse optical lever deflection, was measured from the slope of the repulsive part of the force curves obtained on the Mica surface with high setpoint. For triangular cantilevers, a deflection sensitivity correction factor (χ) of 1.12 was used instead of

the default value of 1.08.¹⁰³ The cantilever stiffness was determined from its thermal noise spectrum using Sader's method.¹⁰⁴

Recent advances in electronics of AFM controllers have made it possible to employ two or three feedback mechanisms to measure multiple properties of a sample simultaneously. One such imaging mode that has been used for our studies is called the AM-FM mode, which employs dual feedback mechanism to maintain constant cantilever amplitude for the first harmonic and constant frequency for the second harmonic resonance of the AFM cantilever.

2.5. Circular Dichroism

Some molecules have a unique property to interact with light based on its state of polarization, i.e., the direction of electric field vector. DNA is one such molecule that can interact with left and right circularly polarized light, and absorb differently based on its secondary structure. Thus, by measuring the difference in absorption of left and right circularly polarized light, it is possible to make assumptions about the structure of nucleic acids. For instance, an unfolded single stranded DNA would give zero differential absorption. The parameter that is usually plotted as CD spectra against wavelength is called "ellipticity" that can be determined using the relation:

$$\theta \text{ (in radians)} = \tan^{-1} \left[\frac{\sqrt{I_R} - \sqrt{I_L}}{\sqrt{I_R} + \sqrt{I_L}} \right]$$

Where I_R and I_L are the intensities of right and left circularly polarized light respectively measured after passing through the sample.

Figure 2.7 CD spectra of different types of G-quadruplexes

d[GGTTGGTGTGGTTGG] adopts a chair type anti-parallel structure resulting in a positive band near 290 nm and a negative band near 260 nm. d[G₄T₄G₄] dimeric G-quadruplexes with a symmetric basket type anti-parallel structure resulting in a positive band near 290 nm and pronounced negative band near 260 nm. d[TGGGGT] forms four-stranded parallel G-quadruplex resulting in a pronounced positive band near 260 nm. d[TAGGGUTAGGGT] forms dimeric parallel G-quadruplex with anti-parallel external loops resulting in a CD spectrum with positive maxima near both 260 nm and 290 nm.¹⁰⁵

2.6. NMR spectroscopy

Nucleons have finite magnetic moment ($2.793 \mu_N$ for protons and $-1.913 \mu_N$ for neutrons), where $\mu_N = 5.051 \times 10^{-27}$ Joules/Tesla, and is called the nuclear magneton. When a nucleon is subjected to an external magnetic field (B), the degeneracy in its spin magnetic moment is broken and split into two spin states ($\pm 1/2$). The $+1/2$ state is less energetic and is thus more populated than the $-1/2$ state in equilibrium. However, it can transition to the $-1/2$ state upon absorption of a photon that matches exactly to the energy difference (γB) between the two states. γ is the

gyromagnetic ratio which is specific for each atomic nucleus (42.578 MHz/T for proton and 29.165 MHz/T for neutron). Nuclear magnetic resonance (NMR) thus refers to the phenomenon in which a nucleus absorbs electromagnetic radiation corresponding to the energy difference between the different spin states, when subjected to an external magnetic field.

At equilibrium, the majority of nuclear magnetic moments in a sample (solution) align along the direction of external magnetic field, resulting in a bulk magnetization in the same (say z) direction. On application of a radiofrequency pulse along the x-direction, an oscillating magnetic field is generated which leads to the precession of the magnetization vector about the yz-plane, at the Larmor frequency. The precession of the magnetization vector induces a current in the detection coil which is recorded as a function of time. The nuclear magnetic moments in the sample revert back to equilibrium over time, leading to a gradual decay in the current referred to as free induction decay (FID). The FID is used to generate NMR spectrum by Fourier transformation. The NMR frequency is usually converted to chemical shift given in ppm (parts per million), which is a ratio of frequencies and thus a dimensionless quantity that can be universally compared. Two popular compounds used as zero reference for chemical shift are 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) and tetramethylsilane (TMS).

The nucleus of an atom is surrounded by a cloud of electrons that can alter the effect of an external magnetic field. Thus, the magnetic field felt by the nucleus of different atoms is usually different as it is highly unlikely that the electron cloud distribution around any two nuclei is exactly the same. Based on this, it is possible to identify specific atoms of a molecule, which for our experiments is $^1\text{H}_1$ in nucleic acids. The different protons of DNA have different characteristic chemical shifts (Figure 2.8) which can help deduce its structure and chemical composition. In a G-tetrad, the imino proton (H1) is hydrogen-bonded to an oxygen atom, resulting in a characteristic chemical shift between 10 – 12.5 ppm, which is of our main interest.

Figure 2.8 Proton chemical shifts in nucleic acids

Protons of nucleic acids have unique chemical shifts that can be used to identify a particular proton. The atom numbering is based on IUPAC convention.

Chapter 3: Development of solution AFM imaging

3.1. Introduction

Scanning Probe Microscopy (SPM) has been used for studying DNA for nearly 30 years.^{106, 107} As of now, it is the only technique with which it is possible to directly visualize individual atoms/molecules in liquid with sub-nm resolution. Imaging in aqueous solution is very crucial for most biological samples as dehydration may adversely affect their structure and properties. Imaging in solution is also useful when high-resolution structural information is desired from biomolecules.

Atomic Force Microscopy is one of several sub-techniques in Scanning Probe Microscopy, which has been extensively used to address several biological questions involving DNA. Our motivation for using AFM was to observe intricate structural features (if any) of higher-order structures formed by biologically relevant G-rich DNA and RNA. However, to achieve sufficiently high-resolution, it is of utmost importance to perfectly immobilize the molecules of interest within the imaging timeframe. In addition to the challenge of immobilization, several impediments were encountered in attempts to imaging in solution, especially with high-resolution. This chapter attempts to shed light on overcoming some of the crucial impediments faced in visualizing unknown structures formed by G-rich DNA and RNA in solution.

3.2. Tip-sample interaction

A major hurdle for imaging soft samples in air is large attractive tip-sample interaction forces that can squish the sample underneath the tip. Most dry surfaces are covered with water molecules from moisture in the air, which can lead to high attractive forces due to surface tension at the tip-sample junction. In water, the attractive van-der-Waals force is very low and may be cancelled by repulsive forces (Figure 3.1). It is thus possible to image with purely the user-defined deflection setpoint which can be as low as ~ 10 pN using very soft AFM cantilevers. Low force imaging is extremely crucial for high-resolution imaging as it not only prevents the delicate features from being deformed under pressure, but it also minimizes the tip-sample contact area, thus improving the lateral resolution.

Figure 3.1 *Force-time curve in air vs. water*

Tip-sample interaction force in air vs. liquid with same PeakForce setpoint of ~ 0.2 nN. The attractive tip-sample interaction force in air is ~ 1 nN, while in water it is only few pN. The same AFM tip was used in both cases so that the effect of tip could be neglected. It can be clearly seen the tip-sample interaction force in solution is just the user-defined repulsive force as attractive forces are practically absent.

3.3. Overcoming the “forest of peaks”

In liquids, tapping mode is extremely challenging as the cantilever dynamics is affected by changes in the volume of the liquid droplet due to evaporation. Also, the liquid droplet vibrates with the oscillating probe and creates spurious resonances that show up as a “forest of peaks”, making it difficult to identify the resonant frequency of the cantilever. It can however be determined from the thermal noise spectrum, but may not correspond to the optimal resonant frequency (Figure 3.2).

Figure 3.2 *Cantilever tuning*

Cantilever excitation spectra in liquid from (A) photothermal and (B) mechanical excitation for FastScan-D AFM cantilever. The thermal noise spectrum is shown in red, while the phase of cantilever excitation is shown in cyan. With photothermal excitation, the cantilever’s resonant peak matches with that obtained from the thermal noise, but with mechanical excitation no distinct resonance peak can be observed at the expected position.

Both the problems faced in liquid with conventional tapping mode was overcome using Cypher ES. It uses a closed pressure-controlled fluid cell to prevent

liquid evaporation and a pulsed laser (BlueDrive™) to directly oscillate the AFM cantilever. BlueDrive™ gives an extremely clean excitation spectrum for the cantilever where the resonant peak matches exactly with that determined from thermal noise spectra (figure 3.2).

BlueDrive™ is a commercialized version of photothermal excitation of AFM cantilever using pulsed laser of appropriate wavelength that can be absorbed by the cantilever. A pulsed violet (405 nm) laser of tunable power (0.01 – 10 mW) is passed through intensity filter and is incident at the base of the cantilever. Cantilevers that can absorb blue light (e.g.: uncoated silicon or gold coated cantilevers) can be heated locally at the base to bend (oscillate) the cantilever. This is based on the property that most materials expand on heating, and can thus bend in presence of a thermal gradient across its thickness/depth.

The intensity of the excitation laser could be modulated using filters labelled as 1x, 0.3x, 0.1x and 0.01x. The choice of the filter depends on desired imaging amplitude and the cantilever. Using the 1x filter, excitation energy of 0.1 – 7 mW could be obtained, while it was preferable to use the 0.1x filter that allowed excitation energy of 10 μ W to 0.7 mW. Typically, for cantilever amplitudes of less than 1 nm with the AFM probes that were used, few μ watts of power was required.

Apart from direct cantilever excitation, it is possible bypass the issue of “forest of peaks” and unstable cantilever dynamics using PeakForce tapping (Section 2.4.).

3.4. Using the Cypher ES

The Cypher ES AFM is a sample scanning AFM like the Multimode 8, i.e., the sample moves in XYZ direction during scanning while the AFM tip is either static or oscillating at a fixed position. A continuous wave laser shines on the free-end of the AFM cantilever, and the movement of the reflected beam is recorded by a position sensitive photodiode (PSD). The PSD is extremely sensitive and has dark current noise of less than 10 nA in a bandwidth from 0.1 Hz to 10 kHz under typical

conditions. The PSD is thus able to register extremely small movements of the cantilever by as low as 10^{-13} m. The photodiode current is sent to pre-amplifiers in the electronic controller to be used as feedback signal to control the Z-position of the sample stage. The AFM has the option to oscillate the cantilever either using a piezoelectric transducer attached to the probe holder or using the pulsed laser. Figure 3.3 shows a simplified schematic of the working principle of the Cypher ES, while Figure 3.4 shows the actual setup. Both the detection and excitation laser pass through an objective lens with NA = 0.45, before reaching the AFM cantilever. The view through the objective lens is captured using a 3.1-megapixel CMOS camera with FireWire interface. The Cypher Top-View optics offers a fixed optical zoom of 20x and a maximum field of view of $690\text{ }\mu\text{m} \times 920\text{ }\mu\text{m}$ which can be digitally zoomed up to 32x.

Figure 3.3 AFM schematic

Schematic representation of Cypher ES AFM with both acoustic (piezoelectric) and photothermal excitation options for the AFM cantilever.

Figure 3.4 *AFM setup*

The Cypher ES setup being used for solution imaging. The syringe is used to inject imaging buffer or water into the closed fluid cell before imaging. The water meniscus usually has a volume of 70 – 100 μL , and the sample is submerged in it throughout the entire duration of imaging.

The Cypher ES employs Peltier heater/cooler in the sample stage to control the temperature of the sample. The fluid cell is ensured to be air-tight before changing the temperature of the sample to prevent condensation/evaporation. With the current module, the temperature of the sample can be varied from 0°C to 100°C. The pressure inside the closed imaging cell can also be controlled to a limited extent by adjusting the position of the piston in a tight syringe attached to the air outlet of the cell. This is helpful in controlling the vapor pressure of the liquid to shift the equilibrium between condensation and evaporation.

The AFM software offers easy-to-use interface for setting up of the experiment. AC mode was used for simple imaging, while AM-FM mode was used for simultaneous imaging of topography with mechanical properties. The cantilever resonant frequency was determined from thermal noise spectra, and the cantilever was subsequently driven using BlueDrive. Free cantilever amplitude of ~250 mV and amplitude setpoint of ~200 mV is usually sufficient to reach near the sample. After engaging onto surface, the setpoint was further decreased (to ~170 mV) until the tip was able to track the sample topography perfectly. For high-resolution imaging, both the free amplitude and amplitude setpoint had to be further decreased to 10 – 20% of initial values in order to see fine structural features. However, imaging with low cantilever amplitudes is challenging as the tip is so close to the sample that it can get stuck to it thus giving rise to streaks in the AFM image. An improper ratio of free amplitude and amplitude setpoint could crush the tip and sample due to high force. In addition to cantilever parameters, the imaging speed and integral gain (of pre-amplifiers) need to be optimized. Even though the AFM allows imaging speeds of up to 30,000 pixels/s, practically the fastest imaging speed would be limited by the response time of the cantilever and the roughness of the sample surface. For instance, with Biolever Mini, MSNL-F and ScanAsyst Fluid+ probes, maximum imaging speeds of 5000 pixels/s were found to be optimal for high-resolution imaging of DNA.

3.5. Ultra-short cantilevers

It was extremely difficult to image G-wires in resonating modes (AM/FM) with Cypher ES using standard cantilevers. It was likely due to movement of G-wires due to large hydrodynamic drag forces created by the oscillating cantilever. The problem was overcome by using ultra-short cantilevers (USCs), which compared to traditional cantilevers, offered ~ 100 times less interaction cross-sectional area with the surrounding fluid. In our Cypher ES AFM, the detection laser module was interchangeable and a SLD (super-luminescent diode) laser with very small incident spot size of $3 \times 10 \mu\text{m}$ was used for imaging with USCs. SLD laser was chosen as it has much lower interference noise compared to standard RF lasers, because of its high degree of spatial coherence.

FastScan-D and USC-F1.2-k0.15 probes (Figure 3.5) were used, however it was not possible to get decent images with the latter. The Fastscan-D cantilevers have a spring constant of $\sim 0.2 \text{ nN/nm}$ and a silicon tip with end-radius of $8 - 12 \text{ nm}$ as per company specifications. Even though the tip-sharpness in FastScan-D is similar to that of standard probes, using the FastScan-D, the minor and major grooves of duplex DNA was resolved much easily compared to using standard probes. This could be attributed to the fact that the FastScan-D cantilever shares very little contact area with surrounding liquid and supposedly generate less hydrodynamic disturbance compared to standard cantilevers, which in turn causes the DNA molecules to be disturbed less while scanning. This allowed imaging of G-wires with very high-resolution and without any observable displacement of the molecules. All solution AFM images in this thesis, unless specifically mentioned, have been obtained using FastScan-D probe in either AM mode or AM-FM mode.

Figure 3.5 *Ultra-short cantilevers*

Optical image of USC AFM cantilevers captured using Top-view optics of Cypher ES with same digital zoom of 16x. The bright elliptical spot has dimensions of $3 \times 10 \mu\text{m}$ which is casted by diffused reflection of the red detection laser. The violet excitation laser casts a circular purple spot of $\sim 5 \mu\text{m}$ diameter, which gets masked under the bright spot in case of the $7 \mu\text{m}$ cantilever.

The resonant frequency and quality factor of AFM cantilevers are reduced in water (Figure 3.6). Thus, the FastScan-D offers very low response time (τ), which was estimated using the formula $\tau = (\pi \Delta f)^{-1}$, where Δf is the full-width half maximum width of the frequency response of the cantilever. With a typical response time of $2 \mu\text{s}$, imaging speeds of up to 500,000 pixels/s could be used. Standard sized soft cantilevers usually have $\Delta f \approx 10 \text{ kHz}$, thus imaging speeds of up to 30,000 pixels/s could be used theoretically. In practice however, optimal image quality is obtained with 5 – 20 times less imaging speed as several other factors are involved, especially the response time of the electronic feedback system of the AFM and the

response time of the piezoelectric sample stage. High-speed imaging with ultra-short cantilevers help to overcome imaging issues due to movement of the molecules within the scanning timeframe and drifting of the scan area.

Figure 3.6 *Thermal noise spectra in air vs. liquid for FastScan-D cantilever*

The thermal noise spectra were obtained in air (black) and water (blue) for the FastScan-D cantilever. In water, it has 1st harmonic resonance at ~100 kHz and 2nd harmonic resonance at ~1 MHz, both of which are nearly 1/4th of the corresponding resonant frequencies in air. The 1st harmonic in liquid has a very low thermal noise of ~5 pm, as can be seen from the logarithmic vertical scale of the graph. The bandwidth of 1st and 2nd harmonic resonances is nearly 20 times wider compared to that in air, thus offering lower response time and faster imaging speeds in liquid.

3.6. DNA immobilization

Sample preparation is usually the most crucial step in any imaging experiment. To capture a sharp (non-blurry) image using a common photographic camera, the object must be in the focal plane and not move within the imaging timeframe. The same is true with Atomic Force Microscopy, where the focal plane is the surface of the sample. It is thus necessary to anchor molecules to the surface for

proper visualization. Since DNA is extremely thin (~2 nm), the background against which it is imaged, should be atomically smooth or have less than 1 nm roughness. There are not a lot of cost-effective options for materials that can readily give atomically smooth surface. Mica is one of the cheapest of the limited options available. A small piece of Mica can be used several times for imaging different samples, simply by cleaving the top layer using a scotch tape.

However, both DNA and Mica are negatively charged in aqueous solution at neutral pH, and thus DNA can be immobilized either by making the Mica surface positively charged or by using multivalent cations that can form a bridge between DNA and Mica. Several such options have been explored by me: functionalizing Mica with APS (Aminopropyl silatrane), Poly-L-Lysine, Glutaraldehyde, or using multivalent cations like Mg^{++} , Ni^{++} and Spermidine.

It was found that without rinsing freshly cleaved Mica surface with DI water prior to application of sample, G-wires could not be immobilized (Figure 3.7). This is likely due to monovalent ions (especially K^+) and other cationic impurities from the surface of natural Mica that competes with divalent cations like Mg^{++} and Ni^{++} . To prepare a sample for imaging, about 30–60 ng of G-quadruplex DNA/RNA is added to 5–10 μ L buffer containing 10–20 mM $NiCl_2$ and 20 mM HEPES (pH 7). After about 5 minutes, 15–30 μ L water is added to it so that $NiCl_2$ concentration is 2–3 mM, and the dilute sample is instantly put on the Mica surface. After loading the sample in the AFM, about 50 μ L DI water was injected using a syringe to form a water meniscus as shown in figure 3.4. Instead of pure water, buffer containing 1 mM of $NiCl_2$ and HEPES could also be used. This protocol was used for imaging both duplex and G-quadruplex DNA.

Figure 3.7 *Floating G-wires on non-pretreated Mica*

Image of the same scan area containing d[G₄TTG₄] wires in solution on Mica, captured at 3.9 Hz for 256 x 256 pixels (left) and 1.95 Hz for 512 x 512 pixels (right). The mica surface was not rinsed prior to application of the sample. It can be seen that attempting to image with higher pixel density, G-wires fly away. The image was obtained in ScanAsyst mode using MSNL-F. The length scale (white bar) is 100 nm.

However, despite using similar protocol and NiCl₂ concentration (2.5 mM), the choice of AFM probe revealed significant difference in the immobilization of DNA. Using standard sized cantilevers of ~80 μm length (MSNL-F, ScanAsyst Fluid+) in tapping mode, we were unable to immobilize G-wires. Using short cantilevers of ~35 μm length, the quality of immobilization was satisfactory, but upon repeated imaging of the same region with high pixel density, most of the G-wires flew away from the scanning region (Figure 3.8). However, using ultra-short cantilevers of ~15 μm length and ~4 μm width (FastScan-D), the same region could be scanned reproducibly with high pixel density without causing any G-wires to float away (Figure 3.9).

Figure 3.8 *Disappearance of weakly immobilized G-wires*

Wires of d[G₄TTG₄] in solution on Mica that was rinsed prior to application of the sample. The area inside the dashed rectangle was imaged several times before zooming out. The density of G-wires is less in this area, which indicated that they flew away due to repeated imaging. The image was obtained in AM-FM mode using Biolever mini. The length scale (white bar) is 100 nm.

Figure 3.9 *Reproducible imaging of G-wires*

Images obtained from successive scanning of the same area containing wires of d[G₄TTG₄] in solution on Mica. The images represent one of the first success in imaging G-wires without making them fly away. The purple background is likely due to excessive NiCl₂ salt precipitated on the Mica surface. The image was obtained in AM-FM mode using FastScan-D. The length scale (white bar) is 100 nm.

Using the FastScan-D probe, it was possible to image duplex DNA alone with just 0.2 mM NiCl₂ (Figure 3.10), which was extremely beneficial for imaging protein-DNA complexes de-stabilized by divalent ions. By imaging with standard cantilevers, it is not possible to image duplex DNA with such good immobilization even with 1 mM NiCl₂.

Figure 3.10 Reproducible imaging of duplex DNA with just 0.2 mM NiCl_2

Images obtained from successive scanning of the same area containing duplex DNA (0.2 ng/ μL). With just 0.2 mM NiCl_2 , it was possible to immobilize duplex DNA for high-speed imaging using FastScan-D probe in AM mode. The length scale (white bar) is 100 nm.

As an alternative to using divalent ions for immobilization of DNA, the mica surface was functionalized with biotinylated poly-lysine (BPL). The initial goal for using such a functionalized surface was however to study protein-DNA interactions, which is a completely different story and outside the scope of this thesis. With BPL coated mica G-wires could be satisfactorily immobilized (Figure 3.11), but upon repeated scanning of a specific region with high pixel density using the FastScan-D probe, both duplex DNA and G-wires flew away (Figure 3.12).

Figure 3.11 *Duplex DNA and G-wires imaged on BPL mica*

Images obtained from successive scanning of the same area containing d[G₄TTG₄] wires and duplex DNA in solution on Mica coated with 10 μ L of biotinylated poly-lysine ($\sim$ 1 ng/ μ L). The purple flexible wires are duplex DNA of 320 bp. The length scale (white bar) is 100 nm.

Figure 3.12 *Disappearance of weakly immobilized DNA on BPL mica*

Molecules on biotinylated Poly-L-Lysine fly away in solution due to repeated imaging of the same area inside the dashed rectangle. The length scale (white bar) is 200 nm.

3.7. Immobilization of long G-wires

The length distribution of G-wires shown in the above solution images is slightly on the lower side than what was observed in air. It was very rare to observe G-wires longer than 100 nm, while in air imaging G-wires of up to 300 nm in length could be observed. It was suspected that longer G-wires are difficult to immobilize compared to shorter wires as large objects are more susceptible to hydrodynamic drag forces in liquid, and are thus easily swept away while scanning. So, in order to immobilize longer G-wires, it was necessary to have stronger attachment with higher density of anchor points. This was achieved using Spermidine ($C_7H_{19}N_3$), which is a polyamine offering three positively charged sites at neutral pH. So, compared to divalent cations, it has 50% more charge per attachment site, thus allowing tighter electrostatic binding between DNA and mica. It was not possible to bind DNA to mica with just Glutaraldehyde (Figure 3.13), but in conjunction with Spermidine, excellent immobilization of G-wires was observed (Figure 3.14). The reason is not clear, but is likely due to crosslinking of the amines (-NH) in spermidine.

Figure 3.13 *DNA/RNA do not attach with just glutaraldehyde*

Failed attempt to immobilize RNA G-wires of r[G₄CCG₄] (0.3 μ M) on Mica using glutaraldehyde (~0.033%). Image was obtained with FastScan-D probe in AM mode. The length scale (white bar) is 200 nm.

Figure 3.14 Duplex DNA and G-wires immobilized with Spermidine

In addition to short DNA G-wires, very long wires of up to 300 nm length can be observed in solution by using spermidine ($\sim 16.7 \mu\text{M}$) and glutaraldehyde (0.0167%). Duplex DNA of 492 bp was imaged alongside for comparison. The image was obtained in AM mode using FastScan-D. The length scale (white bar) is 100 nm.

Increasing the concentration of glutaraldehyde by 100%, helped getting rid of streakiness in the AFM image (Figure 3.15). Surprisingly, the height of G-wires imaged with spermidine and glutaraldehyde is 0.5 – 1.0 nm higher than what is typically observed with NiCl_2 or BPL-mica. This could be due to a coating of spermidine and glutaraldehyde on the G-wires.

Figure 3.15 *Clean imaging of G-wires with spermidine*

In addition to short DNA G-wires, very long wires of up to 300 nm length can be observed in solution by using spermidine ($\sim 16.7 \mu\text{M}$) and glutaraldehyde (0.0334%). The image was obtained in AM mode using FastScan-D. The length scale (white bar) is 200 nm.

3.8. High-resolution imaging

When the objects of interest are properly immobilized in solution, it is possible to image with high pixel density. However, the true resolution of an AFM image is not just limited by the pixel density, but also the tip-sample contact area. It is easier to obtain atomic resolution images on hard surfaces compared to soft surfaces because even at the lowest possible imaging forces, multiple atoms of a soft sample are in contact with the AFM tip. In order to obtain the highest resolution image possible for DNA/RNA, extremely low imaging forces need to be used. For

this it is necessary to image with extremely small cantilever amplitudes and use AFM probes with very soft cantilevers. The right combination of cantilever softness and low amplitude is very difficult to achieve as tips attached to soft cantilevers tend to stick to the surface when oscillating at very small amplitudes close to the sample surface.

The stickiness issue can be solved with sample preparation optimizations. The following precautions were taken to minimize the tip-sample stickiness:

- (i) Use low concentration of $\text{MgCl}_2/\text{NiCl}_2$ (<2.5 mM) for binding.
- (ii) Rinse freshly cleaved mica surface with plentiful DI water (2 – 3 mL) prior to loading of sample.
- (iii) Ensure that the mica surface do not become dry at any point after loading of the sample, so that the concentration of salts never get too high.
- (iv) Use little or no glutaraldehyde.
- (v) Use spermidine with the right quantity of glutaraldehyde.
- (vi) Do not use tris or phosphate based buffer.

Even after proper immobilization of the molecules and being able to image with very small cantilever amplitudes, it still might not be possible to obtain the desired high-resolution because of a lot of factors. One of the most crucial factors is thermal drifting of the sample with time due to temperature fluctuations in the environment. Due to thermal expansion/contraction of piezoelectric material used in the sample stage, the molecules are constantly moving in random directions while imaging, thus leading to blurriness. The AFM software comes with the option for drift compensation which relies on determining the drift vector from cross-correlation between two consecutive images. This is helpful only if the same region of interest is being scanned multiple times, but it might reduce the resolution due to error in cross-correlation between two images that are not exactly the same. So, the most practical way of resolving the thermal drift was to wait patiently for several

hours, or try on different days as we had no control over the centralized air-conditioning system.

It was very pitiful to start imaging with a new AFM tip for high-resolution imaging only to find out that any of the aforementioned issues remained. The AFM tip-end gets blunt with imaging, so extreme care was taken so that the tip never hits the surface too hard (>1 nN). Usually, the AFM tip loses its maximum sharpness at the very first time it touches the surface. So, it was brought close to the surface with extreme precaution by suitable choice of free amplitude, amplitude setpoint, engage speed and step size. Despite all the efforts, a new tip eventually lost its sharpness with imaging.

It was not possible to obtain high-resolution images of G-wires or resolve the minor and major grooves of duplex DNA with either ScanAsyst™ or Tapping mode using the Multimode 8 AFM. The reason was likely the inability to use sufficiently low imaging force due to high environmental noise. The AFM being completely exposed, was also subject to never-ending thermal drifting. With the same protocols developed with the Multimode 8 AFM, it was possible to resolve the grooves of duplex DNA with Cypher ES AFM (from Asylum Research, Oxford Instruments). However, since ScanAsyst™ is exclusive to Bruker AFMs, AM-AC (tapping) mode had to be used with Cypher ES.

For statistical analysis and increasing the chances of obtaining the best high-resolution image, several images were captured from different regions by using an automatic routine that was setup with the help of “Macro builder” in Asylum research’s AFM software (Figure 3.16). Thus, it was possible to capture several tens of high-resolution images, but it was quite tedious to analyze them all in search for the best well-resolved molecules. One such molecule has been shown in Figure 3.17 where the minor and major grooves of duplex B-DNA are clearly observed. The minor groove widths are 1.5 ± 0.2 nm, while the major groove widths are 2.0 ± 0.2 nm. The molecule reveals a periodicity of 3.4 ± 0.2 nm, which is exactly as expected and similar to that reported previously¹⁰⁸ using AFM.

Figure 3.16 Programmed setup for unattended imaging

AFM imaging was automated by selecting suitable routines from the MacroBuilder palette and by defining the function to adjust the x and y offset values after every scan. The number of images to be captured could be controlled by the iteration number in “For” loop.

Figure 3.17 *Resolving the minor and major grooves of double helix*

The AFM topography image shows a section of a 320-bp duplex B-DNA that was obtained by PCR amplification. Well-resolved periodic features can be clearly seen across the entire length of the molecule. The minor groove widths are 1.5 ± 0.2 nm, while the major groove widths are 2.0 ± 0.2 nm. The image was captured in AM-mode with FastScan-D probe at imaging speed of 10,000 pixels/s and using cantilever amplitudes of less than 1 nm.

3.9. Conclusions and future work

We find that AFM imaging of G-wires in aqueous solution offer unanticipated challenges which were overcome with the development of new strategies for sample preparation. We could not obtain high-resolution images of either duplex DNA or G-wires using Multimode 8 AFM, but could achieve that very easily with the Cypher ES using photothermal cantilever excitation. Tapping mode was found to be very beneficial for high-resolution imaging and high-speed imaging, while ScanAsyst was found to be very user-friendly and preferred choice for imaging in air. Use of ultra-short cantilevers is extremely beneficial for imaging weakly immobilized samples in liquid. The sample preparation protocols for imaging in aqueous solution as discussed in this chapter has been summarized in the following table:

Surface/Substrate	Anchor	Sample	Immobilization quality
Cleaved Mica	Mg ⁺⁺	Duplex B-DNA	Decent
Cleaved Mica	Mg ⁺⁺	G-wires	Very poor
Rinsed Mica	Mg ⁺⁺	Duplex B-DNA	Good
Rinsed Mica	Mg ⁺⁺	G-wires	Poor
Cleaved Mica	Ni ⁺⁺	Duplex B-DNA	Good
Cleaved Mica	Ni ⁺⁺	G-wires	Poor
Rinsed Mica	Ni ⁺⁺	Duplex B-DNA	Excellent
Rinsed Mica	Ni ⁺⁺	G-wires	Good
BPL Mica	-	Duplex B-DNA	Good
BPL Mica	-	G-wires	Decent
Rinsed Mica	Spermidine	Duplex B-DNA	Good
Rinsed Mica	Spermidine	G-wires	Decent
Rinsed Mica	Glutaraldehyde (GA)	Duplex B-DNA	None
Rinsed Mica	Glutaraldehyde (GA)	G-wires	None
Rinsed Mica	Spermidine + GA	Duplex B-DNA	Good
Rinsed Mica	Spermidine + GA	G-wires	Good

From the above table, it can be seen that only two protocols allowed for good attachment of G-wires on Mica. However, it was only with Ni⁺⁺ that intricate periodic features of G-wires were resolved. This could be due to slightly smaller cationic radius and hydration enthalpy of Ni⁺⁺ compared to Mg⁺⁺. Excellent immobilization of duplex DNA is achievable with Ni⁺⁺, with which it is possible to resolve its intricate structural details even using standard cantilevers oscillated mechanically. However, the quality of immobilization required can vary with the experimental needs, and thus any of the protocols mentioned above could be useful. For example, with our protein-DNA experiments with histones, TRF1, TRF2 and TRF2/Rap1, it was observed that <1 mM MgCl₂ was greatly beneficial. At higher divalent ion concentrations, fixation of the protein-DNA complex using Glutaraldehyde was of utmost necessity, otherwise we could mainly see free DNA on the Mica surface and some isolated proteins that were mobile.

It should be noted that the above listed imaging protocols in aqueous solution are physiologically relevant to limited extents. The protocols that call for the need

of tighter immobilization, make the environment less physiologically relevant, especially due to significantly low concentration of monovalent ions. This is primarily because monovalent cations compete with divalent and trivalent cations to bind with negatively charged DNA and Mica, and thus reduce the anchoring sites. The concentration of divalent ions and spermidine is however within the range of physiological relevance. Using BPL mica, it is however possible to use 10 – 50 times higher concentration of K^+ than possible using other immobilization protocols. This is possible because BPL-mica surface is positively charged and thus do not provide nucleation sites for monovalent cations, but instead repel them and maintain a very clean background for prolonged periods. It has thus been the preferred choice for imaging TERRA which demanded at least 5-10 mM K^+ for G-quadruplex formation (section 7.4, 7.5).

Chapter 4: DNA nanomechanical properties

4.1. Introduction

Atomic Force Microscopy provides a unique opportunity to measure mechanical properties of nanomaterials. Conventionally, the stiffness or elastic modulus of a nanomaterial is determined from force-distance curves, in which the deflection of AFM cantilever is recorded as the tip deforms/indents a sample. This method works really well as long as the deformation is large enough but not comparable to the height of the material, in which case the AFM tip starts to feel the effect of the substrate beneath the sample. DNA is extremely thin (1.5 – 3 nm) and so it is very likely that measurement of its stiffness/elastic modulus will be largely inaccurate due to proximity of the AFM tip to the substrate. Thus, the widely used single-point force curve approach^{109, 110} is not optimal for measurement of DNA deformation mainly because of the inaccuracy in measurement of deformed height.

A way around the single-point force curve approach is to measure of the average height of DNA molecules by imaging with a user-defined force. By changing the imaging force, small changes in the average height of DNA molecules could be statistically analyzed. Since the median height from several points on the sample is considered, the error in height difference is less compared to measurement from a single point as in traditional indentation measurements. This method can be used to probe both the elastic and plastic properties of a sample. It is a very straightforward method which is surprisingly rarely mentioned in literature.

With the aforementioned method, the lateral resolution and speed is very limited. When a sample is deformed by the tip, the tip-sample contact area is much more compared to what is typical in simple imaging. The large contact area and deformation reduces the imaging resolution compared to standard AFM imaging. The imaging speed has to be low enough so that the sample gets enough time to respond to the deformation force in a single tap-cycle, and yield the expected behavior. Owing to viscoelastic nature of most soft samples, the elasticity of the material is a function of the time over which the sample experiences the deformation force. When the AFM tip hits the sample with high speed, the sample behaves like a stiffer material compared to what it would have behaved if it was subject to very slow deformation. Thus, for most biological materials, it makes more sense to characterize their viscoelasticity rather than elasticity which does not account for their time varying response to mechanical deformation.

In the past decade, a significant amount of foundational research has been pursued on the higher harmonics of resonating AFM cantilevers. The bending of a rectangular cantilever beam (homogenous), under the action of external forces at its free end can be explained by the Euler-Bernoulli equation¹¹¹:

$$EI \frac{\partial}{\partial x^4} [w(x, t) + a_1 \frac{\partial w(x, t)}{\partial t}] + a_0 \frac{\partial w(x, t)}{\partial t} + \rho b h \frac{\partial^2 w(x, t)}{\partial t^2} = \delta(x - L) [F_{ts} + \sum_i \frac{k_i}{Q_i} A_i \cos \omega_i t]$$

Where E is the elastic modulus, I is the area moment of inertia, a_1 is the internal damping coefficient, a_0 is the hydrodynamic damping due to surrounding fluid. L, b, w and ρ are the length, width, thickness and mass density of the cantilever respectively. $w(x, t)$ is the vertical displacement of the differential element of the beam at a distance x from the fixed end of the cantilever. F_{ts} is the tip-sample interaction force. The last term represents the cantilever excitation force; where A_i , ω_i , k_i and Q_i are the amplitude, resonant frequency, force constant and quality factor of the i^{th} harmonic of the cantilever.

It is important to note from the above equation that the cantilever dynamics is related not only to its physical attributes, but also to the tip-sample interaction forces and mode of cantilever excitation. This can be understood from the contrast observed in higher harmonic amplitudes over purple membrane vs. Mica (Figure 4.1). The best contrast in cantilever amplitude was observed for the 9th harmonic, while almost no contrast was observed for 1st, 2nd and 4th harmonics.¹¹²

Figure 4.1 *Cantilever amplitude over Mica vs. Purple membrane*

Higher harmonic imaging of Purple membrane on Mica using amplitude modulation reveals that the cantilever amplitude could vary with sample for higher harmonics. Figure adapted from Xin Xu et. al, PRL, 102, 060801 (2009).

AM-FM mode was used for high-resolution mechanical characterization of G-quadruplexes, duplex DNA and protein-DNA complexes. In AM-FM, the cantilever is simultaneously driven with two resonant frequencies corresponding to the 1st and 2nd harmonics of the cantilever. Amplitude modulation of the 1st harmonic is used for amplitude feedback as in Tapping/AM mode, while frequency modulation of the 2nd harmonic is used for frequency feedback. A stiffer sample shifts the 2nd harmonic oscillation frequency higher compared to a softer sample. The frequency shift is thus directly related to the stiffness/elastic modulus of the sample, while amplitude shift of the first harmonic is directly related to the topography and energy dissipated/stored by the sample.

Apart from the fluctuations in 2nd harmonic resonance, another parameter that can be used to characterize the viscoelasticity of a material is called the loss tangent which is the ratio of energy dissipated to energy stored in a tip-sample interaction cycle. If a sample is more viscoelastic, it dissipates less energy. The loss tangent can be related to the phase (φ_1) of the first harmonic of the cantilever using the relation:¹¹³

$$\text{Tan } \delta = \frac{\sin \varphi_1 - \sum_i i^2 \frac{A_i^2}{A_1 A_{\text{free}}}}{\cos \varphi_1 + Q \sum_i \frac{A_i^2}{A_1 A_{\text{free}}} (i^2 - 1)}$$

Where Tan δ is the loss tangent and can vary from $-\infty$ to $+\infty$, but usually lies in the range of 0 to 10 for soft matter. Q is the quality factor of the first harmonic of the cantilever. A_{free} is the free cantilever amplitude while A_i is the amplitude (A_i) for the i^{th} harmonic of the cantilever. A_1 is the first harmonic amplitude used for topography feedback. Our measurements of loss tangent was limited to just the first harmonic of the cantilever, so the above equation reduced to:

$$\text{Tan } \delta = \frac{\sin \varphi_1 - \frac{A_1}{A_{\text{free}}}}{\cos \varphi_1}$$

The loss tangent makes physical sense only when imaging is done in repulsive mode ($\varphi_1 < 90^\circ$), which is practically very challenging. At $\varphi_1 = 90^\circ$, loss tangent is infinite, and at $\varphi_1 > 90^\circ$, loss tangent is negative, both of which are absurd as the ratio of stored and dissipated energies can't be negative or infinite.

It is possible to characterize the material property of long DNA molecules based on their axial rigidity, by measurement of persistence length. However, it requires the molecule to freely equilibrate on the mica surface which cannot be assumed for granted. If the molecules are kinetically trapped on the surface, as was the case for our experiments in solution, the extent of kinetic trapping must be known. The issue of unknown degree of kinetic trapping of DNA molecules is further worsened by the over-simplicity of the worm-like-chain model. Significant theoretical work need to be done to use a proper model to correlate the bending of

DNA to its persistence length, with proper considerations for torsion and twisted nature of DNA. DNA is very far from an isotropic rigid rod at 0 K. Our preliminary results for persistence length indicated that for short duplex DNA (320 bp) the persistence length is nearly half of the expected value of ~50 nm, while for G-wires it could not be determined due to smaller length of G-wires observed, but it is likely to be more than 200 nm.

4.2. Variation of attractive force

The force felt by a point on the sample is the sum of attractive forces plus repulsive forces. Using ScanAsyst, the repulsive force can be controlled by specifying a parameter called PeakForce setpoint. The attractive forces however cannot be specified by the user, and is thus a variable. In order to use the method based on studying the average height of molecules, it is of primordial importance to understand what parameters can affect the attractive force between the tip and sample. It was found that the attractive force in air depends on the repulsive force, tip-sample interaction time, nature of the sample and the tip-sample contact area. The repulsive force and tip-sample interaction time can be controlled by the user. However, the tip-sample contact area cannot be controlled as it depends on several other factors, including the attractive force. For instance, the tip-sample contact area would be different when the tip is right in the middle of the cross-section of DNA and when the tip is barely touching an edge of the DNA. Also, the attractive force on DNA is less compared to that on Mica. The best solution to overcoming the issue of variable attractive force is to perform the experiments in solution.

Figure 4.2 *Attractive forces in air vs. solution*

Force curves were obtained in air and water over DNA and Mica. The x-axis shows the distance between the surface and AFM tip determined from the voltage (V_z) applied to the piezoelectric sample stage, while the y-axis shows the interaction force between the AFM tip and surface determined from the photodiode voltage (V_P) that registers cantilever deflection.

From the force curves (Figure 4.2), it can be seen that the attractive forces in water are significantly small and are nearly independent of the surface. The waviness observed in the force curve obtained in solution, is likely due to different hydration layers of water, each of which present certain barrier for the AFM tip to penetrate and reach the surface. The thickness of hydration layers on Mica and DNA are slightly different.

4.3. Variable force imaging of G-wires in air

In ScanAsyst mode, the imaging force can be controlled by specifying a parameter called PeakForce setpoint that controls the maximum bending of the cantilever before the tip starts to retract from the sample. Thus, it is possible to

capture topography images at different imaging forces, which could give different height distribution of the molecules depending on their deformation.

From Figure 4.3 it can be seen that the height distribution of five G-wires change with the PeakForce setpoint. A Gaussian fit was applied to the histogram to estimate the deviation of the height distribution from a Gaussian nature. At very low setpoints of -50 pN to 250 pN, the height distribution is nearly the same. From 250 pN onwards and up to 2 nN, the average height of the G-wires reduced with increasing PeakForce setpoint. In addition to the decrease in average height, the increase in standard deviation of height should also be noted. Some parts of the G-wires are deformed more compared to some other parts, which is not purely due to the inhomogeneity of the molecule itself, but also partly due to the variations in the attractive force experienced by different parts of the molecule.

Figure 4.3 *Height variation with variable imaging force*

The height histogram was computed for just the spline of the G-wires, from each AFM image. The number in blue shows the fitted Gaussian mean, while the number at the top right corner of the histogram shows the average height and standard deviation. The imaging force can be obtained by adding ~ 1 nN to the PeakForce setpoint. The length scale (white bar) is 100 nm. The height color scale is 0 – 3 nm.

By selectively imaging a certain G-wire with increasing imaging force, it was observed that with PeakForce setpoints of over ~ 8 nN, they underwent plastic deformation, i.e., they did not retain their crushed height even after reducing the imaging force (Figure 4.4).

Figure 4.4 *Selective plastic deformation of G-wires in air*

A single G-wire (shown inside dashed rectangle) was selectively deformed gradually until it underwent plastic deformation. The PeakForce setpoint was then reduced to ~ 0.1 nN to image a larger area that showed clear contrast between the height of non-deformed G-wires and the selectively deformed G-wire. The same process was repeated for another neighboring G-wire.

The observed plastic behavior is not expected of a typical viscoelastic material, unless some significant changes in the structure of matter are involved. Assuming that G-wires are composed of stacked G-tetrads, the permanent collapse of G-wires probably suggest that the G-tetrads were forcefully dislocated by the AFM tip to such an extent that they were flattened and could not re-stack side-by-side even long after the loading force was reduced.

However, when a large area was chosen to collectively deform several G-wires simultaneously, it was observed that different G-wires responded differently to the deformation force (Figure 4.5). By increasing the imaging force beyond 2 nN, G-wires collapsed to either 2.4 nm height or 0.8 nm height. However, not all G-wires collapse simultaneously. In fact, even different parts of the same G-wire do not collapse simultaneously. This could be due to variability in the viscoelasticity of different G-wires or different parts of the same G-wire.

Figure 4.5 *Collapsing G-wires in air*

By increasing the imaging force beyond 2 nN, G-wires collapse to either 2.4 nm height or 0.8 nm height. Height histograms were determined from the splines of detected molecules from the above image. The number at the top left corner shows the mean height and standard deviation, while the number (in blue) shows the Gaussian mean.

4.4. Deformation of G-wires in solution

G-wires of d[G₄T₂G₄] were successfully immobilized on Mica using NiCl₂ and imaged with ScanAsyst in solution using standard cantilevers (MSNL-E, MSNL-F and ScanAsyst Fluid+). However, the range of PeakForce setpoint that could be used for variable force imaging with ScanAsyst was much less than what was possible in air. At imaging forces of over ~400 pN, some G-wires would disintegrate and fly away, thus making it nearly impossible to image with forces high enough to plastically deform G-wires in solution. By repeated imaging even at low imaging forces of ~150 pN, some G-wires or their parts would fly away. Thus, it was not possible to investigate the mechanical properties of G-wires in the same way as discussed in the previous section.

By imaging the same region twice with ~160 pN and ~400 pN, it was observed that the average height of G-wires reduced by ~0.1 nm (Figure 4.6). Such small decrease in height is probably less significant compared to the decrease in pixel count by ~30%. From our imaging experiments in solution, we failed to statistically analyze the mechanical properties of G-wires, but gained valuable insight on the nature of G-wire assembly. The ease with which G-wires could be disintegrated in solution, hinted towards modular assembly of G-quadruplexes blocks in a Lego fashion to form a long wire. In solution, it was much easier to break a G-wire into pieces compared to deforming the entire G-wire.

Figure 4.6 *Variable force imaging in solution*

(A) The same region containing G-wires was imaged in solution with ScanAsyst at 162 pN and 404 pN. The length scale (white bar) is 100 nm, while the height scale is 3 nm. (B) Mathematica output for the above figure showing the objects selected for analysis and their splines detected (red curves). (C) Histogram of height measured from detected splines in the above image. The number at the top left corner shows the mean height and standard deviation, while the number (in blue) shows the Gaussian mean.

4.5. High-resolution nanomechanical mapping in solution

The Cypher ES AFM can use two feedback systems (lock-in amplifiers) to track two cantilever parameters simultaneously. AM-FM mode was used for determining dissipation energy, stiffness and loss tangent of DNA. The resonant frequencies (f_1 , f_2) and quality factors of the 1st and 2nd harmonic of the cantilever was determined from the thermal noise spectra. Prior to imaging, the deflection sensitivity and force constant (k_1) of the cantilever was measured. The force constant of the 2nd harmonic is automatically determined by the software using the relation: $k_2 = k_1 * (f_1/f_2)$.

The free amplitude for 2nd harmonic was set to 25 mV, while for the 1st harmonic it was initially set to 250 mV. During imaging, the amplitude setpoint of the 1st harmonic was gradually reduced until intricate surface features on DNA showed up. The 2nd harmonic amplitude was left untouched, and it was ensured that it never reduced to less than 20 mV. It was also ensured that the 1st harmonic amplitude is much higher than the 2nd harmonic amplitude. The imaging parameters were suitably chosen so as to image in repulsive mode (phase <90°). Initially, no features were observed in the mechanical property channels, but as the amplitude and frequency gain for 2nd harmonic were increased to several tens of thousands, features started to show up. Imaging parameters had to be very carefully and painstakingly optimized to obtain decent contrast in the mechanical property channels.

Duplex DNA was chosen for benchmark and practice prior to experiments with G-quadruplexes. The helical pitch of B-DNA was observed in the height, phase, amplitude, dissipation energy, loss tangent and stiffness images (Figure 4.7). A small straight section on the duplex DNA was arbitrarily chosen to trace the profile and correlate with mechanical properties. The width of the minor and major grooves is 1.6 and 1.9 nm respectively. The height peaks correspond directly to dissipated energy, loss tangent and stiffness. In few cases, the maxima/minima in height did not coincide exactly with the maxima/minima of mechanical properties.

Figure 4.7 Mechanical mapping of minor and major grooves of duplex DNA

Resolving the minor and major grooves of supercoiled plasmid DNA in height, dissipation energy, loss tangent and stiffness channels simultaneously. The image was obtained in AM-FM mode using FastScan-D. The length scale (white bar) is 30 nm. A small section was chosen in the image and the profile was obtained from all channels to correlate different mechanical properties and topography.

G-wires of d[G₄T₂G₄] were subsequently imaged with AM-FM. Initially, in the height channel, no periodic surface features were observed on the G-wires. But surprisingly, in the phase channel a surface feature with 4 nm periodicity was observed. Initially, it was thought to be a periodic artifact, but eventually the same periodicity could be observed in both topography and mechanical property channels, and that too for G-wires oriented along different directions (Figure 4.8).

Figure 4.8 *Height vs. Loss tangent for G-wires*

One of the first AFM images of G-wires reproducibly showing periodic pattern in both height and loss tangent. The length scale (white bar) is 10 nm.

Profile analysis of a selected section on the G-wire was done for different channels corresponding to the topography and mechanical properties (Figure 4.9). It was observed that the peaks in height correlated directly with peaks in loss tangent and dissipation energy. In each of the three channels, a periodic pattern of 4 nm was observed. The 2nd harmonic frequency channel, which directly related to stiffness, also revealed periodic features, but not with well-defined 4 nm periodicity. The additional peaks might be indicative of additional structural details with variable stiffness, but unrelated height and loss tangent.

Figure 4.9 *4 nm periodicity in topography and mechanical properties*

A small section of a G-wire was chosen from the above image and the profile was obtained from all channels to correlate different mechanical properties and topography. The dotted lines are all separated by 4 nm and correlate directly with peaks in height, loss tangent and dissipation energy. For the frequency and stiffness channel, the pattern is a bit more complicated, but still there is an underlying 4 nm pattern. The individual peaks in height was found to split into two or three peaks in frequency/stiffness.

4.6. Discussions and future work

From the AM-FM experiments in aqueous solution, it is evident that G-wires have higher value of loss tangent and stiffness compared to duplex B-DNA. Even though the value of dissipated energy reported here is not an absolute parameter that can be compared, higher value of loss tangent for G-wires indicate that they dissipate more energy per tip-sample interaction cycle, which is typically expected of a stiffer material. Thus, it can be concluded that G-wires are stiffer than duplex B-DNA. This is also intuitive since G-quadruplexes are composed of tetrads instead of base-pairs, and thus have twice the mass density compared to duplex DNA of the same length.

The indentation experiments in solution using ScanAsyst mode sheds light on a very interesting fact that G-wires are modular and could be easily cut into pieces with the slightest indentation force. However, further work needs to be done to quantify exactly how much vertical and lateral force exerted by the AFM tip is causing the disassociation.

The indentation experiments in air using ScanAsyst mode reveal that DNA G-wires are extremely resistant to deformation, however they collapse suddenly and permanently beyond a certain threshold of ~ 8 nN. This sheds light on the irreversibility of G-tetrad stacking, albeit not in solution. Performing plastic deformation experiments in aqueous solution would be extremely challenging as the G-wires would rather disassociate and fly off than be crushed.

In future, a proper theoretical model for quantification of persistence length of DNA could be developed. Otherwise, in order to measure the axial rigidity of DNA, direct measurement of DNA bending by nano-manipulation is a more viable choice, if the ends could be clamped. The Cypher ES software interface allows for programming in Igor Pro required to precisely move the AFM tip on the surface, while recording the lateral forces felt by the AFM cantilever. Porous surfaces are also available commercially which can be used as a substrate instead of Mica. Simple indentation experiments on DNA over a pit of the porous surface can be used to quantify its axial rigidity.

Chapter 5: Solution AFM structure of DNA G-wires

5.1. Introduction

With structural insights gained from the 4-nm periodicity in topography and mechanical properties of G-wires, we dug deeper into their structure by combining high-resolution AFM with ensemble techniques and molecular simulations. High-resolution topography images of hundreds of G-wires formed by d[G₄T₂G₄] were analyzed across different sample batches. The statistics was necessary for hypothesizing anything about a completely unknown structure.

Our objective was to build a structure from bottom-up using molecular modelling and simulation with insights from NMR and CD spectroscopy. The generated structure(s) could then be simulated to look like an AFM image for direct comparison with real AFM images.

5.2. Ensemble characterization of G-wires

Using 1D proton NMR spectroscopy (Figure 5.1), it was found that most of the structures formed by d[G₄T₂G₄] in 70 mM KCl, 30 mM KPi and 10 mM MgCl₂ were G-quadruplexes. The observed proton peaks in the region 10 – 12.5 ppm are likely due to imino (H1) protons of Guanines of G-tetrads. The aromatic (H6) protons of Thymines show up in the region 7 – 8 ppm, while the amino (H21, H22) protons of Guanines show up at around 9 ppm (Figure 2.8). Upon annealing the

sample (fast heating and slow cooling), the sharp imino proton peaks merged into a broad peak, which indicated the formation of G-quadruplexes that were either highly polymorphic or higher-order.

Figure 5.1 Proton NMR spectra of d[GGGGTTGGGG]

NMR Spectra was acquired using 600 MHz NMR system at 25°C both before and after annealing in 70 mM KCl, 30 mM KPi and 10 mM MgCl₂.

Using circular Dichroism (CD) spectroscopy, maxima around 260 nm and minima around 240 nm was observed (Figure 5.2) both before and after annealing of d[G₄T₂G₄] in 70 mM KCl, 30 mM KPi and 10 mM MgCl₂. Usually this kind of CD spectra for G-quadruplexes indicate parallel topology, i.e., unique strand directionality. This could suggest that it is unlikely for a single strand of d[GGGGTTGGGG] to form a G-quadruplex, as then the strand has to turn in the opposite direction to form a 2-layer G-Quadruplex with its 8 Guanine residues. Though there could be several possibilities of parallel strand arrangement for inter-molecular G-quadruplexes, we prioritized the assumption that four parallel strands of d[G₄T₂G₄] form an intermolecular parallel G-quadruplex.

By varying the temperature while obtaining CD spectra, it was observed that intensity of the positive peak at 260 nm barely reduced by 30% at 95°C (Figure 5.3). This indicated that the parallel G-quadruplexes are extremely stable, which is quite uncommon for G-quadruplexes with 2 – 4 layers of G-tetrads. Thus, it seemed that the 260-nm peak was due to higher-order parallel G-quadruplexes.

Figure 5.2 CD spectra of $d[G_4T_2G_4]$

Molar ellipticity as a function of wavelength of circularly polarized light at 25°C before and after annealing in 70 mM KCl, 30 mM KPi and 10 mM $MgCl_2$.

Figure 5.3 Decreased ellipticity at high temperature

Molar ellipticity as a function of wavelength of circularly polarized light at 25°C and 95°C while heating $d[G_4T_2G_4]$ in 70 mM KCl, 30 mM KPi and 10 mM $MgCl_2$.

5.3. Role of annealing

The NMR and CD observations indicated that annealing of d[G₄T₂G₄] in physiologically relevant conditions led to the formation of higher-order parallel DNA G-quadruplexes. AFM images obtained in air for d[G₄T₂G₄] revealed that after annealing, long wires are formed (Figure 5.4).

Figure 5.4 d[G₄T₂G₄] imaged before and after annealing

Topography images of d[G₄T₂G₄] before and after annealing, acquired with ScanAsyst Mode in air using ScanAsyst Fluid+ probe.

The height of G-wires formed by d[G₄T₂G₄] after annealing is 2.33 ± 0.14 nm ($n=1085$ molecules) on mica in air (Figure 5.5), which agrees with the value reported earlier by Tapping mode imaging in air on Mica¹¹⁴. The observed height prior to annealing is 2.24 ± 0.18 nm ($n=798$ molecules), which is indicative of either short G-quadruplexes that are lying axially on mica or 8-layered G-quadruplexes that are sitting upright on Mica.

Figure 5.5 Height of $d[G_4T_2G_4]$ before and after annealing

Histogram obtained from the median height of molecules in the topography image in air.

As can be clearly seen from the AFM image (Figure 5.4) and the length histogram (Figure 5.6), G-wires are primarily formed after annealing. Prior to annealing, a few short wires however exist, indicating that wire formation is an inherent tendency of $d[G_4T_2G_4]$, but only a very small fraction of molecules can overcome the energy barrier needed to form G-wires. Heating allows the molecules to overcome the energy barrier, and merge together into long wires upon cooling.

Figure 5.6 Length histogram of $d[G_4T_2G_4]$ before and after annealing

Histogram of approximate length estimate of molecules imaged in air.

5.4. Height of G-wires in solution

It is important to know the width of G-wires in fully hydrated and undeformed state. However, since the height values measured from AFM topography images can vary with imaging parameters, a 320-bp duplex DNA was imaged with G-wires as standard reference. The duplex DNA sample (in DI water) was added to the d[G₄T₂G₄] solution just before immobilization on the mica substrate. The topography images obtained in aqueous solution containing 2-3 mM NiCl₂, revealed duplex DNA, isolated G-quadruplex blobs and G-wires (Figure 5.7, Figure 5.8). Duplex DNA were identified as long flexible molecules of uniform length of 109 ± 4 nm ($n=7$ molecules) with periodical groove-like features resembling minor and major grooves of duplex DNA as reported in previous AFM studies^{108, 115}.

21 AFM images were analyzed containing both G-wires and duplex B-DNA in the same frame. One such example has been shown in Figure 5.7. The height histogram of the molecules was extracted after thinning them to a contour (spline) of single-pixel width (Figure 5.8). This ensured that the set of pixel values obtained from a G-wire would not overlap with those of duplex DNA. Otherwise, the lower edges of G-wires would give the same height values as that obtained from the central spline of duplex DNA, and then it would not be possible to distinguish from which molecule the height values in the histogram originally came from (Figure 5.9).

Figure 5.7 *Height of duplex DNA vs. G-wires*

(A) High-resolution topography image of G-wires and duplex DNA showing G-wires of two different periodic features highlighted with a box and dashed rectangle. The G-wire highlighted in dashed rectangle is rarely observed and only two such molecules can be seen in the whole image. Duplex B-DNA has been indicated by asterisk (*), and four such molecules can be seen. Objects with length < 7 nm and height > 1.5 nm were identified as isolated G-quadruplex blobs and highlighted with a circle. The length scale shown at the bottom (white bar) is 50 nm and the coloring of the image corresponds to the height scale shown on the right. (B) Histogram showing the height distribution of all pixels in the AFM image. (C) Height histogram obtained after skeletonization (thinning) of all molecules in the image. The histogram was fitted with a double Gaussian, which shows two distinct peaks at 2.1 nm and 3 nm.

Figure 5.8 Molecules and their splines from which height was analyzed

High-resolution topography images of G-wires and duplex DNA used for determination of height distribution of the molecules. The central spline of the molecules was detected after removing the background using pixel count threshold of 1000 and Otsu's intensity threshold.

Figure 5.9 Height histogram of images with duplex and G-quadruplex DNA

Figure 5.10 Height histogram from splines of duplex and G-quadruplex DNA

The values shown in blue are the gaussian fitted mean and standard deviation respectively.

In all 21 height histograms (Figure 5.10) obtained from the thinned contours (splines), two distinct peaks were observed, which supposedly corresponded to duplex DNA and G-quadruplexes. The values obtained from the histograms by fitting a double gaussian are summarized below:

Gaussian mean for Peak 1 (in nm) = 2.2, 2.18, 2.07, 2.06, 2.18, 2.18, 2.05, 2.11, 2.04, 2.07, 2.08, 2.18, 2.09, 2.09, 2.05, 2.16, 2.01, 1.99, 2.15, 2.03, 2.13

Gaussian mean Peak 2 (in nm) = 3.07, 3.06, 2.97, 2.99, 3.07, 3.03, 2.98, 3.04, 2.98, 2.98, 3.02, 3.03, 3., 3.01, 2.93, 3.02, 2.9, 2.87, 3.04, 2.96, 2.97

Gaussian S.D for peak 1 (in nm) = 0.28, 0.32, 0.33, 0.36, 0.43, 0.39, 0.35, 0.37, 0.27, 0.27, 0.36, 0.47, 0.37, 0.38, 0.4, 0.46, 0.33, 0.37, 0.38, 0.35, 0.33

Gaussian S.D for peak 2 (in nm) = 0.26, 0.25, 0.26, 0.25, 0.28, 0.25, 0.25, 0.24, 0.28, 0.28, 0.27, 0.25, 0.26, 0.25, 0.28, 0.28, 0.29, 0.26, 0.24, 0.26, 0.25

Height difference between peaks (in nm) = 0.87, 0.88, 0.9, 0.93, 0.89, 0.85, 0.93, 0.93, 0.94, 0.91, 0.94, 0.85, 0.91, 0.92, 0.88, 0.86, 0.89, 0.88, 0.89, 0.93, 0.84

Maximum error for height differences between peaks (in nm) = 0.54, 0.57, 0.59, 0.61, 0.71, 0.64, 0.6, 0.61, 0.55, 0.55, 0.63, 0.72, 0.63, 0.63, 0.68, 0.74, 0.62, 0.63, 0.62, 0.61, 0.58

Height of duplex DNA from 21 images = 2.1 ± 0.063 nm

Height of G-quadruplex DNA from 21 images = 2.996 ± 0.053 nm

Height difference between peaks from 21 images = 0.896 ± 0.031 nm

From the values reported above, it is clear that there is negligible variation in the mean height of duplex DNA and G-wires in the 21 images obtained consecutively with similar imaging parameters. The standard deviation of the gaussian fit in the histograms represent the real variation in height of molecules, rather than the uncertainty in measurement of their height.

5.5. High-resolution imaging of G-wires

With the same protocol and AFM probe as used for solution imaging in AM-FM mode (as in the previous chapter), high-resolution images of G-wires could be reproducibly obtained for nearly 100% of the molecules in the imaging frame (Figure 5.11). An advantage in AM mode compared to AM-FM mode is that there is no lower limit to the cantilever amplitude in AM mode, which otherwise needs to be maintained higher than the 2nd harmonic amplitude in AM-FM mode. Imaging with ultra-small amplitudes is challenging, but worth the effort.

Figure 5.11 Reproducible high-resolution imaging of G-wires in solution

High-resolution images obtained from two different regions on the mica surface containing G-wires formed by d[G₄T₂G₄]. Ridge-like undulating periodic features can be seen on almost every G-wire.

It wasn't quite sufficient to image G-wires alone, as it is an unknown structure showing unexpectedly new results. So duplex B-DNA was used as a reference to image alongside G-wires. We found that it is relatively easier to obtain high-resolution images of duplex DNA. The minor and major grooves of duplex DNA could be resolved almost along the entire length of the molecule allowing estimation of the number of base-pairs from the helical pitch (Figure 5.12). To the best of our knowledge, such high-resolution images of well isolated short linear duplex DNA have not been reported in literature. Usually, it is preferable to use longer DNA molecules, or at high concentration to limit its freedom of movement, and thus achieve high-resolution.

Figure 5.12 *Grooves of duplex DNA resolved throughout its entire length*

(A) AFM image containing both G-wires and duplex DNA. A section of duplex DNA in the highlighted rectangle was selected for periodicity analysis. (B) Duplex DNA and its spline highlighted in red. (C) Height profile along the highlighted spline with peaks identified. Distance between consecutive peaks are shown below the profile.

5.6. Periodical features on G-wires

From several high-resolution AFM images (4+28+42) obtained on 3 different days from 3 different sample batches, the pattern of the periodical surface features was visually inspected for thousands of G-wires and thoroughly analyzed

for (100+10+12) G-wires, with each image containing 30 – 100 molecules. This process involved careful visual inspection of the AFM images followed by cropping of selected molecules and extraction of their height profiles. In this way, three classes of AFM images of G-wires were identified based on periodicity of the surface features: Type I, Type II and Hybrid type.

The surface features of Type I and Type II G-wires comprised of parallel ridges spaced by 4.2 ± 0.2 nm and 2.1 ± 0.2 nm respectively (Figure 5.13). The periodicity of Type I G-wires is exactly double that of Type II G-wires, which initially led to the doubt that 4 nm periodicity was the low-resolution version of 2 nm periodicity. But when we built molecular models and found a unique structure with 4.2 nm periodicity, we realized that it was a different class of molecules altogether. Another factor that helped overcome the suspicion was that some of the Type I G-wires revealed a distinct 1 nm pattern in addition to the 4-nm pattern, which indicated that the images had sub-nm resolution (see next section). It should however be noted that Type II G-wires were extremely rare to such an extent that out of thousands of G-wires that were visually inspected, only 15 Type II G-wires were identified.

Figure 5.13 *Type I and Type II G-wires*

(A) Two different types of periodic features observed on G-wires formed by $d[G_4T_2G_4]$. The major class of G-wires are referred to as Type I. It has left-handed periodic features separated by ~ 4.2 nm, (B) which is exactly double of the separation between periodic ridges on Type II G-wires.

The surface features of the Hybrid Type G-wire are a combination of the periodic features of Type I and Type II G-wires. However, only three such rare observations were encountered with. In a Hybrid Type G-wire, four different kinds of periodicities could be observed in the same molecule (Figure 5.14). The observation of Hybrid G-wires is intriguing as it indicates topological similarity and compatibility between different structures in terms of strand orientation and stacking interface that allow the two different structures to combine into one.

Figure 5.14 3 types of periodical features in the same G-wire

(A) AFM images of Hybrid G-wires. The length scale (white bar) is 5 nm. (B) Height profile extracted from the above images and superimposed to show reproducibility of the 3 different periodicities. (C) AFM image of a Hybrid G-wire with 4 nm pattern on either side of the 2-nm pattern. The length scale (white bar) is 10 nm.

5.7. 1 nm sub-features

In certain portions of a few ($n = 50$) Type I G-wires (Figure 5.15), a very interesting sub-feature was observed which had a periodicity of 0.99 ± 0.22 nm (Figure 5.16).

Figure 5.15 *Sub-periodic features in Type I G-wires*

The above 12 images were cropped after careful visual inspection of 10 AFM images each containing over 30 well-resolved Type I G-wires. The length scale (white bar) is 5 nm for all images.

Figure 5.16 Height profiles obtained from selected sections of Type I G-wires

The height profiles were obtained from the sections shown in the previous image, arranged in the same order. The red dots are automatically identified peaks that were used to measure the periodicity of the pattern.

5.8. Handedness of periodic features

It was observed that the periodic features of the G-wires have a well-defined inclination, which was indicative of the handedness of the periodic features. For type I G-wires, the periodic features appeared to be either left-handed or zig-zag (Figure 5.17). While for type II G-wires, the 2 nm-features were mostly left-handed except for just two molecules showing right-handed pattern (Figure 5.18). It is important to note that the periodic feature of duplex DNA reveal a right-handed pattern (Figure 3.17, Figure 4.7) as expected from right-handedness of the double helix.

Figure 5.17 *Handedness of type I G-wires*

Schematic and AFM image representing the class of molecules where the features with 4 nm periodicity are tilted in the positive x-direction (**A**) purely or (**B**) with interconnect.

Figure 5.18 *Handedness of type II G-wires*

Schematic and AFM image representing the class of molecules where the features with 2 nm periodicity are tilted in either purely (**A**) the positive or (**B**) negative x-direction.

We do not emphasize on angle of inclination of the periodic features mainly due to its variability which can be expected from a molecule that is not perfectly straight and tightly immobilized on a surface. The maximum inclination angle for the left-handed features of 4 nm and 2 nm periodicity were determined to be $\sim 55^\circ$, and $\sim 35^\circ$ respectively after correcting for the finite tip dilation (Figure 5.19).

Figure 5.19 *Correcting tip dilation for G-wire*

A hybrid type G-wire was rotated to align it parallel to the horizontal (x-axis). The rotated image was then compressed in the y-direction so that the width of the G-wire became the same as its maximum height. The corrected image revealed inclination angles of 52° and 35° for the 4 nm and 2 nm patterns respectively.

A very simple method for correction of tip-sample dilation was used to determine the real inclination angle of the periodic features. If the actual width of the molecule is known, as is the case with duplex B-DNA, the AFM image could be compressed along the perpendicular direction to the axis of the molecule, so that the new full-width-half-maxima of the cross-sectional height profile reduces to the expected width of 2.4 nm.

Figure 5.20 *Correcting tip dilation for duplex DNA*

A relatively straight section of duplex DNA was rotated to align it parallel to the horizontal (x-axis). The rotated image was then compressed in the y-direction so that its width reduced to the expected value of 2.4 nm. The corrected image revealed an inclination angle of 36° for both the minor and major grooves.

5.9. Statistical analysis of periodicities

To systematically analyze the periodicities of the surface features of hundreds of G-wires, a program was developed using Mathematica. With the AFM software (Gwyddion), it was possible to extract height profiles only by drawing straight lines along a section of the molecule. This severely limited the periodicity analysis as perfectly straight G-wires were usually very small (< 25 nm). Most G-wires had an intrinsic curvature, albeit G-wires were not as flexible as duplex DNA. With the program, the height profile could be obtained from the spline of the entire molecule which could be either straight or curved. Analyzing periodicities from the entire length of the molecule helped to overcome any bias that could have been caused by selective choice of sections on G-wires. The entire process of periodicity analysis has been depicted below in a flow chart (Figure 5.21).

Figure 5.21 Flowchart for periodicity analysis

16-bit grayscale AFM images were stored in a folder, whose address was fed to the program, along with the height and length scale of the images. The program would automatically crop G-wires from the AFM images and store it in an array (Figure 5.22). Each of the cropped images would then be individually analyzed with a program to obtain a spline along the molecule. The molecule was rotated automatically by fitting it crudely to a straight line, so as to align it primarily along the x-axis (Figure 5.23). If the molecule was relatively straight, the spline was fitted with a straight line or polynomial of up to 5th order. If the molecule was too bent or curved to be fitted with a standard polynomial, interpolation was used to determine the fitting polynomial. Once the fitting polynomial was determined, the spline could be translated along the cross-section of the molecule by varying the intercept.

The intensity variation along the spline was then obtained using the code, and was multiplied to the height scale to obtain the height profile. A single spline might not give the most periodic height profile, thus several height profiles were analyzed and compared to find the one with the least standard deviation in periodicity. The periodicity was determined by identifying the peaks in a height profile and measuring the distances between successive peaks. Peak detection was based on 1st and 2nd order derivatives of the height profile Gaussian smoothened with 4-pixel radius (~1 nm). The smoothening was essential to get rid of sudden unexpected sharp peaks in the height profile, and to ignore peaks that were not distinct enough.

Figure 5.22 *Collage of a subset of G-wires cropped from AFM images*

An output of the Mathematica program that automatically cropped G-wires from different AFM images. An index is printed on each of the molecule specifying the serial number of the molecule and the image from which it was cropped. The index is used for selective analysis of the molecule.

Figure 5.23 *Splines of 100 G-wires with least standard deviation in periodicity*

An output of the Mathematica program that made a collage of all 100 molecules that were analyzed. The red curve on each G-wire represent the spline of the molecule.

Figure 5.24 Obtaining periodicity histogram for type I G-wires

(A) Few representative G-wires, across the entire length of which a periodic pattern of ~4nm could be observed. (B) The histogram of all peak-to-peak distances measured from 100 height profiles of G-wires that were sequentially selected without any bias that were captured during a single experimental session.

Figure 5.25 Obtaining periodicity histogram for type II G-wires

(A) Selected Type II G-wires that were searched across thousands of G-wire images captured from different experimental sessions and sample batches. (B) The histogram of all peak-to-peak distances.

5.10. Building structural models

Molecular models of short G-wires of ~5 nm length were built with XPLOR-NIH using standard dihedral constraints for Guanines and Thymines. Artificial constraints like planarity of G-tetrads and distance between carbon atoms of consecutive G-tetrads were enforced for the structure to look like a typical parallel interlocked G-quadruplex as per our design. In order to develop these constraints, it was important to know which residue follows which one, and which residue is adjacent to which one. For this, the residue arrangement needed to be pre-built in mind and a schematic was drawn for each of the imagined structures.

One of the first structure built was the one proposed by Marsh et. al. in 1993, which we refer to as the 0° rotamer of the (2,2) adjacent slip model (Figure 5.26). This is one of the simplest structures to imagine, in which half of two adjacent DNA strands pair with half of another two adjacent DNA strands. The 4 strands run in parallel always leaving two hanging strands on both ends for more adjacent strands to pair up. Since, the DNA strand of d[GGGGTTGGGG] contain two G4 tracts separated by two thymines, 4-layered G-quadruplexes form the building blocks of the long wire-like interlocked G-quadruplex.

Due to the flexible Thymine loop connecting two G4 tracts, it is possible that the individual 4-layered G-quadruplex blocks rotate relative to each other. Such rotation may give rise to 3 different rotational conformers referred to as -90° , 0° , and $+90^\circ$ based on their relative rotation from a natural backbone progression in the $5' \rightarrow 3'$ strand direction. So, three variations of the (2,2) adjacent slip model was investigated (Figure 5.26). Rotation by more than $\pm 90^\circ$ was not investigated as the TT linker would be unfavorably stretched.

Figure 5.26 *Guanine arrangement for (2,2) adjacent slip model*

Three possible rotamers that were built using XPLOR with constraints generated from the Guanine arrangement as shown in the schematic (blue dots). The dashed lines correspond to stretched thymine loops. Thymines have not been shown in the schematic for simplicity. The numbers 'x.y' just over the blue dots represent the residue number, where 'x' denotes the strand and 'y' denotes the guanine position on the 'x' strand from 5' end.

In addition to the (2,2) adjacent slip model, we investigated alternative slip configurations namely (2,2) diagonal slip model and (3,1) single slip model. In the former, instead of two adjacent DNA strands, the diagonally opposite strands run in parallel and can allow 3 possible rotamers (Figure 5.27). In the (3,1) single slip model, instead of 2 slipping strands, only one strand forms the slipping connection and it can also exist as 3 rotamers (Figure 5.28). In this configuration, there is only one sticky end at both sides, thus its propensity to form long wires might be less.

Figure 5.27 *Guanine arrangement for (2,2) diagonal slip model*

Three possible rotamers that were built using XPLOR with constraints generated from the Guanine arrangement as shown in the schematic (blue dots). The dashed lines correspond to stretched thymine loops. Thymines have not been shown in the schematic for simplicity. The numbers 'x.y' just over the blue dots represent the residue number, where 'x' denotes the strand and 'y' denotes the guanine position on the 'x' strand from 5' end.

Figure 5.28 *Guanine arrangement for (3,1) single slip model*

Three possible rotamers that were built using XPLOR with constraints generated from the Guanine arrangement as shown in the schematic (blue dots). The dashed lines correspond to stretched thymine loops. Thymines have not been shown in the schematic for simplicity. The numbers ‘x.y’ just over the blue dots represent the residue number, where ‘x’ denotes the strand and ‘y’ denotes the guanine position on the ‘x’ strand from 5’ end.

Though our investigation has been mainly focused on the 9 structures discussed above, we were also open to other possible strand arrangements which seemed to be less likely. For instance, instead of half-strand slipping, it might be also possible to form a continuous structure by slipping of just one, two or three residues, each leaving two short flanking strands at both ends (Figure 5.29).

Figure 5.29 *Guanine positions for less than half-strand slipping arrangement*

Three possible rotamers that were built using XPLOR with constraints generated from the Guanine arrangement as shown in the schematic (blue dots). Thymines have not been shown in the schematic for simplicity. The numbers 'x.y' just over the blue dots represent the residue number, where 'x' denotes the strand and 'y' denotes the guanine position on the 'x' strand from 5' end.

Once the residue placements were finalized, Dev-C++ was used to generate constraint files for Guanines using which the structure was to be computed. General anti-glycosidic dihedral constraints were used (Figure 5.30). It was the same for all of the 12 structures. Planarity constraints were used depending on which residues form the G-terad (Figure 5.31). This varied with the models but was the same for all rotamers of the same slipping configuration.


```

assi (resi 1 and name O4' and segi DN1)
      (resi 1 and name C1' and segi DN1)
      (resi 1 and name N9 and segi DN1)
      (resi 1 and name C4 and segi DN1) 1.0 240.0 70.0 2
assi (resi 1 and name C4' and segi DN1)
      (resi 1 and name C3' and segi DN1)
      (resi 1 and name O3' and segi DN1)
      (resi 2 and name P and segi DN1) 1.0 225.0 75.0 4

```

Figure 5.30 *Anti-glycosidic dihedral constraints for Guanines*

A snapshot of the text file showing the dihedral constraints for the 1st Guanine residue of the 1st DNA strand.

```

group selection=((not (store1) and residue 1 and segi DN1)
or (not (store1) and residue 7 and segi DN2)
or (not (store1) and residue 1 and segi DN3)
or (not (store1) and residue 7 and segi DN4)) weight = 1.0 end

group selection=((not (store1) and residue 2 and segi DN1)
or (not (store1) and residue 8 and segi DN2)
or (not (store1) and residue 2 and segi DN3)
or (not (store1) and residue 8 and segi DN4)) weight = 1.0 end

group selection=((not (store1) and residue 3 and segi DN1)
or (not (store1) and residue 9 and segi DN2)
or (not (store1) and residue 3 and segi DN3)
or (not (store1) and residue 9 and segi DN4)) weight = 1.0 end

group selection=((not (store1) and residue 4 and segi DN1)
or (not (store1) and residue 10 and segi DN2)
or (not (store1) and residue 4 and segi DN3)
or (not (store1) and residue 10 and segi DN4)) weight = 1.0 end

```

Figure 5.31 *Planarity constraints for Guanines*

A snapshot of the text file showing the planarity constraints for the very first G-tetrad layer of the G-wire. The G-tetrad is comprised of 1st Guanine residue of the 1st and 3rd DNA strand, and the 7th Guanine residue of the 2nd and 4th DNA strand. The placement of the residues is in accordance to the schematic shown in Figure 5.27.

There were also distance constraints to specify distances between Guanines in a G-tetrad (Figure 5.32) and distances between carbon atoms of consecutive G-tetrads (Figure 5.33).

```

assi (resi 1 and name H21 and segi DN1) (resi 7 and name N7 and segi DN2) 2.00 0.20 0.20
assi (resi 1 and name N2 and segi DN1) (resi 7 and name N7 and segi DN2) 2.90 0.30 0.30
assi (resi 1 and name H1 and segi DN1) (resi 7 and name O6 and segi DN2) 2.00 0.20 0.20
assi (resi 1 and name N1 and segi DN1) (resi 7 and name O6 and segi DN2) 2.90 0.30 0.30
assi (resi 1 and name O6 and segi DN1) (resi 7 and name O6 and segi DN2) 3.15 0.15 0.15

assi (resi 7 and name H21 and segi DN2) (resi 1 and name N7 and segi DN3) 2.00 0.20 0.20
assi (resi 7 and name N2 and segi DN2) (resi 1 and name N7 and segi DN3) 2.90 0.30 0.30
assi (resi 7 and name H1 and segi DN2) (resi 1 and name O6 and segi DN3) 2.00 0.20 0.20
assi (resi 7 and name N1 and segi DN2) (resi 1 and name O6 and segi DN3) 2.90 0.30 0.30
assi (resi 7 and name O6 and segi DN2) (resi 1 and name O6 and segi DN3) 3.15 0.15 0.15

assi (resi 1 and name H21 and segi DN3) (resi 7 and name N7 and segi DN4) 2.00 0.20 0.20
assi (resi 1 and name N2 and segi DN3) (resi 7 and name N7 and segi DN4) 2.90 0.30 0.30
assi (resi 1 and name H1 and segi DN3) (resi 7 and name O6 and segi DN4) 2.00 0.20 0.20
assi (resi 1 and name N1 and segi DN3) (resi 7 and name O6 and segi DN4) 2.90 0.30 0.30
assi (resi 1 and name O6 and segi DN3) (resi 7 and name O6 and segi DN4) 3.15 0.15 0.15

assi (resi 7 and name H21 and segi DN4) (resi 1 and name N7 and segi DN1) 2.00 0.20 0.20
assi (resi 7 and name N2 and segi DN4) (resi 1 and name N7 and segi DN1) 2.90 0.30 0.30
assi (resi 7 and name H1 and segi DN4) (resi 1 and name O6 and segi DN1) 2.00 0.20 0.20
assi (resi 7 and name N1 and segi DN4) (resi 1 and name O6 and segi DN1) 2.90 0.30 0.30
assi (resi 7 and name O6 and segi DN4) (resi 1 and name O6 and segi DN1) 3.15 0.15 0.15

```

Figure 5.32 Distance constraints for Guanines in a G-tetrad

A snapshot of the text file specifying the distances between atoms of Guanines in the same plane. The snapshot here is specific to the very first G-tetrad layer of the (2,2) diagonal slip model of G-wire. The 3 numbers on the right specify the distance and tolerance. For instance, the first line of the file is supposed to keep a distance of 2 Å between the H21 atom of 1st Guanine residue of 1st strand and N7 atom of 7th Guanine residue of 2nd strand, with a tolerance of ± 0.2 Å.

```

assi (resi 1 and name C5' and segi DN1) (resi 2 and name C5' and segi DN1) 6.51 0.15 0.15
assi (resi 1 and name C5' and segi DN1) (resi 2 and name C4' and segi DN1) 7.68 0.15 0.15
assi (resi 1 and name C5' and segi DN1) (resi 2 and name C3' and segi DN1) 8.62 0.15 0.15
assi (resi 1 and name C5' and segi DN1) (resi 2 and name C2' and segi DN1) 8.48 0.15 0.15
assi (resi 1 and name C5' and segi DN1) (resi 2 and name C1' and segi DN1) 8.25 0.15 0.15
assi (resi 1 and name C5' and segi DN1) (resi 2 and name C8 and segi DN1) 7.02 0.15 0.15
assi (resi 1 and name C5' and segi DN1) (resi 2 and name C5 and segi DN1) 8.57 0.15 0.15
assi (resi 1 and name C5' and segi DN1) (resi 2 and name C6 and segi DN1) 9.59 0.15 0.15
assi (resi 1 and name C5' and segi DN1) (resi 2 and name C2 and segi DN1) 10.8 0.15 0.15
assi (resi 1 and name C5' and segi DN1) (resi 2 and name C4 and segi DN1) 8.87 0.15 0.15
assi (resi 1 and name C4' and segi DN1) (resi 2 and name C5' and segi DN1) 5.09 0.15 0.15
assi (resi 1 and name C4' and segi DN1) (resi 2 and name C4' and segi DN1) 6.23 0.15 0.15
assi (resi 1 and name C4' and segi DN1) (resi 2 and name C3' and segi DN1) 7.26 0.15 0.15
assi (resi 1 and name C4' and segi DN1) (resi 2 and name C2' and segi DN1) 7.21 0.15 0.15
assi (resi 1 and name C4' and segi DN1) (resi 2 and name C1' and segi DN1) 6.91 0.15 0.15
assi (resi 1 and name C4' and segi DN1) (resi 2 and name C8 and segi DN1) 5.98 0.15 0.15
assi (resi 1 and name C4' and segi DN1) (resi 2 and name C5 and segi DN1) 7.60 0.15 0.15
assi (resi 1 and name C4' and segi DN1) (resi 2 and name C6 and segi DN1) 8.71 0.15 0.15
assi (resi 1 and name C4' and segi DN1) (resi 2 and name C2 and segi DN1) 9.70 0.15 0.15
assi (resi 1 and name C4' and segi DN1) (resi 2 and name C4 and segi DN1) 7.73 0.15 0.15
assi (resi 1 and name C3' and segi DN1) (resi 2 and name C5' and segi DN1) 4.16 0.15 0.15
assi (resi 1 and name C3' and segi DN1) (resi 2 and name C4' and segi DN1) 5.30 0.15 0.15
assi (resi 1 and name C3' and segi DN1) (resi 2 and name C3' and segi DN1) 6.11 0.15 0.15
assi (resi 1 and name C3' and segi DN1) (resi 2 and name C2' and segi DN1) 5.99 0.15 0.15

```

Figure 5.33 Distance constraints for carbon atoms in consecutive G-tetrads

A snapshot of the text file specifying the distances between carbon atoms of Guanines in consecutive G-tetrad layers. The 3 numbers on the right specify the distance and tolerance. For instance, the first line in the snapshot is supposed to keep a distance of 6.51 Å between the C5' atoms of 1st and 2nd Guanine residues of 1st strand, with a tolerance of ± 0.15 Å.

With the four constraint files, thousands of structures for each model were computed using XPLOR, out of which only a handful looked like what was expected (Figure 5.34). This was likely because no long-range distance constraints were used which is very important for building large structures. Another major reason was that the short-range distance constraints were obtained from an arbitrary G-quadruplex structure and was not very fitting for the structures we were trying to build.

Figure 5.34 *Half-slipping structures computed using XPLOR*

Pymol generated images of the half-slipped structures computed using XPLOR. Without considering the flanking part of DNA strands, each of the structures have a length of 5 nm and consist of 16 layers of G-tetrads with 3 interfaces of Thymine loops. The nucleobases in cyan are Guanines, while those in orange are Thymines.

Figure 5.35 *Partial slipping structures computed with XPLOR*

Pymol generated images of the half-slipped structures computed using XPLOR. The nucleobases in cyan are Guanines, while those in orange are Thymines.

All the 12 structures that were investigated have all-parallel topology and should conform well to the observed CD spectra. However, it has been reported that d[G₄T₂G₄] can form G2-structures by folding upon itself, similar to a hairpin.¹¹⁶ Such an antiparallel structure would give rise to a peak at 290 nm in CD spectra, but was never observed. This would imply that either antiparallel structures do not exist,

or they exist in such a small proportion that they are not detectable. Two such possible structures are the (2,1) adjacent and (2,1) diagonal slip models (Figure 5.36).

Figure 5.36 *Guanine arrangement for an anti-parallel G-wire*

Two possible anti-parallel G-wire models that were not investigated due to complexity of the constraints and lack of experimental evidence. Thymine has not been shown in the schematic for simplicity. The numbers ‘x.y’ just over the blue dots represent the residue number, where ‘x’ denotes the strand and ‘y’ denotes the guanine position on the ‘x’ strand from 5’ end. The red dashed connections represent the thymine loops for the hairpin structure.

5.11. Simulations and comparison with experimental data

The lowest energy structures obtained from Xplor calculations, were used as building blocks to form wires by manual translation and rotation using Pymol. The new structure was 11 – 12 nm in length and comprised of 14 full and 4 truncated ssDNA strands. This length was deemed sufficient to reveal any periodic pattern. Potassium (K^+) ions were added along the axis of the wire in between G-tetrads and on the exterior of terminally exposed G-tetrads.

The long G-wire models were simulated using Molecular Dynamics (MD) for relaxation in an explicit solvent environment. The input files were prepared using xleap (AmberTools15) and simulated using Amber 10¹¹⁷ with the ff99 DNA force field¹¹⁸. Models underwent unrestrained energy minimization of 1000 steps of steepest decent followed by 9000 steps of conjugant gradient method to remove clashes generated by manual model building. More K⁺ ions were added to neutralize the system. The structures were then solvated with water (truncated octahedron, TIP3P).

Systems underwent an initial restrained minimization involving 500 steps of steepest decent followed by 500 steps of conjugated gradient while maintaining 25 kcal.mol⁻¹Å⁻² harmonic potential restraints on solute atoms. Systems were next heated from 100 to 300K over 10ps under constant volume while maintaining solute constraints. Systems further underwent a gradual relaxation of constraints in which systems underwent cycles of minimization and restrained dynamics with constraints of 5, 4, 3, 2, 1 and 0.5 kcal.mol⁻¹Å⁻². Finally, simulations were run unrestrained at 1 atm and 300K for a duration of 1 ns. Molecules were visualized using VMD and Pymol. Molecular structures were averaged over the 1ns unrestrained simulation, and used for AFM simulations.

Figure 5.37 *The starting structures for MD simulation*

Pymol generated image of all 9 models of 11-12 nm length, manually built in Pymol from 5 nm computed structures, and minimized in vacuum with K^+ ions at the core using Amber. The surrounding water and ions have not been shown. The Thymine have been highlighted as orange surfaces to get a hint of the periodic pattern observed in AFM images.

Simulated Atomic Force Microscopy images were generated by a custom script written in MATLAB. In brief, AFM height images are produced for an input set of molecular coordinates by scanning a virtual paraboloid AFM tip over the molecule. The radius of the virtual tip was subjectively chosen based on the resemblance of the simulated image to real AFM image. Its closest distance of approach without colliding with the molecule is registered as the contact height, and stored as intensity in the simulated AFM image. Orientation of the molecule was varied as -90° , 0° and 90° ; thus, generating a total of 27 simulated AFM images (Figure 5.38). The 0° orientation of the molecule is arbitrary, as generated by the

MD simulation average of 1 ns. The images were saved as grayscale image and opened with the Gwyddion program for AFM color rendering and further analysis.

Figure 5.38 *Simulated AFM images*

Simulated AFM images of slip-strand G-wire models. Images of the (A) (3+1) (B) (2+2) *adjacent* and (C) (2+2) *diagonal* arrangements adopting (*left*) -90° , (*center*) 0° , (*right*) $+90^\circ$ rotamers. Three orientations of the G-wire molecule are shown for each model. Only the (2+2) diagonal arrangement with a -90° rotamer displays clear *Type I* features. Each of the 9 images are 16 nm in width.

Particularly for the -90° rotamer of the (2,2) *diagonal* model, longer structures were simulated which were built using the structure averaged over 1 ns of the MD simulation (Figure 5.39). In this case, a suitable orientation of the molecule was chosen so that it could be phased/aligned well with the AFM image.

Figure 5.39 Agreement between experimental data with predicted structure

Two different orientations of the molecular model based on -90° rotamer of the (2,2) *Diagonal* configuration to explain two different AFM observations of *Type I* G-wires: left-handed and zigzag. **(A)** Pymol generated image of the same molecule in two different orientations. Thymines have been highlighted as orange surfaces to give a hint of the periodic pattern. **(B)** VMD generated image where the grayscale intensity of the surface is directly proportional to its distance from the plane over which the molecule is lying. Van-der-Waals radii have been altered for enhanced visualization of the periodic pattern traced by Thymines. **(C)** AFM images simulated from the above molecular orientation and colored using Gwyddion to have same color scale as the real AFM images. **(D)** Experimentally observed AFM images of *Type I* G-wires showing bulged surface features separated by 4.2 nm and mild undulating features in the background with periodicity of 1.1 nm. **(E)** The proposed model overlaid with the actual AFM image, showing excellent overall fit and particularly the periodicity and handedness of the 4.2 nm and 1.1 nm surface features.

5.12. Discussions and future work

As per our AFM data, the majority of both short and long DNA G-wires have undulating left-handed or zig-zag surface pattern with 4.2 ± 0.2 nm periodicity, running along a subtle undulating right-handed pattern with 1 ± 0.2 nm periodicity. As per our proposed molecular model, a G-wire of ~ 20 nm length, which is composed of 32 strands of d[G₄T₂G₄] with the overall molecular weight of ~ 100 kDa, should be detectable as a broad hump in solution proton NMR spectrum. Significantly longer wires, as observed by AFM, might be beyond the detection limit of solution NMR, and an increase in the population of such long wires after annealing (Figure 5.6) may explain the observed reduction in the integral area under the proton NMR spectrum (Figure 5.1). We have strong experimental evidence to support the proposed model for formation of interlocked parallel G-quadruplexes of d[GGGGTTGGGG]. In future further work could be done to explain the structure of Type II and Hybrid Type G-wires.

Understanding of the structure of G-wires is important for interpretation of its biological role and for designing of nanoscale devices based on G-wires. The solution AFM structure observed in this case could apply to interlocked parallel G-quadruplexes formed by several other DNA and RNA G-rich repeat motifs, which could hint towards biological relevance due to universality. The groove-like pattern on G-wires could provide molecular grip for a certain unknown class of proteins (may be helicases) to bind and act on the G-wires. If there was no undulating pattern on the surface of G-wires, it would be difficult for proteins/ligands to dock in a specific position on the G-wire, and not slip. Just like it is important for a rope to offer friction and grip to make it usable, similar concept can be applied to proteins/ligands docking on the surface of G-wires.

Chapter 6: Higher-order structures of GGGGCC

6.1. Introduction

During experiments with alternative DNA sequences that might be able to form G-wires, we found that if the Thymines in d[G₄TTG₄] were replaced by Cytosines, it still formed DNA G-wires. And it couldn't have been a coincidence that GGGGCC repeats in the C9orf72 gene have been associated to neurodegenerative diseases like ALS and FTD.

Few high-resolution images of both DNA and RNA G-wires of G₄CCG₄ were obtained, both of which showed similar periodic pattern as observed in case of Type I DNA G-wires formed by the *Tetrahymena* telomeric sequence d[G₄TTG₄]. So, we were straightforward in building a model for the DNA G₄CC-wires based on the -90° rotamer of the (2,2) diagonal configuration. The model was simulated using molecular dynamics and Mathematica to generate an AFM-like image, which turned out to be an excellent match with the real experimental data.

6.2. NMR spectroscopy

In 1D proton NMR spectroscopy, both DNA and RNA show the signature proton peaks in the same region (Figure 2.8). NMR spectra was acquired for r[G₄CCG₄] in 90 mM KCl, 20 mM KPi (pH 7) and 10 mM MgCl₂ before and after annealing (Figure 6.1). Proton peaks due to imino (H1) protons of Guanines of G-

tetrads are observed in the region 10 – 12.5 ppm. The aromatic (H6) protons of Cytosines show up in the region 7 – 8 ppm, while the amino (H21, H22) protons of Guanines and (H41, H42) protons of Cytosines show up in the range 7 – 10 ppm (Figure 2.8). Upon annealing the sample (fast heating and slow cooling), the sharp imino protons were replaced by 3 broad peaks at 10.5, 11 and 11.5 ppm (Figure 6.1). This is indicative of the possible formation of multiple and/or higher-order G-quadruplex structures, which were detectable by solution NMR.

Figure 6.1 1-D proton NMR spectra for *r*[G₄CCG₄]

NMR Spectra was acquired using 700 MHz NMR system at 25°C both before and after annealing in 90 mM KCl, 20 mM KPi (pH 7) and 10 mM MgCl₂.

The area under the proton peaks in 1-D NMR spectra were calculated using Mathematica. DSS was used as a reference for which the integral area under the peak at 0 ppm was made the same before and after annealing. After standardization of the NMR spectra, 40% reduction in area was observed after annealing for protons in the region 10 – 12 ppm and 7 – 8.1 ppm (Figure 6.2). This could be indicative of the fact that 40% of the initial RNA structures become too large to be detectable by solution NMR.

Figure 6.2 Integral area ratio before and after annealing

After standardization of the NMR spectra with respect to the **(A)** DSS peak, the integral area was computed for **(B)** 3 different parts of the NMR spectra corresponding to imino, amino and aromatic protons.

1D Proton NMR spectra was also obtained for d[G₄CCG₄] in 90 mM KCl, 20 mM KPi and 10 mM MgCl₂. It revealed existence of imino protons (H1) of Guanines in G-tetrad, just like any typical G-quadruplex. After annealing the sample, 60% reduction of integral area in imino proton region was observed (Figure 6.3). This could possibly indicate that 60% of the initial DNA structures became too large to be detectable by solution NMR. The merger of the imino proton peaks into a single broad peak looks very similar to that of d[G₄TTG₄] (Figure 5.1).

Figure 6.3 *Imino proton spectra for d[G₄CCG₄]*

(A) NMR Spectra was acquired using 700 MHz NMR system at 25°C both before and after annealing in 90 mM KCl, 20 mM KPi (pH 7) and 10 mM MgCl₂. (B) Standardized NMR peak for DSS before and after annealing with same integral area.

6.3. CD spectroscopy

For estimation of G-quadruplex folding topology and its stability, Circular Dichroism (CD) spectroscopy was used while heating/cooling the sample. For both DNA and RNA G₄CCG₄, 20 mM HEPES buffer (pH 7) with 90 mM KCl and 10 mM MgCl₂ was used. CD spectra of r[G₄CCG₄] showed a positive peak at 260 nm whose intensity reduced to about 25% by heating from 25°C to 95°C, but it never reduced to zero (Figure 6.4). At about 88°C, 50% of the G-quadruplex structures were melted, so this can be considered the melting temperature for r[G₄CCG₄].

Figure 6.4 UV-CD heating spectra for *r*[G₄CCG₄]

(A) Molar ellipticity and (B) absorption of *r*[G₄CCG₄] as a function of wavelength (x-axis) and temperature. The grayscale intensity from dark to light corresponds to increasing temperatures from 35°C to 96°C.

CD spectra of *d*[G₄CCG₄] was quite complicated. It showed a positive peak predominantly at 260 nm, whose intensity never reduced even slightly, but instead started increasing beyond 80°C (Figure 6.5). In the range of 290 – 420 nm, reduction in ellipticity to negative values was observed with increase in UV absorption, which was previously zero for <80°C. Even upon cooling, this negative ellipticity and positive UV absorption in the range of 290 – 420 nm remained intact. The experimental results were reproducible with two different sample batches, one of which was manufactured by IDT USA and the other from IDT Singapore. The results were also unaffected by concentration as 1/5th of the concentration reported here was also investigated.

Figure 6.5 UV-CD heating spectra for d[G₄CCG₄]

(A) Molar ellipticity and (B) absorption of d[G₄CCG₄] as a function of wavelength (x-axis) and temperature. The grayscale intensity from dark to light corresponds to increasing temperatures from 35°C to 96°C.

As with many G-rich oligonucleotides, heating/cooling of d[G₄CCG₄] resulted in milky solution with white precipitates at both low and high concentrations. When the quartz cuvette containing d[G₄CCG₄] was taken out after the CD measurements, a white translucent residue was observed at the bottom of the cuvette (Figure 6.6). Since a sharp decrease in UV absorption at 260 nm was observed at ~80°C during heating (Figure 6.7), it is most likely that the white residue is mainly DNA. A significant amount of DNA was precipitated in the heating process thus reducing the amount of UV light absorbed.

Figure 6.6 White residue formation by d[G₄CCG₄] after CD experiment

The picture was captured in ambient light with rear camera of Galaxy S7 edge with manual focus. The sample is in a transparent quartz cuvette typically used for UV measurements.

Figure 6.7 Sharp decline in DNA concentration at high temperature

Concentration was measured as a function of temperature from the variation in ellipticity. It was calculated using Beer-Lambert's law assuming molar extinction coefficient of ssDNA of d[GGGGCCGGG] which was $96000 \text{ cm}^{-1}\text{M}^{-1}$ as obtained from IDT website.

Despite the complicated spectroscopic observation for d[G₄CCG₄] and complete absence of any melting, it can be concluded that both d[G₄CCG₄] and r[G₄CCG₄] predominantly form highly stable parallel G-quadruplexes.

6.4. RNase treatment of RNA G-wires

From the CD melting experiments, it was quite evident that the RNA G-quadruplexes formed by r[G₄C₂G₄] are extremely stable at high temperatures. So, we ventured to test the stability of these structures against enzymatic degradation which is biologically more relevant.

A freshly synthesized G-wire sample of r[G₄C₂G₄] was incubated with equimolar mix of RNase A and RNase T1. The duration of incubation as well as the ratio of RNase to RNA was varied, but we did not observe any degradation to single nucleotides. From the PAGE results, it was observed that the smallest structures migrated to the same extent as 20 bp duplex DNA (Figure 6.8). Surprisingly, upon prolonged incubation with RNase for a week, a well-defined higher order species appeared at the top of the gel.

Figure 6.8 Incubation of RNA G-wires with RNase

12% native PAGE showing RNA G-wires before (Lane 3) and after (Lane 4 – 8) treatment with mix of RNase A and RNase T1. The 1st lane shows duplex DNA ladder. The 2nd lane shows a supposedly monomeric 3-layered G-quadruplex of d[TT(GGGT)₄] consisting of 18 residues. The last lane shows RNA G-wires that were incubated at room temperature for a week with 10 times more RNase compared to that shown in lanes 4 – 7.

The findings about the resistance of RNA G₄CC structures to enzymatic degradation has significant biological relevance. Being nuclease resistant, higher-order RNA G-quadruplex structures might accumulate and aggregate over time in cells containing abortive transcripts of C9orf72 gene with GGGGCC repeats.

6.5. Formation of G-wires

To investigate the size distribution of the DNA and RNA structures, native polyacrylamide gel electrophoresis was used. It was observed that both d[G₄CCG₄] and r[G₄CCG₄] formed significant population of higher-order structures, some of which migrated even slower compared to 300 bp duplex DNA (Figure 6.9). The observed size distribution was similar to that of G-wires formed by *Tetrahymena* telomeric sequence d[G₄TTG₄]. This indicated that if the structures formed by G₄CCG₄ and G₄TTG₄ had similar shapes, i.e., wires of similar diameter, the length distribution should also be the same. The smallest structures migrated to the same extent as 20 bp duplex DNA, but marginally less compared to 3-layered parallel G-quadruplexes¹¹⁹ formed by d[TT(GGGT)₄].

Figure 6.9 Native PAGE of G₄CC DNA and RNA G-wires

The 1st lane shows duplex DNA ladder. The 2nd and 3rd lane shows 3-layered monomeric and dimeric DNA G-quadruplexes¹¹⁹ formed by d[TT(GGGT)₄] and d[(GGGT)₄] respectively. Lane 4 shows G-wires formed by d[G₄T₂G₄]. Lane 5 and 6 shows formation of higher-order structures by d[G₄C₂G₄] and r[G₄C₂G₄] respectively. The samples were stained with SYBR gold prior to loading.

From Figure 6.9, it is quite evident that both $d[G_4C_2G_4]$ and $r[G_4C_2G_4]$ self-assemble to form higher-order structures. If these structures turn out to be wires with similar diameter as G-wires of $d[G_4TTG_4]$, then the structures formed by $d[G_4C_2G_4]$ and $r[G_4C_2G_4]$ should have similar size distribution as those formed by self-assembly of $d[G_4TTG_4]$.

Atomic Force Microscopy was used to investigate the shape and size of the higher-order structures formed by $d[G_4C_2G_4]$ and $r[G_4C_2G_4]$. It was extremely challenging to immobilize them on Mica in solution with $NiCl_2$, so Spermidine was used instead. The solution AFM images of $d[G_4C_2G_4]$ (Figure 6.10) and $r[G_4C_2G_4]$ (Figure 6.11) revealed isolated blobs and short G-wires. It was however surprising that no long G-wires could be seen like those of $d[G_4T_2G_4]$ (Figure 6.12). It seemed most likely that the long G-wires of $d[G_4C_2G_4]$ and $r[G_4C_2G_4]$ could not be immobilized on the Mica surface. However, it is also possible that the structures that migrated the slowest in the polyacrylamide gel were not wires, but some other big structures as could be seen from some of the white clusters in the AFM images.

The lengths of G-wires were computed using Mathematica's default method of measurement of the length of the major axis of an ellipse fitted to a G-wire. This method underestimates the length curved molecules, but works well for rough estimation of the length of straight G-wires. The longest 500 G-wires formed by $d[G_4T_2G_4]$ had lengths in the range of 100 – 250 nm, while it was just 30 to 80 nm for the wires formed by $d[G_4C_2G_4]$ and $r[G_4C_2G_4]$.

Figure 6.10 *d[G₄CCG₄] immobilized with Spermidine in solution*

(A) AFM image showing short DNA G-wires and few isolated blobs formed by d[G₄C₂G₄]. The image was acquired in aqueous solution with 600 nM DNA, 0.033% glutaraldehyde and 16.7 μ M spermidine. The length scale (white bar) is 200 nm. (B) Length histogram was determined for all the molecules detected from several AFM images. The histogram is of non-Gaussian nature and beyond 30 nm, the length decreases exponentially. (C) Length histogram for 500 of the longest G-wires formed by d[G₄C₂G₄]. The longest wires are typically 70 – 80 nm.

Figure 6.11 *r[G₄CCG₄] immobilized with Spermidine in solution*

(A) AFM image showing short DNA G-wires, few isolated blobs and few large blobs formed by r[G₄C₂G₄]. The image was acquired in aqueous solution with 450 nM DNA, 0.025% glutaraldehyde and 16.7 μ M spermidine. The length scale (white bar) is 200 nm. (B) Length histogram was determined for all the molecules detected from several AFM images. The histogram is of non-Gaussian nature and beyond 40 nm, the length decreases exponentially. (C) Length histogram for 500 of the longest G-wires formed by r[G₄C₂G₄]. The longest wires are typically 80 – 90 nm.

Figure 6.12 *d[G₄TTG₄] immobilized with Spermidine in solution*

(A) AFM image showing both short and long DNA G-wires along with few isolated blobs formed by d[G₄T₂G₄]. The image was acquired in aqueous solution with 300 nM DNA, 0.033% glutaraldehyde and 16.7 μ M spermidine. The length scale (white bar) is 200 nm. (B) Length histogram was determined for all the molecules detected from several AFM images. The histogram is of non-Gaussian nature and beyond 40 nm, the length decreases exponentially. (C) Length histogram for 500 of the longest G-wires formed by d[G₄T₂G₄]. The longest wires are typically 250 – 270 nm.

6.6. Purification of G-wires

Since majority of the G₄CC-structures observed by AFM were small, it was thought that if only the very higher-order structures could be specifically purified, longer wires could be consistently observed. Using native PAGE, the top portion of the gel corresponding to lanes containing d[G₄T₂G₄] and d[G₄C₂G₄] were excised out. The gel was visualized both before and after the excision by UV shadowing; no fluorescent dyes were used (Figure 6.13). The excised gel was finely crushed in an Eppendorf tube and 1 mL of buffer was added to every ~0.3 mg of gel. The buffer contained 50 mM KCl, 10 mM MgCl₂ and 10 mM HEPES. The tubes were stirred at 1000 rpm in a shaker at 55°C for ~18 hours. It was then centrifuged at 16,000g and the supernatant was taken out into a new Eppendorf tube. The supernatant was

then passed through a desalting column, which yielded very dilute DNA solution of ~ 3 ng/ μ L. The recovery efficiency was low ($\sim 33\%$) because out of 6 μ g of DNA that was loaded into the gel, about 1 μ g was recovered from half of the excised portion which was $\sim 50\%$ of the total smear in gel lane. No lyophilization was done to concentrate the samples as the concentration was sufficient for AFM imaging.

Figure 6.13 Native PAGE purification of DNA G-wires

Gel visualization of d[G₄T₂G₄] and d[G₄C₂G₄] (**A**) before and (**B**) after cutting the bands to be purified. Only those structures that migrated slower than 300 bp duplex DNA were purified. The 1st lane of the gel consists of duplex DNA markers of length 100 bp to 1000 bp at regular intervals of 100 bp.

The purified higher-order structures of d[G₄T₂G₄] and d[G₄C₂G₄] were visualized by AFM in air (Figure 6.14). Contrary to our expectations, the AFM images showed G-wires of all possible lengths from less than 10 nm to longer than 200 nm, as if the PAGE purification hardly made any difference in the length distribution of G-wires. After purification, the average length increases barely by ~ 20 nm, but the longest G-wires are only about 150 nm, while it was up to 450 nm before purification (Figure 6.15). Previous attempts to purify G-wires of d[G₄T₂G₄] using size exclusion HPLC also yielded similar results. This indicated that formation of G-wires is a dynamic process in which long G-wires can disassociate into short G-wires and vice-versa. However, if the purification process is optimized, it might be possible to shift the equilibrium of the length distribution of G-wires to either the shorter or longer side.

Figure 6.14 *AFM images of PAGE purified DNA G-wires*

Topography images of $d[G_4T_2G_4]$ and $d[G_4C_2G_4]$ after PAGE purification, acquired with ScanAsyst Mode in air using ScanAsyst Fluid+ probe. The length scale (white bar) is 300 nm.

Figure 6.15 *G-wires of $d[G_4T_2G_4]$ before and after purification*

The length histogram was computed using Mathematica's default method from AFM images obtained in air before and after PAGE purification of DNA G-wires formed by $d[G_4T_2G_4]$. Some of the longest G-wires of over 150 nm length were observed only before purification.

6.7. Duplex DNA vs. G4CC wires

AFM was used to determine the width/height of the G-wires formed by d[G₄C₂G₄] and r[G₄C₂G₄] in aqueous solution. Since the height measured in AFM is not an absolute parameter but depends on several factors and is relative to the background, duplex B-DNA was used chosen as a reference height standard. From the AFM images it was quite clear that G-wires are strikingly different from duplex DNA in terms of height as can be seen from the color contrast (Figure 6.16). The reference duplex DNA consisted of 492 bp and had uniform length of 160±10 nm as expected. Duplex DNA appeared to be much longer and highly flexible compared to G-wires.

Figure 6.16 *Solution images of G4CC structures with duplex DNA*

(A) AFM image showing duplex DNA along with structures formed by d[G₄C₂G₄]. The image was acquired in aqueous solution with 1 nM duplex DNA, 600 nM d[G₄C₂G₄], 0.0167% glutaraldehyde and 16.7 μ M spermidine. (B) AFM image showing duplex DNA along with structures formed by r[G₄C₂G₄]. The image was acquired in aqueous solution with 1 nM duplex DNA, 300 nM r[G₄C₂G₄], 0.033% glutaraldehyde and 16.7 μ M spermidine. Both the stock DNA and RNA samples were synthesized with initial oligo concentration of 180 μ M in buffer containing 90 mM KCl, 10 mM MgCl₂, 20 mM HEPES (pH 7). The length scale (white bar) of both AFM images is 200 nm.

In order to quantify the height of G-wires, height histograms were obtained from several AFM images. Structures formed by both $d[G_4C_2G_4]$ and $r[G_4C_2G_4]$ were found to be higher than duplex B-DNA by ~ 1.6 nm (Figure 6.17) and ~ 1.3 nm (Figure 6.18) respectively. Previously it has been observed that G-wires formed by $d[G_4T_2G_4]$ are 0.9 nm higher than duplex B-DNA (Figure 5.10). This suggests that G-wires formed by $d[G_4T_2G_4]$, $r[G_4C_2G_4]$ and $d[G_4C_2G_4]$ have marginally different heights. However, such small height variations could be due to Glutaraldehyde-spermidine or due to difference in mechanical rigidity and structural integrity of different G-wires. The strength of attachment of the molecules to the mica surface could determine the gap between the molecule and the substrate, which in turn could also lead to different height of the molecules.

Figure 6.17 Height histograms for $d[G_4CCG_4]$ with duplex B-DNA

Histograms were plotted for the heights extracted from the spline of molecules detected in the solution AFM images containing both DNA G4CC wires and duplex B-DNA. Two distinct peaks at ~ 1.9 nm and ~ 3.3 nm could be clearly seen which corresponds to duplex and G-quadruplex DNA respectively.

Figure 6.18 Height histograms for $r[G_4CCG_4]$ with duplex B-DNA

Histograms were plotted for the heights extracted from the spline of molecules detected in the solution AFM images containing both RNA G4CC wires and duplex B-DNA. Two distinct peaks at ~ 1.9 nm and ~ 3.2 nm could be clearly seen which corresponds to duplex DNA and G-quadruplex RNA respectively.

6.8. Periodical features on G₄CC-wires

After finding out that G-wires formed by $d[G_4C_2G_4]$ and $r[G_4C_2G_4]$ looked very similar to G-wires formed by $d[G_4T_2G_4]$ except for the length distribution, we ventured to look for differences in their high-resolution structures. Using $NiCl_2$ for immobilization, high-resolution AFM images were acquired in aqueous solution. DNA G-wires formed by $d[G_4C_2G_4]$ revealed left-handed surface features with periodicity of 4.1 ± 0.2 nm (Figure 6.19). While RNA G-wires formed by $r[G_4C_2G_4]$ revealed left-handed surface features with periodicity of 4.5 ± 0.36 nm (Figure 6.20). Considering the standard deviation in periodicity, it can be said that the periodic features of G-wires formed by $d[G_4C_2G_4]$ and $r[G_4C_2G_4]$ were exactly similar to Type I G-wires formed by *Tetrahymena* telomeric sequence $d[G_4T_2G_4]$. In fact, we could also see some 1 nm features on few DNA G-wires formed by $d[G_4C_2G_4]$ (Figure 6.21).

Figure 6.19 *High-resolution images of DNA G4CC-wires in solution*

(A) High-resolution solution AFM images of d[G4CCG4] showing periodic features. The length scale in each image is 5 nm. (B) Height profile showing well-defined peaks separated by 4.1 ± 0.2 nm. The individual peak-to-peak distances are shown in blue. The horizontal and vertical scale represent length and height in nm respectively.

Figure 6.20 *High-resolution images of RNA G4CC-wires in solution*

(A) High-resolution solution AFM images of r[G4CCG4] showing periodic features. The length scale in each image is 5 nm. (B) Height profile showing well-defined peaks separated by 4.5 ± 0.36 nm. The individual peak-to-peak distances are shown in blue. The horizontal and vertical scale represent length and height in nm respectively.

Figure 6.21 *Sub-nm resolution showing 1 nm periodicity*

(A) AFM height image of a typical DNA G-wire formed by d[G₄C₂G₄] revealing features along the edges with 1 nm periodicity. The length scale (white bar) is 5 nm. (B) Height profile of the indication section showing peaks separated by ~ 1 nm.

6.9. Modelling and simulation

With valuable insights gained from the high-resolution AFM images of G-wires formed by three different molecules and our previously predicted structure for Type I G-wires of d[G₄T₂G₄], it seemed very obvious that all G-wires (excluding Type II and Hybrid Type G-wires) had similar molecular configuration of interlocked G-quadruplexes. So, we went straight to building a molecular model for DNA G-wires of d[G₄C₂G₄] based on the -90° rotamer of the (2,2) diagonal configuration for half-slipping strands (Figure 6.22).

Figure 6.22 *Computed molecular structure of d[G₄CCG₄]*

(A) Molecular structure of G-wire computed using XPLOR in vacuum and equilibrated using Amber 14 in ionic solution. (B) Molecular structure of G-wire after 1 ns of completely unrestrained MD simulation. The ionic solution simulation contained 45,023 residues of water (TIP3P), 208 atoms of K⁺, 50 atoms of Mg⁺⁺ and 200 atoms of Cl⁻.

12-stranded G-wire models computed using XPLOR were simulated using Molecular Dynamics (MD) for relaxation in an explicit solvent environment. The

input files were prepared using xleap (AmberTools14) and simulated using Amber 14¹¹⁷ with the ff99 DNA force field¹¹⁸. Systems underwent an initial restrained minimization involving 500 steps of steepest decent followed by 500 steps of conjugated gradient while maintaining 25 kcal.mol⁻¹Å⁻² harmonic potential restraints on solute atoms. Systems were then heated from 100 to 300K over 10ps under constant volume while maintaining solute constraints. Systems further underwent a gradual relaxation of constraints in which systems underwent cycles of minimization and restrained dynamics with constraints of 5, 4, 3, 2, 1 and 0.5 kcal.mol⁻¹Å⁻². Finally, simulations were run unrestrained at 1 atm and 300K for a duration of 1 ns.

The DNA structures generated in the 1 ns trajectory of the MD simulation were averaged using CPPTRAJ¹²⁰ to generate a single structure that could be simulated to generate an AFM image using a program written in Mathematica. The pdb structure was manually rotated and aligned parallel to the x-axis, depending on which face of the structure will be scanned by the AFM tip. In the program, the AFM tip was raster scanned over three orientations of the averaged G-wire model. At every point in a user defined grid, the minimum of all Euclidean distances between atoms of the AFM tip and all atoms exactly beneath the AFM tip was determined. A parabolic tip was chosen as it is mathematically simple and is usually considered a better representation of the AFM tip-end compared to a sphere, cone or pyramid. The program utilizes multiple cores of a processor, which made it possible to compute three simulated images for three different orientations of the same molecule in ~16 hours on a Ryzen 1700 CPU overclocked to 4.9 GHz.

Figure 6.23 Few snapshots showing the process of AFM simulation

As a parabolic AFM tip-end (in orange) hovers over the molecule (in green), the atoms of the sample just below the AFM tip are highlighted in blue. In all the six snapshots, the y-position of the tip is constant = 4 nm, while the x-position changes in intervals of 1 nm. The number (in red) shown along the vertical z-axis is the height of the sample that would be measured by the AFM at different points.

The simulated AFM images were colored using Gwyddion¹²¹ to correlate the color with height of the molecule. It was observed that the maximum height of the simulated G-wire was 3.7 nm for a particular orientation, while it was 3.6 nm and 3.5 nm for two other orientations that differed by $\pm 90^\circ$ (Figure 6.24). Thus, the simulated height for G4CC-wire is similar to that of G4TT-wire (Figure 5.38). A periodicity of ~ 4 nm can be clearly seen from different orientations of the molecule.

Figure 6.24 Simulated AFM images for G-wire formed by $d[G_4CCG_4]$

Simulated AFM images of three orientations of an averaged molecular structure of G-wire formed by $r[G_4C_2G_4]$ from the MD simulation, based on -90° rotamer of the (2,2) Diagonal configuration. The length scale (white bar) is 5 nm.

6.10. Discussions and future work

We found the both G4CC DNA and RNA can form G-wires which appear to be alike those formed by d[G₄TTG₄] in terms of height and left-handed periodic features. DNA and RNA G-wires of G₄CC have undulating left-handed surface pattern with ~4.1 nm and ~4.5 nm periodicity respectively. A subtle undulating right-handed pattern with ~1 nm periodicity was observed for few G-wires of d[G₄CCG₄]. So, we propose a structure of interlocked parallel G-quadruplexes similar to what was determined for d[G₄TTG₄] in the previous chapter.

The higher-order species observed after prolonged incubation of RNA G-wires with RNase, could be further investigated with both PAGE and AFM. Preliminary attempts to image these structures using AFM in air failed, likely due to high concentration of glycerol from the stock RNase. By ensuring that the concentration of glycerol in the stock RNase mix is very little or negligible, the issues with immobilization of the higher-order RNA species could be overcome.

To the best of our knowledge, formation of RNA G-wires has never been proposed or reported earlier. Our discovery of RNA G-wires could have significant biological implications. However, it should be noted that till date, existence of G-wires or long G-quadruplexes inside the living cell has never been reported, but it is very likely for mRNA (especially abortive transcripts) which are usually chopped into smaller pieces in the cells. RNA G-wires are exceptionally stable and could possibly accumulate in the cell over extended periods and disrupt nucleocytoplasmic transport,^{122, 123} eventually leading to death of the cell. In fact, G-4 RNA foci observed in fibroblasts and astrocytes of C9 ALS patients could be RNA G-wires rather than 3-4-layered RNA G-quadruplexes. The structural information obtained from the high-resolution AFM images, combined with our proposed model for interlocked G-quadruplexes, could help design drugs/targets to tackle the issue of cytotoxicity due to formation of highly stable multi-layered RNA-quadruplexes, which we call “RNA G-wires”.

Chapter 7: Solution AFM imaging of TERRA

7.1. Introduction

Until now, our investigation has been limited to short G-rich DNA and RNA. In this chapter, we report our investigation of long G-rich RNA transcribed from mammalian telomeric repeats. Our goal was to investigate the structure(s) formed by G-rich TERRA in aqueous solution using Atomic Force Microscopy.

7.2. The pRST5 plasmid

Our lab was provided with a pGEM-based plasmid called pRST5 by our collaborator Prof. Jack D. Griffith from UNC Chappel Hill (USA). It contained 576 bp of the Mammalian telomeric sequence d[TTAGGG] and a T7 promoter sequence ~58 bp ahead of the TTAGGG repeats towards the 5' direction (Figure 7.1). TERRA could be generated from it by in vitro transcription using T7 RNA polymerase. However, since the plasmid had no T7 terminator sequence, it was necessary to either cut the plasmid just at the end of the telomeric repeats towards the 3' direction or obtain a 0.6 – 1 kbp DNA template from the plasmid containing just the T7 promoter sequence and telomeric repeats.

Figure 7.1 Plasmid map of pRST5 containing d[TTAGGG]₉₆

The plasmid contains 96 repeats of TTAGGG (highlighted in red), placed strategically between suitable restriction sites and slightly ahead of a T7 promoter. The plasmid confers Ampicillin resistance to the host *E.Coli* for selective growth in a cell culture. The forward and reverse primers shown in purple can be used for PCR amplification of the telomeric region with the T7 promoter.

Initially the pRST5 plasmid was extracted from *E.Coli* after overnight amplification of bacteria at 37°C in liquid Lysogeny broth (LB) containing Ampicilin (antibiotic). Except for the control LB with no *E.coli*, liquid media in the other falcon tubes turned translucent from transparent (Figure 7.2). The absence of cell growth in the control liquid media even after incubation for several days, indicated that the extracted plasmids were only from the desired *E.coli* strain, and not any other bacteria with Ampicilin resistance from uncontrolled sources.

Figure 7.2 *Cell culture for pRST5 plasmid extraction*

LB media after bacterial growth became translucent, while without bacterial growth (last tube) it was still transparent as earlier.

The following extraction protocol for QIAGEN miniprep kit was used for plasmid DNA purification:

- A. Centrifuged the 50 mL falcon tubes containing the cell culture at ~10,000g for 10–15 minutes until the solution became completely transparent.
- B. Whitish-brown residue/pellet could be seen at the bottom of the falcon tubes. The liquid media was discarded while ensuring the bottom residue remained intact.
- C. 750 μ L of buffer P1 (with RNase) was added to the falcon tubes and mixed well with to re-suspend the bacterial pellet.
- D. 750 μ L of lysis buffer P2 was added and mixed gently for ~5 minutes.
- E. Added 1 mL of buffer N3 to stop the lysis reaction. The solution became transparent blue from milky white.
- F. The reaction mixture was then equally distributed among several 1.5 mL Eppendorf tubes, and centrifuged at ~14,000 g for 15 minutes.
- G. A white pellet was deposited at the bottom of the tubes while the supernatant contained the plasmid DNA.
- H. The supernatant was carefully transferred to new Eppendorf tubes, and then carefully applied to spin columns.

- I. It was centrifuged at ~14,000 g for 30 seconds and the flow-through was discarded.
- J. 750 μ L of wash buffer PE was applied to the spin columns and centrifuged at ~14,000 g for 1 minute.
- K. The flow through was discarded and the spin column were centrifuged again at ~14,000 g for 1 minute.
- L. The plasmids trapped on the silica membrane of the spin columns were eluted with 40 μ L of 10 mM HEPES buffer (pH 7).

Figure 7.3 Linear digestion of *pRST5* plasmid

SYBR gold-stained 1.3% agarose gel in 0.4x TAE buffer run at 80 V for 1 hour. Lane 1 shows GeneRuler 1 kb DNA Ladder with 14 markers from 250 bp to 10 kbp. Lane 2 shows plasmid cut with NotI-HF. Lane 3 shows XhoI digested product of DNA shown in lane 2. Lane 4 shows the undigested pRST5 plasmid.

The purified plasmids were cut with XhoI and/or NotI and visualized using native Agarose gel electrophoresis (Figure 7.3), from which several conclusions were drawn. Before linear digestion, the plasmids migrate into 5 different bands. The lowest two bands correspond to ~6 kbp and ~10 kbp, which appear to be a minority species. The remaining 3 bands are all over 10 kbp. After linear digestion using NotI-HF, 5 well-defined bands could be observed, of which the most prominent bands correspond to 3 kbp and 6 kbp. The 3 kbp band actually lies between 3 kbp and 3.5 kbp and consists of two closely spaced bands (Figure 7.4), which likely correspond to two classes of the full length pRST5 plasmid. A small fraction of plasmid seems to be non-digested which is expected since half amount of enzyme was used than required. The two new bands near 10 kbp are likely due to concatemers containing 2, 3 and 4 full-length linearized DNA linked with sticky ends generated by NotI. After digestion of the sample in lane 2 with XhoI, only 3 bands can be seen. The most prominent band lies between 2.5 kbp and 3 kbp, which is in line with the expected length of 2888 bp without the telomeric part. Two faint bands appear with lengths corresponding to ~250 bp and ~400 bp. If the plasmid contained 96 repeats, the result of double digestion would have given 640 bp linear DNA instead of these. 250 bp and 400 bp should correspond to ~26 and ~52 repeats of TTAGGG. The species corresponding to 400 bp could be some kind of dimer of the 26-repeat telomeric DNA formed by intermolecular G-quadruplexes.

Figure 7.4 *Length variation in linearized pRST5*

A cropped region from the gel image shown in previous figure, showing existence of two closely spaced bands when the pRST5 plasmid is digested with NotI-HF.

For verification of the number of telomeric repeats, the full-length plasmid was sent to 1st BASE for Sanger-based DNA sequencing. The sequencing results were aligned using commercial DNA alignment software (SnapGene) with the plasmid DNA sequence we previously received from Prof. Griffith's lab. Using T7 promoter as the sequencing primer, three different sequencing results were generated from three different batches of plasmid, one of which was extracted from *E.Coli* more than a year ago. The alignment results clearly indicated that a major part of the telomeric repeat region was missing (Figure 7.5). Using two other sequencing primers it was found that except for the telomeric-repeat part, the remaining DNA sequence of pRST5 plasmid was intact. Careful investigation of the sequencing chromatogram generated using T7 sequencing primer confirmed the existence of 26 repeats of TTAGGG as was predicted from the Agarose gel results (Figure 7.6). We suspect that the loss of telomeric repeats is a consequence of mutation introduced by several generations of *E.Coli* over 7 years since we originally received the plasmid with 96 repeats of TTAGGG. So recently, we requested Prof. Griffiths for a fresh batch of pRST5 plasmid.

Figure 7.5 Multiple sequence alignment

The sequencing results were aligned with the original plasmid DNA sequence. The alignments in dark red are from the T7 sequencing primer, while the ones in green and blue are from two opposite sequencing primers chosen from the extreme ends of the Ampicillin resistant gene. A significant portion of the mammalian telomere has gone missing.

Figure 7.6 Reproducible sequencing chromatogram

Fluorescence chromatogram obtained using T7 promoter sequencing primer, generated by reading of fluorescent nucleotides that get incorporated via polymerase chain reaction during Sanger-based sequencing. The excellent signal-to-noise ratio and absence of spectral overlap is a direct indication of the reliability of the three sequencing results, all of which reveal 26 repeats of TTAGGG.

7.3. Transcription with T7 RNA polymerase

The telomeric DNA template for the transcription reaction was generated using polymerase chain reaction with Pfu polymerase and primers as indicated in the plasmid map (Figure 7.1). 50 μ L of the transcription reaction was setup separately for both the 26 and 96 repeat telomeric DNA template, containing 2 μ g of template DNA, 0.5 mM of rNTPs and 100 units of T7 RNA polymerase. The reaction also contained 1x RNA polymerase reaction buffer from NEB (pH 7.9 @ 25°C) comprised of 40 mM Tris-HCl, 6 mM MgCl₂, 2 mM Spermidine and 1 mM dithiothreitol. The reaction was carried out for 6 hours at 37°C and then stored overnight at 4°C. The next day, 5 μ L (20 units) of DNaseI (RNase free) was added to the 50 μ L reaction volume, and was incubated at 37°C for 45 minutes. It was then

purified using RNase free spin columns (purchased from Epoch Life Science) by the following steps developed/modified intuitively by me:

- A. Added 45 μ L of RNase free water to 55 μ L of unpurified transcripts.
- B. Added 600 μ L of buffer PB (Qiagen) and applied to spin columns.
- C. Centrifuged it at 10,000 g for 30 seconds.
- D. Re-applied the flow-through to the same spin column.
- E. Centrifuged it at 10,000 g for 30 seconds.
- F. Discarded the flow-through.
- G. Washed twice with 500 μ L buffer RPE (Qiagen) by centrifugation at 10,000 g for 30 seconds.
- H. Centrifuged the empty spin columns at 16,000 g for 2 minutes to dry it.
- I. Added 40 μ L of RNase free water to each spin column.
- J. After 5 minutes, centrifuged it at 16,000 g for 2 minutes to elute RNA.

The RNA concentration of purified transcripts was estimated using Nanodrop™. About 5 μ g and 9 μ g of RNA was obtained from 2 μ g of 96-repeat and 26-repeat telomeric DNA templates respectively. Extreme care was taken at all steps to prevent RNase contamination. The quality of purified transcripts was checked using native Agarose gel electrophoresis in which RNA was denatured. TERRA was denatured in 7.3 M urea by heating in boiling water for about 10 minutes. It was then cooled rapidly and stored in 4°C for use in gel electrophoresis. The RNA ladder was denatured in 1.8X RNA loading dye by heating in boiling water for about 10 minutes. 1X RNA loading dye contains 47.5% Formamide, 0.01% SDS, 0.01% bromophenol blue, 0.005% Xylene Cyanol and 0.5mM EDTA. Formamide and Urea are excellent denaturing agents for both DNA and RNA and prevents hydrogen bonding between nucleobases once they are separated by heating. With information from several scientific sources from the internet, it was found that even though it is better to use Formamide agarose gel for denaturing gel electrophoresis, it usually

takes at least a few hours for RNA to renature back in non-denaturing conditions if it was previously denatured in high concentrated solution of Urea or Formamide. This was found to be true at least for the RNA ladder with six well-distinguishable bands as expected from a denaturing gel, which would otherwise be heavily smudged since long RNA can adopt various secondary structures in non-denaturing conditions.

Figure 7.7 *Gel visualization of telomeric-repeat containing DNA/RNA*

SYBR gold-stained 1.7% agarose gel in 0.4x TAE buffer run at 80 V for 45 minutes. Lane 1 shows dsDNA ladder of 250 bp to 10 kbp. Lane 2 shows dsDNA ladder of 100 bp to 1 kbp. Both ladders have 0.5 kbp and 1 kbp markers in common. Lane 3 shows PCR product of 96-repeat telomeric DNA. Lane 4 shows PCR product of 26-repeat telomeric DNA. Lane 5 shows denatured transcribed product using DNA template shown in lane 3. Lane 6 shows denatured transcribed product using DNA template shown in lane 4. Lane 7 shows denatured ssRNA ladder with markers at 50, 80, 150, 300, 500 and 1000 bases.

From the gel image (Figure 7.7), it can be seen that the DNA templates used for transcription consist of varying DNA lengths, but a band near 1 kbp and 0.5 kbp can be identified for the 96-repeat and 26-repeat telomeric DNA respectively. As per our primer design, the DNA template is supposed to be of 972 bp with 96 telomeric repeats and 552 bp with 26 telomeric repeats. This agrees well with the observed bands in gel. The smudging is very unlikely due to non-specificity of the PCR primers as no well-defined bands are observed other than those expected at 0.5 kbp and 1 kbp. Also, the PCR primers were designed with extreme care using a program written in Mathematica (section 2.2) that was specially developed to ensure high melting temperature and uniqueness. The smudging was most likely due to the complications caused by the telomeric repeat region as when a 469 bp non-telomeric segment of the pRST5 plasmid was amplified using PCR with primers of similar melting temperatures ($\sim 62^{\circ}\text{C}$), no smudging was observed (Figure 7.8).

The expected lengths of transcribed RNA for 96 and 26 telomeric repeats were 835 and 415 bases respectively. From the gel image, the two bands corresponding to 300 and 500 bases of ssRNA for both 96 and 26 telomeric repeats, are unexpectedly shorter. This could be due to several reasons out of which, incomplete transcription or incomparable migration rate of ssRNA ladder seem to be the most likely. In any case, the company (NEB) from which the RNA ladder was bought warned of the limited accuracy for determination of RNA lengths. At this point, we were unsure of what exactly the transcription generated, but it had to be RNA or very low concentrations of DNA that were somehow resistant to DNase degradation.

Figure 7.8 *Non-telomeric PCR amplification of pRST5 plasmid*

Lane 1 shows dsDNA ladder of 250 bp to 10 kbp. Lane 2 shows pRST5 plasmid linearized with XhoI. Lane 3 shows the PCR product of non-telomeric segment of pRST5 plasmid.

7.4. AFM imaging of TERRA in solution

The structures obtained from transcription of the PCR product using the DNA template containing 26 repeats of TTAGGG, were visualized using AFM in aqueous solution. The AFM images revealed wires, rods and blocks with specific heights but variable lengths (Figure 7.9). Some of the highly flexible wires (shown in purple color) had height of 1.3 ± 0.3 nm; let's call them TERRA I. The remaining molecules that could be seen in the AFM images were much higher (Figure 7.10) indicating that none of them were duplex DNA, ssDNA or ssRNA. Most of these molecules (shown in red-orange) had a height of 3.0 ± 0.4 nm, similar to the height of DNA and RNA G-wires investigated in the previous chapters. Let's refer to this major class of molecules as TERRA II.

Figure 7.9 *Solution AFM images of TERRA*

Topography images of TERRA obtained by in vitro transcription of PCR amplified 26-repeat telomeric DNA template. The images show two different types of objects that differ significantly in height, length and flexibility. Length scale (white bar) of all the four images is 70 nm. The images were obtained in AM-FM mode in aqueous solution with 5 mM KCl on BPL mica.

Figure 7.10 Height histograms for TERRA

Height histograms obtained from spline of all detected objects in several AFM images of TERRA in aqueous solution. In each histogram, two distinct species can be clearly identified with consistent height difference of ~ 1.3 nm. The species with ~ 1.3 nm is supposed to correspond to TERRA I.

Apart from differences between TERRA I and TERRA II observed in topography images in solution, striking differences were also observed in mechanical property contrast like phase and dissipation energy (Figure 7.11). Such observable differences helped to re-affirm that TERRA I and TERRA II are two different species formed by G-rich telomeric transcripts.

The cantilever phase is directly related to the loss tangent of the sample. The phase image can thus give an indirect estimation of the viscoelasticity of the TERRA molecules. With information from the phase and dissipation energy channels, it can be said that TERRA II is significantly stiffer than TERRA I.

Figure 7.11 *Mechanical property contrast between TERRA I and TERRA II*

Phase and dissipation energy images for TERRA obtained by in vitro transcription of PCR amplified 26-repeat telomeric DNA template. The mechanical property contrast of TERRA I and TERRA II is clearly identifiable in these images. Length scale (white bar) of all the four images is 70 nm. The images were obtained in AM-FM mode in aqueous solution with 5 mM KCl on BPL mica.

7.5. Intricate structural details of TERRA

In some of AFM images of TERRA obtained in solution, intricate periodic features were observed which differed for TERRA I and TERRA II (Figure 7.12). TERRA I showed very prominent bead-like features with 1.3 ± 0.3 nm height separated by 9 ± 2.7 nm. The distances between the beads were neither completely random nor very systematic, as can be expected from a single strand of RNA stretched by varying amounts but within certain limits. TERRA II molecules had a height of 3.0 ± 0.4 nm and appeared to be more rigid with intricate surface features that were not very deep like those of TERRA I, but were more periodic. The pattern resembled the minor and major grooves of duplex DNA (Figure 5.12) except for the exact periodicity (Figure 3.17). The minor grooves of TERRA II have a width of 1.8 ± 0.3 nm, while the major grooves have a width of 2.2 ± 0.5 . We do not use the term “grooves” in the same sense as the actual minor and major grooves of duplex DNA. It is just that TERRA II has two kinds of consecutive periodicities.

Figure 7.12 Periodical features of TERRA

High-resolution solution AFM images of (A) TERRA I and (B) TERRA II showing periodic features. The length scale (white bar) in each image is 10 nm. (C) Height profile for the highlighted section of TERRA I in the above AFM image, showing well-defined peaks separated by 6 nm to 12 nm. (D) Height profile for the highlighted section of TERRA II in the above AFM image, showing well-defined peaks that seem to be like minor and major grooves which are slightly wider than those of duplex B-DNA. The individual peak-to-peak distances are shown in blue. The horizontal and vertical scale represent length and height in nm respectively. The length scale (white bar) is 10 nm.

Our single-stranded TERRA construct is supposed to consist of 26 repeats of UUAGGG with two arms of 58 nt and 201 nt on either side. Usually, full-length (or the longest) TERRA I molecules are about ~250 nm in contour length as measured with a special program developed with Mathematica. Such a long contour length suggests that the nucleobases are stretched to ~0.6 nm on average.

The distance between the beads in TERRA I usually varies between 6 to 12 nm, which suggests that the linkers are longer than just r[UUA], because it is highly unlikely for 3 nucleobases to be stretched to 6 nm. So, even Guanines are involved in the linkers that connect the beads.

If a single bead is a 3-layered G-quadruplex of r[(GGGUUA)₄], we should see a maximum of 6 beads in AFM images of TERRA I, but in reality, much more beads are observed. This would indicate that the beads are not composed of 3-layered G-quadruplexes. It is possible that the beads are either a mixture of 2 and 3-layered G-quadruplexes, or not G-quadruplexes at all. The beads could also be GGG hairpins with Hoogsteen bonding, which could explain why we see so many beads from just 26 telomeric repeats.

As far as TERRA II is concerned, it is likely composed of 3 or 4 single strands of TERRA that form intermolecular parallel triplexes or G-quadruplexes via non-canonical hydrogen bonding between similar or different nucleobases. Even though TERRA II has the height of a G-wire, its periodic features are nowhere close to the uniform 4 nm left-handed ridge-like or zigzag periodic features we saw for G-wires. In fact, its periodic features resemble more to duplex B-DNA except for the slightly larger gaps, but it is not dsDNA.

7.6. Discussions and future work

We find that long mammalian telomeric repeats are difficult to amplify using in-vitro PCR. *E.Coli* is supposed to do a better job in replication compared to Pfu polymerase, however we found that a significant number of telomeric repeats were lost over several years. With DNA sequencing and DNA length estimation, it was

found that there existed only 26 repeats of TTAGGG, instead of 96 repeats as was expected.

Transcription of PCR amplified template DNA yielded RNA containing 26 repeats of UUAGGG. When the transcribed products were visualized using denaturing Agarose gel, only one prominent band was observed corresponding to 300 bases. However, AFM imaging in solution containing K^+ , revealed two different types of structures that differed drastically in terms of height, length, flexibility and periodic features. We broadly classified these as TERRA I and TERRA II.

In future, more high-resolution AFM images could be analyzed with the program I have developed for statistical analysis of bead-to-bead distances of TERRA I, and the length of the 58 nt and 201 nt arms of ssRNA. Better high-resolution AFM images of TERRA II could help narrow down the standard deviation in measured groove widths. A control RNA of similar length as TERRA, but not containing any G-quadruplexes (for sure) should be imaged with AFM to see if the “pearls” in the “necklace” are indeed G-quadruplexes.

Our findings on the periodic features on two types of TERRA could have significant biological implications. The groove-like pattern on TERRA II could provide molecular grip for a certain unknown class of proteins (may be RNA helicases) to bind and act on the TERRA G-wires. Without an undulating pattern as observed for TERRA I and TERRA II, it would be difficult for proteins/ligands to dock in a specific position, and not slip. If a structure/molecular model could be predicted for TERRA I and TERRA II, it would pave the way for smart drug design for resolving telomere-related diseases.

Chapter 8: Scope of AFM in G-quadruplex research

Different molecules can offer different types of challenges for serious investigation using AFM. We clearly see that even though both duplex DNA and G-quadruplex DNA are composed of the same four nucleobases, the difference in their secondary structure calls for development of different protocols for their immobilization on the Mica surface in water. The solution AFM imaging techniques that were developed and discussed in this thesis, lay the foundations of imaging large DNA and RNA molecules which are difficult to immobilize in aqueous solution. Even though in principle, any molecule having sufficient charge density should be easy to immobilize on a charged surface, the strength of attraction between the molecule and surface would depend on the distance between the charge centers, which can vary with the structure and orientation of the molecule. In cases where the electrostatic attraction is not strong enough, the molecules have to be imaged very gently which is extremely challenging with standard cantilevers oscillated mechanically. However, this has been recently made possible with the commercialization of high-speed AFM, photothermal excitation method and compatible ultra-short cantilevers. AFM is not just supposed to be used for measurement of shape and size, but so much more.

The AFM protocols discussed here can be generalized to any DNA/RNA, and not just G-quadruplexes. Usually for large molecules that cannot be crystallized or investigated with X-ray diffraction and NMR, Cryo-EM is the preferred choice for structure determination. My PhD work has opened a new realm in structural

biophysics and provided a promising alternative/supplement to Cryo-EM for studying polymorphic DNA/RNA structures. With some preliminary idea about the structure and by using computational techniques, it is possible to build molecular models which can then be rendered to generate AFM-like images and be compared with high-resolution AFM images at sub-nanometer level. This is a new and novel method to precisely estimate structure of higher-order DNA/RNA. With this method, it is possible to identify different minor species with population of even less than 0.1%, which would be nearly undetectable using ensemble techniques. This is extremely beneficial for investigating polymorphic molecules. However, the available force field parameters for molecular dynamics (MD) simulations are not optimized for G-quadruplex DNA/RNA, making it quite challenging and unrealistic to fold arbitrary DNA/RNA molecule(s) into G-quadruplexes.¹²⁴ With high-resolution AFM imaging and molecular modelling, it is possible to predict the structure(s) of highly polymorphic G-quadruplexes with very high accuracy. It is very costly and time consuming to build hundreds of molecular models to find what is the best fit to experimental observation. However, if such a bottom-up approach for structural prediction is feasible, then it possible to determine several unknown biomolecular structures. Once one or more structural candidates have been found to match with the simulated AFM images, further computational work could be done on simulating their mechanical properties (by indentation, pulling, etc.) which could be verified with AFM results to double-confirm the hypothesized molecular structure.¹²⁵⁻¹³⁸

Biotinylated poly-lysine (BPL) was initially chosen to functionalize mica with streptavidin while ensuring a clean background, but we were unable to tightly immobilize streptavidin on the BPL-mica surface. In future if it is possible to tightly bind streptavidin to the BPL-mica surface, then it should be possible to perform nano-manipulation and force spectroscopy studies on individual G-rich molecules, and also use it for regular imaging.

Most of the scientific questions investigated with AFM, do not exploit the advantage of solution imaging mainly because the challenges often outweigh the

benefits. With Atomic Force Microscopy in solution, several kinds of scientific questions related to DNA, RNA and proteins can be answered:

- (A) Determination of physical aspects like shape and size.
- (B) Determination of high-resolution structure.
- (C) Investigation of structural rigidity and integrity.
- (D) Study effect of small molecules like drugs, peptides, ATP, etc.
- (E) Observation of nanoscale changes due to self-dynamics or interaction with other molecules.
- (F) Study electrochemical interactions at nanoscale.
- (G) Study fundamental electronic properties.
- (H) Determination of surface charge distribution.
- (I) Search for pockets/sites with binding affinity for another object.
- (J) Investigate fluctuations in interaction forces between individual molecules or dynamics of a single molecule with μs temporal resolution.
- (K) All the above listed studies in different chemical environments, some of which may mimic a real-world situation better than the other.

In this thesis, our focus has been mainly on the first three kinds of studies that has been listed above, viz., determination of shape, size, structure and mechanical properties. But as per the list goes, so many other kinds of studies are possible. It is only left to the imagination what scientific questions can be addressed using AFM.

References

1. Ratmeyer L, Vinayak R, Zhong YY, Zon G, Wilson WD. Sequence specific thermodynamic and structural properties for DNA.RNA duplexes. *Biochemistry* **33**, 5298-5304 (1994).
2. Wang S, Kool ET. Origins of the large differences in stability of DNA and RNA helices: C-5 methyl and 2'-hydroxyl effects. *Biochemistry* **34**, 4125-4132 (1995).
3. Sarma MH, Umemoto K, Gupta G, Luo J, Sarma RH. In search of a Hoogsteen base paired DNA duplex in aqueous solution. *J Biomol Struct Dyn* **8**, 461-482 (1990).
4. Miyoshi D, Karimata H, Sugimoto N. Hydration Regulates Thermodynamics of G-Quadruplex Formation under Molecular Crowding Conditions. *Journal of the American Chemical Society* **128**, 7957-7963 (2006).
5. Mandelkern M, Elias JG, Eden D, Crothers DM. The dimensions of DNA in solution. *Journal of molecular biology* **152**, 153-161 (1981).
6. Wing R, *et al.* Crystal structure analysis of a complete turn of B-DNA. *Nature* **287**, 755-758 (1980).
7. Cook PR. The organization of replication and transcription. *Science* **284**, 1790-1795 (1999).
8. Kuryavyi VV, Jovin TM. Triad-DNA: a model for trinucleotide repeats. *Nat Genet* **9**, 339-341 (1995).
9. Guschlbauer W, Chantot J-F, Thiele D. Four-Stranded Nucleic Acid Structures 25 Years Later: From Guanosine Gels to Telomer DNA. *Journal of Biomolecular Structure and Dynamics* **8**, 491-511 (1990).

10. Kettani A, *et al.* A dimeric DNA interface stabilized by stacked A · (G · G · G · G) · A hexads and coordinated monovalent cations¹¹ Edited by P. E. Wright. *Journal of molecular biology* **297**, 627-644 (2000).
11. Phan AT, Kuryavyi V, Luu KN, Patel DJ. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K⁺ solution. *Nucleic Acids Res* **35**, 6517-6525 (2007).
12. Lim KW, Ng VC, Martin-Pintado N, Heddi B, Phan AT. Structure of the human telomere in Na⁺ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Res* **41**, 10556-10562 (2013).
13. Nagesh N, Chatterji D. Ammonium ion at low concentration stabilizes the G-quadruplex formation by telomeric sequence. *Journal of Biochemical and Biophysical Methods* **30**, 1-8 (1995).
14. Pedroso IM, Duarte LF, Yanez G, Baker AM, Fletcher TM. Induction of parallel human telomeric G-quadruplex structures by Sr(2+). *Biochem Biophys Res Commun* **358**, 298-303 (2007).
15. Andersen O. Cellular Electrolyte Metabolism. In: *Encyclopedia of metalloproteins* (ed[^](eds Permiakov RHKVNUA). Springer (2013).
16. Gellert M, Lipsett MN, Davies DR. Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America* **48**, 2013-2018 (1962).
17. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* **33**, 2908-2916 (2005).
18. Todd AK, Johnston M, Neidle S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res* **33**, 2901-2907 (2005).
19. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* **35**, 406-413 (2007).

20. Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature chemistry* **5**, 182-186 (2013).
21. Do NQ, Lim KW, Teo MH, Heddi B, Phan AT. Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity†. *Nucleic Acids Research* **39**, 9448-9457 (2011).
22. Cáceres C, Wright G, Gouyette C, Parkinson G, Subirana JA. A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Tl⁺/Na⁺ ions. *Nucleic Acids Research* **32**, 1097-1102 (2004).
23. Heddi B, Cheong VV, Martadinata H, Phan AT. Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide–quadruplex complex. *Proceedings of the National Academy of Sciences* **112**, 9608-9613 (2015).
24. Palumbo SL, Memmott RM, Uribe DJ, Krotova-Khan Y, Hurley LH, Ebbinghaus SW. A novel G-quadruplex-forming GGA repeat region in the c-myc promoter is a critical regulator of promoter activity. *Nucleic Acids Res* **36**, 1755-1769 (2008).
25. Morgan RK, Batra H, Gaerig VC, Hockings J, Brooks TA. Identification and characterization of a new G-quadruplex forming region within the kRAS promoter as a transcriptional regulator. *Biochimica et biophysica acta* **1859**, 235-245 (2016).
26. Onel B, *et al.* A New G-Quadruplex with Hairpin Loop Immediately Upstream of the Human BCL2 P1 Promoter Modulates Transcription. *J Am Chem Soc* **138**, 2563-2570 (2016).
27. Renciuk D, *et al.* G-quadruplex formation in the Oct4 promoter positively regulates Oct4 expression. *Biochimica et biophysica acta* **1860**, 175-183 (2017).
28. Langley AR, Graf S, Smith JC, Krude T. Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq). *Nucleic Acids Res* **44**, 10230-10247 (2016).

29. Valton AL, Prioleau MN. G-Quadruplexes in DNA Replication: A Problem or a Necessity? *Trends Genet* **32**, 697-706 (2016).
30. Higa M, Fujita M, Yoshida K. DNA Replication Origins and Fork Progression at Mammalian Telomeres. *Genes (Basel)* **8**, (2017).
31. Wu CG, Spies M. G-quadruplex recognition and remodeling by the FANCI helicase. *Nucleic Acids Res* **44**, 8742-8753 (2016).
32. Heddi B, Cheong VV, Martadinata H, Phan AT. Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 9608-9613 (2015).
33. Chatterjee S, *et al.* Mechanistic insight into the interaction of BLM helicase with intra-strand G-quadruplex structures. *Nat Commun* **5**, 5556 (2014).
34. Sanders CM. Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochem J* **430**, 119-128 (2010).
35. Hanaoka S, Nagadoi A, Nishimura Y. Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities. *Protein Sci* **14**, 119-130 (2005).
36. Hwang H, Buncher N, Opresko PL, Myong S. POT1-TPP1 regulates telomeric overhang structural dynamics. *Structure* **20**, 1872-1880 (2012).
37. Opresko PL, Shay JW. Telomere-associated aging disorders. *Ageing Res Rev* **33**, 52-66 (2017).
38. Griffith JD, *et al.* Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503-514 (1999).
39. Yoshimura SH, Maruyama H, Ishikawa F, Ohki R, Takeyasu K. Molecular mechanisms of DNA end-loop formation by TRF2. *Genes Cells* **9**, 205-218 (2004).

40. Huffman KE, Levene SD, Tesmer VM, Shay JW, Wright WE. Telomere shortening is proportional to the size of the G-rich telomeric 3'-overhang. *Journal of Biological Chemistry* **275**, 19719-19722 (2000).
41. Zeng S, *et al.* Telomerase-mediated telomere elongation from human blastocysts to embryonic stem cells. *J Cell Sci* **127**, 752-762 (2014).
42. Hahn WC. Telomere and telomerase dynamics in human cells. *Curr Mol Med* **5**, 227-231 (2005).
43. Yang CW, *et al.* Telomere shortening triggers a feedback loop to enhance end protection. *Nucleic Acids Res* **45**, 8314-8328 (2017).
44. Draskovic I, Londono-Vallejo A. Telomere recombination and the ALT pathway: a therapeutic perspective for cancer. *Curr Pharm Des* **20**, 6466-6471 (2014).
45. Grobelny JV, Kulp-McEliece M, Broccoli D. Effects of reconstitution of telomerase activity on telomere maintenance by the alternative lengthening of telomeres (ALT) pathway. *Hum Mol Genet* **10**, 1953-1961 (2001).
46. A Jayaraman KGK, P Muthusamy. Telomere and Telomerase in Cancer, Telomere - A Complex End of a Chromosome. (2016).
47. Marsh TC, Vesenska J, Henderson E. A new DNA nanostructure, the G-wire, imaged by scanning probe microscopy. *Nucleic Acids Res* **23**, 696-700 (1995).
48. Zhang XY, Cao EH, Zhang Y, Zhou CQ, Bai CL. K⁺ and Na⁺-induced self-assembly of telomeric oligonucleotide d(TTAGGG)(n). *Journal of Biomolecular Structure & Dynamics* **20**, 693-701 (2003).
49. Miyoshi D, Nakao A, Sugimoto N. Structural transition from antiparallel to parallel G-quadruplex of d(G(4)T(4)G(4)) induced by Ca²⁺. *Nucleic Acids Research* **31**, 1156-1163 (2003).
50. Vesenska J, Marsh T, Henderson E, Vellandi C. The diameter of duplex and quadruplex DNA measured by Scanning Probe Microscopy. *Scanning Microscopy* **12**, 329-342 (1998).

51. Caceres C, Wright G, Gouyette C, Parkinson G, Subirana JA. A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Tl⁺/Na⁺ ions. *Nucleic Acids Res* **32**, 1097-1102 (2004).
52. Rodrigues Pontinha AD, Chiorcea-Paquim AM, Eritja R, Oliveira-Brett AM. Quadruplex nanostructures of d(TGGGGT): influence of sodium and potassium ions. *Anal Chem* **86**, 5851-5857 (2014).
53. Liu SP, Weisbrod SH, Tang Z, Marx A, Scheer E, Erbe A. Direct Measurement of Electrical Transport Through G-Quadruplex DNA with Mechanically Controllable Break Junction Electrodes. *Angew Chem Int Edit* **49**, 3313-3316 (2010).
54. Livshits GI, *et al.* Long-range charge transport in single G-quadruplex DNA molecules. *Nat Nanotechnol* **9**, 1040-1046 (2014).
55. Cohen H, *et al.* Polarizability of G4-DNA observed by electrostatic force microscopy measurements. *Nano Lett* **7**, 981-986 (2007).
56. Calzolari A, Di Felice R, Molinari E. Electronic properties of guanine-based nanowires. *Solid State Commun* **131**, 557-564 (2004).
57. Gao ZF, Huang YL, Ren W, Luo HQ, Li NB. Guanine nanowire based amplification strategy: Enzyme-free biosensing of nucleic acids and proteins. *Biosens Bioelectron* **78**, 351-357 (2016).
58. Ren W, Zhang Y, Chen HG, Gao ZF, Li NB, Luo HQ. Ultrasensitive Label-Free Resonance Rayleigh Scattering Aptasensor for Hg(2+) Using Hg(2+)-Triggered Exonuclease III-Assisted Target Recycling and Growth of G-Wires for Signal Amplification. *Anal Chem* **88**, 1385-1390 (2016).
59. Miyoshi D, Karimata H, Wang ZM, Koumoto K, Sugimoto N. Artificial G-wire switch with 2,2'-bipyridine units responsive to divalent metal ions. *J Am Chem Soc* **129**, 5919-5925 (2007).
60. Holste C, Sondermann A, Moller R, Fritzsche W. Coupling G-wires to metal nanoparticles. *DNA-Based Molecular Electronics* **725**, 53-57 (2004).

61. Marsh TC, Piras BA, Frisbie CD, Taton TA. Self-assembly and streptavidin decoration of biotinylated G-wires. *Abstr Pap Am Chem S* **227**, U439-U440 (2004).
62. Calzolari A, Di Felice R, Molinari E, Garbesi A. G-quartet biomolecular nanowires. *Appl Phys Lett* **80**, 3331-3333 (2002).
63. Zhou J, Rosu F, Amrane S, Korkut DN, Gabelica V, Mergny JL. Assembly of chemically modified G-rich sequences into tetramolecular DNA G-quadruplexes and higher order structures. *Methods* **67**, 159-168 (2014).
64. DeJesus-Hernandez M, *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245-256 (2011).
65. Renton AE, *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257-268 (2011).
66. Byrne S, *et al.* Cognitive and clinical characteristics of patients with amyotrophic lateral sclerosis carrying a C9orf72 repeat expansion: a population-based cohort study. *The Lancet Neurology* **11**, 232-240 (2012).
67. Hewitt C, *et al.* Novel FUS/TLS mutations and pathology in familial and sporadic amyotrophic lateral sclerosis. *Archives of neurology* **67**, 455-461 (2010).
68. Tripolszki K, *et al.* Genetic analysis of the SOD1 and C9ORF72 genes in Hungarian patients with amyotrophic lateral sclerosis. *Neurobiology of aging* **53**, 195 e191-195 e195 (2017).
69. Belzil VV, *et al.* Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. *Acta Neuropathol* **126**, 895-905 (2013).
70. Ciura S, *et al.* Loss of function of C9orf72 causes motor deficits in a zebrafish model of amyotrophic lateral sclerosis. *Annals of neurology* **74**, 180-187 (2013).

71. Haeusler AR, *et al.* C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195-200 (2014).
72. Brcic J, Plavec J. Solution structure of a DNA quadruplex containing ALS and FTD related GGGGCC repeat stabilized by 8-bromodeoxyguanosine substitution. *Nucleic acids research* **43**, 8590-8600 (2015).
73. Kanekura K, *et al.* Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. *Hum Mol Genet* **25**, 1803-1813 (2016).
74. Shi KY, *et al.* Toxic PRn poly-dipeptides encoded by the C9orf72 repeat expansion block nuclear import and export. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E1111-E1117 (2017).
75. Donnelly CJ, *et al.* RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron* **80**, 415-428 (2013).
76. Gendron TF, Cosio DM, Petrucelli L. c9RAN translation: a potential therapeutic target for the treatment of amyotrophic lateral sclerosis and frontotemporal dementia. *Expert Opin Ther Targets* **17**, 991-995 (2013).
77. Lee YB, *et al.* Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell reports* **5**, 1178-1186 (2013).
78. Mori K, *et al.* hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta neuropathologica* **125**, 413-423 (2013).
79. Cooper-Knock J, *et al.* Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain : a journal of neurology* **137**, 2040-2051 (2014).
80. Brcic J, Plavec J. ALS and FTD linked GGGGCC-repeat containing DNA oligonucleotide folds into two distinct G-quadruplexes. *Biochimica et biophysica acta*, (2016).

81. Fratta P, *et al.* C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci Rep* **2**, 1016 (2012).
82. Zhou B, Liu C, Geng Y, Zhu G. Topology of a G-quadruplex DNA formed by C9orf72 hexanucleotide repeats associated with ALS and FTD. *Scientific reports* **5**, 16673 (2015).
83. Conlon EG, *et al.* The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *eLife* **5**, (2016).
84. Nergadze SG, *et al.* CpG-island promoters drive transcription of human telomeres. *Rna* **15**, 2186-2194 (2009).
85. Schoeftner S, Blasco MA. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nature cell biology* **10**, 228-236 (2008).
86. Yu HQ, Miyoshi D, Sugimoto N. Characterization of structure and stability of long telomeric DNA G-quadruplexes. *J Am Chem Soc* **128**, 15461-15468 (2006).
87. Feuerhahn S, Iglesias N, Panza A, Porro A, Lingner J. TERRA biogenesis, turnover and implications for function. *FEBS letters* **584**, 3812-3818 (2010).
88. Horard B, Gilson E. Telomeric RNA enters the game. *Nature cell biology* **10**, 113-115 (2008).
89. Luke B, Lingner J. TERRA: telomeric repeat-containing RNA. *The EMBO journal* **28**, 2503-2510 (2009).
90. Luke B, Panza A, Redon S, Iglesias N, Li Z, Lingner J. The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in *Saccharomyces cerevisiae*. *Molecular cell* **32**, 465-477 (2008).

91. Redon S, Reichenbach P, Lingner J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res* **38**, 5797-5806 (2010).
92. Azzalin CM, Reichenbach P, Khoriatuli L, Giulotto E, Lingner J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**, 798-801 (2007).
93. Caslini C, Connelly JA, Serna A, Broccoli D, Hess JL. MLL associates with telomeres and regulates telomeric repeat-containing RNA transcription. *Molecular and cellular biology* **29**, 4519-4526 (2009).
94. Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Molecular cell* **35**, 403-413 (2009).
95. Schoeftner S, Blasco MA. A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. *The EMBO journal* **28**, 2323-2336 (2009).
96. Randall A, Griffith JD. Structure of long telomeric RNA transcripts: the G-rich RNA forms a compact repeating structure containing G-quartets. *J Biol Chem* **284**, 13980-13986 (2009).
97. Menter P. http://www.bio-rad.com/LifeScience/pdf/Bulletin_1156.pdf.
98. Tuma RS, *et al.* Characterization of SYBR Gold nucleic acid gel stain: a dye optimized for use with 300-nm ultraviolet transilluminators. *Anal Biochem* **268**, 278-288 (1999).
99. Binnig G, Quate CF, Gerber C. Atomic Force Microscope. *Physical Review Letters* **56**, 930-933 (1986).
100. Van Noort SJT, Van der Werf KO, De Grooth BG, Van Hulst NF, Greve J. Height anomalies in tapping mode atomic force microscopy in air caused by adhesion. *Ultramicroscopy* **69**, 117-127 (1997).
101. Moreno-Herrero F, Colchero J, Baro AM. DNA height in scanning force microscopy. *Ultramicroscopy* **96**, 167-174 (2003).

102. Naeem S, Liu Y, Nie H-Y, Lau W, Yang J. Revisiting atomic force microscopy force spectroscopy sensitivity for single molecule studies. *Journal of Applied Physics* **104**, 114504 (2008).
103. Ohler B. Practical Advice on the Determination of Cantilever Spring Constants. *Veeco Instruments Inc AN94, Rev A0*.
104. Sader JE, *et al.* Spring constant calibration of atomic force microscope cantilevers of arbitrary shape. *Rev Sci Instrum* **83**, 103705 (2012).
105. Paramasivan S, Rujan I, Bolton PH. Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. *Methods* **43**, 324-331 (2007).
106. Lindsay SM, *et al.* STM and AFM images of nucleosome DNA under water. *J Biomol Struct Dyn* **7**, 279-287 (1989).
107. Hansma PK, Elings VB, Marti O, Bracker CE. Scanning tunneling microscopy and atomic force microscopy: application to biology and technology. *Science* **242**, 209-216 (1988).
108. Ido S, *et al.* Beyond the Helix Pitch: Direct Visualization of Native DNA in Aqueous Solution. *Acs Nano* **7**, 1817-1822 (2013).
109. Burnham N, Colton R, Pollock H. Interpretation of force curves in force microscopy. *Nanotechnology* **4**, 64 (1993).
110. Ma C, Chen Y, Wang T, Chu J. Atomic force microscopy force-distance curves with small amplitude ultrasonic modulation. *Scanning* **37**, 284-293 (2015).
111. Lozano JR, Garcia R. Theory of phase spectroscopy in bimodal atomic force microscopy. *Physical Review B* **79**, 014110 (2009).
112. Xu X, Melcher J, Basak S, Reifenberger R, Raman A. Compositional contrast of biological materials in liquids using the momentary excitation of higher eigenmodes in dynamic atomic force microscopy. *Physical review letters* **102**, 060801 (2009).

113. Proksch R, Yablon DG. Loss tangent imaging: Theory and simulations of repulsive-mode tapping atomic force microscopy. *Applied Physics Letters* **100**, 073106 (2012).
114. Marsh TC, Vesenka J, Henderson E. A New DNA Nanostructure, the G-Wire, Imaged by Scanning Probe Microscopy. *Nucleic Acids Res* **23**, 696-700 (1995).
115. Pyne A, Thompson R, Leung C, Roy D, Hoogenboom BW. Single-Molecule Reconstruction of Oligonucleotide Secondary Structure by Atomic Force Microscopy. *Small* **10**, 3257-3261 (2014).
116. Balagurumoorthy P, Brahmachari SK, Mohanty D, Bansal M, Sasisekharan V. Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Res* **20**, 4061-4067 (1992).
117. Case DA, *et al.* The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668-1688 (2005).
118. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of computational chemistry* **21**, 1049-1074 (2000).
119. Do NQ, Phan AT. Monomer-dimer equilibrium for the 5'-5' stacking of propeller-type parallel-stranded G-quadruplexes: NMR structural study. *Chemistry* **18**, 14752-14759 (2012).
120. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **9**, 3084-3095 (2013).
121. Nečas D, Klapetek P. Gwyddion: an open-source software for SPM data analysis. *Open Physics* **10**, 181-188 (2012).
122. Kim HJ, Taylor JP. Lost in Transportation: Nucleocytoplasmic Transport Defects in ALS and Other Neurodegenerative Diseases. *Neuron* **96**, 285-297 (2017).

123. Freibaum BD, *et al.* GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525**, 129-133 (2015).
124. Sponer J, *et al.* Folding of guanine quadruplex molecules-funnel-like mechanism or kinetic partitioning? An overview from MD simulation studies. *Biochimica et biophysica acta* **1861**, 1246-1263 (2017).
125. Best RB, Li B, Steward A, Daggett V, Clarke J. Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys J* **81**, 2344-2356 (2001).
126. Best RB, Fowler SB, Herrera JL, Steward A, Paci E, Clarke J. Mechanical unfolding of a titin Ig domain: structure of transition state revealed by combining atomic force microscopy, protein engineering and molecular dynamics simulations. *Journal of molecular biology* **330**, 867-877 (2003).
127. Bizzarri AR, Bonanni B, Costantini G, Cannistraro S. A combined atomic force microscopy and molecular dynamics simulation study on a plastocyanin mutant chemisorbed on a gold surface. *Chemphyschem* **4**, 1189-1195 (2003).
128. Zhao J, Davis JJ, Sansom MS, Hung A. Exploring the electronic and mechanical properties of protein using conducting atomic force microscopy. *J Am Chem Soc* **126**, 5601-5609 (2004).
129. Godzi MG, Tolstova AP, Oferkin IV. [Application of molecular dynamics simulation to the interpretation of atomic force microscopy data]. *Biofizika* **55**, 415-423 (2010).
130. Rico F, Su C, Scheuring S. Mechanical mapping of single membrane proteins at submolecular resolution. *Nano Lett* **11**, 3983-3986 (2011).
131. Doi K, Takeuchi H, Nii R, Akamatsu S, Kakizaki T, Kawano S. Self-assembly of 50 bp poly(dA).poly(dT) DNA on highly oriented pyrolytic graphite via atomic force microscopy observation and molecular dynamics simulation. *J Chem Phys* **139**, 085102 (2013).

132. Ido S, *et al.* Beyond the helix pitch: direct visualization of native DNA in aqueous solution. *ACS Nano* **7**, 1817-1822 (2013).
133. Rico F, Gonzalez L, Casuso I, Puig-Vidal M, Scheuring S. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations. *Science* **342**, 741-743 (2013).
134. Ido S, Kimiya H, Kobayashi K, Kominami H, Matsushige K, Yamada H. Immunoactive two-dimensional self-assembly of monoclonal antibodies in aqueous solution revealed by atomic force microscopy. *Nat Mater* **13**, 264-270 (2014).
135. Kobayashi K, *et al.* Molecular Dynamics Simulation of Atomic Force Microscopy at the Water-Muscovite Interface: Hydration Layer Structure and Force Analysis. *Langmuir* **32**, 3608-3616 (2016).
136. Rangl M, Miyagi A, Kowal J, Stahlberg H, Nimigean CM, Scheuring S. Real-time visualization of conformational changes within single MloK1 cyclic nucleotide-modulated channels. *Nat Commun* **7**, 12789 (2016).
137. Colom A, Redondo-Morata L, Chiaruttini N, Roux A, Scheuring S. Dynamic remodeling of the dynamin helix during membrane constriction. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 5449-5454 (2017).
138. Ruan Y, Miyagi A, Wang X, Chami M, Boudker O, Scheuring S. Direct visualization of glutamate transporter elevator mechanism by high-speed AFM. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 1584-1588 (2017).

Appendix

Figure A1: Native PAGE results for different DNA oligos checked for higher-order structure formation.

Figure A2: CD spectra at 25°C for different DNA oligos checked for parallel G-quadruplex topology, which is considered crucial for continuity of a wire-like structure.

Figure A3: Solution AFM image of G-wires formed by d[G4TG4]. Length scale is 30 nm.

Figure A4: Solution AFM image of G-wires formed by d[G4TCG4]. Length scale is 100 nm.