Nucleic acid structures containing g-quadruplex and duplex motifs : a structural study and potential targeting approach

Nguyen, Thi Quynh Ngoc

2018

Nguyen, T. Q. N. (2018). Nucleic acid structures containing g-quadruplex and duplex motifs : a structural study and potential targeting approach. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/73601>

<https://doi.org/10.32657/10356/73601>

**NUCLEIC ACID STRUCTURES CONTAINING G-
QUADRUPLEX AND DUPLEX MOTIFS: A
STRUCTURAL STUDY AND POTENTIAL
TARGETING APPROACH**

NGUYEN THI QUYNH NGOC

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

2017

**NUCLEIC ACID STRUCTURES CONTAINING G-
QUADRUPLEX AND DUPLEX MOTIFS: A
STRUCTURAL STUDY AND POTENTIAL
TARGETING APPROACH**

NGUYEN THI QUYNH NGOC

School of Physical and Mathematical Sciences

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

2017

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Phan Anh Tuan, for his guidance, motivation and generous support throughout my four years of PhD research. His enthusiasm, advice, and passion for research have motivated me to overcome various challenges and allowed me to grow as a researcher.

I would like to express my gratitude to Kah Wai for the suggestion, discussion and collaboration I have received from him. I would like to express my appreciation to Vee Vee for help me in review this thesis. My thanks extend to the members and alumni of the Biophysics lab Brahim, Dung, Chris, Nerea, Ranjani, Vineeth, Wanjun, Anjali, Kim, Krishnashi, Hong Anh, Fernaldo, Blaz, Derrick, Arijit, Tuan Anh, Huy, Poulomi, Gitali, Jun Kee, Yan Lin, Sin Yi for their discussion and wonderful time we had together. I thank the internship student, Nico, for his contribution to telomere junction project.

I would like to thank Nanyang Technological University for offering me the scholarship during my PhD study.

Many thanks to my close friends, Nhung and Truc, for sharing with me the joyful moments and difficulty during my PhD.

I would like to express my deepest gratitude to my parent, Nguyen Thai and Nguyen Thi Bich, for their unconditional love, understanding, and encouragements for all the time of my life. I am also grateful to my two brothers who have supported me along the way. I also want to thank my parents-in-law, for their understanding and support throughout my PhD study.

Last but not least, I thank my husband, Do Xuan Phu, for his love, understanding, support, and encouragements.

Finally, I want to thank my little son, Minh Khoi, for being cooperative and growing well during the time I write this thesis.

Table of Contents

ACKNOWLEDGEMENTS	I
TABLE OF CONTENTS	II
LIST OF FIGURES	V
ABBREVIATIONS	XI
ABSTRACT	XIII
CHAPTER 1 GENERAL INTRODUCTION	1
1.1 INTRODUCTION TO DEOXYRIBOSE NUCLEIC ACIDS (DNA).....	1
1.2 DOUBLE HELICAL DNA (DUPLEX).....	1
1.2.1 B-DNA	2
1.2.2 A-DNA	2
1.2.3 Z-DNA	3
1.3 G-QUADRUPLIX	3
1.3.1 Structural features of G-quadruplex	4
1.3.2 Biological Relevance of G-quadruplex	10
1.3.3 G-quadruplex Binding Ligands	21
1.4 QUADRUPLIX-DUPLEX HYBRIDS	24
1.5 RESEARCH OBJECTIVES.....	26
CHAPTER 2 METHODOLOGY	29
2.1 DNA SAMPLES PREPARATION.....	29
2.2 NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY	31
2.2.1 NMR Basics	31
2.2.2 Two-Dimension NMR.....	38
2.2.3 Scalar Coupling and Through-Bond Correlation	40
2.2.4 Nuclear Overhauser Effect and Through-space Correlation	41
2.3 NMR STRUCTURE COMPUTATION.....	44
2.4 CIRCULAR DICHROISM SPECTROSCOPY	44
2.5 ULTRAVIOLET ABSORPTION	46
CHAPTER 3 JOINING OF MULTIPLE DUPLEX STEMS AT A SINGLE QUADRUPLIX LOOP.....	47
3.1 INTRODUCTION.....	48

3.2	RESULT AND DISCUSSION.....	49
3.2.1	A Two-Stem Quadruplex-Duplex Hybrid Construct, <i>2S-QDHL2</i>	49
3.2.2	NMR Validation of <i>2S-QDHL2</i> Folding Topology	52
3.2.3	A Three-Stem Quadruplex-Duplex Hybrid Construct, <i>3S-QDHL2</i>	55
3.2.4	Thermal Stability of <i>2S-QDHL2</i> and <i>3S-QDHL2</i>	56
3.2.5	Structural Contexts at the Quadruplex-Duplex Junction.....	57
3.2.6	Diverse Structures of Multi-Stem Quadruplex-Duplex Hybrid Complexes.....	58
3.2.7	Biological Relevance of Multi-Stem Quadruplex-Duplex Hybrids.....	59
3.3	CONCLUSION	60
3.4	SUPPORTING INFORMATION.....	61
CHAPTER 4	DUPLEX FORMATION IN BULGES OF A G-QUADRUPLEX.....	65
4.1	INTRODUCTION.....	66
4.2	RESULT AND DISCUSSION.....	68
4.2.1	Inclusion of Duplex Hairpin within The Bulges of G-quadruplex.....	68
4.2.2	NMR Spectral Assignments of <i>B4-dx2</i>	69
4.2.3	Solution Structure of <i>B4-dx2</i>	72
4.2.4	Effects of Duplex Size on the Structure and Stability of Duplex-Bulged G-quadruplex.....	75
4.2.5	Insertion of Duplex Bulge at Different Positions within the Core of a G-quadruplex.....	78
4.2.6	Broadening the Definition of G-quadruplex Forming Sequences.....	80
4.2.7	A Unique Structure for G-quadruplex Targeting	80
4.3	CONCLUSION	81
CHAPTER 5	TELOMERIC G-QUADRUPLEX STRUCTURE FORMED AT 3' INTERFACE OF TELOMERIC DUPLEX.....	82
5.1	INTRODUCTION.....	83
5.2	RESULT AND DISCUSSION.....	85
5.2.1	Effect of Terminal Bases in Telomeric C-rich Strand.....	85
5.2.2	Effects of Duplex Hairpin Length.....	86
5.2.3	NMR Assignment of <i>TJ-5</i>	88
5.2.4	Effects of Mutation on <i>TJ-5</i> Sequence.....	89
5.2.5	NMR Elucidation of <i>TJ-5</i> Folding Topology.....	92
5.2.6	Truncated four-repeat telomere sequence adopts two-layer G-quadruplex	94
5.2.7	Potential Molecular Target for Small Molecules	96
5.3	CONCLUSION	96

CHAPTER 6 A DUAL-SPECIFIC TARGETING APPROACH BASED ON THE SIMULTANEOUS RECOGNITION OF DUPLEX AND QUADRUPLEX MOTIFS.....	97
6.1 INTRODUCTION	99
6.2 RESULT AND DISCUSSION.....	100
6.2.1 Binding of Quadruplex-Duplex Hybrid with Duplex Ligand	100
6.2.2 Binding of Quadruplex-Duplex Complex with Quadruplex Ligand.....	102
6.2.3 Simultaneous Binding of Quadruplex-Duplex Hybrid with Duplex and Quadruplex Ligands	103
6.2.4 Potential Targeting Strategy to Achieve High Specificity toward G- quadruplex.....	109
6.3 CONCLUSION	113
CHAPTER 7 SUMMARY AND FUTURE PERSPECTIVES	114
BIBLIOGRAPHY	122

List of Figures

Chapter 1 General Introduction

Figure 1-1	: Structure of nucleotides.	1
Figure 1-2	: C2'-endo (left) and C3'-endo (right) sugar pucker conformation.	2
Figure 1-3	: Double helical DNA structure.	3
Figure 1-4	: Schematic showing a G-tetrad formed by association of four Hoogsteen hydrogen-bonded guanines and a monovalent cation coordinated with O6 oxygen. Stacking of G-tetrads gives rise to a G-quadruplex.	4
Figure 1-5	: <i>anti</i> and <i>syn</i> glycosidic conformation of guanine.	5
Figure 1-6	: Strand stoichiometry of G-quadruplex DNA.	5
Figure 1-7	: Strand orientation within the G-quadruplex core.	6
Figure 1-8	: Three general types of loop in the G-quadruplex.	6
Figure 1-9	: Schematic showing different groove dimensions of a G-quadruplex.	7
Figure 1-10	: Schematic showing G-quadruplex containing a snapback.	8
Figure 1-11	: Schematic showing the formation of bulges within different G-quadruplexes.	9
Figure 1-12	: Schematic structure and ribbon view of Z-G4 adopted by a variant of AGRO100.	9
Figure 1-13	: BG4 antibody staining on the chromosomes at metaphase showed G-quadruplex formation across chromosomes isolated from HeLa cervical cancer cells.	11
Figure 1-14	: G-quadruplex formation is shown by staining cells with 1H6 antibody.	11
Figure 1-15	: Shelterin complex of human telomere.	12

Figure 1-16	: Schematic showing how the G-quadruplex formation inhibits telomerase activity	13
Figure 1-17	: Folding topology of human telomeric G-quadruplexes adopted by different sequences in different monovalent ionic conditions.	15
Figure 1-18	: Chemical structure of ³ H-360A small molecule.	15
Figure 1-19	: Schematic showing G-quadruplex formation at the promoter region regulates gene transcription.	16
Figure 1-20	: Schematic showing that G-quadruplexes form during replication and impede the movement of DNA polymerase.	19
Figure 1-21	: Schematic showing that G-quadruplex formation is necessary for replication initiation.	21
Figure 1-22	: Stick representation showing π - π stacking between G-tetrads and ligands in G4-ligand complexes.	22
Figure 1-23	: Stick representation showing examples on hydrogen bonds formed between G-quadruplexes and ligands in G4-ligand complexes.	23
Figure 1-24	: Schematic showing five quadruplex-duplex hybrids constructs.	25
Figure 1-25	: Two different orientation between a duplex and a quadruplex.	25
Chapter 2 Methology		
Figure 2-1	: The cycle of oligonucleotide synthesis.	30
Figure 2-2	: The net magnetic vector of nuclei at equilibrium state, after a pulse is applied in x direction and precession to equilibrium state.	33
Figure 2-3	: Schematic illustrates the rotation of magnetization vector after 90° and 180° pulse.	34
Figure 2-4	: Fourier transform is a mathematical process which converts a function of time (FID signal, left) to a function of frequency (NMR spectrum, right)	35

Figure 2-5	: Dephasing of transverse magnetization.	36
Figure 2-6	: Chemical shift pattern of nucleic acid protons.	37
Figure 2-7	: Schematic of a 2D homonuclear NMR spectrum for two protons A and B	38
Figure 2-8	: General scheme of 2D NMR pulse sequence.	39
Figure 2-9	: Different energy states of a two-spin system.	40
Figure 2-10	: A typical pulse sequence of NOESY experiment.	42
Figure 2-11	: Typical NOE signal of G-tetrad and Watson-Crick base pairs.	42
Figure 2-12	: Schematic showing characteristic NOE cross peaks of Watson-Crick base pairs and G-tetrads.	43
Figure 2-13	: Examples of CD spectra pattern of different G-quadruplex conformations.	45

Chapter 3 Joining of Multiple Duplex Stems at a Single Quadruplex Loop

Figure 3-1	: Coordination of multiple duplex stems through a single loop of a G-quadruplex scaffold.	50
Figure 3-2	: 1D imino proton NMR spectra of <i>2S-QDHL2</i> , <i>1S-QDHL2</i> , <i>dx1</i> , and <i>dx2</i> in K ⁺ solution.	50
Figure 3-3	: CD spectra of <i>2S-QDHL2</i> (blue) and <i>3S-QDHL2</i> (red) at 25 °C in buffer containing 20 mM KCl and 20 mM KPi.	51
Figure 3-4	: Guanine imino proton assignments of two-stem quadruplex-duplex hybrid construct <i>2S-QDHL2</i> in K ⁺ solution.	53
Figure 3-5	: G-quadruplex folding topology of two-stem quadruplex-duplex hybrid construct <i>2S-QDHL2</i> in K ⁺ solution, as determined by site-specific ² H-labeling.	54
Figure 3-6	: Folding topology of three-stem quadruplex-duplex hybrid construct <i>3S-QDH</i> in K ⁺ solution.	55

Figure 3-7	: CD melting of <i>2S-QDHL2</i> DNA at different concentrations strands (2–100 μ M) in solution containing 50 mM K^+ , as monitored through the CD spectra at 262 nm.	57
Figure 3-8	: 1D imino proton NMR spectra of <i>2S-QDHL1</i> , <i>2S-QDHL3</i> , <i>3S-QDHL1</i> , and <i>3S-QDHL3</i> .	59
Figure S3-1	: NMR structural characterization of <i>dx2</i>	61
Figure S3-2	: NOESY spectrum (mixing time, 300ms) of <i>dx2</i> in K^+ solution showing H8/H6-H1' connectivity	62
Figure S3-3	: Solution structure of <i>dx2</i> in K^+ solution	62
Figure S3-4	: CD melting of <i>dx1</i> , <i>dx2</i> and <i>T10</i> in solution containing 1 mM KPi and 100 mM LiCl showing that the melting temperature of quadruplex <i>T10</i> is lower than melting temperature of duplex <i>dx1</i> and <i>dx2</i> .	63
Chapter 4	Duplex Formation in the Bulges of G-quadruplex	
Figure 4-1	: Formation of bulge contains single nucleotide and a duplex motif.	68
Figure 4-2	: 1D imino proton NMR and CD spectra of <i>B4-dx2</i> , <i>B4-T</i> and <i>dx2</i> in K^+ solution.	69
Figure 4-3	: Guanine imino and aromatic protons assignment of <i>B4-dx2</i> construct in K^+ solution using 4% ^{15}N -labeled samples.	71
Figure 4-4-	: NOESY spectrum (mixing time, 300ms) of <i>B4-dx2</i> in K^+ solution showing H8/H6-H1' connectivity.	72
Figure 4-5	: NMR structural characterization of <i>B4-dx2</i> .	73
Figure 4-6	: Solution structure of <i>B4-dx2</i> in K^+ solution: ten superimposed lowest energy structures and ribbon view of a representative structure.	75
Figure 4-7	: Stacking between T19 and G20.	75
Figure 4-8	: 1D imino proton NMR spectra of G-quadruplex containing duplex bulge of different size in K^+ solution.	76

Figure 4-9	: CD spectra of <i>B4-dx2</i> , <i>B4-dx3</i> and <i>B4-dx4</i> in buffer containing 30 mM KCl and 20 mM KPi.	77
Figure 4-10	: CD melting of G-quadruplexes containing duplex bulges of different size: <i>B4-dx2</i> (blue), <i>B4-dx3</i> (red), <i>B4-dx4</i> (green), and <i>B4-15T</i> (pink) in 60 mM K ⁺ solution.	77
Figure 4-11	: Schematic showing the formation of duplex bulge at different positions.	78
Figure 4-12	: 1D imino proton NMR spectra of quadruplex-duplex hybrids in K ⁺ solution with a duplex inserted at different bulge positions: <i>B3-dx2</i> , <i>B4-dx2</i> , <i>B7-dx2</i> , <i>B8-dx2</i> , <i>B12-dx2</i> , and <i>B16-dx2</i> .	79
Figure 4-13	: CD spectra of <i>B3-dx2</i> , <i>B4-dx2</i> , <i>B7-dx2</i> , <i>B8-dx2</i> , <i>B12-dx2</i> , and <i>B16-dx2</i> .	79
Chapter 5	Telomeric G-quadruplex Structure Formed at 3' Interface of Telomeric Duplex	
Figure 5-1	: Schematic structure of five reported human telomeric G-quadruplexes adopted by different sequences	84
Figure 5-2	: 1D imino proton NMR spectra of <i>TJ-1</i> , <i>TJ-2</i> , <i>TJ-3</i> , and <i>TJ-4</i> in K ⁺ solution.	86
Figure 5-3	: 1D imino proton NMR spectra of <i>TJ-7</i> , <i>TJ-6</i> , <i>TJ-5</i> , <i>TJ-4</i> and strand A + Strand B in K ⁺ solution.	87
Figure 5-4	: NMR spectral assignments of <i>TJ-5</i> in K ⁺ solution.	88
Figure 5-5	: 1D imino proton NMR spectra of single G-to-T mutations on <i>TJ-5</i> sequence in K ⁺ solution.	90
Figure 5-6	: 1D imino proton NMR spectra of single G-to-I mutations on <i>TJ-5</i> sequence in K ⁺ solution: <i>TJ-5</i> , I17, I25, I29, and I37.	91
Figure 5-7	: Determination of G-quadruplex folding topology of <i>TJ-5</i> in K ⁺ solution by site-specific ² H-labeling.	92
Figure 5-8	: Schematic of the quadruplex-duplex hybrid structure adopted by <i>TJ-5</i> sequence.	93
Figure 5-9	: 1D imino proton NMR spectra of <i>TJ-5</i> , T17, I17 and <i>Telo22-TT</i> in K ⁺ solution.	95

Figure 5-10	: CD spectra of <i>Telo22-TT</i> (blue) and ^{Br} G7-form3 (red) in K ⁺ solution.	95
Chapter 6	A Dual-Specific Targeting Approach Based on the Simultaneous Recognition of Duplex and Quadruplex Motifs	
Figure 6-1	: Schematic diagram of <i>QDHI</i> .	100
Figure 6-2	: NMR spectra show the binding of <i>QDHI</i> to netropsin.	101
Figure 6-3	: NMR spectra show the binding of <i>QDHI</i> to Phen-DC ₃ .	102
Figure 6-4	: NMR spectra show the binding of <i>QDHI</i> to netropsin and Phen-DC ₃ .	103
Figure 6-5	: 2D NMR NOESY spectrum of <i>QDHI</i> :netropsin:Phen-DC ₃ complex.	104
Figure 6-6	: Structural model of <i>QDHI</i> :netropsin:Phen-DC ₃ complex.	105
Figure 6-7	: Melting curve of dx hairpin in the absence and presence of ligands detected by CD signal at 267 nm.	105
Figure 6-8	: CD spectra and quadruplex melting curves detected by CD at 256nm of <i>QDHI</i> in the absence and presence of ligands.	106
Figure 6-9	: NMR spectra show the binding of <i>QDH2</i> to netropsin and Phen-DC ₃ .	107
Figure 6-10	: NMR spectra show the binding of <i>QDH3</i> to netropsin and Phen-DC ₃ .	108
Figure 6-11	: NMR spectra show the binding of <i>QDHI</i> to netropsin and BRACO-19.	109
Figure 6-12	: NMR spectra show binding of <i>QDHI</i> to netropsin and pyridostatin.	110
Figure 6-13	: Examples of chemical linker that can be utilized to join a duplex-binding ligand and a quadruplex-binding ligand.	111
Figure 6-14	: Examples of structural and biological contexts in which both duplex and quadruplex elements exist in close spatial proximity.	112

Abbreviations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
A	adenine
^{Br} G	8-bromoguanine
bp	base pairs
CD	circular dichroism
COSY	correlation spectroscopy
C	cytosine
DAPI	4',6-diamidino-2-phenylindole
DGSA	distance geometry simulated annealing
DNA	deoxyribonucleic acid
ds	double-stranded
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
FID	free induction decay
G	guanine
G4	G-quadruplex
G-tetrad	G•G•G•G tetrad
HMBC	heteronuclear multiple-bond correlation
HPLC	High performance liquid chromatography
HSQC	heteronuclear single-quantum correlation
I	inosine

mRNA	messenger RNA
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
ppm	parts per million
PQS	putative quadruplex sequence
RNA	ribonucleic acid
T	thymine
T _m	melting temperature
TMS	tetramethylsilane
TOCSY	total correlation spectroscopy
TSS	transcription start site
UTR	untranslated region
UV	ultraviolet

Abstract

Guanine-rich nucleic acid sequences have the propensity to assume four-stranded structures known as G-quadruplex. G-quadruplex-forming sequences were identified at important regulatory regions in the human genome such as the telomere, gene promoter, and 5'-untranslated region of mRNA. Important biological processes such as telomere maintenance, transcription, translation, and replication can be influenced by the presence of G-quadruplex. Quadruplex-duplex hybrid motifs, which contain both G-quadruplex and duplex segments, are also gaining increasing interest as potential therapeutic targets, with potential application in nanotechnology. This thesis explores structural features of different types of quadruplex-duplex motifs to provide important structural knowledge for G-quadruplex-based drug design and development. Two novel quadruplex-duplex hybrids including the multi-stem-loop quadruplex and a duplex-bulged quadruplex are presented in this thesis. This study also investigated a telomeric quadruplex-duplex structure in which the G-quadruplex adopts a basket-type G-quadruplex with two G-tetrads. Additionally, a specific targeting approach toward quadruplex-duplex hybrid structures is proposed based on the simultaneous recognition of duplex and quadruplex motifs. These findings provide structural insights and potential strategies which could be useful for the future design and development of drugs for G-quadruplex targeting. In addition, novel hybrid structures discovered in this thesis could have potential application in DNA nanotechnology.

Chapter 1 General Introduction

1.1 Introduction to Deoxyribose Nucleic Acids (DNA)

Genetic information which directs the development and function of living organisms are contained within the DNA. A DNA strand is a biological polymer constituted by repeating monomer units called nucleotides. Each nucleotide is composed of three fragments: a deoxyribose, a phosphate group and a nucleobase (adenine, thymine, cytosine, and guanine) (Figure 1-1). DNA typically adopts a double helix in double-stranded DNA, and are held together by Watson-Crick hydrogen bonding.¹ Besides the double helix, DNA can adopt other non-canonical structures such as the four-stranded G-quadruplex.

1.2 Double helical DNA (duplex)

Figure 1-1. Structure of nucleotides (a) Schematic diagram of a dinucleotide chain. Each nucleotide comprises of a deoxyribose ring, a nitrogenous base, and a phosphate group, (b) adenine-thymine (A-T) base pair, and (c) guanine-cytosine (G-C) base pair.

The double helix model of DNA (duplex), first proposed in 1953 by James Watson and Francis Crick,¹ consists of two anti-parallel strands held together by Watson-Crick hydrogen bonds between bases adenine-thymine (A-T) (Figure 1-1b), and guanine-cytosine (G-C) (Figure 1-1c). The two strands are coiled around the

helical axis and are twisted around each other. Double helix can exist in various structural forms: A-form, B-form, and Z-form.

Figure 1-2. C2'-endo (left) and C3'-endo (right) sugar pucker conformation.

1.2.1 B-DNA

B-form DNA (Figure 1-3a) is a right-handed double helical structure which predominates in cells.^{2, 3} In this structural form, base pairs are stacked 3.3 to 3.4 Å apart with right-handed rotation about 36° between adjacent base pairs. This yields a double helix with 10 base-pairs per helical turn. The sugar puckering mode in B-DNA is C2'-endo and the distance between adjacent phosphates of the same strand is 7 Å.

1.2.2 A-DNA

In dehydrated condition, DNA can adopt the A-form (Figure 1-3b), a shorter and more compact helical structure with 11 base-pairs per helical turn.⁴ In A-DNA, sugar adopt C3'-endo puckering mode, and the distance between adjacent phosphates of same strand is 5.9 Å. A-form is the major conformation of ribonucleic acid (RNA)⁵ and DNA/RNA hybrids⁶. In B-DNA, the base pairs locate near the helical axis. However, in A-DNA, the helical axis tilt towards the major groove, giving rise to a deeper, narrower major groove but shallower, wider minor groove.

Figure 1-3. Double helical DNA structure: (a) Right-handed B-DNA (PDB code :1BNA) (b) Right-handed A-DNA (PDB code: 2D47) (c) Left-handed Z-DNA (PDB code: 2ZNA)

1.2.3 Z-DNA

Unlike A- and B-DNA, Z-DNA (Figure 1-3c) is a left-handed double helical structure in which the phosphate backbones winds to the left in a zig-zag fashion⁷ with 12 base-pairs per helical turn. In Z-DNA, purines adopt *syn* glycosidic conformation and C3'-endo sugar pucker while pyrimidines adopt *anti* glycosidic conformation and C2'-endo sugar pucker. Z-DNA formation can be promoted by conditions such as alternating purine-pyrimidine sequences, high salt condition, negative DNA supercoiling or protein binding.⁸

1.3 G-quadruplex

G-rich oligonucleotides have the ability to adopt four-stranded helical structures, G-quadruplex (G4) which exhibit high thermal stability and structural difference from the canonical double helix structures.⁹⁻¹² A G-quadruplex consists of multiple vertically stacked G-tetrads, with each G-tetrad containing four guanines arrange in square planar and interact with each other by Hoogsteen hydrogen bonding (Figure 1-4) which involves the hydrogen bonds between N7 and O6 with amino proton H2 and imino proton H1, respectively.¹³ The formation of G-tetrads usually occurs in the presence of monovalent cations including K^+ , Na^+ , and NH_4^+ which interact with electronegative carbonyl oxygen O6 of guanine bases in G-

tetrads.¹⁴ Stability of G-quadruplexes mainly come from the Hoogsteen hydrogen bonding of guanines, stacking interaction between G-tetrads and the coordination of cations which reduces the repulsion among carbonyl oxygen O6 of guanines. More than 370,000 potential G-quadruplex sequences have been identified across the human genome. There have been growing evidences showing the regulatory role of G-quadruplex in biological processes such as telomere maintenance, replication, transcription, translation. Due to their potential biological importance, G-quadruplexes have attracted an increasing number of research attentions in the past few decades.

Figure 1-4. Schematic showing a G-tetrad formed by association of four Hoogsteen hydrogen-bonded guanines and a monovalent cation coordinated with O6 oxygen. Stacking of G-tetrads gives rise to a G-quadruplex.

1.3.1 Structural features of G-quadruplex

The G-quadruplex structure is highly polymorphic where variations in the constituting nucleotide sequence, type and concentration of ion, temperature, and presence of co-solutes may affect the overall folding topology. Such diversity observed in the G-quadruplex structures arises from differences in the strand stoichiometry, strand directionality, glycosidic conformation of guanines in the G-tetrad, groove width and loop (the linker connect two strands of G-quadruplex) type.

Guanine glycosidic conformation

The relative orientation between a guanine base and its constituting sugar is characterized by the dihedral angle O4'–C1'–N9–C4 or chi (χ) torsion angle. Guanines within a G-tetrad can adopt two different glycosidic conformations: *syn* conformation in which the six ring member of guanine base is oriented toward the

sugar ring or *anti* glycosidic conformation in which the six ring member of guanine base is oriented away from the sugar ring (Figure 1-5).

Figure 1-5. (a) chi (χ) torsion angle, atoms and bonds involved in χ torsion angle are labeled in red color (b) *anti* glycosidic conformation of guanine and (c) *syn* glycosidic conformation of guanine.

Strand stoichiometry

The G-quadruplex structure can assemble from different G-tracts of a single oligonucleotide sequence, or from the multiple sequences. Depending on the number of oligonucleotide participating in a G-quadruplex, a G-quadruplex can be classified as monomeric, dimeric, trimeric or tetrameric consisting of one, two three or four strands, respectively (Figure 1-6).

Figure 1-6. Strand stoichiometry of G-quadruplex DNA: (a) monomer G-quadruplex, (b) dimer G-quadruplex, (c) and tetramer G-quadruplex.

Strand directionality

There are four possible strand arrangement in G-tetrad core: all-parallel (4+0) in which all four strands have the same directionality; hybrid (3+1) in which three strands orient in one direction and one strand has opposite directionality; antiparallel (2+2) in which two adjacent strands point to one direction and the rest point to the opposite direction or two diagonal strands have the same directionality while the rest have the opposite directionality (Figure 1-7). Glycoside conformations of guanines participating in G-tetrad are associated with the relative orientation of G-strand. Glycoside conformations of guanines within a G-tetrad can be (a) *anti.anti.anti.anti* (b) *syn.syn.syn.anti* or *anti.anti.anti.syn* (c) *anti.anti.syn.syn* and (d) *anti.syn.anti.syn*.

Figure 1-7. Strand orientation within the G-quadruplex core: (a) Parallel “(4+0)” in which all strands align in the same direction. (b) Hybrid (3+1) in which three strands orient in the same direction and one strand in the opposite. (c,d) Anti-parallel (2+2) in which 2 strands orient in one direction, the remaining two strands point to the opposite direction. *syn* and *anti* guanines are shown in magenta and cyan, respectively.

Loop arrangement

Figure 1-8. Three general types of loop in the G-quadruplex: (a) diagonal loop (b) edge-wise loop (c) propeller loop.

The four backbone strands (G-column) to support G-tetrads core are connected by loops. The loop of a G-quadruplex can vary in length, sequence and

geometry. Structure data available of G-quadruplexes shows that there are three common occurring loop types: diagonal loop, propeller (or double chain reversal) loop and edgewise (or lateral) loop (Figure 1-8). A diagonal loop connect two strands in opposite direction and located diagonally across the G-quadruplex. A propeller loop connect two adjacent strands point to the same direction. An edgewise loop connect two adjacent strands point to opposite directions. A loop of one nucleotide prefers to adopt propeller type and is especially stable.^{15, 16} An increase in loop length usually lead to a decrease in thermal stability of G-quadruplex.¹⁶ Stable G-quadruplexes typically contain loops with length of less than 7 nucleotides.¹⁷

Types of groove

Figure 1-9. Schematic showing different groove dimensions of a G-quadruplex. Groove dimension depends on the strand arrangement of G-tetrads. (●) indicates the 5'-3' strand directs out of the plane of paper, (+) indicates the 5'-3' strand direct into the plane of paper (This figure is taken from Prof. Phan Anh Tuan, unpublished).

A groove is the cavity between two neighboring phosphate backbones within G-tetrad core. Groove dimension is defined by the distance between matching pairs of phosphate groups. The orientation of the strands in G-tetrad core defines specific patterns of groove dimension which can be classified as wide with strand separation of ~ 19 Å, narrow with strand separation of ~ 12 Å and medium with strand orientation of ~ 16 Å (Figure 1-9).

Non-canonical G-quadruplex

A typical G-quadruplex forming sequence is defined with the following sequence motif, $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$, which requires four runs of at least three continuous guanines and the two adjacent guanine runs are separated by one to seven nucleotide. Numerous G-quadruplexes reported in literature are right handed G-quadruplexes following this sequence motif.¹⁸⁻²⁰ However, increasing number of G-quadruplex structures with unusual scaffold deviating from the classical G-quadruplex sequence motif have been observed such as G-quadruplex structures with discontinuous G-tracts and left-handed G-quadruplex.

Discontinuous G-tracts

Figure 1-10. Schematic showing G-quadruplexes containing a snapback (a) G-quadruplex from c-myc promoter and (b) G-quadruplex from c-kit promoter in K^+ solution. *Anti* and *syn* guanines are shown in blue and magenta, respectively. Figure is adapted from ref ¹².

A common feature observed in most G-quadruplex is the presence of continuous G-tracts which form continuous G-column to support a G-tetrad core. However, several non-canonical G-quadruplex structures have been reported with G-column made of discontinuous guanines. In G-quadruplex structure of c-myc and c-kit promoter, one G-column of G-tetrad core is interrupted, resulting in a snap-

back at 3'-end to complete G-tetrad formation^{21, 22} (Figure 1-10). G-quadruplexes containing bulges which connect two adjacent guanines of the same G-column were also reported (Figure 1-11).²³

Left-handed G-quadruplex

Most G-quadruplex structures are right-handed in which the 5'-3' phosphate backbone winds in clockwise direction. Recently, a left-handed G-quadruplex structure in which the 5'-3' phosphate backbone winds in anticlockwise direction was first reported by Chung et al. The left-handed G-quadruplex was formed by a variant of AGRO100 sequence which is a popular aptamer to inhibit proliferation of a wide range of cancer²⁴.

Figure 1-11. Schematic showing the formation of bulges within different G-quadruplexes. *syn* and *anti* guanines are shown in magenta and cyan, respectively. Figure is adapted from ref ²³.

Figure 1-12. Schematic structure and ribbon view of Z-G4 adopted by a variant of AGRO100. *Anti* and *syn* guanines are shown in cyan and magenta, respectively. Figure is adapted from ref ²⁴.

1.3.2 Biological Relevance of G-quadruplex

By using the consensus sequence ($G_3+N_{1-7}G_3+N_{1-7}G_3+N_{1-7}G_3$), up to 376,000 G-quadruplex forming sequences were found across the human genome.²⁵ In 2015, a high-resolution sequencing-based method to detect G-quadruplex in the human genome in presence of G4-stabilizing ligand pyridostatin (PDS) was presented, in which 716,310 distinct G-quadruplex structures were identified.²⁶ About 70% of these G-quadruplex structures were not predicted by the consensus sequence ($G_3+N_{1-7}G_3+N_{1-7}G_3+N_{1-7}G_3$) in the earlier studies. The location of G4 forming sequences are not random but tend to localize in functional regions of the genome such as the human telomere²⁷, oncogenic promoters²⁸, 5'-untranslated region of mRNA¹² and human intronic sequences²⁹. Recently, a G4 ChIP-seq protocol employing an antibody specific to G4 structure (BG4) to map endogenous G4 structures found 10,000 G4 structures in the human chromatin of HaCaT cells, predominantly located in regulatory, nucleosome-depleted regions in proximity to the transcription start sites of highly transcribed genes.³⁰ The location of G-quadruplex forming sequences at functional regions of genome suggests that G-quadruplex could play an important role in biological processes such as telomere maintenance, replication, transcription, and translation.

Detection of G-quadruplex in Cells

Although a large amount of G-quadruplex forming sequences were mapped in the human genome, it is a question where and when the PQS can assemble into G-quadruplex *in vivo*. The first direct visualization of G-quadruplex formation *in vivo* was shown in 2001 by using specific antibodies (Sty3 and Sty49) directed against telomeric DNA G-quadruplexes in *Stylonychia lemnae*.³¹ Indirect immunofluorescence staining of *Stylonychia lemnae* nuclei with these antibodies showed the formation of antiparallel telomeric G-quadruplex in the nuclei.

Figure 1-13. BG4 antibody staining on chromosomes at metaphase showed G-quadruplex formation (red foci) across chromosomes isolated from HeLa cervical cancer cells: (i-iii) non-telomeric regions and (iv-v) telomeric regions. Chromosomes were isolated from HeLa cervical cells. Chromosomes are counterstained with DAPI (blue). White scale bar represents 20 μm . Figure is adapted from ref ³².

Figure 1-14. G-quadruplex formation is shown by staining cells with 1H6 antibody. Immunofluorescence analysis of nuclear (a) and chromosome, white scale bar represents 50 μm (b) of human cells with DAPI (blue) and 1H6 (red) demonstrating the presence of G-quadruplex in human genome, red scale bar represents 10 μm . Figure is adapted from ref ³³.

Later, the presence of G-quadruplex structures in the human genome was visualized by antibody BG4³² (Figure 1-13) and 1H6³³ (Figure 1-14) which exhibited high binding affinity and specificity to G-quadruplex structures. G-quadruplexes were found at both human chromosomal ends and chromosomal interstitial sites. The quantity of G-quadruplex structures was shown to depend on the cell cycle, the presence of G-quadruplex binding ligands and the presence of G-quadruplex helicases. The number of G-quadruplex structures was at the maximum during S phase, the period when DNA is replicated. The presence of G-quadruplex binding ligand and the absence of G-quadruplex unwinding helicase led to an increase in the number of G-quadruplex structures detected in the human genome. These observations demonstrated the effect of G4 binding ligands in stabilizing the G4 structures formed in cells and the role of G4 helicases in resolving the G4 structure in cells.

G-quadruplex in Telomeres

Figure 1-15. Shelterin complex of human telomere. Shelterin complex contains six telomeric DNA binding proteins: TRF1, TRF2, POT1, TIN2, TPP1 and Rap1. (a) "open" state of telomere. Telomerase can bind and elongate the 3' overhang of telomere in "open" state (b) "closed" state of telomere.

Telomeres are nucleoprotein complexes located at the ends of the chromosome (Figure 1-15) and they play a vital role in preventing fusion and degradation at chromosome ends.^{34, 35} The human telomeric DNA is composed of a

3-15 kilobases long guanine-rich duplex and 100-200 nucleotides long single-stranded overhang segments which both consist tandem repeats of (TTAGGG).³⁶ Telomere is protected by a protein complex called shelterin containing six proteins, namely TRF1, TRF2, TIN2, TPP1, rap1, and POT1.³⁷ While TRF1 and TRF2 preferentially bind to telomeric double strand, POT1 prefers binding to single strand guanine-rich over the double strand. The telomere can exist in two states: “closed” state in which the single strand telomeric overhang invades the double strand segment to create T-loop or “open” state in which the single strand telomeric overhang is accessible.³⁸ The T-loop configuration provides the end protection since the telomere end is prevented from being recognized by the DNA damage response pathway and the DNA repair machinery. In the “open” state, the single strand overhang can interact with other proteins such as telomerase. It has been proposed that the “open” state appears during S phase of cell cycle.³⁹ In somatic cells, telomere length gradually decreases with cell division due to the difficulty in replicating the lagging telomeric DNA strand by endogenous DNA polymerase.⁴⁰ After each round of cell division, DNA telomere length is estimated to be reduced by about 50-200 bases.⁴¹ When a critical length of telomeres is reached, cells enter replicative senescence phase and ultimately apoptotic cell death.^{42, 43}

Figure 1-16. Schematic showing how the G-quadruplex formation inhibits telomerase activity.

G-rich strand of human telomeric DNA has been shown to adopt G-quadruplexes under physiological ionic conditions¹⁸ with many high resolution structures of telomeric G-quadruplex in crystal and solution well described in the

literature (Figure 1-17). Analyses of these G-quadruplex structures found the human telomeric G-quadruplex to be highly polymorphic, and the overall conformation adopted is highly dependent on their flanking sequence and buffer conditions.

In cancer cells, telomere shortening is suppressed by a reverse transcriptase named telomerase which utilizes its associated RNA template (hTR) to catalyze the addition of multiple copies of TTAGGG repeat to the 3'-end of the G-strand of the telomere in the open state.^{39, 44} Telomerases are over-expressed in majority of cancer cells, but not detected in somatic cells.⁴⁵ The RNA template (hTR) of telomerase complex recognizes the telomeric single strand overhang as a substrate. It was found that telomerase activity was inhibited by the assembly of single stranded telomeric DNA into G-quadruplexes (Figure 1-16),^{46, 47} making telomeric G-quadruplex a potential target for cancer treatment. Consequently, ligands that accelerate the formation of telomeric G-quadruplex may inhibit the telomerase activity or interfere with the telomere conformation and thereby inhibiting cancer growth. Hundreds of G-quadruplex binding ligands have been shown to bind to G-quadruplex *in vitro*.⁴⁸ *In vivo* binding of G-quadruplex ligand with telomere was shown by a study using radiolabeled G-quadruplex ligand, ³H-360A (Figure 1-18)⁴⁹ which was found to preferentially bind to telomere of human cells. Other telomeric G-quadruplex ligands including BRACO-19,⁵⁰ RHSP4⁵¹ and telomestatin⁵² have also been demonstrated to inhibit anti-cancer effect *in vivo*. These ligands have been shown to induce telomere shortening and/or telomere dysfunction.⁵³ Since inefficient length of telomeres and dysfunctional telomeres are inefficient to act as chromosome end-protection, DNA damage pathway will be activated, which leads to cell senescence and apoptosis. Although targeting telomeric G-quadruplex by small molecules provides a potential strategy for anticancer therapeutics, the presence of a significant number of G-quadruplex motif across the human genome makes it difficult to specifically target to G-quadruplexes in the telomere. To effectively target telomeric G-quadruplexes, it is necessary to study the structures adopted by telomeric sequences and more efforts should be concentrated to develop highly specific ligands toward telomeric G-quadruplex.

Figure 1-17. Folding topology of human telomeric G-quadruplexes adopted by different sequences in different monovalent ionic conditions (a) d[AG₃(TTAGGG)₃] in Na⁺ solution, PDB code: 143D⁵⁴ (b) d[AG₃(TTAGGG)₃] in crystalline state containing K⁺, PDB code: 1KF1⁵⁵ (c) d[TA(GGGTTA)₃GGG] in K⁺ solution, PDB code: 2JSM⁵⁶ (d) d[TA(GGGTTA)₃GGGTT] in K⁺ solution, PDB code: 2JSQ⁵⁶ (e) d[(GGGTTA)₃GGGT] in K⁺ solution, PDB code: 2KF8⁵⁷ (f) (TTAGGG)₄TTA in Na⁺ solution, PDB code :2MBJ⁵⁸. *syn* and *anti* guanines are shown in magenta and cyan, respectively.

Figure 1-18. Chemical structure of ³H-360A small molecule

G-quadruplex in Transcription

G-quadruplex motifs are over represented in regions that are close to the transcription start sites of protein-coding genes in human,^{28, 59} chimpanzee,⁵⁹ rat,⁵⁹ mouse,⁵⁹ yeast⁶⁰ and prokaryotic genomes.⁶¹ In the human genome, more than 40% of the gene promoter regions which is 1 kilobase upstream of the transcription start site contain at least one G-quadruplex motif.²⁸ The prevalence of G-quadruplex at the regions near transcription start sites suggests that G-quadruplex may be involved in the transcription regulation. Interestingly, in a study to correlate gene function with G-quadruplex motifs, it was found that the occurrence of G-quadruplex motifs is high in proto-oncogene but significantly lower in tumour suppressor genes.⁶²

These observations suggested that G-quadruplex structures associated with proto-oncogenes are potential therapeutic targets for transcription-induced destabilization. G-quadruplex formation in promoter regions was proposed to down regulate gene expression by inhibiting the binding of transcription factors leading to the inhibition of gene expression (Figure 1-19).^{63, 64} On the other hand, G-quadruplex can also lead to an increase in gene expression when DNA damage lesion 8-oxoguanine occurs in G-quadruplex forming sequence in promoter-coding strands.⁶⁵

Figure 1-19. Schematic showing G-quadruplex formation at the promoter region regulates gene transcription. Figure is adapted from ref ⁶³.

The interest in targeting promoter G-quadruplex emerged from a study showing that the transcription repression of c-MYC proto-oncogene can be attained by targeting a G-quadruplex motif presents in the nuclease hypersensitive element (NHE) within the promoter region of MYC with a small molecule.⁶⁶ C-MYC is a transcription factor which is associated with cell proliferation and over-expressed in 80% of human cancer cells including gastrointestinal, ovarian and breast cancer tumours. The nuclease hypersensitive element III₁ (NHE III₁) of the c-MYC promoter contains the 27-nucleotide guanine-rich sequence d(TGGGGAGGGTGGGGAGGGTGGGGAAGG) which was shown to form a mixture of G-quadruplex structures,^{21, 67, 68} Stabilization of these G-quadruplex structures by a G-quadruplex binding ligand, TMPyP4, led to the inhibition of c-myc transcription.⁶⁶ Besides the TMPyP4 ligand, other G-quadruplex ligands including actinomycin D,⁶⁹ berberine derivative,⁷⁰ furan based tricyclic homo-oligopeptides⁷¹ and telomestatin derivative⁷² were similarly found to inhibit the synthesis of c-myc mRNA.

Table 1-1. NMR structures of promoter G-quadruplexes

Gene	Sequence	Topology	PDB code	Ref
C-myc	d[$\text{TGAG}_3\text{TTG}_3\text{TAG}_3\text{TG}_3\text{TAA}$]	Parallel	1XAV	⁶⁷
C-myc	d[$\text{TGAG}_3\text{TGGIGAG}_3\text{TG}_4\text{AAGG}$]	Parallel	2A5P	²¹
VEGF	d[$\text{CGG}_3\text{CG}_3\text{CCTTG}_3\text{CG}_3\text{GT}$]	Parallel	2M27	¹⁹
BCL-2	d[$\text{G}_3\text{CGCG}_3\text{AG}_2\text{A}_2\text{T}_2\text{G}_3\text{CG}_3$]	Hybrid	2F8U	²⁰
C-kit	d[$\text{AG}_3\text{AG}_3\text{CGCTG}_3\text{AG}_2\text{AG}_3$]	Parallel	2O3M	²²
C-kit	d[$\text{CG}_3\text{CG}_3\text{CGCTAG}_3\text{AG}_3\text{T}$]	Parallel	2KYP	⁷³
C-kit	d[$\text{CG}_3\text{CG}_3\text{CGCGAG}_3\text{AG}_3\text{T}$]	Parallel Dimer	2KYO	⁷³
C-kit	d[$\text{CG}_3\text{CG}_3\text{CACGAG}_3\text{AG}_3\text{T}$]	Parallel	2KQG	⁷⁴
C-kit	d[$\text{CG}_3\text{CG}_3\text{CGCTAG}_3\text{AG}_3\text{T}$]	Parallel	2KYP	⁷⁴
<i>TERT</i>	d[$\text{AGGGIAGGGGCTGGGAGGGC}$]	Hybrid	2KZD	⁷⁵
<i>TERT</i>	d[$\text{AIGGGAGGGICTGGGAGGGC}$]	Parallel	2KZE	⁷⁵

Besides c-myc, G-quadruplex formation by promoter sequences of other proto-oncogenes have been studied, including the receptor tyrosine kinase (KIT),⁷⁶ B cell lymphoma 2 (BCL-2),^{20, 77} vascular endothelial growth factor (VEGF),^{19, 78} hypoxia-inducible factor 1 α (HIF1 α),⁷⁹ the transcription factor MYB,⁸⁰ KRAS⁸¹ and human telomerase reverse transcriptase (TERT).⁸² Among these, a few high resolution structures of promoter G-quadruplex have been reported in the literature (Table 1-1). Most of these quadruplex structures adopted all-parallel-stranded topology.

G-quadruplex in Translation

Bioinformatics analysis revealed that about 3000 G-quadruplex forming motifs are present in 5'-UTR RNA across the human genome, suggesting the involvement of RNA G-quadruplexes in protein translation regulation.⁸³ The regulation role of G-quadruplex in translation has been demonstrated *in vitro* and *in vivo*.⁸⁴ RNA G-quadruplexes can affect translation in two ways: (i) acting as an inhibitor to translation by interfering with translation initiation and/or perturbing the

scanning of the ribosome or (ii) promoting translation initiation. The first evidence showing the ability of G-quadruplex to regulate protein expression was presented by Moine and co-workers.⁸⁵ By using *in vitro* translation, they showed that efficiency of protein synthesis was decreased by 1.5 fold when a G-quadruplex motif identified from FMR1 mRNA was inserted into 5'-UTR of a luciferase reporter gene. This observation was explained by the ability to impede translation initiation by a stable G-quadruplex structure. Balasubramanian's group identified a G-quadruplex motif d(GGGAGGGGCGGGUCUGGG) in the 5'-UTR of human NRAS proto-oncogene mRNA.⁸³ This RNA motif was shown to cause translation to decrease by 80% *in vitro* when inserted downstream of the promoter region.

The regulation role of RNA G-quadruplex in translation was also observed in the human cell. A 27-nucleotide G-quadruplex motif was identified within the 5'-UTR of the mRNA of the human Zic-1 zinc-finger protein.⁸⁶ It was shown that this 5'-UTR RNA G-quadruplex motif repressed 80% protein expression when inserted upstream of the luciferase gene. Quantitative RT-PCR assays validated the decrease in protein expression as a result of the regulation of the translation rather than a result of reduced transcription. The regulation role of 5'-UTR RNA G-quadruplex in human cell was also demonstrated for other genes including the telomere shelterin protein TRF2,⁸⁷ matrix metalloproteinase MT3-MMP,⁸⁸ and disintegrin and metalloprotease ADAM10.⁸⁹

While most studies in literature demonstrated that 5'-UTR RNA G-quadruplexes are inhibitory to translation, there are also studies showing RNA G-quadruplexes to promote protein synthesis.^{90, 91} Morris et al provided evidence that a 17-nucleotide G-quadruplex motif d(GGAGGAGGGGGAGGAGGA) within the 5'-UTR of human vascular endothelial growth factor (hVEGF) mRNA was essential for translation initiation in HeLa cells.⁹¹ They showed that the mutation sequences of this 5'-UTR RNA G-quadruplex motif, which were unable to form G-quadruplex, almost completely suppressed protein translation.

The discovery of the regulatory role of mRNA G-quadruplex structure in translation provides the chance to target mRNA G-quadruplex by small molecules.

Small molecules can affect mRNA translation through two manners: stabilizing the RNA G-quadruplex and impeding the scanning process of ribosomal complex or interfering with protein/RNA G-quadruplex binding events. Several studies have shown the effect of small molecules including Phen-DC₃, Phen-DC₆, 360A^{87, 92} and RGB-1⁹³ on inhibiting mRNA translation of genes containing 5'-UTR RNA G-quadruplex structures. These studies demonstrated that G-quadruplex binding ligand can target RNA G-quadruplex and interfere with translation process, suggesting a new strategy in RNA-directed drug design.

G-quadruplex in Replication and Genomic Instability

The Watson-Crick hydrogen bonds within the double helix in genome prevent the formation of G-quadruplex. During replication, the transient open state of double helix provides an opportunity for the assembly of G-quadruplexes from the G-quadruplex forming sequences. By visualization G-quadruplex by the specific G-quadruplex binding antibody (BG4),³² it was shown that the number of G-quadruplex significantly increased during S-phase in which replication occurs.

Figure 1-20. Schematic showing that G-quadruplexes form during replication and impede the movement of DNA polymerase. Figure is adapted from ref ⁹⁴.

The occurrence of G-quadruplex structures during replication may hinder the progression of replication fork and induce DNA mutations (Figure 1-20). Therefore, there exists a need for the resolution of G-quadruplex structures during replication to remove these steric blocks and allow the progression of the replication fork. Many helicases including DOG-1,⁹⁵ FANCI,⁹⁶ WRN,⁹⁷ BLM,⁹⁸ and Pif1 have been reported to play important role in resolving G-quadruplex structures during

replication to prevent genetic instability. Mutation/absence of several helicases that unwind G-quadruplex structures have been associated with genomic instability disease in human including Werner syndrome (associated with WRN),⁹⁹ Bloom syndrome (associated with BLM),¹⁰⁰ Fanconi anemia (associated with FANCD1),¹⁰¹ and also in increase of cancer risks (associated with Pif1).¹⁰²

The role of G-quadruplex structure and G-quadruplex helicase in replication was first demonstrated by genomic instability study in *Caenorhabditis elegans*.⁹⁵ In *C.elegans* containing a loss of function mutation or deletion of DOG-1, deletion was observed throughout the genomic DNA at the regions containing polyguanine tracts, implicating the formation of G-quadruplex structures in polyguanine tracts and the role of DOG-1 helicase in resolving these structures. Later, Nicolas's group showed that genomic instability was induced by the formation of stable G-quadruplexes in the *Saccharomyces cerevisiae* genome when the cells were treated with a G-quadruplex stabilizing ligand, Phen-DC₃ or when Pif1 helicase was deleted.¹⁰³⁻¹⁰⁵ Replication fork stalling caused by G-quadruplex forming sequence was also demonstrated *in vivo* by Sarkies et al.¹⁰⁶ In their study, it was shown that the site of replication fork stalling was mapped to regions containing G-quadruplex forming sequence in an avian DT40 cell line and the presence of a translesion polymerase REV1 was necessary for replication of G-quadruplex forming sequences.

G-quadruplex motifs are essential for replication initiation

Recent developments in the mapping of replication origins have revealed a new role of G-quadruplex in DNA replication. A genome-wide study discovered > 350,000 potential replication origins in four human cell lines with G-quadruplexes found to be strongly associated with the majority of these replication origins.¹⁰⁷ In another study, 90% of replication origins in mouse and human cells and 70% in *Drosophila* cells contained a specific G-rich sequence, OGRE, which had the propensity to form into G-quadruplex. The initiation site of replication located at 160bp (*Drosophila*) and 280 base pairs (mouse) downstream of OGRE, suggesting the important role of G-quadruplex in the initiation and localization of replication in metazoans (Figure 1-21). The strong evidence showing the role of G-quadruplex in

replication initiation comes from a genetic study in chicken DT40 cell line.¹⁰⁸ In this study, it was shown that G-quadruplex was essential to activate replication initiation as the deletion of the G-quadruplex motif or mutation resulting in destabilization of the G-quadruplex led to a decrease in replication efficiency. The role of G-quadruplex in replication initiation was further supported by the observation that human origin recognition complex (ORC) preferentially binds to RNA/DNA G-quadruplex.¹⁰⁹

Figure 1-21. Schematic showing that G-quadruplex formation is necessary for replication initiation and determines the location of initiation. Figure is adapted from ref ⁹⁴.

1.3.3 G-quadruplex Binding Ligands

The observation that G-quadruplex structures occur at genomic region related to cancer proliferation such as telomere, oncogenomic promoter has started the interest in targeting these structure as an anticancer therapy strategy. There has been increasing interest in the development of small molecules that exhibit high affinity and specific binding to G-quadruplex structures with numerous G-quadruplex binding ligands described in the literature to date.⁴⁸ Most G-quadruplex binding ligands have two properties in common: planar aromatic rings for π stacking on planar G-tetrads, and cationic side chains for electrostatic interaction with negatively charged phosphate backbone.

Figure 1-22. Stick representation showing π - π stacking between G-tetrads and ligands in G4-ligand complexes (a) 2mgn¹¹⁰, (b) 2L7V¹¹¹, (c) 1L1H¹¹², and (d) 2mb3¹¹³.

Structural studies of quadruplex-ligand complexes have revealed the main binding mode between small molecules and G-quadruplex as π - π interactions. End-stacking binding mode serves as a unique feature of G-quadruplex recognition since G-tetrads provide a large aromatic area surface (Figure 1-22) which is not commonly found in other DNA structures.

Beside end-stacking, small molecules can interact with G-quadruplex through hydrogen bonds. In the crystal structure of trisubstituted acridine complexed with *Oxytricha* telomeric quadruplex d(G₄T₄G₄), two hydrogen bonds were observed between acridine with a loop thymine base (Figure 1-23a).¹¹⁴ In the complex of quindoline and C-myc G-quadruplex, a hydrogen bond also appeared between quindoline and the 23rd thymine base stacking on the 3' end (Figure 1-23b).¹¹¹ Another common DNA-ligand interaction is via electrostatic interaction. Since DNA is negatively charged, small ligands for targeting DNA usually have positive charges.

In G-quadruplex ligands, positive charges can be achieved by protonated amine groups such as L2H2-6M(2)OTD ligand,¹¹³ Phen-DC₃ ligand¹¹⁰ or incorporation of metal ions such as Zn-DIGP ligand¹¹⁵ and [Pt(Dip)₂](PF₆)₂.¹¹⁶

Figure 1-23. Stick representation showing examples on hydrogen bonds formed between G-quadruplexes and ligands in complex of (a) Acridine/ G-quadruplex (pdb code: 3ET8)¹¹⁴ (b) Quindoline/G-quadruplex complex (pdb code: 2L7V).¹¹¹

For a small molecule to be a good G-quadruplex ligand, two requirements must be met: the small molecule must exhibit high binding affinity and high selectivity to G-quadruplex over duplex DNA and other DNA structures. Current development of G-quadruplex binding small molecules has allowed numerous ligands with high affinity and high selectivity to G-quadruplex over duplex DNA to be developed. Notable examples of small molecules that showed high affinity and selectivity toward G-quadruplex are telomestatin,¹¹³ and BRACO-19⁵⁰ with the selectivity toward G-quadruplex over duplex of 70 and 100 fold, respectively. However, the current generation of G-quadruplex binding ligands lack good selectivity between different G-quadruplex structures.^{117, 118} Therefore, the future of

G-quadruplex ligand development should be focused to increase the selectivity toward a specific G-quadruplex structure.

1.4 Quadruplex-Duplex Hybrids

Interest on quadruplex-duplex junction has emerged recently. Quadruplex-duplex hybrids, which contain both G-quadruplex and duplex segments, have been applied in designing nanodevices,^{119, 120} RNA aptamer for imaging application^{121, 122} and were shown to be a promising target for anticancer therapy.¹²³ Previously, Lim and Phan have extensively studied different ways of connecting duplexes to a quadruplex.¹²⁴ In this study, two relative arrangements of duplex and quadruplex including coaxial and orthogonal orientation (Figure 1-25) were demonstrated through five quadruplex-duplex hybrid constructs (Figure 1-24). A thermal stability study of quadruplex-duplex hybrid has shown that G-quadruplexes containing duplex stems are more stable than of G-quadruplexes containing nonstructural loops of the same length, and the increase in the length of stem duplex increases the thermal stability of the quadruplex-duplex hybrid.¹²⁵ This observation changed the view about loop length in G-quadruplex as it was generally accepted that stable G-quadruplex should not contain loops of more than 7 nucleotides.¹⁷ Since the presence of a stem-loop enhances the stability of quadruplex-duplex hybrid, this allowed the loop length of a G-quadruplex to be extended to a much larger limit. For example, a G-quadruplex with 37 nucleotide loop was shown to exhibit a high melting temperature of 66.9 °C in 10 mM K⁺.¹²⁵ A bioinformatics study has revealed about 80,000 quadruplex-duplex motifs across the human genome, many of which are present within regulatory important regions such as oncogene PIM1, MYBL1, transcription factor MYT1L, member of RAS oncogene family (RAB12 and RAD3D), cell cycle regulator cyclin Y (CCNY), proto-oncogene RET, GTPase activation protein 31 ARHGAP31 (Table 1-2).¹²⁶ To date, most of quadruplex-duplex hybrids were studied are quadruplex structures containing a duplex stem in the loop. It is interesting to explore other possibilities of quadruplex-duplex hybrids and structural feature of quadruplex-duplex motif which could be present in the human genome. Structural features of quadruplex-duplex hybrids may provide a

platform for designing specific binding ligands and designing DNA nanoarchitectures.

Table 1-2. Examples of quadruplex-duplex hybrid found in the human genome.¹²⁶

Sequence (5'-3') ^{a,b}	Gene	Gene Segment
A GGG <u>AGAGGAGAGCTCT</u> GGG TT GGG T GGG	MYT1L	Intron 22
CA GGG AA GGG <u>ACCTGATAGGTA</u> GGG A GGGGC	RET	Promoter
GC GGG A GGG <u>CGCGCCAGCG</u> GGG TC GGG C	PIM1	5'-UTR
GA GGG C GGGG CT GGGG <u>AGCTGGAAGCT</u> GGG A	MYBL1	Promoter
GA GGG C GGG <u>CGCCGCTGGCGA</u> GGG A GGG C	CCNY	Promoter
GT GGG <u>TGCAGTCAGAGCTGCT</u> GGG AGT GGG	ARHGAP31	Intron 1
GA GGG <u>GCCGGAGGC</u> GGG C GGG T GGG C	RAB12	Promoter
TT GGG C GGG <u>GCTTTCAGAGAGC</u> GGG T GGG A	RAB3D	Promoter

^aG-tracts are shown in bold.

^bSelf-complementary tracts are underlined.

Figure 1-24. Schematic showing five constructs of quadruplex-duplex hybrids: (a-d) a duplex is connected to a quadruplex through wide groove (a); medium groove (b); diagonal corner (c); narrow groove (d), (e) Sideway connectivity of a duplex to a quadruplex. Figure is adapted from ref ¹²⁴.

Figure 1-25. Two different orientation between a duplex and a quadruplex : (a) coaxial orientation in which duplex is joined to quadruplex through two different strand (b) orthogonal orientation in which duplex is incorporated in the loop of quadruplex. Figure is adapted from ref ¹²⁴.

1.5 Research Objectives

The past few decades have witnessed the growing evidence to propose G-quadruplexes as attractive targets for anticancer therapy. Understanding structure of G-quadruplex is important for drug design for targeting G-quadruplex as well as for the prediction of G-quadruplex forming sequences. In the cell context, G-quadruplexes may form in close proximity to the duplex. Therefore, it is important to understand the possible connectivity between duplexes. Knowledge on the quadruplex-duplex hybrid structures would provide useful information for the prediction of quadruplex forming sequences in the human genome as well as the development for targeting G-rich sequence. In addition, understanding on quadruplex-duplex connectivity would be also helpful for the implementation of G-quadruplex elements in the design of DNA nanoconstructs which has mainly relied on duplex motifs. The overall aims of the study presented in this Ph.D. thesis are to study structural features of novel quadruplex-duplex hybrid structures as well as discover a new strategy for targeting quadruplex-duplex hybrid structures. This goal can be separated into three main parts:

- 1. To investigate the different possible quadruplex-duplex hybrids.** Chapter 3 and chapter 4 of this thesis present two different connectivity between quadruplex segments and duplex segments: incorporation of multiple duplex stems into a single loop of a G-quadruplex and incorporation of a duplex stem in different bulge positions of a G-quadruplex.

Chapter 3 investigates the incorporation of multiple duplex stems into a single loop of an all-parallel-stranded G-quadruplex. Folding topology of a two-stem quadruplex-duplex hybrid is validated by NMR. Thermal stability study of multi-stem quadruplex-duplex hybrids is examined by CD spectroscopy.

Chapter 4 investigates the incorporation of a duplex stem into bulges of a G-quadruplex. High resolution structure of a duplex-bulge G-quadruplex is elucidated by NMR. The effect of duplex length on structural and physical property of the G-quadruplex and the possibility of incorporation of the duplex bulge at different positions are also examined by NMR and CD spectroscopy.

Knowledge obtained from these two chapters expands the definition of G-quadruplex forming sequences. In addition, these stable quadruplex-duplex hybrid motifs could have potential application in DNA nanostructure design. The potential existence of these quadruplex-duplex hybrid motifs in the human genome could give rise to exciting opportunities for therapeutic targeting, since the interface between quadruplex and duplex segments and the minor groove of duplex segments could provide unique structural features for specific ligand binding.

2. To study structural features of telomeric quadruplex-duplex hybrid which could occur in human chromosomes.

Chapter 5 studies the structure of the human telomeric duplex-quadruplex junction. This junction could exist at the transition between telomeric duplexes and single stranded segments, in which the telomeric duplex is connected to the 5' end of the telomeric G-quadruplex. Folding topology of G-quadruplex segment of the telomeric duplex-quadruplex junction is elucidated by NMR. Structural insights gained from this study could be helpful toward the rational design of drug for targeting telomere in cancer cells.

3. To explore a new targeting approach to achieve specific targeting quadruplex-duplex hybrid structure.

Chapter 6 explores a simultaneous targeting of the duplex and quadruplex structure with duplex-specific ligands and quadruplex-specific ligands. The simultaneous

binding of a quadruplex ligand (Phen-DC₃) and a sequence specific duplex ligand (netropsin) to quadruplex-duplex hybrid structures is validated by CD and NMR spectroscopy. A suitable linker to connect these two ligands would lead to a single entity conjugate ligand which could provide synergistic binding of the two modules to the quadruplex-duplex hybrid target and overcome the poor selectivity within different G-quadruplex structures of the current generation of G-quadruplex ligands. The proposed targeting strategy could contribute to the future development of compounds for targeting specific genomic quadruplex-duplex sites.

Chapter 2 Methodology

2.1 DNA Samples Preparation

DNA oligonucleotides were synthesized based on solid phase chemical synthesis using phosphoramidite chemistry. In solid phase DNA synthesis, nucleotides are sequentially added onto a starting nucleotide immobilized on a solid support in 3' to 5' direction. Nucleoside phosphoramidite was used to enhance the selectivity and the rate of the formation of internucleosidic linkages. All other functional groups on nucleoside phosphoramidite are protected to prevent undesirable reactions. Each cycle of nucleotide addition involves four steps: detritylation, coupling, capping, and oxidation (Figure 2-1).

Step 1: Detritylation (or de-blocking)

The protecting group DMT is first removed by a solution of acid such as 2% trichloroacetic acid. The DMT cation emits orange color and its intensity can be used to monitor the yield of the coupling reaction. After detritylation step, 5'-OH groups of the solid support-immobilized oligonucleotides are free to react with the incoming nucleoside in the coupling step.

Step 2: Coupling

The appropriate nucleoside phosphoramidite is activated by an activator (tetrazole or its derivatives). Activated nucleoside with excess amount is then flowed to the solid support-immobilized oligonucleotides whose free 5' OH groups quickly react with the incoming activated nucleoside to form a phosphite triester linkage.

Step 3: Capping

The coupling step may not be 100% efficient, therefore the 5' OH group of the unreacted immobilized oligonucleotide (failure sequences) must be permanently blocked in the capping step. Two capping solutions including acetic anhydride and N-methylimidazole are utilized to acetylate the unreacted 5' OH group. After acetylation, failure sequences are inert to subsequent reactions.

Figure 2-1. The cycle of oligonucleotide synthesis includes four steps: detritylation, coupling, capping, and oxidation.

Step 4: Oxidation

The phosphite triester (P(III)) linkage is unstable and needs to be converted to stable phosphate triester (P(V)) linkage before the next detritylation step. This is done by treatment with iodine solution in presence of water and a weak base such as pyrimidine during the oxidation step.

After the oxidation step, the immobilized oligonucleotides are ready for the next cycle of nucleoside phosphoramidite addition. The cycle of addition keeps repeating until the oligonucleotides are synthesized to their full length. Upon the completion of DNA synthesis, the oligonucleotide needs to be cleaved off the solid support.

DNA de-protection

After the completion of DNA synthesis, the immobilized oligonucleotide contains base- and phosphate-protecting groups. These protecting groups as well as the linkage between oligonucleotides and solid supports are cleaved by treatment with ammonium hydroxide.

DNA purification

The final product from DNA synthesis contained both desired and failure oligonucleotide sequences. Several methods can be employed to purify DNA including polyacrylamide gel electrophoresis, high performance liquid chromatography (HPLC), and purification by oligonucleotide purification cartridges (OPC). In this study, oligonucleotides are purified using OPC method. In this method, DNA oligonucleotide sample is loaded onto the cartridge. The full length sequences with DMT group on are trapped into hydrophobic solid support of cartridge while the failure sequences are eluted out by ammonium hydroxide (1:20). The DMT group is subsequently cleaved via treatment with 2% trifluoroacetic acid. Finally, oligonucleotides are eluted from the cartridge by 20% acetonitrile. The efficiency of OPC method depends on the efficiency of synthesis process. When the synthesis yield is good (95%-97% average yield per coupling cycle), the purity of DNA after purification by OPC method is typically more than 90%.

2.2 Nuclear Magnetic Resonance Spectroscopy

2.2.1 NMR Basics

Nuclear spin (I) is the total angular momentum of a nucleus and comes with multiples of $\frac{1}{2}$. Each individual unpaired of proton, neutron, and electron possesses a spin of $\frac{1}{2}$. The nuclear spin defines the magnetic moment of nucleus. For a given

I, the nucleus can adopt $2I+1$ different energy levels. The most biologically relevant nuclei are ^1H , ^{13}C , ^{15}N and ^{31}P which all possess a spin of $\frac{1}{2}$. Hence, they can adopt two different spin states: a lower-energy $+\frac{1}{2}$ (or α) state and a higher-energy $-\frac{1}{2}$ (or β) state when subjected to an external magnetic field (B_0). When the magnetic moment of nucleus aligns in the same direction of the external magnetic field, the nucleus is at low energy state $+\frac{1}{2}$. On the other hand, when the magnetic moment of nucleus aligns in the opposite direction of the external magnetic field, the nucleus is at high energy state $-\frac{1}{2}$. At equilibrium, most of the nuclei would adopt the state $+\frac{1}{2}$, following the Boltzmann's equation:

$$\frac{N^-}{N^+} = e^{\frac{-E}{k_b \times T}}$$

Where: E is the energy gap between two spin states; k_b is Boltzmann's constant, 1.380×10^{-23} J/Kelvin; and T is the temperature in Kelvin. N^- is the population of nuclei at α state, N^+ is the population of nuclei at β state.

Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) is a phenomenon which allows the nucleus at $+\frac{1}{2}$ state to undergo a transition to $-\frac{1}{2}$ state upon irradiation with a photon which exactly matches the energy difference between the two states. The frequency of the radiation depends on the strength of the external magnetic field and the property of the nucleus called gyromagnetic ratio (γ), as described by the equation: $f = \gamma \times B$, where f is the resonance frequency and B is the strength of external magnetic field. Among the nuclei, proton (^1H) possesses the highest γ (42.576 MHz/T), hence proton is the most sensitive nuclei for NMR experiment.

Magnetization Vector

The sum of magnetic moments of nuclear spins which experience the same magnetic field strength is called the magnetization vector (M). At equilibrium, magnetization vector would align in the same direction in z axis with an external magnetic field (B_0) and is called M_0 (Figure 2-2a). Magnetization vector can be changed when a radio wave of frequency equal to the energy difference between two

spin states is applied. When the magnetization vector tilts away from the z axis (Figure 2-2b), the magnetization vector will return to the equilibrium state over time while precession about the z-axis at the Larmor frequency which is equal to the resonance frequency (Figure 2-2c). The magnetization vector includes two components: longitudinal magnetization (M_z) which is the projection of magnetization vector on z-axis and transverse magnetization (M_{xy}) which is the projection of magnetization vector on transverse plane (xy plane) (Figure 2-2c).

Figure 2-2. The net magnetic vector of nuclei (a) at equilibrium state (b) after a pulse is applied in x direction (c) precession to equilibrium state. B_0 is the external magnetic field, M is the magnetization vector, M_0 is the magnetization vector at equilibrium, M_z is the longitudinal magnetization, M_{xy} is the transverse magnetization.

Pulses.

At equilibrium, the magnetization vector aligns in the z direction. The application of a radiofrequency field (B_1 field) which creates an oscillating magnetic field at lamor frequency along the x axis during the time t_p is called a radiofrequency pulse. After the pulse, the magnetization vector would rotate in zy plane around the x axis by an angle φ :

$$\varphi = \omega_1 \times t_p = \gamma \times B_1 \times t_p$$

Where:

φ (rad) is the flip angle of the pulse

ω_1 (rad.s⁻¹) is lamor frequency about the B_1 field

t_p (s) is the duration of the pulse (the time when radiofrequency field is turned on)

γ is gyromagnetic ratio (rad.s⁻¹.T⁻¹)

A $\frac{\pi}{2}$ (90°) pulse (Figure 2-3a) or π (180°) pulse (Figure 2-3b) rotates the magnetization vector by 90° or 180°, respectively.

Figure 2-3. Schematic illustrates the rotation of magnetization vector after 90° and 180° pulse. B_0 is the external magnetization field, B_1 is the radiofrequency magnetization field.

Relaxation and Detection

At equilibrium, magnetization vector would align in the same direction with the external magnetic field (B_0) and is called M_0 (Figure 2-2a). At this equilibrium state, $M_z = M_0$ and $M_{xy} = 0$. A radio frequency pulse applied along the x axis will lead to the rotation of the magnetization vector in the yz-plane (Figure 2-2b). If a 90° pulse is applied along x axis, the magnetization vector rotates 90° onto y axis. After the pulse, $M_z = 0$ and $M_{xy} = M_{xy0}$. The magnetization vector will try to return to its equilibrium state over time while precessing about the z-axis at the Larmor frequency (Figure 2-2c). This process is called relaxation. The relaxation of magnetization vector induces a current (free induction decay or FID signal) which is gradually lost over time in a detection coil (Figure 2-2c) placed along the xy plane. The FID signal (which has time domain) is converted to NMR signal (which has frequency domain) by a Fourier transform (Figure 2-4).

Figure 2-4. Fourier transform is a mathematical process which converts a function of time (FID signal, left) to a function of frequency (NMR spectrum, right)

T₁ and T₂ Relaxation

There are two kinds of relaxation: longitudinal relaxation and transverse relaxation

Longitudinal relaxation (T₁ relaxation) describes the return of M_z component to its equilibrium state (M₀). The function describes T₁ relaxation is:

$$M_z = M_0 \times (1 - e^{\frac{-t}{T_1}}).$$

Transverse relaxation (T₂ relaxation) describes the decay of transverse magnetization. Only the relaxation of transverse leads to the FID signal in detection coil. Each of the spin making up the net magnetization vector experiences slightly different magnetization field due to the effect of its neighbor spins and the inhomogeneity of the external magnetization field. Therefore, they rotate at different Larmor frequencies. As a result, transverse magnetization starts to dephase during its rotation about z axis (Figure 2-5). The longer the elapsed time results in the greater the phase difference. The time constant which describes the decay of M_{xy} (the return of M_{xy} to 0) is called the spin-spin relaxation time, T₂. The function describes T₂ relaxation is: $M_{xy} = M_{xy_0} \times e^{\frac{-t}{T_2}}.$

The relaxation constant T₂ mainly depends on the homogeneity of the magnetic field, the strength of the dipolar interaction with other nuclei spin and the overall tumbling time of the molecule which is related to its size.

The transverse relaxation time T_2 determines the line widths of the NMR signal. The longer T_2 is, the sharper the NMR signal. The line width at the half height of the signal is described as: $\Delta\nu_{1/2} = \frac{1}{\pi \times T_2}$.

Figure 2-5. Dephasing of transverse magnetization.

Chemical Shift

As described above, the resonance frequency depends on the strength of the magnetic field and the gyromagnetic ratio of the nucleus: $f = \gamma \cdot B$. Nuclei that experience different magnetic field B will resonate at different frequencies. When an atom is placed in an external magnetic field, B_0 , its electron circulates about the external magnetic field and causes a small magnetic field, $\delta \cdot B_0$, at the nucleus, which has the opposite direction to the external magnetic field. Therefore, the nucleus will experience the effective magnetic field, B , which is different from the external magnetic field, B_0 : $B = B_0 \times (1 - \delta)$. This difference factor varies according to the local electronic environment at the nucleus. The greater the electron density around the nucleus, the smaller the effective magnetic field experienced by the nucleus. Consequently, the nucleus resonates at a lower resonance frequency. This phenomenon is called shielding. Conversely, nuclei with weaker surrounding electron density would feel the stronger the effective magnetic field and therefore resonate at the higher resonance frequency. It is called de-shielding.

Since the resonance frequencies vary with the strength of magnetic field, nuclei placed in different magnetic field strength will resonate at different frequencies. Thus, it is challenging to compare NMR spectra performed at different field strengths. To avoid this problem, the term chemical shift is used. The chemical

shift (δ) of a nucleus is the difference in the resonance frequency of the nucleus and a reference relative to the frequency of the reference. The chemical shift is normally reported in ppm:

$$\delta = \frac{(f - f_{\text{ref}}) \times 10^6}{f_{\text{ref}}}$$

Where f and f_{ref} are the resonance frequencies of the nucleus and of the reference, respectively.

Figure 2-6. Chemical shift pattern of nucleic acid protons (a) chemical structure of guanine nucleoside, cytosine base, adenine base and thymine base; (b) NMR chemical shift distribution of different proton types in nucleotides.

The two universally accepted references are 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) and tetramethylsilane (TMS). DNA is composed of many

repeating nucleotide units: adenine; cytosine; guanine and thymine, and these nucleotides have common features of a ribose sugar and a nucleobase. As illustrated in Figure 2-6, protons on the DNA with similar chemical environment will display similar chemical shifts. In addition, DNA secondary structure can also be indicated by chemical shift. The imino protons of guanine in G-tetrad involved in Hoogsteen hydrogen bonding with oxygen show characteristic chemical shifts at 10.5-12.5 ppm while imino protons of nucleobases involved in Watson-Crick base-pairing show characteristic chemical shifts at 12.5-14.5 ppm. Four imino protons of four guanines participating in a G-tetrad exhibit four imino proton peaks at 10.5-12.5 ppm. A typical G-quadruplex consisting of three G-tetrads exhibits 12 imino proton peaks at 10.5-12.5 ppm.

2.2.2 Two-Dimension NMR

Figure 2-7. Schematic of a 2D homonuclear NMR spectrum for two protons A and B. The appearance of cross peaks between two protons (B-A and A-B cross peak) indicates the existence of a correlation between them. In 2D homonuclear NMR spectra, the diagonal peak A and B are always present.

In one dimensional (1D) NMR experiment, the FID signals are recorded with respect to a single time variable. Fourier transform of the resulting signal generates a spectrum where intensity is a function of a single frequency variable. In two-dimensional (2D) NMR, the FID signals are recorded with respect to two time variables (t_1 and t_2). The obtained FID signals are Fourier transformed two times to

generate a two-dimensional spectrum which is a plot of intensity as a function of two frequency variables (F1 and F2) (Figure 2-7). There are two types of 2D correlation experiments: through-space correlation such as nuclear overhauser effect spectroscopy (NOESY) and through-bond correlation such as heteronuclear single-quantum correlation (HSQC), heteronuclear multiple-bond correlation (HMBC), total correlation spectroscopy (TOCSY), and correlation spectroscopy (COSY). The appearance of cross peaks between chemical shifts f_A and f_B in 2D spectra (Figure 2-7) indicates the existence of dipolar coupling (through-space correlation) or scalar coupling (through-bond correlation) between two protons at these chemical shifts. 2D NMR spectra are especially useful in the 3D molecule structure elucidation.

Figure 2-8. General scheme of 2D NMR pulse sequence.

The general pulse sequence of a 2D NMR experiment is illustrated in Figure 2-8: The sample is excited by one or more pulses during the preparation period, and the magnetization is then allowed to evolve during evolution time t_1 . The subsequent period is called mixing time which consists of one or more pulses. After that, the signal is recorded over the detection time t_2 . This sequence of event repeats many times with varying t_1 . First, t_1 is set to zero, the pulse sequence is applied and the resulting signal is recorded. The system is allowed to obtain the equilibrium state. The sequence of action is repeated with increment t_1 by Δt (the sampling interval of t_1). The corresponding FID signal of each sequence is recorded and saved separately. The repetition is kept continuously until sufficient signal is accumulated, typically requiring 50 to 500 increments of t_1 .

The time scale of Δt and t_2 depends on spectral width (sw, Figure 2-7) and the required resolution of the 2D spectra. Normally, Δt and t_2 are defined as following:

$$\Delta t = \frac{1}{2 \times \text{sw}}$$

$$t_2 = \Delta t \times \text{td}_2$$

Where: sw (hz) is spectral width of the spectra (the width of the acquired spectrum).

td₂ is the number of acquisition point in t₂ time domain. In our study, usually

td₂ is set at 2048.

2.2.3 Scalar Coupling and Through-Bond Correlation

Figure 2-9. Different energy states of a two-spin system (a) four energy levels of a two-spin system (b) splitting of signal in two-spin system. J_{12} is the scalar coupling between spin 1 and spin 2, quoted in Hz. $f_{0,1}$ and $f_{0,2}$ is the Larmor frequencies of spin 1 and spin 2, respectively.

Scalar couplings (J-coupling) are mediated via electron establishing chemical bonds connecting two nuclei. Scalar coupling leads to a splitting of NMR signal. The effective magnetic field experienced by the nucleus is affected by the other nuclei in the vicinity of the subject nucleus through shared electrons between them. If there are two nuclei in the molecule, each of the spin can be in the α or β state. There are four possibilities of energy state of the system: $\alpha_1\alpha_2$, $\alpha_1\beta_2$, $\beta_1\alpha_2$ and $\beta_1\beta_2$ (Figure 2-9a). The effective magnetic field at proton 1 at the state $\alpha_1\alpha_2$ is different from that at proton 1 at the state $\alpha_1\beta_2$. Proton 1 can experience two transitions: 12 and 34 (Figure 2-9a). Therefore, proton 1 can resonate at two different resonance frequencies, and consequently the NMR signal of nuclei 1 is split into two peaks (Figure 2-9b). Similarly, proton 2 can experience two transitions: 13 and 24 (Figure 2-9b). The difference in resonance frequency 12-34 and 13-24 transitions is equal

and is called scalar coupling between proton 1 and proton 2. The magnitude of scalar coupling (J) depends on the overlap amount of electron orbitals of the involved nuclei. Consequently, J depends on the hybridization of the nuclei as well as the number of bonds between them. For example, $J_{1H,13C}$ coupling constants are 140 Hz, 160 Hz, and 250 Hz for Sp^3 , Sp^2 and Sp orbital, respectively. The magnitude of J -coupling are important information for designing of through-bond correlation NMR experiments including heteronuclear single-quantum correlation (HSQC), heteronuclear multiple-bond correlation (HMBC), total correlation spectroscopy (TOCSY), and correlation spectroscopy (COSY). The appearance of cross peaks between two protons in these experiments indicates that the two protons are correlated through chemical bonds.

2.2.4 Nuclear Overhauser Effect and Through-space Correlation

Nuclear overhauser effect refers to the signals arising from the spin-spin cross relaxation of spatially closed nuclei. The spectrum obtained from this effect contains diagonal peaks and cross peaks which connect resonances from spatially closed nuclei. The strength of NOE signal is dependent on the distance between nuclei (R), according to the relationship $NOE \propto (1 / R^6)$. Cross peak between resonances of two nuclei indicates that the two nuclei are close in space. NOE can detect cross peaks between two nuclei whose inter-nuclei distance is less than or equal to 5 Å. In structure elucidation, NOE signals are used to measure the distance between two nuclei.

As illustrated in Figure 2-10, the typical pulse sequence of NOESY experiment first requires the application of a 90° pulse which rotates magnetization vector 90° from equilibrium position onto xy plane (transverse plane). This pulse creates transverse magnetization (M_{xy}), which evolves during evolution time t_1 . During this time, the magnetization is labeled with chemical shifts of a first nucleus (f_1). Then, the second pulse turns the transverse magnetization component perpendicular to the pulse direction into longitudinal magnetization. The spin-spin cross relaxation of two spatially closed nuclei occurs during mixing time t_m , which is constant throughout the experiment. During mixing time, the magnetization is

transferred between a first nucleus to a second nucleus. The third pulse is the read out pulse which rotates the longitudinal magnetization to the transverse plane, which is followed by the acquisition phase to record the signal of transverse magnetization as a function of time t_2 . During the time t_2 , the magnetization is labeled with the chemical shift of the second nucleus (f_2). This series of events is repeated multiple times with varying t_1 and the data undergo double Fourier transform with respect to t_1 and t_2 to generate the 2D NOESY spectrum (f_1, f_2).

Figure 2-10. A typical pulse sequence of NOESY experiment.

NOE signature of Watson-Crick base pairs and G-tetrad

Figure 2-11. Typical NOE signal of G-tetrad and Watson-Crick base pairs. Base arrangement of (a) A-T base pair, (b) G-C base pair and (c) G-tetrad. Red arrow connecting two protons indicates these two protons are close in space and give a corresponding NOE peak.

In Watson-Crick A-T base pairs, thymine imino proton H3 and adenine proton H2 are close to each other (Figure 2-11a), giving rise to a signature NOE cross peak T(H3)-A(H2) (Figure 2-12). Similarly, in Watson-Crick G-C base pairs, guanine imino proton H1 and two cytosine amino protons H4 are spatially closed (Figure 2-11b), giving rise to two NOE cross peaks G(H1)-C(H41) and G(H1)-C(H42) (Figure 2-12) .

In a G-quadruplex, the folding topology can be determined from NOE patterns between guanine imino proton and H8 proton. Within a G-tetrad, each guanine imino proton H1 is closed to a neighboring guanine H8 proton (Figure 2-11c), and therefore exhibit a NOE cross peak H1-H8. A cyclic connectivity between imino H1 proton and H8 proton of four guanines would establish a G-tetrad (Figure 2-12).¹²⁷ By assigning all the signals participating in the G-tetrads through this method, the folding topology of a G-quadruplex can be determined.

Figure 2-12. Schematic showing characteristic NOE cross peaks of Watson-Crick base pairs and G-tetrads. Characteristic NOE cross peaks of A-T base pairs, G-C and G.G.G.G tetrads usually appear in green, red and blue region, respectively. Four cyclic G(H1)-G(H8) NOE peaks in blue region establish G_α . G_β . G_γ . G_δ tetrad.

2.3 NMR Structure Computation

Structures of duplex–quadruplex constructs in this study were calculated using the program XPLOR-NIH.¹²⁸ Structure computation involves two steps: distance geometry simulated annealing and distance restrained molecular dynamics refinement. Inter-proton distance constraints, Hydrogen-bond constraints, planarity constraints, dihedral constraints, and repulsive constraints were imposed during structure calculations. Initially, an extended conformation of the sequence was generated randomly. In distance geometry simulated annealing (DGSA) step, atoms are brought together based on the topological information and NMR constraints provided to build a family of molecular structures, and then the system is subjected to simulated annealing regularization and refinement. During this step, the system is heated up to high temperature and subsequently cooled down to search for global minimum energy states. Structures obtained from DGSA calculation were subjected to NOE-restrained simulated annealing refinement in XPLOR in which the system is heated up, allowed to reach equilibrium and then cooled down. NOE-restraint molecular dynamic calculation is carried out after each step of heating, reaching equilibrium and cooling. 10 best structures with the lowest overall energy were selected.

2.4 Circular Dichroism Spectroscopy

Circular Dichroism is a phenomenon originating from the difference in absorbance of the left- and right-handed polarized light by chiral molecules. When a coherent left- and right-handed circularly polarized light passes through a sample consisting of chiral molecules, it would gain ellipticity which varies with the wavelength of light. A plot of ellipticity across different wavelengths could be used to characterize the folding topology of DNA structure.^{129, 130}

Various topologies of G-quadruplexes are different in strand arrangement, and *syn/anti* glycoside angle of guanines in G-tetrads which can affect the stacking pattern of G-tetrads. As a result, each of the three common G-quadruplex topologies exhibits a unique CD signature. All-parallel-stranded G-quadruplexes give CD spectra with a positive peak at around 260 nm and a negative trough at 240 nm.

Antiparallel G-quadruplexes, comprising of two strands pointing to the same direction and two other strands of the opposite directionality, produce CD spectra with a positive peak at around 295 nm and a negative peak close to 260 nm. Hybrid G-quadruplex, consisting of three strands arranged in the same direction and one strand in the opposite direction, would give two positive peaks at ~295 nm, ~260 nm and a negative peak at around 240 nm.¹³¹ However, it should be noted that the CD profile is determined mainly by the stacking of the bases rather than the strand orientation of the G-quadruplex. Therefore, the use of CD spectra is insufficient to determine the folding topology of the G-quadruplex and should be supported by other biophysical techniques.

Figure 2-13. Examples of CD spectra pattern of different G-quadruplex conformations (a) parallel G-quadruplex formed by d[TG4T] (b) hybrid (3+1) G-quadruplex formed by d[TT(G3TTA)₃GGGA] (c) anti parallel G-quadruplex formed by d[G4T4G4] in potassium solution. Figures is adapted from ref ¹³².

CD spectroscopy is also used in this study to monitor the conformation transitions with the change in solvent or temperature. CD spectroscopy can also be

applied to determine the stability of G-quadruplexes by monitoring the change of CD signal at characteristic wavelength across different temperatures. For example, CD signal at ~260 nm and ~290 nm are recorded at different temperatures to monitor the unfolding or folding process of parallel and antiparallel G-quadruplex, respectively.

2.5 Ultraviolet Absorption

DNA bases absorb ultraviolet (UV) radiations at maximum close to 260 nm. To measure the concentration of a DNA sample, the sample is heated in the quartz cuvette to temperature of 90-95 °C, then the absorbance at 260 nm is recorded. The sample concentration can be calculated based on the formula:

$$A = \epsilon \times C \times l$$

Where

A (a.u) is the UV absorbance at 260 nm,

ϵ ($M^{-1}cm^{-1}$) is the molar extinction coefficient of the oligonucleotide

C (M) is oligonucleotide strand concentration,

l (cm) is the path length of the quartz.

In addition, change in absorbance at 295 nm represents the formation of G-quadruplex. Hence, through monitoring change in the absorbance signal at 295 nm, facilitates the monitoring of G-quadruplex formation over time.

Chapter 3 Joining of Multiple Duplex Stems at a Single Quadruplex Loop

In this chapter, quadruplex-duplex hybrid structures in which multiple duplex stems are incorporated into a single loop of G-quadruplex are presented. The folding topology of a two-stem quadruplex-duplex hybrid construct was validated using nuclear magnetic resonance spectroscopy. This multi-stem incorporation principle is applicable at different loop positions of the same quadruplex construct, and could be extended to three or more duplex stems, giving rise to a diverse range of possible structures. These multi-stem complexes offer new design principles for the assembly of DNA nanoarchitectures. The potential existence of such complex motifs in genomic sequences could have biological implications and would represent novel targets for drug development

This chapter was reproduced with permission from: Lim, K.W.[†], Nguyen, T. Q. N.[‡], and Phan, A. T. (2014) Joining of multiple duplex stems at a single quadruplex loop, *J. Am. Chem. Soc.*, *136*, 17969–17973, doi: [dx.doi.org/10.1021/ja5078816](https://doi.org/10.1021/ja5078816) ([†] These authors contributed equally). Copyright [2017] American Chemical Society.

Author contributions: T. Q. N. Nguyen performed experiments under the supervision of K.W. Lim and A.T. Phan. All authors designed experiments, analysed data and co-wrote the paper.

3.1 Introduction

A nucleic acid junction arises with the confluence of two or more distinct helical segments. It can exist in the form of a branched architecture,¹³³ joining together three or more duplex stems. Branched nucleic acid junctions such as three-way and four-way junctions have been implicated in DNA recombination¹³⁴ and repair,¹³⁵ and they also serve as RNA architectural elements.¹³⁶ On the other hand, a nucleic acid junction can also mediate the transition between two separate helical domains, as in the case of the DNA B-Z junction.¹³⁷ Variants of branched nucleic acid junctions are widely utilized as basic units in the construction of DNA nanomaterials.¹³⁸ To date, most of DNA nanoconstructs have been built based on branched duplex DNA. As compared to duplex DNA, G-quadruplex DNA has higher nuclease resistance, rigidity and thermal stability.¹³⁹ Functional group can be easily incorporated into G-quadruplex scaffold by chemical modification of guanines.¹⁴⁰ In addition, the folding of G-quadruplex can be controlled by monovalent cations. These features make G-quadruplexes become attractive objects to be integrated into DNA nanotechnology design.

A quadruplex-duplex junction¹²⁴ is formed upon the connection of a duplex helix onto a G-quadruplex, a four-stranded helical structure^{11, 12} made up of multiple G•G•G•G tetrads.¹³ The connectivity between a single duplex and quadruplex has been established across the various edges of a quadruplex core, in both coaxial and orthogonal manners,¹²⁴ illustrating the facile juxtaposition of these two structural elements. These quadruplex-duplex hybrid complexes have been shown to display high mechanical¹⁴¹ and thermal stability¹²⁵ and diverse functionalities.^{119, 120, 142} In addition, multiple duplex stems can be coordinated onto a single quadruplex scaffold through their separate connection at distinct edges of the tetrad core.¹²⁴ These observations point to the feasibility application of the tetrad core in serving as an organizational platform on which diverse nucleic acid structural elements can be coordinated. This feasibility application and the novel features of G-quadruplexes as compared to duplexed have inspired us to develop the branched nucleic acid junctions based on the G-quadruplex.

Herein, I reported branched nucleic acid junctions based on the incorporations of multiple duplex stems into a single loop of a G-quadruplex. Two-stem (dx1; dx2, Table 3-1) and three-stem (dx1; dx2; dx3, Table 3-1) hairpin duplex are plugged in the loop of an all-parallel-stranded G-quadruplex (*T25-2T*, Table 3-1),¹⁴³ giving rise to sequences: *2S-QDHL1*, *2S-QDHL2*, *2S-QDHL3*, *3S-QDHL1*, *3S-QDHL2*, *3S-QDHL3* (Table 3-1). The formation of multi-stem quadruplex-duplex hybrid structures by these sequences was validated by NMR. These multi-stem quadruplex-duplex hybrid complexes are attractive candidates to be utilized in the construction of DNA nanomaterials. In addition, the possibility of formation of such stable quadruplex–duplex hybrid structures would further expand the sequence space of quadruplex-forming motifs beyond the well accepted definition of G-quadruplex forming sequences $G_{3+N_1-7}G_{3+N_1-7}G_{3+N_1-7}G_{3+}$ which limits the length of loops to be no more than 7 nucleotides.

Table 3-1. Sequences of Quadruplex and Duplex Stems Used in This Study

Name	Sequence (5'–3') ^{a,b}
<i>2S-QDHL1</i>	TT GGG T (CGCGAA GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T GGG T GGG T
<i>2S-QDHL2</i>	TT GGG T GGG T (CGCGAA GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T GGG T
<i>2S-QDHL3</i>	TT GGG T GGG T GGG T (CGCGAA GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T
<i>3S-QDHL1</i>	TT GGG T (CGCGAA GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) (<u>AGTCGC</u> GTA <u>GCGACT</u>) T GGG T GGG T GGG T
<i>3S-QDHL2</i>	TT GGG T GGG T (CGCGAA GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) (<u>AGTCGC</u> GTA <u>GCGACT</u>) T GGG T GGG T
<i>3S-QDHL3</i>	TT GGG T GGG T GGG T (CGCGAA GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) (<u>AGTCGC</u> GTA <u>GCGACT</u>) T GGG T
<i>1S-QDHL2</i>	TT GGG T GGG (CGCGAA GCA <u>TTCGCG</u>) GGG T GGG TT
<i>T95-2T</i>	TT GGG T GGG T GGG T GGG T
<i>dx1</i>	CGCGAA GCA <u>TTCGCG</u>
<i>dx2</i>	<u>ATCTGA</u> GAA <u>TCAGAT</u>
<i>dx3</i>	<u>AGTCGC</u> GTA <u>GCGACT</u>

^aG-tracts shown in bold. ^bIndividual duplex stems enclosed by parentheses, with self-complementary tracts underlined

3.2 Result and Discussion

3.2.1 A Two-Stem Quadruplex-Duplex Hybrid Construct, *2S-QDHL2*

Starting from all-parallel-stranded G-quadruplex scaffold, two duplex stems were incorporated across a single propeller loop of an all-parallel-stranded G-tetrad scaffold (Figure 3-1a). One-dimensional imino proton NMR spectrum of this two-stem constructs in buffer containing 20 mM KCl and 20 mM KPi (pH ~7), d[TT

GGG T GGG T (CGCGAA GCA TTCGCG) (ATCTGA GAA TCAGAT) GG T GGG T] (G-tracts in bold; individual duplex stems enclosed by matching pairs of parentheses, in which self-complementary tracts are underlined) (henceforth denoted as *2S-QDHL2*; Table 3-1) supported the formation of a single predominant quadruplex–duplex hybrid structure (Figure 3-2a). Twelve major sharp peaks at ~ 11–12 ppm (Figure 3-2a, labeled with filled circles) indicated the formation of a three-layered G-tetrad core, whereas at least nine peaks at ~ 12.6–14 ppm indicated the formation of Watson–Crick base pairs of the two duplex stems.

Figure 3-1. Coordination of multiple duplex stems through a single loop of a G-quadruplex scaffold. Attachment of (a) two and (b) three duplex stems onto the same edge of a G-tetrad core. G-tracts are colored in blue while duplex stems are colored in red. G-tetrad layers are represented as blue planes.

Figure 3-2. 1D imino proton NMR spectra of (a) *2S-QDHL2*, (b) *1S-QDHL2*, (c) *dx1*, and (d) *dx2* in K^+ solution. Drop lines show the proposed assignment of duplex imino protons for *2S-QDHL2* with respect to those for the individual duplex stem components.

Figure 3-3. CD spectra of *2S-QDHL2* (blue) and *3S-QDHL2* (red) at 25 °C in buffer containing 20 mM KCl and 20 mM KPi.

Chemical shift patterns of the duplex imino proton peaks closely resembled those of the individual duplex stem components (Table 3-1 and Figure 3-2), suggesting successful establishment of the two duplex stems. The structures of the two individual duplex stems, which are closed off by a three-nucleotide hairpin loop (GCA and GAA, respectively), were validated by two-dimensional NMR experiments¹²⁴ (supporting information 1). Chemical shift patterns of the quadruplex imino proton peaks were similar to those of single-stem quadruplex-duplex hybrid constructs (Table 3-1 and Figure 3-2) reported previously,^{124, 125} in agreement with its adoption of the desired G-tetrad core topology (see validation below). This was also supported by the CD spectrum of *2S-QDHL2* (Figure 3-3), which was similar to those of single-stem quadruplex-duplex hybrid constructs.¹²⁵ The intramolecular folding nature of *2S-QDHL2* was investigated by analysis of the CD-melting profiles across different DNA strand concentrations (see below).

3.2.2 NMR Validation of 2S-QDHL2 Folding Topology

Table 3-2. Site-Specific Labeled DNA Sequences of 2S-QDHL2 Used in This Study

	Name	Sequence ^{a,b,c,d}
¹⁵ N-labeled	2S-QDH [N3]	TT * GGG T GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T GGG T
	2S-QDH [N4]	TT G*GG T GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T GGG T
	2S-QDH [N8]	TT GGG T G*GG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T GGG T
	2S-QDH [N12]	TT GGG T GGG T (<u>C*GCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T GGG T
	2S-QDH [N14]	TT GGG T GGG T (<u>CGC*GAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T GGG T
	2S-QDH [N23]	TT GGG T GGG T (<u>CGCGAA</u> GCA <u>TTC*GCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T GGG T
	2S-QDH [N30]	TT GGG T GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCT*GA</u> GAA <u>TCAGAT</u>) T GGG T GGG T
	2S-QDH [N38]	TT GGG T GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCA*GAT</u>) T GGG T GGG T
	2S-QDH [N42]	TT GGG T GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T * GGG T GGG T
	2S-QDH [N43]	TT GGG T GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T G*GG T GGG T
	2S-QDH [N44]	TT GGG T GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GG*G T GGG T
	2S-QDH [N46]	TT GGG T GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T * GGG T
	2S-QDH [N47]	TT GGG T GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T G*GG T
	2S-QDH [N48]	TT GGG T GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T GG*G T
² H-labeled	2S-QDH [D3]	TT # GGG T GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T GGG T
	2S-QDH [D7]	TT GGG T # GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T GGG T
	2S-QDH [D42]	TT GGG T GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T # GGG T GGG T
	2S-QDH [D46]	TT GGG T GGG T (<u>CGCGAA</u> GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) T GGG T # GGG T

^aG-tracts shown in bold. ^bIndividual duplex stems enclosed by parentheses, with self-complementary tracts underlined. ^cIndividual guanine nucleotides 2–4% ¹⁵N-labeled are marked by asterisks (*G). ^dIndividual guanine nucleotides ²H-labeled at the H8 position are marked by hash signs (#G).

We proceeded to elucidate the quadruplex-duplex hybrid topology of 2S-QDHL2. Guanine imino protons were unambiguously assigned through ¹⁵N-labeling (2%–4%) at individual positions,¹⁴⁴ one residue at a time (Figure 3-4a; Table 3-2). Solvent exchange experiment indicated a slower rate of exchange for the imino protons of G4, G8, G43, and G47 as compared to the rest (Figure 3-4b), in accordance with their central placement in the G-tetrad core. A previous reported site-specific deuteration strategy¹⁴⁵ was utilized to define the alignments of the G-tetrad layers for 2S-QDHL2. Specifically, gyromagnetic ration of deuterium (²H) is 6.536 MHz/T which is different from that of ¹H (42.576 MHz/T), therefore ²H resonates at different frequency from that of ¹H and does not give signal in proton

(^1H) NMR spectra. Consequently, site-specific ^2H -labeling at a guanine H8 proton would lead to the disappearance of the corresponding imino-H8 NOE cross-peaks (Figure 3-5a,b) with the neighboring guanine residue in the G-tetrad (Figure 3-5c-f), hence confirming the assignment of the NOE cross peak between these two guanine residues. Unambiguous assignment of four cyclic NOE cross peaks by this approach would establish the hydrogen-bond directionality of the respective tetrad. Using this approach, the G-tetrad core topology of 2*S*-*QDHL2* was verified to correspond to the expected all-parallel-stranded arrangement (Figure 3-5g), with the three G-tetrads (G3•G7•G42•G46, G4•G8•G43•G47, and G5•G9•G44•G48) all aligned in the same hydrogen-bond directionality.

Figure 3-4. Guanine imino proton assignments of two-stem quadruplex-duplex hybrid construct 2*S*-*QDHL2* in K^+ solution. (a) Examples of imino proton assignments in ^{15}N -filtered spectra of samples, 2–4% ^{15}N -labeled at the indicated positions. The reference spectrum is shown at the top. G-tetrad imino proton peaks are indicated with filled circles. Regions corresponding to imino protons taking part in G-tetrad and Watson-Crick base pair formation are demarcated. (b) 1D imino proton NMR spectrum after 1 h in D_2O at 25 °C.

Figure 3-5. G-quadruplex folding topology of two-stem quadruplex-duplex hybrid construct 2S-QDHL2 in K⁺ solution, as determined by site-specific ²H-labeling.¹⁴⁶(a) Cyclic imino-H8 NOE connectivity patterns within the top G-tetrad layer (G3•G7•G42•G46), as indicated with arrows. (b) NOESY spectrum (mixing time, 200 ms) showing the imino-H8 NOE connectivity patterns of 2S-QDHL2 across the G3•G7•G42•G46 tetrad. Cross-peaks corresponding to the cyclic connectivity patterns are framed and labeled (in blue) with the residue number of imino protons in the first position and that of H8 protons in the second position. Site-specific ¹H8-to-²H8 substitutions of (c) G3, (d) G7, (e) G42, and (f) G46 led to the disappearance of the corresponding imino-H8 cross-peaks (boxed in red), establishing the hydrogen-bond directionality of the G-tetrad. (g) Schematic structure of the quadruplex-duplex hybrid adopted by 2S-QDHL2. G-tetrad guanine bases are colored in blue. The two duplex stems are colored separately in orange and green. Watson-Crick and non-canonical base pairs of the duplex stems are represented as solid and dotted lines, respectively. 5'- and 3'-terminal residues are shown as red circles.

3.2.3 A Three-Stem Quadruplex-Duplex Hybrid Construct, *3S-QDHL2*

Figure 3-6. Folding topology of three-stem quadruplex-duplex hybrid construct *3S-QDH* in K^+ solution. (a) 1D imino proton NMR spectrum of *3S-QDH* in K^+ solution, with ^{15}N -filtered spectra of samples, doubly labeled (2–4% ^{15}N -labeled at the indicated positions) within the same duplex stems, shown below. G-tetrad imino proton peaks are indicated with filled circles. (b) Schematic structure of the quadruplex-duplex hybrid adopted by *3S-QDH*. G-tetrad guanine bases are colored in blue. The three duplex stems are colored separately in orange, green, and purple. Watson-Crick and non-canonical base pairs of the duplex stems are represented as solid and dotted lines, respectively. 5'- and 3'-terminal residues are shown as red circles

To probe a possible incorporation of multiple duplex stems across the same loop of the all-parallel-stranded G-quadruplex scaffold (Figure 3-1b), a three-stem quadruplex-duplex hybrid complex, d[TT **GGGTGGGT** (CGCGAAGCATTCGCG) (ATCTGAGAATCAGAT) (AGTCGC GTA GCGACT) TGGGTGGGT] (henceforth denoted as *3S-QDHL2*; Table 3-1), was constructed. In comparison to *2S-QDHL2*, additional Watson–Crick base-paired

imino proton peaks were observed in the one-dimensional imino proton NMR spectrum of *3S-QDHL2* at ~ 12.6 – 14 ppm (Figure 3-6a), supporting the formation of three duplex stems. Furthermore, *3S-QDHL2* displayed highly similar G-tetrad imino proton peak patterns as *2S-QDHL2* at ~ 11 – 12 ppm (Figure 3-4, labeled with filled circles), suggesting successful adoption of the desired G-tetrad core topology. Establishment of the three duplex stems was corroborated by double guanine/thymine ^{15}N -labeling (2–4%) within each stem (Figure 3-6, Table 3-3). The CD spectrum of *3S-QDHL2* (Figure 3-3) was also similar to those of single-stem quadruplex-duplex hybrid constructs reported previously.¹²⁵

Table 3-3. Site-Specific Labeled DNA Sequences of *3S-QDHL2* Used in This Study

Name	Sequence ^{a,b,c,d}
3S-QDHL2 [N14/N23]	TT GGG T GGG T (CGC*GAA GCA <u>TTC*GCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) (<u>AGTCGC</u> GTA <u>GCGACT</u>) T GGG T GGG T
3S-QDHL2 [N29/N35]	TT GGG T GGG T (CGCGAA GCA <u>TTCGCG</u>) (<u>ATC*TGA</u> GAA <u>*TCAGAT</u>) (<u>AGTCGC</u> GTA <u>GCGACT</u>) T GGG T GGG T
3S-QDHL2 [N42/N52]	TT GGG T GGG T (CGCGAA GCA <u>TTCGCG</u>) (<u>ATCTGA</u> GAA <u>TCAGAT</u>) (<u>A*GTCGC</u> GTA <u>GC*GACT</u>) T GGG T GGG T

^a G-tracts shown in bold. ^b Individual duplex stems enclosed by parentheses, with self-complementary tracts underlined. ^c Individual guanine/thymine nucleotides 2–4% ^{15}N -labeled are marked by asterisks (*G/*T). ^d

3.2.4 Thermal Stability of *2S-QDHL2* and *3S-QDHL2*

Thermal stability of *2S-QDHL2* in the presence of ~ 50 mM K^+ was measured based on the CD signal transition at 262 nm—a wavelength at which the quadruplex exhibited large variation while the duplex segments exhibited a small variation over the temperature range,¹²⁵ giving a melting temperature (T_m) of ~ 56 °C across different DNA strand concentrations (2 μM : 54.3 °C, 4 μM : 54.3 °C, 50 μM : 57.4 °C, 100 μM : 58.4 °C; Figure 3-7). Small changes in the T_m were observed for strand concentrations from 2 to 100 μM consistent with the intramolecular nature of the major conformation. The T_m of *3S-QDHL2*, on the other hand, was found to be ~ 47 °C. These observations attested to the robust nature of such multi-stem quadruplex-duplex hybrid complexes, which would make them attractive candidates to serve as stable scaffolds in the construction of DNA nanomaterials.¹³⁸

Here, our measurement based on the CD signal transition at 262 nm provides an estimate on the melting process of the quadruplex segment alone.¹²⁵ Interplay between the duplex and quadruplex components are important considerations for the implementation of multi-stem quadruplex–duplex hybrid motifs in DNA nanotechnology, as well as their potential formation in a biological context. The presence of duplex stems within the loop of the quadruplex led to a positive influence on the stability of the quadruplex, relative to that of a non-structured loop of similar lengths (T_m of 2*S-QDHL2* was ~54 °C, compared to ~50 and ~48 °C for T₂₇ and T₃₇, respectively).¹²⁵ However, melting transitions of the duplex and quadruplex components do not necessarily occur simultaneously; T_m of the quadruplex was lower than that of the duplexes under low K⁺, high Li⁺ conditions (supporting information 2), whereas the reverse should apply for weak duplexes (e.g. comprising mismatched base pairs) under high K⁺ conditions.

Figure 3-7. CD melting of 2*S-QDHL2* DNA at different concentrations strands (2–100 μM; color-coded on the top right corner) in solution containing 50 mM K⁺, as monitored through the CD spectra at 262 nm.

3.2.5 Structural Contexts at the Quadruplex-Duplex Junction

Contexts at the junction between the quadruplex and duplex components is highly influential toward the stability and structural integrity of quadruplex-duplex hybrid complexes.¹²⁵ In the case of two-stem and three-stem constructs examined

here, guanine/thymine imino proton peaks were not observed for the potential immediate base pairs at the junction (C11/G25 and A26/T40 for *2S-QDHL2*; C11/G25, A26/T40, and A41/T55 for *3S-QDHL2*), suggesting that their formation might have been disrupted. On the other hand, the next base pairs down the stem are more stably established, supported by sharp imino proton peaks for G12 (for *2S-QDHL2*) and G42 (for *3S-QDHL2*).

In previous instances of single-stem quadruplex-duplex hybrids, the base pair at the junction was found to be extruded due to local geometry.¹²⁴ It was also shown that these bulged nucleotides bridging the quadruplex-duplex junction can be extended up to three nucleotides without negatively affecting the stability of quadruplex-duplex hybrids for which the duplex stem projects out as a propeller loop.¹²⁵ Similar behavior could be expected of the multi-stem constructs examined here because the duplex stems are also projected out of the tetrad core in a side way manner. We note that the size of these two bulges at the junction would affect global orientation of the various duplex stems within the loop.

3.2.6 Diverse Structures of Multi-Stem Quadruplex-Duplex Hybrid Complexes

We have shown that quadruplex-duplex hybrid complexes harboring two (*2S-QDHL2*) and three (*3S-QDHL2*) duplex stems within the same loop of a quadruplex can exist. In both constructs, the duplex stems were incorporated within the middle propeller loop of an all-parallel-stranded G-tetrad core.^{143, 147} Incorporation of these duplex stems in the first (i.e. *2S-QDHL1* and *3S-QDHL1*; Table 1) or third (i.e. *2S-QDHL3* and *3S-QDHL3*; Table 1) propeller loop of the same quadruplex construct also led to the establishment of the expected quadruplex-duplex hybrid structures (Figure 3-8). We note that this multi-stem incorporation strategy could also be applicable to other G-tetrad core topologies (e.g. antiparallel-stranded basket-type or (3+1) G-tetrad scaffold) and/or different loop configurations (i.e. edgewise or diagonal loop).

In a structural context, the two-stem construct *2S-QDHL2* can be regarded as a variation of the three-way junction, while the three-stem construct *3S-QDHL2* a

variation of the four-way junction, wherein the quadruplex unit takes over the place of one duplex arm in each case. In principle, incorporation of four or more duplex stems within the same quadruplex loop could also be possible. Previously, 8-arm and 12-arm branched DNA junctions had been successfully constructed.¹⁴⁸ It would be interesting to find out if any number of these arms can be substituted with quadruplex units.

Figure 3-8. 1D imino proton NMR spectra of (a) *2S-QDHL1*, (b) *2S-QDHL3*, (c) *3S-QDHL1*, and (d) *3S-QDHL3*.

3.2.7 Biological Relevance of Multi-Stem Quadruplex-Duplex Hybrids

The joining of multiple duplex stems at a single quadruplex loop brings forward a myriad of possibilities in which a quadruplex structure (either intramolecular or intermolecular in nature) can arise, wherein complementary bases of the duplex stems facilitate the congregation of G-tracts. Sequence motifs capable of forming intramolecular single-stem quadruplex-duplex hybrids have been identified in the human genome.¹²⁶ The potential existence of more complex sequence motifs in the human genome could reveal as yet unknown biological functions, and would give rise to exciting opportunities for therapeutic targeting,

since the interface between quadruplex and duplex segments could provide unique structural features for specific ligand binding.

3.3 Conclusion

We showed here the simultaneous incorporation of two and three duplex stems within the same loop of a quadruplex to generate stable quadruplex-duplex hybrid structures. The multi-stem quadruplex-duplex hybrid complexes presented offer new design principles for the construction of DNA nanoarchitectures, in which the G-tetrad scaffold serves as a robust scaffold for the coordination of multiple duplex arms. In addition, the potential existence of such complex motifs in genomic sequences could have biological implications, and would represent unique targets for the design of chemical ligands.

3.4 Supporting Information

Supporting Information 1

The supporting information 1 provides the structure elucidation of *dx2* stem by NMR. NMR solution structure of *dx1* was validated in reference ¹²⁴.

Figure S3-1. NMR structural characterization of *dx2*. (a) sequence of the hairpin *dx2* (b) Imino proton NMR spectrum of *dx2* with the assignment (c) NOESY spectrum (mixing time, 200ms) showing the cross-peaks that identify Watson-Crick base pairs. Cross-peaks between thymine imino proton and adenine H2 proton, and between guanine imino proton and cytosine amino protons, are boxed and labeled with the residue number of thymine/guanine in the first position and that of adenine/cytosine in the second position. (d) NOEs from thymine imino proton to adenine H2 proton (e) NOEs from guanine imino proton to cytosine amino proton (f) Schematic of *dx2* hairpin.

Figure S3-2. NOESY spectrum (mixing time, 300ms) of *dx2* in K^+ solution showing H8/H6-H1' connectivity, intra-residue H8/H6-H1' cross peaks, which are labeled with their respective residue number.

Figure S3-3. Solution structure of *dx2* in K^+ solution (a) Ten superimposed lowest energy structures (b) Ribbon view of a representative structure.

Table S3.1. Statistics of the Computed Structures of the *dx2* structure.

	D ₂ O	H ₂ O
A. NMR Restraint		
Distance restraints		
Intraresidue	267	12
Sequential (i, i+1)	170	15
Long-range (i, $\geq i+2$)	9	18
Other restraints		
Hydrogen-bond		24
dihydral		11
repulsive		0
B. Structure Statistics		
NOE violation		
Number (>0.2)	0.100 $\pm$ 0.316	
Maximum violation ($\text{\AA}$)	0.180 $\pm$ 0.025	
Rmsd of violation ($\text{\AA}$)	0.021 $\pm$ 0.002	
Deviations from the ideal covalent geometry		
Bond length ($\text{\AA}$)	0.003 $\pm$ 0.000	
Bond angles ($^\circ$)	0.690 $\pm$ 0.006	
Impropers ($^\circ$)	0.357 $\pm$ 0.006	
Pairwise all heavy atom rmsd ($\text{\AA}$)		
All heavy atoms in duplex hairpin	0.359 $\pm$ 0.271	
All heavy atoms	0.421 $\pm$ 0.167	

Supporting Information 2

The supporting information 2 provides the CD melting of a quadruplex *T10* (sequence: 5'-TTGGGTGGGTTTTTTTTTTGGGTGGGT) and *dx1*, *dx2* hairpin in buffer containing 1 mM KPi and 100 mM LiCl

Figure S3-4. CD melting of *dx1*, *dx2* and *T10* in solution containing 1 mM KPi and 100 mM LiCl showing that the melting temperature of quadruplex *T10* is lower than melting temperature of duplex *dx1* and *dx2*. The signal is detected at 262 nm and 270 nm for the quadruplex and the hairpins, respectively. T_m of *dx1*, *dx2* and *T10* are ~72, 57 and 43 °C, respectively.

Chapter 4 Duplex Formation in

Bulges of a G-Quadruplex

Besides the connection of duplexes to the quadruplex through the loop which has been discussed in chapter 3, the possibility of connection of a duplex to a quadruplex through the bulge is also investigated in this chapter. It is well accepted that a G-quadruplex is formed by four guanine tracts which form continuous G-column to support G-tetrad core. However, contrary to canonical G-quadruplexes, the recent identified bulged G-quadruplex has its bases projected out of G-column.²³ Expanding on the study of the bulged G-quadruplexes, we report the incorporation of a duplex hairpin within a bulge of a G-quadruplex. The influence of the size and position of the duplex bulge on the G-quadruplex structure and stability was investigated using NMR and CD spectroscopy. The high-resolution solution structure of a G-quadruplex containing a duplex bulge was resolved, and it revealed that the duplex bulge is stacked below the G-tetrad at the 3'-end. A duplex bulge can be incorporated at different positions of a G-quadruplex scaffold. In contrast to non-structured bulged G-quadruplex, thermal stability of duplex-bulged G-quadruplex does not decrease with the increase in the size of the bulge. With G-quadruplexes containing duplex bulges representing new structural motifs with potential biological significance, our findings broaden the definition of potential putative G-quadruplex forming sequence in the genome.

The work in this chapter is being written for publication: Nguyen, T. Q. N., Lim, K. W., and Phan, A. T. (2017) Duplex Formation in Bulges of G-Quadruplexes, to be submitted.

Author contributions: T. Q. N. Nguyen performed experiments under the supervision of K.W. Lim and A.T. Phan. All authors designed experiments, and analysed data.

4.1 Introduction

Using the consensus sequence $G_{3+N_1-7}G_{3+N_1-7}G_{3+N_1-7}G_{3+}$ which is comprised of four adjacent runs of guanine separated by 1 to 7 nucleotides, more than 370,000 putative G-quadruplex sequences were found across the human genome.²⁵ However, recent studies have been shown that some sequences did not strictly conform to this algorithm but are capable of forming stable G-quadruplex. For example, G-quadruplex motifs from human c-myc promoter²¹ and c-kit promoter²² were shown to form an unusual G-quadruplex structures which contain a G-column made of discontinuous guanine residues. Various motifs of long loop G-quadruplex structures containing duplex stem-loops (with loop size up to 35 nucleotides) were demonstrated to assemble into quadruplex-duplex hybrids which displayed high thermal stability.^{124, 125} These stem-loop-containing G-quadruplex motifs were found at regulatory important regions of the human genome.¹²⁶

Recently, the formation of bulged G-quadruplex was reported by Mukundan et al.²³ In this study, it was shown that a G-quadruplex can tolerate the presence of bulges (between adjacent guanine residues) within the G-column of the tetrad core wherein the size of the bulge ranges from one to seven nucleotides. The formation of bulge in G-quadruplex extended significantly the number of G-quadruplex forming sequence. Recently, by using a high resolution sequencing-based method, more than 200,000 distinct bulged G-quadruplex structures were detected in the human genome, suggesting the biological relevance of bulged G-quadruplexes.²⁶ The bulges investigated in the study by Mukundan et al. are non-structured and do not contain any explicit stabilization interactions.²³ It was shown that increasing the size of non-structured bulge led to a significant decrease in thermal stability of G-quadruplex. This destabilization effect of non-structured bulge led to a limit in the size of the bulge.

Herein, we investigate the possibility of incorporation of a duplex stem in the bulge of a G-quadruplex. Starting from the bulged G-quadruplex reported in the previous study (*B4-T*, Table 4-1),²³ the T-bulge is replaced by a duplex hairpin motif (*dx2*, Table 4-1), giving rise to a duplex bulged G-quadruplex structure (*B4-*

dx2, Table 4-1). The high resolution NMR structure of this quadruplex-duplex hybrid was elucidated. Thermal stability study by CD reveals that increase the size of duplex hairpin bulge did not have negative effects on the thermal stability of G-quadruplex, allowing the size of the duplex bulge can be extended to 33 nucleotides in this study. The formation of duplex bulge at different positions was also investigated by incorporating the duplex hairpin motif at different bulge positions of the G-tetrad, giving rise to the sequences: *B3-dx2*, *B7-dx2*, *B8-dx2*, and *B12-dx2* (Table 4-1). The NMR and CD data of these sequences showed that a duplex bulge can be formed at various positions of a G-tetrad scaffold. Our finding should broaden the definition of the G-quadruplex forming sequence. Besides, quadruplexes containing duplex bulges represent new structural duplex-quadruplex hybrid motifs with potential biological significance.

Table 4-1. G-quadruplex and duplex sequences used in this study

Name	Sequence (5'-3') ^{a, b}	T _m (°C) ^c
<i>dx2</i>	<u>ATCTGA</u> GAA <u>TCAGAT</u>	
<i>B4-T</i>	TT GGT G T GGG T GGG T GGG T	
<i>B4-15T</i>	TT GG TTTTTTTTTTTTTTTTT G T GGG T GGG T GGG T	33.5
<i>B4-dx2</i> (6bp)	TT GG (<u>ATCTGAGAATCAGAT</u>) G T GGG T GGG T GGG T	46.2
<i>B4-dx3</i> (11bp)	TT GG (<u>ATCTGATATTAGAATAATATCAGAT</u>) G T GGG T GGG T GGG T	49.2
<i>B4-dx4</i> (15bp)	TT GG (<u>ATCTGATATTAGCAGGAAGCTAATATCAGAT</u>) G T GGG T GGG T T GGG T	50.5
<i>B3-dx2</i>	TT G (<u>ATCTGAGAATCAGAT</u>) GG T GGG T GGG T GGG T	54.5
<i>B7-dx2</i>	TT GGG T G (<u>ATCTGAGAATCAGAT</u>) GG T GGG T GGG T	40.9
<i>B8-dx2</i>	TT GGG T GG (<u>ATCTGAGAATCAGAT</u>) G T GGG T GGG T	47.2
<i>B12-dx2</i>	TT GGG T GGG T GG (<u>ATCTGAGAATCAGAT</u>) G T GGG T	45.6

^aG-tracts are shown in bold-face. ^bIndividual duplex stems are enclosed by parentheses, with self-complementary tracts underlined, ^cG-quadruplex melting temperature of samples in buffer containing 60 mM K⁺ detected by CD signal at 263nm.

4.2 Result and Discussion

4.2.1 Inclusion of Duplex Hairpin within The Bulges of G-quadruplex

Previously, it was shown that the sequence TTGGTGTGGGTGGGTGGGT (henceforth denoted as *B4-T*, bulge underlined; Table 4-1) could form a stable all-parallel-stranded G-quadruplex structure with a bulged T5 residue positioned between G4 and G6 of the first G-column (Figure 4-1, left).²³ To investigate the potential expansion of the bulge for the inclusion of a duplex stem, T5 was replaced with a duplex hairpin-forming motif, giving rise to the sequence TTGG(ATCTGAGAATCAGAT)GTGGGTGGGTGGGT (henceforth denoted as *B4-dx2*, duplex hairpin enclosed by parentheses, with self-complementary sequences underlined; Table 4-1).

Figure 4-1. Formation of bulge containing a nucleotide (left) and a duplex motif (right).

1D imino proton NMR spectrum of *B4-dx2* in buffer containing 30 mM KCl and 20 mM KPi showed twelve major peaks at ~10.5–12 ppm (indicated by filled circle, Figure 4-2a), which is characteristic of guanine imino protons involved in the formation of G-tetrads. Five additional imino proton peaks were observed at ~12.5–14 ppm (indicated by open square, Figure 4-2a), indicative of duplex formation. Distribution patterns of these imino proton peaks closely resembled those of the component G-quadruplex (*B4-T*; Figure 4-2b) and duplex hairpin (*dx2*; Figure 4-2c and Table 4-1), indicative of the formation of a quadruplex-duplex hybrid with a duplex located within the bulge of the all-parallel-stranded G-tetrad scaffold (Figure 4-1, right). CD spectrum of *B4-dx2* showed an intense positive peak at ~260 nm and a negative trough at ~240 nm, which is close to the approximated linear sum of the CD spectra of *B4-T* and *dx2* (Figure 4-2d), supporting the formation of all-parallel-

stranded G-quadruplex. Overall, these observations corroborated the hybrid nature of *B4-dx2*.

Figure 4-2. 1D imino proton NMR spectra of (a) *B4-dx2* (b) *B4-T* (c) *dx2* (G-tetrad imino proton peaks are indicated by filled circles, whereas duplex imino proton peaks are indicated by open square) and (d) CD spectra of *B4-T* (pink), *dx2* (blue), *B4-dx2* (red), sum of *B4-T* and *dx2* (green) in K⁺ solution.

4.2.2 NMR Spectral Assignments of *B4-dx2*

We proceeded to elucidate the solution structure of *B4-dx2*. Guanine imino (Figure 4-3a) and aromatic (Figure 4-3b) protons were unambiguously assigned through site-specific ¹⁵N-labeled (4%) samples.¹⁴⁴ Complete assignment of proton resonance of *B4-dx2* was obtained by using through-bond (COSY and TOCSY) and

through-space (NOESY) NMR experiments. The typical sequential NOE connectivity of DNA backbone conformation (n)H8 and (n-1)H1'/H2'/H2''/H3'/H4' are clearly observed along the G-columns and two strands of the hairpin (Figure 4-4). These sequential NOE cross-peaks are absent at the transition from G-column to the bulges (G4-G5 step) and at the three double-chain-reversal loops, T21, T27, and T29. Moderate intensities of the intra-residue H8-H1' NOE cross-peaks indicated *anti* glycosidic conformations for all guanine residues except G20, which adopts a *syn* conformation as seen by its strong intra-nucleotide H8-H1' NOE cross-peak. The imino proton of isolated guanine G20 was clearly detected in a 1D ¹⁵N-filtered experiment at the region of 10.5-12 ppm (Figure 4-3a), confirming the participation of isolated G20 in the G-tetrad formation. Other 11 guanines involved in G-tetrads are G3-G4, G22-G24, G26-G28, and G30-G32.

Within a G-tetrad layer, imino proton of a guanine is in close spatial vicinity to the H8 of the adjacent guanine and therefore would exhibit a NOE cross-peak. By observing these cross-peaks, the folding topology of G-quadruplex can be validated. For example, the NOEs of G3.H1-G22.H8, G22.H1-G26.H8, G26.H1-G30.H8 and G30.H1-G1.H8 point to the formation of the tetrad plane of G3•G22•G26•G30. The other two tetrads planes, G4•G23•G27•G31 and G20•G24•G28•G32, were defined in the similar way (Figure 4-5d). These tetrad arrangements suggest an all-parallel-stranded G-tetrad core topology for *B4-dx2* (Figure 4-5e).

Due to their close proximity, signature NOE cross-peaks are observed in Watson-Crick base pairs between adenine H2 proton and thymine imino proton (Figure 4-5b), as well as between cytosine amino protons and guanine imino proton (Figure 4-5c). These signatures NOE cross-peaks were observed for five base pairs (T6•A18, G17•C7, T8•A16, G9•C15, and T14•A10), pointing to the formation of the duplex stem (Figure 4-5e).

NOEs observed between proton of G20 residue and those of G4 residue, i.e. G20.H8-G4.H8; G20.H8-G4.H1'; G20.H8-G4.H2'; G20.H8-G4.H2''; G20.H8-G4.H3'; G20.H1'-G4.H1'; and G20.H1-G4.H1, indicated that G20 to be in the same G-column as G4 and G3.

Figure 4-3. Guanidine imino and aromatic protons assignment of *B4-dx2* construct in K^+ solution using 4% ^{15}N -labeled samples (a) Guanidine imino proton (b) Guanidine H8 proton

Having determined the arrangement of three G-tetrads and the Watson-Crick base pairs formation, the topology of *B4-dx2* adopts an all-parallel stranded G-quadruplex which involves three continuous tracts of guanine: G22-G24, G26-G28, G30-G31 and one discontinuous tract: G3, G4, and G20 (Figure 4-5e). This discontinuous tract results in a bulge between G4 and G20 which consists of a 5 base pairs duplex stem. Four G-tracts are connected by three 1-nt propeller loops: T21, T25, and T29.

Figure 4-4. NOESY spectrum (mixing time, 300ms) of *B4-dx2* in K^+ solution showing H8/H6-H1' connectivity, intra-residue H8/H6-H1' cross peaks, which are labeled with their respective residue number.

4.2.3 Solution Structure of *B4-dx2*

The solution structure of *B4-dx2* in K^+ (Figure 4-6) was calculated on the basis of NMR restraints (Table 4-2) using the XPLOR-NIH program.¹²⁸ The G-quadruplex core of *B4-dx2* adopts a three-layered all parallel-stranded scaffold, with three single-nucleotide double-chain-reversal loops (T21, T25, and T29). The bases of these loops project outwards from the G-tetrad core.

Figure 4-5. NMR structural characterization of *B4-dx2* (a) NOESY spectrum (mixing time, 200ms) showing the cross-peaks that identify the three G-tetrads and Watson-Crick base pairs. (b) NOEs from thymine imino proton to adenine H2 proton (c) NOEs from guanine imino proton to cytosine amino proton (d) NOEs inter-residue H8-imino proton within G-tetrads (e) Schematic diagram of *B4-dx2*.

The first two thymine residues (T1 and T2) are positioned on top of the G-tetrad core; the last thymine T33 is stacked at the bottom of the last G-tetrad. The duplex at bulge position between G4 and G20 includes five Watson-Crick base pairs, while the predicted base pair A5-T19 is shown to be unpaired. The duplex hairpin showed continuous stacking under the 3'-end of the G-tetrad, allowing the

progressive helical transition between quadruplex and duplex segment. Stacking between the duplex bulge and the quadruplex is also reflected by the generally upfield-shifted imino proton peaks of *B4-dx2* (Figure 4-2a) as compared to that of *B4-T* (Figure 4-2b) in the region of 11-12 ppm. Residue T19 stacks below residue G20 of the G-tetrad in 6 members ring-6 members ring mode (Figure 4-7), consistent with the up-field shift of resonances belonging to T19 and G20. Stacking between T19 and G20 is supported by the NOEs observed between T19.H2'-G20.H1, T19.H2''-G20.H1, T19.C7-G20.H1, T19.H1'-G20.H1, T19.C7-G24.H1, T19.C7-G24.H8, T19.H6-G24.H8, and T19.C7-G24.H1.

Table 4-2. Statistics of the Computed Structures of the *B4-dx2* structure.

	D ₂ O	H ₂ O
C. NMR Restraint		
Distance restraints		
Intraresidue	508	9
Sequential (i, i+1)	194	17
Long-range (i, $\geq i+2$)	25	46
Other restraints		
Hydrogen-bond		72
dihedral		22
repulsive		15
D. Structure Statistics		
NOE violation		
Number (>0.2)	0.500±0.972	
Maximum violation (Å)	0.189±0.063	
Rmsd of violation (Å)	0.015±0.002	
Deviations from the ideal covalent geometry		
Bond length (Å)	0.003±0.000	
Bond angles (o)	0.686±0.009	
Improper (o)	0.340±0.006	
Pairwise all heavy atom rmsd (Å)		
All heavy atoms in G-tetrad core	0.486±0.120	
All heavy atoms in duplex hairpin	0.578±0.154	
All heavy atoms	1.086±0.269

Figure 4-6. Solution structure of *B4-dx2* in K^+ solution (a) Ten superimposed lowest energy structures (b) Ribbon view of a representative structure. Guanines in G-tetrads are colored in cyan; thymine in loops, orange; bases in duplex stem, magenta; backbone and sugar, gray; O4' atoms, red; phosphorus atoms, yellow.

Figure 4-7. Stacking between T19 (magenta) and G20 (cyan).

4.2.4 Effects of Duplex Size on the Structure and Stability of Duplex-Bulged G-quadruplex

To explore the effects of larger duplex bulges on the structure and stability of a G-quadruplex scaffold, we investigated similar sequences containing duplex bulges of 11- and 15-base pairs (denoted as *B4-dx3* and *B4-dx4*, respectively; Table 4-1). In 1D imino proton NMR spectra of *B4-dx2*, *B4-dx3*, and *B4-dx4*, twelve major peaks were observed at 10.5–12 ppm region for all sequences, while the number of peaks at region 12–14 ppm region were increased with the length increase of duplex (Figure 4-8). The similarity of peak patterns for *B4-dx2*, *B4-dx3*, and *B4-dx4* in the

region of 10.5-12 ppm suggested that all of them adopt the same G-quadruplex folding topology. CD spectra of these three structures also show the signature of an all-parallel-stranded G-quadruplex structure with a peak at ~260 nm and a trough at ~240 nm (Figure 4-9).

Figure 4-8. 1D imino proton NMR spectra of G-quadruplex containing duplex bulge of different size in K^+ solution (a) *B4-dx2* (6 base pairs) (b) *B4-dx3* (11 base pairs) and (c) *B4-dx4* (15 base pairs).

Previously, it has been shown that increasing the bulge size within the core of G-quadruplex led to decreasing the stability of the G-quadruplex.²³ In this study, increasing size of duplex bulge within the core of a G-quadruplex did not lead to a decrease in the stability of the G-quadruplex, as observed in the case of *B4-dx2*, *B4-dx3*, and *B4-dx4*. On the contrary, melting temperature slightly increased from 46.2 °C to 50.5 °C as the duplex bulge size was increased from 6 base pairs to 15 base pairs (Figure 4-10). A similar trend was also observed when a duplex of increasing size was incorporated within the loop of a G-quadruplex.¹²⁵ Melting temperature of a G-quadruplex was observed to increase by ~13 °C (Figure 4-10) when a non-structured bulge (*B4-15T*; Table 4-1) is replaced by a duplex bulge with the same length (*B4-dx2*; Table 4-1), indicating that G-quadruplexes with duplex bulge exhibit

significantly higher thermal stability than a G-quadruplex containing non-structured bulge of the same length.

Figure 4-9. CD spectra of *B4-dx2* (green), *B4-dx3* (blue) and *B4-dx4* (red) in buffer containing 30 mM KCl and 20 mM KPi.

Figure 4-10. CD melting of G-quadruplexes containing duplex bulges of different size: *B4-dx2* (blue), *B4-dx3* (red), *B4-dx4* (green), and *B4-15T* (pink) in ~ 60 mM K⁺ solution.

4.2.5 Insertion of Duplex Bulge at Different Positions within the Core of a G-quadruplex

Figure 4-11. Schematic showing the formation of duplex bulge at different positions.

To investigate the effect of different placement of the duplex bulge on the biophysical properties of the quadruplex-duplex hybrid, the duplex bulge was inserted at various positions within the tetrad core (*B3-dx2*, *B4-dx2*, *B7-dx2*, *B8-dx2*, and *B12-dx2*; Table 4-1, Figure 4-11). In all cases, imino proton peaks were observed at both G-quadruplex and duplex characteristic regions (Figure 4-12), which confirmed the formation of quadruplex-duplex hybrids. A single predominant G-quadruplex conformation was observed for *B4-dx2* (Figure 4-12b), *B8-dx2* (Figure 4-12d), and *B12-dx2* (Figure 4-12e) in which duplex bulges are between the second and the third tetrad of G-quadruplex. More than one conformation was observed for *B3-dx2* (Figure 4-12a), *B7-dx2* (Figure 4-12c) in which duplex bulges are between the first and second tetrad. Similar CD spectra were observed for all six sequences (Figure 4-13), suggesting that all major conformations adopt the all-parallel stranded G-quadruplex core topology. Among these structures, *B3-dx2* exhibited the highest thermal stability with T_m of 54.5 °C in presence of ~60 mM K^+ , while the rest exhibited similar T_m (Table 4-1).

Figure 4-12. 1D imino proton NMR spectra of quadruplex-duplex hybrids in K^+ solution with a duplex inserted at different bulge positions (a) *B3-dx2* with duplex at bulge 3, (b) *B4-dx2* with duplex at bulge 4, (c) *B7-dx2* with duplex at bulge 7, (d) *B8-dx2* with duplex at bulge 8, and (e) *B12-dx2* with duplex at bulge 12.

Figure 4-13. CD spectra of *B3-dx2* (black), *B4-dx2* (red), *B7-dx2* (green), *B8-dx2* (blue), and *B12-dx2* (pink).

4.2.6 Broadening the Definition of G-quadruplex Forming Sequences

It was previously shown that disrupted guanines can form bulged G-quadruplexes with bulge size ranging from 1 to 7 nucleotides²³ and 200,000 bulged G-quadruplex structures were found across the human genome.²⁶ In this chapter, we have shown that the size of the bulge can be extended to 33 nucleotides or even longer with the presence of Watson-Crick base pairs within the bulge, allowing isolated guanines participate on the G-tetrad formation. Such duplex bulge can exist at different positions within different types of G-quadruplex scaffolds, which would significantly enhance the flexibility of quadruplex formation beyond the $G_{3+N_{1-7}}G_{3+N_{1-7}}G_{3+}$ sequence motif previously defined for potential G-quadruplex forming sequences. This redefinition of G-quadruplex forming sequence motifs greatly broadens the number of potential G-quadruplex formation sites in the genomes.

4.2.7 A Unique Structure for G-quadruplex Targeting

Targeting G-quadruplex structure was shown to be a promising strategy for anti-cancer therapy.^{63, 64} Numerous G-quadruplex binding ligands have been developed to facilitate G-quadruplex targeting with the current generation of G-quadruplex binding ligands shown to exhibit high affinity toward G-quadruplex and high selectivity between G-quadruplex and other DNA structures. However, these ligands still show poor selectivity between different G-quadruplex structures. As most G-quadruplex binding ligands stack on the top and bottom tetrad of G-quadruplex, it is challenging to achieve specific targeting of a single G-quadruplex amongst 700,000 potential G-quadruplex structures found across the human genome.²⁶ There has been increasing interest in designing specific G-quadruplex binding ligands where additional structural features of G-quadruplex besides the G-tetrad core such as loops and bulges could provide insight to designing a highly specific G-quadruplex ligand. The duplex-bulged G-quadruplex structure in this study presents a new type of quadruplex-duplex hybrid which can serve as a unique structure for specific G-quadruplex targeting. One such duple-bulged G-quadruplex motif was recently identified in bcl-2 promoter,¹⁴⁹ suggesting that this motif is

biologically relevant and could be a potential therapeutic target. Unique structural features of the duplex-bugled G-quadruplex such as broken G-column induced by the bugle can allow the intercalation of ligand between two tetrads, while the minor groove of duplex bugle can serve as a substrate for specific targeting by duplex ligands with sequence selectivity. Polyamide, for example, can be programmed to target a distinguishing sequence with high selectivity. The high specific targeting toward this duplex-bulged G-quadruplex structure can be obtained by simultaneous targeting the duplex part, the G-quadruplex part and the junction between duplex and G-quadruplex.

4.3 Conclusion

In this study, we showed that duplexes can be incorporated into bulges of a G-quadruplex. The high resolution solution structure of one duplex-bulged G-quadruplex was presented, showing the stacking of duplex bulge below the last G-tetrad at 3'-end. In contrast to non-structured bulge, increasing the length of duplex bulge does not lead to a decrease in the thermal stability of G-quadruplex, therefore allowing to extend the size of bulge up to 33 nucleotides in this study which is a lot bigger than 7 nucleotide of non-structured bulge shown previously. Duplex bulges can occur at various positions of a G-quadruplex and increasing the length of duplex bulge does not lead to a decrease in the stability of overall G-quadruplex core. The formation of a long duplex bulge within G-quadruplexes would expand the current knowledge on predicting G-quadruplex forming sequences. Potential existence of such duplex-bulged G-quadruplex structure in the human may serve as unique targets for designing ligands for specific targeting of G-quadruplex.

Chapter 5 Telomeric G-quadruplex

Structure Formed at 3' Interface of

Telomeric Duplex

In this chapter, the structural feature of a quadruplex-duplex hybrid which could occur at the human telomere is investigated. Human telomeric DNA consists thousands of tandem repeats of TTAGGG, ending with a single-stranded G-rich overhang of 100–200 nucleotides at the 3'-end. The telomeric G-strand sequences can adopt a variety of G-quadruplex structures. The assembly of G-rich telomeric sequence into G-quadruplex was shown to inhibit the activity of telomerase enzyme which plays important role in the proliferation of cancer cell, rendering telomeric G-quadruplex as an attractive target for anticancer therapy. This chapter presents a study on the possible structure of a duplex–quadruplex junction that could arise at the point of transition from the double-stranded to the single-stranded segment. Using NMR, we show that four repeats of telomere sequence located immediately downstream of a telomeric double-stranded segment adopt a basket type G-quadruplex with two G-tetrads. The telomeric quadruplex-duplex hybrid structure represents a unique structural scaffold ideal for targeting.

The work in this chapter is being written for publication: Nguyen, T. Q. N., Konig N., Lim, K. W. and Phan, A. T. (2017) Structure of human telomere downstream of telomeric double-stranded segment in K⁺ solution: A basket type G-quadruplex with two G-tetrads, to be submitted.

Author contributions: T. Q. N. Nguyen and N. Konig performed experiments under the supervision of K.W. Lim and A.T. Phan. All authors designed experiments, and analysed data.

5.1 Introduction

Human telomeric DNA consists of thousands of TTAGGG repeats with a single strand overhang of 100-200 nucleotides at 3'-end of the G-rich sequence.³⁶ In somatic cells, telomere's length is gradually shorten after each cell division.⁴⁰ Eventually, when telomeres reach critical length, the cell enter apoptosis and senescence phase.^{42, 43} However, the length of telomere in cancer cells is maintained by a reverse transcriptase enzyme called telomerase which is over-expressed in the majority of cancer cells.^{44, 45} Therefore, inhibition of telomerase activity is one of promising strategy to suppress the growth of cancer cell.^{41, 150-152} Interestingly, folding G-rich sequences of telomere into G-quadruplex has been found to be effective to inhibit telomerase activity in cancer cell and/or induce telomere dysfunction, and consequently lead to cell senescence and apoptosis.^{46, 50, 51, 153} Formation of G-quadruplex at telomere can be induced with small molecules that can bind and stabilize G-quadruplex.^{50, 51} To design effective telomeric G-quadruplex binding ligands, understanding the structures of G-quadruplex adopted by telomeric sequence is necessary.

Over the past decades, G-quadruplex structures adopted by four-repeat human telomeric TTAGGG sequences have been extensively studied under various conditions.¹² At least four high-resolution structures of telomeric G-quadruplex in condition containing potassium have been reported, pointing to the polymorphic nature of telomeric G-quadruplex. The sequence d[A(GGGTTA)₃GGG] was shown to adopt all-parallel-stranded G-quadruplex structure in crystalline state containing K⁺ (Figure 5-1a).⁵⁵ Two hybrid (3+1) G-quadruplex structures (form 1 and form 2) were formed in K⁺ solution by the sequences d[TA(GGGTTA)₃GGG] and d[TA(GGGTTA)₃GGGTT]. Both structures comprise two edgewise and one propeller loops, but varying in the order of loop arrangements (Figure 5-1b-c).⁵⁶ The sequence d[(GGGTTA)₃GGGT] adopted a two-G-tetrad basket-type G-quadruplex (form 3) in solution containing K⁺ (Figure 5-1d).⁵⁷

In telomere, G-quadruplex can form adjacent to a duplex such as quadruplex formation within the telomeric double-stranded region or quadruplex formation at

the point of transition from double-stranded to single-stranded segment. In literature, there is only one structural study about the telomeric quadruplex-duplex interface using X-ray crystallography. In this study, structural detail of a telomeric quadruplex-duplex junction was investigated in a model for which the duplex segment was placed at the 3'-end of the G-quadruplex. A TAT triad was observed to stack at the interface between quadruplex and duplex. However, it should be noted that the three loops TTA were all mutated to T, making it substantially different from the real context.¹²³ With the current difficulty of G-quadruplex ligands to achieve specificity among different G-quadruplex structure, understanding structure of G-quadruplex and adjacent duplex may provide a unique structural scaffold ideal for targeting. Herein, we have designed several sequences to simulate the formation of quadruplex immediately downstream of the telomeric duplex segment. Using NMR, we demonstrated that these sequences form a hybrid duplex-quadruplex structure and the major structure of quadruplex is antiparallel-stranded basket-type involving two layers of G-tetrads.

Figure 5-1. Schematic structure of four reported human telomeric G-quadruplexes adopted by different sequences in condition containing K^+ (a) $d[AG_3(TTAGGG)_3]$ in crystalline state containing K^+ ⁵⁵ (b) $d[TA(GGGTTA)_3GGG]$ in K^+ solution⁵⁶ (c) $d[TA(GGGTTA)_3GGGTT]$ in K^+ solution⁵⁶ (d) $d[(GGGTTA)_3GGGT]$ in K^+ solution⁵⁷. *syn* and *anti* guanines are showed in magenta and cyan, respectively.

5.2 Result and Discussion

5.2.1 Effect of Terminal Bases in Telomeric C-rich Strand

Table 5-1. Sequences of quadruplex-duplex hybrid used in this study.

Name	Sequence (5' - 3') ^{a, b}
<i>TJ-1</i>	(<u>ACCCT</u> GCA <u>AGGGT</u>) TA GGG TTA GGG TTA GGG TTA GGG TT
<i>TJ-2</i>	(<u>AACCCT</u> GCA <u>AGGGTT</u>) A GGG TTA GGG TTA GGG TTA GGG TT
<i>TJ-3</i>	(<u>TAACCCT</u> GCA <u>AGGGTTA</u>) GGG TTA GGG TTA GGG TTA GGG TT
<i>TJ-4</i>	(<u>CTAACCCT</u> GCA <u>AGGGTTAG</u>) GG TTA GGG TTA GGG TTA GGG TT
<i>TJ-5</i>	(<u>CTAACCC</u> GCA <u>GGGTTAG</u>) GG TTA GGG TTA GGG TTA GGG TT
<i>TJ-6</i>	(<u>CTAACCC</u> GCA <u>GGTTAG</u>) GG TTA GGG TTA GGG TTA GGG TT
<i>TJ-7</i>	(<u>CTAAC</u> GCA <u>GTTAG</u>) GG TTA GGG TTA GGG TTA GGG TT
<i>Strand A</i>	C TAA CCC TAA C
<i>Strand B</i>	G TTA GGG TTA GGG TTA GGG TTA GGG TTA GGG TT
<i>Telo22-TT</i>	GG TTA GGG TTA GGG TTA GGG TT
<i>BrG7-form3</i>	GGG TTA BrGGG TTA GGG TTA GGG T

^aG-tracts shown in bold, ^bIndividual duplex stems enclosed by parentheses, with self-complementary tracts underlined

We have designed several quadruplex-duplex constructs for structural analysis using NMR. These constructs consist of three components (i) a quadruplex-forming sequence containing four GGG runs at the 3'-end (ii) a hairpin loop (GCA) (iii) a mixed single-/double-stranded segment of variable length at 5'-end. Four constructs *TJ-1*, *TJ-2*, *TJ-3* and *TJ-4* differ with regards to the length of duplex complementary strand on C-rich strand (Table 5-1). 1D imino proton spectra of these four sequences in solution containing 70 mM KCl and 20 mM KPi (Figure 5-2) show the formation of quadruplex-duplex hybrids as indicated by the presence of imino peaks at 10-12 ppm which is characteristic of G-quadruplex formation and imino peaks at 12-14 ppm which is characteristic of Watson-Crick base pair formation. Multiple G-quadruplex conformations were observed in all four constructs as indicated by the number of imino protons observed in the 10-12 ppm region. Interestingly, one major G-quadruplex conformation is adopted by *TJ-4* sequence, as indicated by 10 imino proton peaks (indicated by dot, Figure 5-2d) in the region 10-12 ppm. The terminal nucleotides on C-strand in *TJ-4* construct is 3'-ATC-5' which were observed in 80% of human telomere DNA.³⁸ Therefore, *TJ-4* construct

represents the majority of telomeric double-stranded to single-stranded transition in a natural context.

Figure 5-2. 1D imino proton NMR spectra of (a) *TJ-1* (b) *TJ-2* (c) *TJ-3* (d) *TJ-4* in K^+ solution.

5.2.2 Effects of Duplex Hairpin Length

Next, we varied the length of duplex hairpin to examine the effects of hairpin length on the structure of G-quadruplex. Construct *TJ-4*, *TJ-5*, *TJ-6*, *TJ-7* have the same terminal base at 5'-end but their duplex hairpin length are 8, 7, 6, and 5, respectively (Table 5-1). 1D imino proton NMR spectra of these four constructs show characteristic peaks of quadruplex and duplex (Figure 5-3a-d), suggesting the adoption of quadruplex-duplex hybrid by these sequences. The similarity in imino proton NMR spectra at the region 10-12 ppm suggests that these four constructs adopt the same G-quadruplex scaffold. We also titrated complementary strand (Strand A) to a single strand (Strand B) to form a duplex structure at the 5'-end of G-quadruplex instead of a duplex hairpin. Imino proton NMR spectrum of the complex (Strand A + Strand B) shows the same peak pattern in the region of 10-12

ppm to those of constructs *TJ-4*, *TJ-5*, *TJ-6*, *TJ-7* (Figure 5-3e). These observations support our view that very long duplex in telomere context will not affect the structure of G-quadruplex form at the point of transition from telomeric double-stranded to single-stranded segment. The quadruplex structure forming at downstream of very long duplex in telomere context may have the same structure as the one in our constructs presented here (*TJ-4*, *TJ-5*, *TJ-6*, and *TJ-7*).

Figure 5-3. 1D imino proton NMR spectra of (a) *TJ-7* (b) *TJ-6* (c) *TJ-5* (d) *TJ-4* (e) Strand A + Strand B in K^+ solution.

5.2.3 NMR Assignment of *TJ-5*

Figure 5-4. NMR spectral assignments of *TJ-5* in K^+ solution. The reference spectrum is shown at the top. Below are examples of imino proton assignments in 2% ^{15}N -filtered samples at the indicated positions. ^{15}N -filtered experiment of G23 was carried out at 313 K, other experiments were carried out at 298 K.

TJ-5 was chosen for further structural analysis to balance the thermal stability of duplex hairpin and the length of the sequence. Short hairpin is less stable and long sequence leads to difficulty in NMR analysis. Guanine imino protons were unambiguously assigned through ^{15}N -labeling (2%) at individual positions, one residue at a time (Figure 5-4).

¹ ³ ⁶ ⁹ ¹² ¹⁵ ¹⁸ ²⁰ ²³ ²⁶ ²⁹ ³² ³⁵
 (CTAACCC GCA GGGTTAG) **GG** TTA **GGG** TTA **GGG** TTA **GGG** TT

10 imino proton peaks in 10-12 ppm region of *TJ-5* were assigned to all guanine residues in the last four GGG runs at the 3'-end of *TJ-5* (highlighted in red in the above sequence). Imino proton peak of G23 (highlighted in blue) showed up at 313K. Imino proton peak of G17 was not observed, suggesting that G17 may not participate in G-tetrad formation.

5.2.4 Effects of Mutation on *TJ-5* Sequence

G-to-T mutation

To confirm the exclusion of G17 residue in the formation of G-tetrads, we systematically performed G-to-T mutation for all the guanines in the four GGG runs at the 3'-end, one at a time. Thymine substitution at G17 (termed as T17) exhibited a similar but much cleaner imino proton spectrum than that of *TJ-5* (Figure 5-5), confirming the exclusion of G17 in the formation of major G-quadruplex conformation by *TJ-5* sequence. In addition, the broadened imino proton peak of G23 in *TJ-5* was observed in the imino proton spectrum of T17. Imino proton NMR spectra of *TJ-5* and T17 exhibited high similarity (8 peaks at 11-12 ppm and two characteristic peaks at 10.4-10.8 ppm) with that of *BrG7-form 3* (Figure 5-5) which was previously reported to adopt an antiparallel-stranded two-layer G-quadruplex (Figure 5-1d).⁵⁷ G-to-T mutation at other positions substantially abolishes the formation of G-quadruplex as indicated by the disappearance of imino proton peaks in the region of 10-12 ppm (Figure 5-5), suggesting their involvement in G-tetrad core formation or in stabilizing G-tetrad core.

G-to-I mutation

We have also performed single G-to-I substitution at four guanine positions including G17, G25, G29 and G37 (denoted as I17, I25, I29 and I37, respectively). We found that these G-to-I substitutions gave rise to spectra similar to that of *TJ-5* (Figure 5-6). It has been shown that the substitution of a guanine to inosine could preclude the nucleotide at that position in G-tetrads formation since inosine lacks the amino group at the C2 position (Figure 5-6a).¹⁵⁴

Figure 5-5. 1D imino proton NMR spectra of single G-to-T mutations on *TJ-5* sequence in K^+ solution. Imino proton peaks of *Br*G7-form 3 are labeled with original residue number of *TJ-5* sequence.

I17 mutation exhibited a similar but much cleaner spectrum than that of *TJ-5* (Figure 5-6c), further supporting that G17 does not participate in G-tetrad formation of major conformation adopted by *TJ-5*. The similarity in imino proton NMR spectra of I25, I29, I37 and *TJ-5* (Figure 5-6d-f) suggest that G25, G29 and G37 may not participate in G-tetrads but play an important role in stabilizing the G-quadruplex core since G-to-T mutation at these positions completely prevented the formation of G-quadruplex. Previously, G-to-I mutation at a corresponding position to G29 position of this sequence was applied in a variant telomeric sequence d[*A*(GGGT_A)₂IGGT_AGGG] to favor the formation of two-tetrad G-quadruplex.¹⁵⁴

Figure 5-6. 1D imino proton NMR spectra of single G-to-I mutations on *TJ-5* sequence in K⁺ solution: (a) chemical structure of guanine and inosine base (b) *TJ-5*, (c) I17, (d) I25, (e) I29, and (f) I37.

From these observations, we predicted that *TJ-5* may adopt antiparallel G-quadruplex with two G-tetrads involving G18, G19, G23, G24, G30, G31, G35, G36 and one triad (G29.G25.G37).

5.2.5 NMR Elucidation of *TJ-5* Folding Topology

Figure 5-7. Determination of G-quadruplex folding topology of *TJ-5* in K^+ solution by site-specific 2H -labeling. (a) Characteristic H1–H8 NOE connectivity patterns within (G18•G30•G36•G24) tetrad, as indicated by arrows. (b) NOESY spectrum (performed at 200 ms mixing time) showing the H1–H8 characteristic NOE patterns of *TJ-5* of the G18•G30•G36•G24 tetrad. NOE cross-peaks between H1–H8 are indicated in blue boxes with the residue number of imino protons in the first position and that of H8 protons in the second position. NOESY spectrum of site-specific 1H8 -to- 2H8 substitution samples led to the vanishing of the corresponding H1–H8 cross-peaks (indicated by red box), confirming the G-tetrad alignment: (c) G18, (d) G30, (e) G36, (f) G24 and (g) G23. (h) Schematic showing G-quadruplex folding topology adopted by *TJ-5*. *Anti* and *syn* guanine bases in G-tetrads are shown in cyan and magenta, respectively. Guanine bases in triad are colored in yellow.

Next, we proceeded to elucidate the quadruplex folding topology of *TJ-5*. We employed the site-specific deuteration strategy reported previously to define the G-tetrad arrangement for *TJ-5*.¹⁴⁵ A series of NOESY experiments were performed on selective deuterated samples. Specifically, the tetrad G18.G30.G36.G24 was chosen for validation. ²H-labeling at selective guanine H8 proton would cause the vanishing of the corresponding H1-H8 cross peak with the adjacent guanine residues in the same G-tetrad. Using this approach, four NOE cross peaks including G18.H1-G30.H8, G30.H1-G36.H8, G36.H1-G24.H8, G24.H1-G18.H8 corresponding to the cyclic connectivity within G18.G30.G36.G24 tetrad were validated (Figure 5-7), thus confirm the alignment of the G18.G30.G36.G24 tetrad which corresponds to the expected antiparallel-stranded G-quadruplex scaffold which is comprised of two tetrads (G18.G30.G36.G24 and G19.G23.G35.G31). The disappearance of cross peak G19.H1-G23.H8 upon substitution of hydrogen H8 in G23 by ²H further confirmed the folding topology of *TJ-5*.

Figure 5-8. Schematic of the quadruplex-duplex hybrid structure adopted by *TJ-5* sequence.

Using NMR, we have shown that the majority population of *TJ-5* adopts a quadruplex-duplex hybrid structure in which the quadruplex adopts a basket-type two-tetrad folding topology (Figure 5-8). However, it should be noted that about 12 peaks are observed in the duplex region of imino proton NMR spectrum of *TJ-5* (indicated by asterisk, Figure 5-5) while the duplex hairpin contains only 7 base pairs which would be expected to give about 6 peaks (the imino proton that belongs to the outermost base pair is usually broadened and cannot be observed). This observation suggested that *TJ-5* also adopts other conformations which could be determined in future studies.

5.2.6 Truncated four-repeat telomere sequence adopts two-layer G-quadruplex

Since it is shown that guanine residue at position 17th in the sequence *TJ-5* does not participate in the formation of G-tetrad, we deleted the whole hairpin motif in *TJ-5* sequence, giving rise to *Telo22-TT* sequence (Table 5-1). This sequence contains four telomeric repeats without the guanine at the 5'-end. 1D imino proton NMR spectrum of this sequence is very clean and similar to those of *TJ-5*, *T17* and *I17* with 9 peaks in 11.2-12ppm region and two peaks in 10.4-10.8 ppm region (Figure 5-9), suggesting the formation of a similar two-layer G-quadruplex scaffold. CD spectrum of *Telo22-TT* in K⁺ solution is similar to that of *^{Br}G7-form 3* (Figure 5-10) which was previously shown to adopt antiparallel G-quadruplex structure containing two G-tetrads, further supporting the adoption of a two-layer G-quadruplex structure by *Telo22-TT*.⁵⁷ In most previous studies, four-repeat telomere sequences usually exhibited conformational heterogeneity in K⁺ solution. To obtain one major conformation, telomere sequences were usually modified by replacing guanines with guanine analogues such as ^{Br}G or inosine. Recently, truncated four-repeat sequence without 3'-end guanine d[TAGGG(TTAGGG)₂TTAGG] exhibited clean 1D imino proton NMR spectrum in K⁺ solution and neutral pH without any mutation on the sequence.¹⁵⁵ However, NMR spectrum of this sequence still showed two set of peaks which belong to two conformations. In this study, the NMR spectrum of *Telo22-TT* only exhibited one set of peaks. This is the first time it has

been shown that native telomere sequence can adopt one major conformation without any mutation.

Figure 5-9. 1D imino proton NMR spectra of (a) *TJ-5* (b) *T17* (c) *I17* (d) *Telo22-TT* in K^+ solution.

Figure 5-10. CD spectra of *Telo22-TT* (blue) and $BrG7$ -form3 (red) in K^+ solution.

5.2.7 Potential Molecular Target for Small Molecules

Although telomeric G-quadruplexes represent potential targets for anticancer therapy, achieving selectivity between telomeric G-quadruplex structures and numerous other G-quadruplex structures that are present in the human genome still remain a challenge. Instead of targeting only G-quadruplex structure, alternative strategies which target quadruplex-duplex junction could be employed to overcome the poor selectivity of G-quadruplex ligands. *TJ-5* construct in this study represents 80% of telomere junction context in the human cells, and thus could be considered as an ideal substrate for designing small molecules which specifically bind to this junction. A detailed study on the structure of *TJ-5* would provide a platform for drug design. Rational design of the ligand should take into consideration the relative orientation between the duplex and G-quadruplex segments as well as the interface between these two motifs.

5.3 Conclusion

We showed the successful design of a hairpin-single stranded construct which represents the majority of the transition between duplex-single stranded segments in human telomere. This sequence was shown to adopt a quadruplex-duplex hybrid structure in which the folding topology of the quadruplex is an antiparallel structure containing two G-tetrads. The telomeric quadruplex-duplex hybrid structure presented in this study represents a unique scaffold ideal for targeting.

Chapter 6 A Dual-Specific Targeting Approach Based on the Simultaneous Recognition of Duplex and Quadruplex Motifs

Small-molecule ligands targeting nucleic acids have been explored as potential therapeutic agents. Duplex groove-binding ligands have been shown to recognize DNA in a sequence-specific manner. On the other hand, quadruplex-binding ligands exhibit high selectivity between quadruplex and duplex, but show limited discrimination between different quadruplex structures. In addition to the large planar surface of G-tetrad, quadruplex-duplex hybrid structures provide the additional structural features such as the minor and major groove of duplex segments and the junction context between the quadruplex and duplex segments. In this chapter, we exploit the DNA-binding sequence specificity of duplex ligands and the high binding affinity of G-quadruplex ligands to propose a dual-specific approach to target quadruplex-duplex hybrid structures. We demonstrated that a quadruplex-specific ligand and a duplex-specific ligand can simultaneously interact at two separate binding sites of a quadruplex-duplex hybrid harboring both quadruplex and duplex structural elements. Such a dual-specific targeting strategy would combine the sequence specificity of duplex-binders and the strong binding affinity of quadruplex-binders, potentially allowing the specific targeting of unique quadruplex structures. Future research can be directed towards the development of conjugated compounds targeting specific genomic quadruplex-duplex sites, for which the linker would be highly context-dependent in term of length and flexibility, as well as the attachment points onto both ligands.

This chapter was reproduced with permission from: Nguyen, T. Q. N., Lim, K. W., and Phan, A. T. (2017) A Dual-Specific Targeting Approach Based on the Simultaneous Recognition of Duplex and Quadruplex Motifs, *Sci. Rep.*, 7, 11969, DOI:10.1038/s41598-017-10583-9.

Author contributions: T. Q. N. Nguyen performed experiments under the supervision of K.W. Lim. All authors designed experiments, analysed data, and co-wrote the paper.

6.1 Introduction

Small-molecule ligands with strong binding affinity to nucleic acids have been explored as potential therapeutic agents.¹⁵⁶⁻¹⁵⁸ These ligands recognize DNA either in its canonical double helical form,^{158, 159} or in alternative forms such as four-stranded G-quadruplex structure.^{64, 160} There are various classes of duplex-binding ligands, including groove-binders,^{158, 159} intercalators,^{161, 162} cross-linking agents,¹⁶³ and triplex-forming oligonucleotides (TFO).^{164, 165} Sequence-specific recognition of duplex DNA was achieved with groove-binders and triplex-forming oligonucleotides through the explicit establishment of hydrogen-bond interactions. Such targeting approach enables selective gene silencing at the target site through down-regulation of transcription.¹⁵⁷ On the other hand, the majority of quadruplex-binding drugs investigated to date mainly recognize G-quadruplex structures through stacking onto terminal G-tetrads.^{110-113, 166} Current generation of quadruplex-binding ligands exhibit very high binding affinity to G-quadruplex, and they show a high level of discrimination of G-quadruplex from all other DNA structural forms. However, selectivity of these ligands between different quadruplex structures is still limited. With more than 700,000 potential G-quadruplex-forming sites within the human genome,²⁶ the challenge is to develop a quadruplex-binding drug specific to a single genomic G-quadruplex with minimal off-target binding. Here we propose a dual-specific targeting strategy based on the simultaneous application of duplex- and quadruplex-binding ligands. Using NMR spectroscopy we demonstrated that a quadruplex-specific ligand and a duplex-specific ligand can simultaneously interact at two separate binding sites of a quadruplex-duplex hybrid harbouring both quadruplex and duplex structural elements. This approach combines the sequence specificity of duplex-binders with the tight binding affinity of quadruplex-binders, and can be directed towards the selective targeting of quadruplex-duplex hybrid-forming sequences.¹²⁶ Chemical linkage of the two ligands could potentially lead to their synergistic recognition of the target quadruplex-duplex hybrid structure.

6.2 Result and Discussion

6.2.1 Binding of Quadruplex-Duplex Hybrid with Duplex Ligand

Figure 6-1. Schematic diagram of *QDH1*. Quadruplex and duplex segments are coloured in cyan and magenta, respectively. The 5'- and 3'-terminal residues are coloured in red. Guanine base G49 in syn glycosidic conformation is shaded.

Table 6-1. Sequences of Quadruplex-Duplex Hybrids Used in This Study (complementary tracts underlined).

Name	Sequence ^{a,b}
<i>QDH1</i>	TT GG TAGGG T GGG T GGG <u>CATGGATATATCCT</u> GCA <u>AGGATATATCCATGG</u>
<i>QDH2</i>	GG TT GGCTGGATATATCCT GCA <u>AGGATATATCCAGGG</u> TT GG
<i>QDH3</i>	TT GGG T GGG TGGATATATCC GCA <u>GGATATATCCT</u> GGG T GGG T
<i>dx</i>	<u>CATGGATATATCCT</u> GCA <u>AGGATATATCCATG</u>

^aG-tracts are shown in bold-face. ^bself-complementary tracts are underlined

To begin with, a quadruplex-duplex hybrid¹²⁴ (*QDH1*; Table 6-1 and Figure 6-1) containing both a duplex segment harbouring six continuous A•T base pairs and a quadruplex was titrated with the duplex-binder netropsin¹⁶⁷ (Figure 6-2d), which was shown to recognize AT-rich regions, and the bisquinolinium quadruplex-binder Phen-DC₃¹⁶⁸ (Figure 6-3d), individually. 1D imino proton NMR spectrum of free *QDH1* is shown in Figure 6-2a: twelve major peaks (indicated with open circles)

were observed at 10.6–11.8 ppm, corresponding to the formation of a three-layered G-tetrad core; fourteen major peaks (indicated with filled squares) were observed at 12.4–14 ppm, corresponding to the formation of the duplex stem. Upon adding half the molar equivalent of netropsin to *QDHI*, additional peaks emerged in the duplex region with a concomitant reduction in the intensity of the original duplex peaks (Figure 6-2b), while the quadruplex peaks remained largely unchanged. This indicated the specific binding of netropsin onto the duplex stem, giving rise to equal populations of bound and unbound duplex stems. At 1:1 DNA to ligand ratio, the duplex stem was fully bound with netropsin at the AT-rich binding site, as evidenced by the disappearance of the unbound duplex peaks (Figure 6-2c), whereas the quadruplex peaks showed minimal change.

Figure 6-2. NMR spectra show the binding of *QDHI* to netropsin. 1D imino proton NMR spectrum of (a) free *QDHI*, (b) *QDHI* bound with half the molar equivalent of netropsin, (c) *QDHI* bound with equimolar ratio of netropsin, and (d) chemical structure of netropsin, a DNA minor groove-binder. G-quadruplex imino proton peaks are labelled with open circles, whereas duplex imino proton peaks of free *QDHI* are labelled with filled squares. Schematic structure of netropsin (in pink) binding to the duplex stem is shown on the right.

6.2.2 Binding of Quadruplex-Duplex Complex with Quadruplex Ligand

Figure 6-3. NMR spectra show the binding of *QDHI* to Phen-DC₃. 1D imino proton NMR spectrum of (a) free *QDHI*, (b) *QDHI* bound with half the molar equivalent of Phen-DC₃, and (c) *QDHI* bound with equimolar ratio of Phen-DC₃, and (d) chemical structure of Phen-DC₃, a quadruplex-binder. Bound and unbound G-quadruplex imino proton peaks are labelled with open circles and filled diamonds, respectively, whereas duplex imino proton peaks of free *QDHI* are labelled with filled squares. Peaks labelled with asterisks originate from the ligand Phen-DC₃. Schematic structure of Phen-DC₃ (in yellow) binding to the G-tetrad is shown on the right.

Titration of *QDHI* with Phen-DC₃ produced comparable results. Addition of half the molar equivalent of Phen-DC₃ to *QDHI* led to the emergence of additional peaks in the quadruplex region (indicated by diamonds), with a concomitant reduction in the intensity of the original quadruplex peaks (Figure 6-3b), while the

duplex peaks showed minimal change. This indicated the specific binding of Phen-DC₃ onto the G-quadruplex. At 1:1 DNA to ligand ratio, the quadruplex was fully bound, as evidenced by the disappearance of the unbound quadruplex peaks (Figure 6-3c). Upon Phen-DC₃ binding, the quadruplex peaks were generally upfield-shifted, reflecting the aromatic stacking effect of Phen-DC₃. Furthermore, these quadruplex peaks were sharp and showed similar line widths as those in the free DNA, indicating the tight binding of Phen-DC₃ to the quadruplex.

6.2.3 Simultaneous Binding of Quadruplex-Duplex Hybrid with Duplex and Quadruplex Ligands

Figure 6-4. NMR spectra show the binding of *QDH1* to netropsin and Phen-DC₃. 1D imino proton NMR spectrum of (a) free *QDH1* and (b) *QDH1* bound with equimolar ratio of netropsin and Phen-DC₃. Bound and unbound G-quadruplex imino proton peaks are labelled with open circles and filled diamonds, respectively, whereas duplex imino proton peaks of free *QDH1* are labelled with filled squares. Peaks labelled with asterisks originate from the ligand Phen-DC₃. Schematic structure of simultaneous netropsin (in pink) and Phen-DC₃ (in yellow) binding to *QDH1* is shown on the right.

Our dual-specific targeting strategy involves the simultaneous application of both duplex and quadruplex ligands for enhanced binding specificity and/or affinity. We demonstrated this approach with the addition of both netropsin and Phen-DC₃ to *QDH1*. The resulting 1D imino proton NMR spectrum displayed features corresponding to those of the respective bound segments; duplex peaks at 12.4–14 ppm showed similar distribution patterns as in *QDH1*:netropsin complex, while tetrad peaks at 10.4–11.7 ppm matched those of *QDH1*:Phen-DC₃ complex (Figure 6-4). These observations indicated that the binding of both netropsin and Phen-DC₃ to *QDH1* is compatible and non-interfering. 2D NOESY spectrum of *QDH1*:netropsin:Phen-DC₃ complex showed signature G(H1)-C(H41)/G(H1)-

C(H42) and T(H3)-A(H2) cross-peaks indicative of Watson-Crick G•C and A•T base pair formation, as well as characteristic guanine imino-H8 cross-peaks consistent with G-tetrad formation (Figure 6-5). In addition, cross-peaks were observed between *QDH1* and Phen-DC₃ (boxed in red; Figure 6-5), further supporting their interactions. A structural model of the *QDH1*: netropsin:Phen-DC₃ complex showing the simultaneous binding of netropsin to the duplex AT-rich region and Phen-DC₃ to the terminal G-tetrad of *QDH1* was built and shown in Figure 6-6.

Figure 6-5. 2D NMR NOESY spectrum of *QDH1*:netropsin:Phen-DC₃ complex.

CD-melting experiments were performed to investigate the binding of netropsin and Phen-DC₃ to duplex and quadruplex motifs, either individually or in concert. The melting temperature (T_m) of the reference duplex hairpin *dx* (Table 6-1), as monitored at 267 nm, was increased by ~19 °C in the presence of netropsin, but was not affected in the presence of Phen-DC₃ (Figure 6-7). This indicated that netropsin binds to the duplex hairpin, whereas Phen-DC₃ does not.

Figure 6-6. Structural model of *QDH1*:netropsin:Phen-DC₃ complex. Phen-DC₃ is coloured in yellow while netropsin is coloured in pink.

Figure 6-7. Melting curve of dx hairpin in solution containing 50 mM K⁺ in the absence and presence of ligands detected by CD signal at 267 nm: free *dx* (blue), *dx*:Phen-DC₃ (red), *dx*:netropsin (green).

Figure 6-8. (a) CD spectra of free *QDH1* (blue), *QDH1*:netropsin (pink), *QDH1*:Phen-DC₃ (green), and *QDH1*:netropsin:Phen-DC₃ (red), (b) Quadruplex melting curves in solution containing 50 mM K⁺ in the absence and presence of ligands detected by CD signal at 256 nm: free *QDH1* (blue), *QDH1*:netropsin (pink), *QDH1*:Phen-DC₃ (green), *QDH1*:netropsin:Phen-DC₃ (red).

The CD spectrum of free *QDH1* was similar to those of *QDH1* bound with either or both of netropsin and Phen-DC₃ (Figure 6-8a), suggesting that ligand

binding did not alter the folding topology of *QDH1*. Melting of the quadruplex component of *QDH1* (Figure 6-8b) was monitored at 256 nm – a wavelength at which the CD signal of the duplex was close to zero and only exhibited a small variation over the temperature range. We observed that netropsin did not affect the T_m of the G-quadruplex, whereas Phen-DC₃ led to an increase in T_m of more than 11 °C (Figure 6-8b). In the presence of both netropsin and Phen-DC₃, the T_m of the quadruplex was close to that in the presence of Phen-DC₃ alone, suggesting that netropsin does not negatively affect the binding of Phen-DC₃ onto the quadruplex. Our data were consistent with previous observations of the high selectivity of Phen-DC₃ between quadruplex and duplex,¹⁶⁸ and indicated that netropsin and Phen-DC₃ do not negatively interfere with their respective binding of duplex and quadruplex motifs. It was reported that distamycin, a minor groove-binder similar to netropsin, could bind to the groove of a G-quadruplex^{169, 170} or stack on a terminal G-tetrad.¹⁷¹ In our case, there exists the possibility that netropsin, aside from binding to the duplex minor groove, could bind to the quadruplex yet do not significantly affect the binding of the high-affinity quadruplex-binder Phen-DC₃.

Figure 6-9. NMR spectra show the binding of *QDH2* to netropsin and Phen-DC₃. 1D imino proton NMR spectrum of (a) free *QDH2*, (b) *QDH2* bound with equimolar ratio of netropsin, (c) *QDH2* bound with equimolar ratio of Phen-DC₃, and (d) *QDH2* bound with equimolar ratio of netropsin and Phen-DC₃. Schematic structure of simultaneous netropsin (in pink) and Phen-DC₃ (in yellow) binding to *QDH2* is shown on the right.

Figure 6-10. NMR spectra show the binding of *QDH3* to netropsin and Phen-DC₃. 1D imino proton NMR spectrum of (a) free *QDH3*, (b) *QDH3* bound with equimolar ratio of netropsin, (c) *QDH3* bound with equimolar ratio of Phen-DC₃, and (d) *QDH3* bound with equimolar ratio of netropsin and Phen-DC₃. Schematic structure of simultaneous netropsin (in pink) and Phen-DC₃ (in yellow) binding to *QDH3* is shown on the right.

To determine the general applicability of this dual-specific targeting strategy, we further titrated various quadruplex-duplex hybrid constructs (Table 6-1) with the same ligands. In all cases, we observed respective binding of both duplex and quadruplex segments by netropsin and Phen-DC₃ (Figure 6-9 and Figure 6-10), as indicated by the change in chemical shift of the imino proton peaks. On the other hand, we have also performed titration of *QDH1* with different quadruplex-binding ligands including BRACO-19¹¹⁷ and pyridostatin⁵³ (Figure 6-11 and Figure 6-12), demonstrating that these ligands too can be utilized in the current approach. Herein, we have employed just the single duplex ligand netropsin, which binds to the minor groove of AT-rich duplexes. In principle, other duplex ligands with different sequence selectivity can also be applied. For instance, the polyamide class of duplex ligands can be designed to target specific sequences of choice.^{157, 159}

Figure 6-11. NMR spectra show the binding of *QDHI* to netropsin and BRACO-19. 1D imino proton NMR spectrum of (a) free *QDHI*, (b) *QDHI* bound with equimolar ratio of netropsin, (c) *QDHI* bound with equimolar ratio of BRACO-19, and (d) *QDHI* bound with equimolar ratio of netropsin and BRACO-19 (e) Chemical structure of quadruplex-binder BRACO-19.

6.2.4 Potential Targeting Strategy to Achieve High Specificity toward G-quadruplex

Previously, various approaches for the two-pronged targeting of DNA structures were proposed. These include the combined use of TFO and intercalator to target duplex,¹⁷² tetrad stacking ligand and groove-binder to target G-quadruplex,¹⁷³ and tetrad stacking ligand and antibody to target G-quadruplex.¹⁷⁴ In these cases, the target elements all arise from the same DNA conformation. On the other hand, the targets in our proposed strategy are duplex and quadruplex structures, two distinct structural conformations.

Figure 6-12. NMR spectra show binding of *QDHI* to netropsin and pyridostatin. 1D imino proton NMR spectrum of (a) free *QDHI*, (b) *QDHI* bound with equimolar ratio of netropsin, (c) *QDHI* bound with equimolar ratio of pyridostatin, and (d) *QDHI* bound with equimolar ratio of netropsin and pyridostatin (e) Chemical structure of quadruplex-binder pyridostatin.

Duplex groove-binding ligands have been shown to exhibit sequence selectivity with a good binding affinity. In contrast, quadruplex-binding ligands have been shown to exhibit exceptional binding affinity and high selectivity for quadruplex over duplex, but limited discrimination between different quadruplex structures. Our present targeting approach would combine the advantages of the two ligand types: sequence specificity of duplex binder and tight binding affinity of quadruplex binder. In certain aspects, we can relate this strategy with bi-specific targeting^{175, 176} and fragment-based drug discovery (FBDD),^{177, 178} which has received mounting interest towards the design of protein-binding ligands. In FBDD,

small fragments were screened for their weak binding to the target protein, and subsequently combined into larger lead molecules for further optimization of target affinity and potency. On the other hand, our proposal aims to combine the use of separate high-affinity quadruplex-binders and specific duplex-binders to target specific genomic sites might address the lack of specificity for G-quadruplex targets. To obtain an effective conjugated ligand, the binding affinity of the quadruplex binder and duplex binder have to be optimized. Quadruplex binder usually exhibit exceptional high binding affinity which can cause the non-specific binding to numerous G-quadruplex structures found in the human genome. To avoid that, duplex binder has to satisfy both sequence specificity and high binding affinity.

Figure 6-13. Examples of chemical linker that can be utilized to join a duplex-binding ligand (in pink) and a quadruplex-binding ligand (in yellow).

Furthermore, we propose that a linker could chemically join the two ligands into a single entity, which could provide synergistic binding of the two modules to the quadruplex-duplex hybrid target. The linker should exhibit a certain extent of

flexibility, and should take into consideration the structural context at the junction between the duplex and quadruplex segments.¹²⁴ Examples of such linkers that can be used are shown in Figure 6-13. Even though the notion of joining the two ligands by a chemical linker is straightforward, the synthesis of an actual construct targeting a genomic quadruplex-duplex hybrid is non-trivial; the design of the linker would be highly context-dependent (e.g. length and flexibility of the linker, as well as the attachment points onto both ligands) and might not apply between different quadruplex-duplex hybrid systems. On a separate note, the duplex-quadruplex junction can also serve as a unique recognition site by a junction binder,^{123, 124} which can be applied independently or in combination with the duplex- and quadruplex-binding ligands.

Figure 6-14. Examples of structural (a-c) and biological (d-g) contexts in which both duplex and quadruplex elements exist in close spatial proximity. (a) A quadruplex immediately up-/downstream of a duplex. (b) A quadruplex-duplex hybrid. (c) A quadruplex and a duplex formed by distinct strands. (d) Formation of quadruplexes at the telomere end. (e) Formation of a quadruplex during replication. (f) Formation of a quadruplex at the transcription start site. (g) Formation of a DNA-RNA quadruplex-duplex hybrid during transcription. The RNA transcript is shown in red.

A conjoined duplex- and quadruplex-targeting ligand would bind to duplex and quadruplex elements which are spatially close. This can arise in one of three arrangements: (i) a quadruplex immediately upstream/downstream of a duplex

(Figure 6-14a), (ii) a quadruplex which harbours within its loop a duplex element (Figure 6-14b), and (iii) a quadruplex and a duplex originating from two distinct strands (or that are sequentially far apart) (Figure 6-14c). In a biological context, these can happen under various scenarios. We have identified previously that diverse quadruplex-duplex hybrid-forming sequences can be found in the human genome, many of which overlap with regulatory important regions.¹²⁶ These sequences, and their RNA counterparts, would be forthright targets for the dual-targeting approach. On the other hand, simultaneous existence of duplex and quadruplex structures can also occur at the end of telomere¹⁸ (Figure 6-14d), during replication⁹⁴ (Figure 6-14e), recombination,¹⁷⁹ transcription^{28, 180} (Figure 6-14f,g), and splicing.¹⁸¹ In this manner, sequence selectivity provided by the duplex-binding ligand would help in discrimination between the myriad G-quadruplex structures,²⁶ thereby enabling the targeting of a unique genomic site.

6.3 Conclusion

In summary, we have proposed a targeting strategy based on the simultaneous application of duplex- and quadruplex-binding ligands. The simultaneous binding of netropsin and Phen-DC₃ to a quadruplex-duplex hybrid construct has been demonstrated, and the approach can be applied to different duplex and quadruplex binders. Hence the targeting approach provides a potential route for the specific targeting of unique quadruplex and quadruplex-duplex hybrid structures in the human genome.

Chapter 7 Summary and Future Perspectives

Summary

G-rich sequences have the propensity to form four-stranded structures known as G-quadruplexes which are comprised of stacking of two or more of Hoogsteen hydrogen bonded G-tetrads. G-quadruplex forming sequences have been identified in many important regions of the human genome including telomere, oncogenic promoter, 5'-UTR RNA, replication origin, and introns. Formation of G-quadruplexes in these regions was demonstrated to affect the biological processes such as telomere maintenance, replication, transcription, and translation. Hence, it is important to understand possible G-quadruplex structures adopted by G-rich sequences and develop ligands which selectively bind to target G-quadruplex. This thesis is set out to study new quadruplex-duplex hybrid structures and proposes a new strategy to selectively target quadruplex-duplex hybrid structures.

In this study, two new quadruplex-duplex hybrid structures including multiple stem-loop G-quadruplex and duplex-bulged G-quadruplex were presented. We have shown that multiple duplex stems can be incorporated into a single loop of G-quadruplex at different loop positions of a quadruplex construct, giving rise to a diverse range of possible structures. Thermal stability study of multi-stem complexes showed that the presence of a duplex stem in the loop has a positive influence on the stability of quadruplex as compared to a non-structured loop of the similar length. These multi-stem quadruplex constructs could offer new principles for assembly of DNA nanostructures in which G-quadruplex serves as a hub for incorporation of multiple duplex motifs. We also showed that a duplex stem can be incorporated into bulges of G-quadruplexes at various bulge positions on different G-quadruplex scaffolds. CD melting showed that the presence of duplex bulge has the positive effect on the thermal stability of a G-quadruplex, relative to a non-structured bulge of the same length. Our study showed that the size of the bulge could be extended to

33 nucleotides without destabilizing the overall structure. Two new quadruplex-duplex hybrids presented in this study and structural knowledge of these hybrids extend the current knowledge on the prediction of G-quadruplex forming sequences.

The 3'-end of the human telomere, comprising of tandem TTAGGG repeats, have the tendency to fold into G-quadruplex structures. Formation of G-quadruplexes at the 3'-end telomeric overhang could inhibit the activity of telomerase, which plays an important role in cell proliferation of cancer cells, rendering telomeric G-quadruplex as a potential anticancer target. To design effective telomeric G-quadruplex binding ligands, understanding the structures of G-quadruplex adopted by the telomeric sequence is necessary to design effective ligands for targeting telomeric G-quadruplexes. In this study, we examined the quadruplex structure could form immediately downstream of the double-stranded telomeric segment. A sequence which represents 80% of the transition between duplex-single stranded segments in human telomere were identified to adopt a quadruplex-duplex hybrid consisting a basket type, two-layer quadruplex and a duplex motif at the 5'-end.

Targeting G-quadruplex structures has been considered as a promising strategy for anticancer therapy. Although numerous G-quadruplex ligands have been developed, the selectivity between different G-quadruplex structures is still poor. In this study, we proposed a dual-specific targeting approach by simultaneous targeting quadruplex and duplex motif. This approach combines the high binding affinity of G-quadruplex ligands and high selectivity of duplex ligands. Our approach was proven via NMR spectroscopy by successful simultaneous binding of quadruplex-duplex hybrids by Phen-DC₃ and netropsin, which exhibit high binding affinity of G-quadruplex ligands and high selectivity of duplex ligands, respectively. CD spectroscopy showed that the binding of two ligands did not interfere with each other. By joining the two ligands into a conjugated ligands, a new class of ligands which exhibit enhanced binding specificity and/or affinity could be developed to target spatially close quadruplex and duplex elements which could occur at telomere, during replication, and transcription. Besides, this approach could be applied to

target quadruplex-duplex hybrid motifs which were identified across the human genome in a previous bioinformatics study.

Future Perspectives

Recent high-throughput sequencing revealed about 700,000 G-quadruplex structures in the human genome, of which G-quadruplexes with extra-long loops and bulges accounts for over half of the predicted structures. Structural knowledge of the two new quadruplex-duplex hybrids presented in this study could be incorporated into existing algorithms of G-quadruplex identification in bioinformatics studies, extending the current knowledge on the G-quadruplex forming sequence in the human genome, and identification of other quadruplex-duplex hybrid motifs.

Telomeric G-quadruplexes have been proposed as potential anti-cancer targets as their presence can inhibit telomerase, which is linked to cancer proliferation. Our study has shown that four telomeric repeats downstream of the telomeric double-stranded segment can adopt a quadruplex-duplex hybrid which can serve as a unique scaffold for telomeric G-quadruplex targeting. A detailed structural study will be performed on this hybrid, especially the relative orientation and the interface between duplex and quadruplex, to facilitate drug design.

A new targeting approach for improved selectivity in G-quadruplex targeting has been successfully proven through simultaneous targeting quadruplex-duplex hybrid with a high affinity G-quadruplex ligand and a high specificity duplex ligand. We proposed that a linker could be used to join the two ligands to form a conjugated ligands which would provide the synergistic binding of the two ligands to the quadruplex-duplex hybrids. This simultaneous targeting approach can be applied to target spatially close quadruplex and duplex elements, and to target quadruplex-duplex hybrid motifs which were identified across the human genome.¹²⁶ Numerous of quadruplex-duplex hybrid motifs were found in the regulatory regions of cancer-associated genes such as oncogene PIM1, transcription factor MYT1L, member of RAS oncogene family (RAB12 and RAD3D), proto-oncogene RET.¹²⁶ Future studies will also include the determination of high resolution structures of these

quadruplex-duplex hybrids, which will be used as a platform for the development of conjugated ligands targeting these hybrids. The successfully designed ligands can be applied to cell studies to investigate their cellular uptake ability and their effect on gene transcription and translation. The effective of the ligands on suppressing tumor growth can also be tested *in vivo*.

Appendices

Experimental protocols

DNA Sample Preparation. DNA oligonucleotides were chemically synthesized on an ABI 394 DNA/RNA synthesizer using phosphoramidites from Glen research. De-protection of the oligonucleotides were carried out by treating with aqueous ammonia. After that, the oligonucleotides were purified using polypak II cartridge following Glen research protocol. Purified DNA samples were successively dialyzed against deionized water, 20 mM KCl solution, and deionized water again. Samples were then frozen and dried in lyophilizer. The oligonucleotide in chapter 3, 4, 5 and 6 were dissolved in buffer containing 20 mM KCl and 20 mM KPi (pH 7.0), 30mM KCl and 20 mM KPi (pH 7.0), 70 mM KCl and 20 mM KPi (pH 7.0), 20mM KCl and 20 mM KPi (pH 7.0) respectively. The DNA concentration was quantified by UV absorbance at 260 nm at 95°C to unfold any secondary structure.

NMR Spectroscopy. DNA strand concentration of NMR samples was typically 0.2–2.0mM. Unless otherwise stated, all NMR spectra were recorded at 25 °C on Bruker AVANCE 600- or 700-MHz spectrometers. All the chemical shifts were calibrated again 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) reference. Spectra analyses were processed by using the program Spinworks and FELIX (Felix NMR, Inc.).

For NMR experiments in H₂O solvent, 10% D₂O was added to buffer and the jump-and-return (JR) pulse sequence was utilized to suppress water signal.¹⁸² Depending on the DNA concentration, the number of scan was set at 32 to 512. More scan was set for sample of lower DNA concentration. Delay time between two scans was 1.1 s. Guanine H1 and H8 protons were assigned by using 2%-4% site-specific low-enrichment ¹⁵N labelling. Spectral assignments were completed by through-bond correlation experiment such as COSY, TOCSY, ¹³C–¹H-HSQC and through space correlation experiment such as NOESY.

Circular Dichroism. Circular dichroism (CD) experiments were performed on a JASCO-815 spectropolarimeter with a reaction volume of 500 µL. CD spectra were

performed over the range of 220–320 nm at 25 °C using a 1 cm path length quartz cuvette. The bandwidth, scan speed and data pitch were set at 2 nm, 200 nm/min and 1nm, respectively. The DNA strand concentration was typically 3-5 μ M. For each measurement, three scans were taken and averaged, the spectral contribution from the buffer was subtracted, and the data were zero-corrected at 320 nm. For CD melting experiments, the CD signal was recorded across the temperature range of 15–95 °C. The melting curves were fitted to two-state model where two baselines corresponding to the folded and unfolded states were drawn at low and high temperature, respectively. Melting temperature is the temperature at which the fraction of folded and unfolded state are equal.

Constraints for Structure Calculation.

NOE distance constraints: Interproton distance constraints obtained from NOESY experiment in H₂O (200 ms mixing time) were classified as very strong (2.50 ± 0.70 Å), strong (3.20 ± 0.80 Å), medium (4.20 ± 1.20 Å), weak (5.00 ± 1.40 Å) and very weak (6.00 ± 1.80 Å). Interproton distance constraints obtained from NOESY experiment in D₂O (100 ms, 200 ms, and 300 ms mixing time) were classified as very strong (2.40 ± 0.60 Å), strong (3.00 ± 0.75 Å), medium (3.80 ± 0.95 Å), weak (4.80 ± 1.20 Å) and very weak (6.00 ± 1.50 Å).

Hydrogen bond distance constraints:

- G-tetrad: Hydrogen bond distance constraints within a G-tetrad were assigned to the distance between H21-N7, N2-N7, H1-O6, and N1-O6 of the two neighboring guanines. The distance constraints for these hydrogen bonds were (2.00 ± 0.20 Å), (2.90 ± 0.30 Å), (2.00 ± 0.20 Å), and (2.90 ± 0.30 Å), respectively.
- A-T base pairs: Hydrogen bond distance constraints for A-T base pairs included A(N1)-T(H3) (2.00 ± 0.20 Å), A(N1)-T(N3) (2.90 ± 0.30 Å), A(H61)-T(O4) (2.00 ± 0.20 Å), and A(N6)-T(O4) (2.90 ± 0.30 Å).
- G-C base pairs: Hydrogen bond distance constraints for G-C base pairs included G(H21)-C(O2) (2.00 ± 0.20 Å), G(N2)-C(O2) (2.90 ± 0.30 Å), G(N1)-C(H3) (2.00 ± 0.20 Å), G(N1)-C(N3) (2.90 ± 0.30 Å), G(O3)-C(H41) (2.00 ± 0.20 Å), G(O6)-C(N4) (2.90 ± 0.30 Å).

Planarity Constraints: Planarity constraints were assigned to each G-tetrad and each Watson-crick base pair.

Dihedral constraints: χ torsion angle constraints for *syn* and *anti* were $60\pm 70^\circ$ and $240\pm 40^\circ$, respectively.

Structure Calculation. Structures of duplex–quadruplex constructs in this study were calculated using the program XPLOR-NIH.¹²⁸ Inter-proton distance constraints, hydrogen-bond constraints, planarity constraints, dihedral constraints, and repulsive constraints were imposed during structure calculations. Inter-proton distance restraints, deduced from NOESY experiments at various mixing times (100 ms, 200 ms and 300 ms), were classified as very strong, strong, medium, weak and very weak. Initially, an extended conformation of the sequence was generated randomly. Structure embedding was performed to produce a family of 100 DG structures. Next, the 100 DG structures were subjected to simulated annealing regularization and refinement. The distance constraints of exchangeable and nonexchangeable protons, G-tetrad hydrogen bonding, Watson-Crick hydrogen bond and negative constraints were incorporated in this step of calculation. 100 structures obtained from DGSA calculation were subjected to NOE-restrained Simulated Annealing refinement in XPLOR. The system is heat up from 300 K to 1000 K in 14 ps, and allowed to reach equilibrium for 6 ps, wherein the force constant were maintained at 2 Kcal.mol⁻¹Å⁻² for exchangeable and non-exchangeable proton distant restraints, and 8 Kcal.mol⁻¹Å⁻² for repulsive distance restraints. Next, the force constants were then scale to 32, 16, 8 and 0 Kcal.mol⁻¹Å⁻² for hydrogen bond, non-exchangeable proton, exchangeable proton and repulsive distance restraints, respectively. Subsequently, the system was allowed to equilibrate at 1000 K for 200 ps then gradually cooled down to 300 K in 42 s. After that, the system was subjected to NOE-restraint molecular dynamic calculation for 18 ps. During the last 10 ps of NOE-restraint molecular dynamic calculation, the coordinates of the systems were saved every 0.5 ps and averaged. The resulting averaged structures were submitted to Powel minimization until the energy gradient is smaller than 0.1 Kcal/mol was reached. Planarity (1 Kcal.mol⁻¹Å⁻² for tetrads; 0.5 Kcal.mol⁻¹Å⁻² for base pairs) and

dihedral ($50 \text{ Kcal.mol}^{-1}\text{\AA}^{-2}$) constraints were maintained during the refinement process. Out of 100 conformers, 10 best structures with the lowest overall energy were selected. Structures were displayed using the program PyMOL.

Ligand preparation. Netropsin, BRACO-19, and pyridostatin were purchased from Sigma Aldrich. Phen-DC₃ was a gracious gift from Marie-Paule Teulade-Fichou. Lyophilized Phen-DC₃ was dissolved in dimethyl sulfoxide. All other ligands in lyophilized form were dissolved in water.

Construction of structural model for *QDHI*:netropsin:Phen-DC₃ complex. The model of a quadruplex was built with the XPLOR-NIH program¹²⁸ using constraints adapted from reported quadruplex-duplex hybrid structures.³³ The quadruplex model was combined with a duplex generated by LEaP¹⁸³ in the PYMOL program to obtain the structural model of *QDHI*. Netropsin and Phen-DC₃ were then aligned onto *QDHI* in PyMOL based on reported duplex:netropsin (PDB code: 2LWH) and quadruplex:Phen-DC₃ (PDB code: 2MGN) structures.

Bibliography

1. Watson, J. D., and Crick, F. H. C. (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid, *Nature* 171, 737–738.
2. Wilkins, M. H. F., Stokes, A. R., and Wilson, H. R. (1953) Molecular structure of deoxypentose nucleic acids, *Nature* 171, 738–740.
3. Franklin, R. E., and Gosling, R. G. (1953) Molecular configuration in sodium thymonucleate, *Nature* 171, 740–741.
4. Verdaguer, N., Aymami, J., Fernandezforner, D., Fita, I., Coll, M., Huynhdinh, T., Igolen, J., and Subirana, J. A. (1991) Molecular structure of a complete turn of A-DNA, *J. Mol. Biol.* 221, 623–635.
5. Arnott, S., Hukins, D. W. L., and Dover, S. D. (1972) Optimized parameters for RNA double-helices, *Biochem. Biophys. Res. Commun.* 48, 1392–1399.
6. Wang, A. H. J., Fujii, S., Vanboom, J. H., Vandermarel, G. A., Vanboeckel, S. A. A., and Rich, A. (1982) Molecular structure of r(GCG)d(TATACGC): A DNA–RNA hybrid helix joined to double helical DNA, *Nature* 299, 601–604.
7. Wang, A. H. J., Quigley, G. J., Kolpak, F. J., Vandermarel, G., Vanboom, J. H., and Rich, A. (1981) Left-handed double helical DNA: Variations in the backbone conformation, *Science* 211, 171–176.
8. Rich, A., Nordheim, A., and Wang, A. H. J. (1984) The chemistry and biology of left-handed Z-DNA, *Annu. Rev. Biochem.* 53, 791–846.
9. Simonsson, T. (2001) G-quadruplex DNA structures – Variations on a theme, *Biol. Chem.* 382, 621–628.
10. Davis, J. T. (2004) G-quartets 40 years later: From 5'-GMP to molecular biology and supramolecular chemistry, *Angew. Chem. Int. Ed.* 43, 668–698.
11. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006) Quadruplex DNA: Sequence, topology and structure, *Nucleic Acids Res.* 34, 5402–5415.
12. Patel, D. J., Phan, A. T., and Kuryavyi, V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: Diverse higher order DNA and RNA targets for cancer therapeutics, *Nucleic Acids Res.* 35, 7429–7455.
13. Gellert, M., Lipsett, M. N., and Davies, D. R. (1962) Helix formation by guanylic acid, *Proc. Natl. Acad. Sci. USA* 48, 2013–2018.
14. Bhattacharyya, D., Mirihana Arachchilage, G., and Basu, S. (2016) Metal Cations in G-Quadruplex Folding and Stability, *Front. Chem.* 4, 38.
15. Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004) Loop-length-dependent folding of G-quadruplexes, *J. Am. Chem. Soc.* 126, 16405–16415.

16. Guédin, A., De Cian, A., Gros, J., Lacroix, L., and Mergny, J. L. (2008) Sequence effects in single-base loops for quadruplexes, *Biochimie* 90, 686–696.
17. Guédin, A., Gros, J., Alberti, P., and Mergny, J. L. (2010) How long is too long? Effects of loop size on G-quadruplex stability, *Nucleic Acids Res.* 38, 7858–7868.
18. Phan, A. T. (2010) Human telomeric G-quadruplex: Structures of DNA and RNA sequences, *FEBS J.* 277, 1107–1117.
19. Agrawal, P., Hatzakis, E., Guo, K., Carver, M., and Yang, D. (2013) Solution structure of the major G-quadruplex formed in the human VEGF promoter in K⁺: insights into loop interactions of the parallel G-quadruplexes, *Nucleic Acids Res.* 41, 10584-10592.
20. Dai, J. X., Chen, D., Jones, R. A., Hurley, L. H., and Yang, D. Z. (2006) NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region, *Nucleic Acids Res.* 34, 5133–5144.
21. Phan, A. T., Kuryavyi, V., Gaw, H. Y., and Patel, D. J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter, *Nat. Chem. Biol.* 1, 167–173.
22. Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S., and Patel, D. J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter, *J. Am. Chem. Soc.* 129, 4386–4392.
23. Mukundan, V. T., and Phan, A. T. (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences, *J. Am. Chem. Soc.* 135, 5017-5028.
24. Chung, W. J., Heddi, B., Schmitt, E., Lim, K. W., Mechulam, Y., and Phan, A. T. (2015) Structure of a left-handed DNA G-quadruplex, *Proc. Natl. Acad. Sci. USA* 112, 2729-2733.
25. Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome, *Nucleic Acids Res.* 33, 2908–2916.
26. Chambers, V. S., Marsico, G., Boutell, J. M., Di Antonio, M., Smith, G. P., and Balasubramanian, S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome, *Nat. Biotechnol.* 33, 877-881.
27. Allshire, R. C., Dempster, M., and Hastie, N. D. (1989) Human telomeres contain at least three types of G-rich repeat distributed non-randomly, *Nucleic Acids Res.* 17, 4611–4627.
28. Huppert, J. L., and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome, *Nucleic Acids Res.* 35, 406–413.
29. Eddy, J., and Maizels, N. (2008) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes, *Nucleic Acids Res.* 36, 1321–1333.
30. Hansel-Hertsch, R., Beraldi, D., Lensing, S. V., Marsico, G., Zyner, K., Parry, A., Di Antonio, M., Pike, J., Kimura, H., Narita, M., Tannahill,

- D., and Balasubramanian, S. (2016) G-quadruplex structures mark human regulatory chromatin, *Nat Genet* 48, 1267-1272.
31. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J., and Plückthun, A. (2001) *In vitro* generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei, *Proc. Natl. Acad. Sci. USA* 98, 8572–8577.
32. Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells, *Nat. Chem.* 5, 182–186.
33. Henderson, A., Wu, Y., Huang, Y. C., Chavez, E. A., Platt, J., Johnson, F. B., Brosh, R. M., Jr., Sen, D., and Lansdorp, P. M. (2014) Detection of G-quadruplex DNA in mammalian cells, *Nucleic Acids Res.* 42, 860-869.
34. Zakian, V. A. (1989) Structure and function of telomeres, *Annu. Rev. Genet.* 23, 579–604.
35. Blackburn, E. H. (1991) Structure and function of telomeres, *Nature* 350, 569–573.
36. Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L., and Wu, J. R. (1988) A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes, *Proc. Natl. Acad. Sci. USA* 85, 6622–6626.
37. de Lange, T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres, *Genes Dev.* 19, 2100–2110.
38. De Cian, A., Lacroix, L., Douarre, C., Temime-Smaali, N., Trentesaux, C., Riou, J. F., and Mergny, J. L. (2008) Targeting telomeres and telomerase, *Biochimie* 90, 131–155.
39. Hug, N., and Lingner, J. (2006) Telomere length homeostasis, *Chromosoma* 115, 413-425.
40. Harley, C. B., Futcher, A. B., and Greider, C. W. (1990) Telomeres shorten during aging of human fibroblasts, *Nature* 345, 458–460.
41. Neidle, S., and Parkinson, G. (2002) Telomere maintenance as a target for anticancer drug discovery, *Nat. Rev. Drug Discov.* 1, 383–393.
42. Greider, C. W. (1996) Telomere length regulation, *Annu. Rev. Biochem.* 65, 337–365.
43. Hemann, M. T., Strong, M. A., Hao, L. Y., and Greider, C. W. (2001) The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability, *Cell* 107, 67-77.
44. Greider, C. W., and Blackburn, E. H. (1985) Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts, *Cell* 43, 405–413.
45. Kim, N. W., Piatyszek, M. A., Prowse, K. R., Harley, C. B., West, M. D., Ho, P. L. C., Coviello, G. M., Wright, W. E., Weinrich, S. L., and Shay, J. W. (1994) Specific association of human telomerase activity with immortal cells and cancer, *Science* 266, 2011–2015.

46. Zahler, A. M., Williamson, J. R., Cech, T. R., and Prescott, D. M. (1991) Inhibition of telomerase by G-quartet DNA structures, *Nature* 350, 718–720.
47. De Cian, A., Cristofari, G., Reichenbach, P., De Lemos, E., Monchaud, D., Teulade-Fichou, M. P., Shin-Ya, K., Lacroix, L., Lingner, J., and Mergny, J. L. (2007) Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action, *Proc. Natl. Acad. Sci. USA* 104, 17347–17352.
48. Monchaud, D., and Teulade-Fichou, M. P. (2008) A hitchhiker's guide to G-quadruplex ligands, *Org. Biomol. Chem.* 6, 627–636.
49. Granotier, C., Pennarun, G., Riou, L., Hoffschir, F., Gauthier, L. R., De Cian, A., Gomez, D., Mandine, E., Riou, J. F., Mergny, J. L., Mailliet, P., Dutrillaux, B., and Boussin, F. D. (2005) Preferential binding of a G-quadruplex ligand to human chromosome ends, *Nucleic Acids Res.* 33, 4182–4190.
50. Burger, A. M., Dai, F., Schultes, C. M., Reszka, A. P., Moore, M. J., Double, J. A., and Neidle, S. (2005) The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function, *Cancer Res.* 65, 1489–1496.
51. Phatak, P., Cookson, J. C., Dai, F., Smith, V., Gartenhaus, R. B., Stevens, M. F., and Burger, A. M. (2007) Telomere uncapping by the G-quadruplex ligand RHPS4 inhibits clonogenic tumour cell growth in vitro and in vivo consistent with a cancer stem cell targeting mechanism, *Br J Cancer* 96, 1223–1233.
52. Tauchi, T., Shin-ya, K., Sashida, G., Sumi, M., Okabe, S., Ohyashiki, J. H., and Ohyashiki, K. (2006) Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: in vitro and in vivo studies in acute leukemia, *Oncogene* 25, 5719–5725.
53. Rodriguez, R., Muller, S., Yeoman, J. A., Trentesaux, C., Riou, J. F., and Balasubramanian, S. (2008) A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres, *J. Am. Chem. Soc.* 130, 15758–15759.
54. Wang, Y., and Patel, D. J. (1993) Solution structure of the human telomeric repeat d[AG₃(T₂AG₃)₃] G-tetraplex, *Structure* 1, 263–282.
55. Parkinson, G. N., Lee, M. P. H., and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA, *Nature* 417, 876–880.
56. Phan, A. T., Kuryavyi, V., Luu, K. N., and Patel, D. J. (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K⁺ solution, *Nucleic Acids Res.* 35, 6517–6525.
57. Lim, K. W., Amrane, S., Bouaziz, S., Xu, W. X., Mu, Y. G., Patel, D. J., Luu, K. N., and Phan, A. T. (2009) Structure of the human telomere in K⁺ solution: A stable basket-type G-quadruplex with only two G-tetrad layers, *J. Am. Chem. Soc.* 131, 4301–4309.

58. Lim, K. W., Ng, V. C., Martin-Pintado, N., Heddi, B., and Phan, A. T. (2013) Structure of the human telomere in Na⁺ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity, *Nucleic Acids Res.* **41**, 10556-10562.
59. Verma, A., Halder, K., Halder, R., Yadav, V. K., Rawal, P., Thakur, R. K., Mohd, F., Sharma, A., and Chowdhury, S. (2008) Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species, *J. Med. Chem.* **51**, 5641-5649.
60. Hershman, S. G., Chen, Q., Lee, J. Y., Kozak, M. L., Yue, P., Wang, L. S., and Johnson, F. B. (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*, *Nucleic Acids Res.* **36**, 144-156.
61. Rawal, P., Kummarasetti, V. B. R., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S. K., and Chowdhury, S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: Role in *Escherichia coli* global regulation, *Genome Res.* **16**, 644-655.
62. Eddy, J., and Maizels, N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome, *Nucleic Acids Res.* **34**, 3887-3896.
63. Balasubramanian, S., and Neidle, S. (2009) G-quadruplex nucleic acids as therapeutic targets, *Curr. Opin. Chem. Biol.* **13**, 345-353.
64. Balasubramanian, S., Hurley, L. H., and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: A novel anticancer strategy?, *Nat. Rev. Drug Discov.* **10**, 261-275.
65. Fleming, A. M., Ding, Y., and Burrows, C. J. (2017) Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair, *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2604-2609.
66. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription, *Proc. Natl. Acad. Sci. USA* **99**, 11593-11598.
67. Ambrus, A., Chen, D., Dai, J. X., Jones, R. A., and Yang, D. Z. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization, *Biochemistry* **44**, 2048-2058.
68. Phan, A. T., Modi, Y. S., and Patel, D. J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter, *J. Am. Chem. Soc.* **126**, 8710-8716.
69. Kang, H. J., and Park, H. J. (2009) Novel molecular mechanism for actinomycin D activity as an oncogenic promoter G-quadruplex binder, *Biochemistry* **48**, 7392-7398.
70. Ma, Y., Ou, T. M., Tan, J. H., Hou, J. Q., Huang, S. L., Gu, L. Q., and Huang, Z. S. (2011) Quinolino-benzo-[5, 6]-dihydroisoquinolium compounds derived from berberine: a new class of highly selective

- ligands for G-quadruplex DNA in c-myc oncogene, *Eur. J. Med. Chem.* **46**, 1906-1913.
71. Agarwal, T., Roy, S., Chakraborty, T. K., and Maiti, S. (2010) Furan based cyclic homo-oligopeptides bind G-quadruplex selectively and repress c-MYC transcription, *Bioorg. Med. Chem. Lett.* **20**, 4346-4349.
 72. Shalaby, T., von Bueren, A. O., Hurlimann, M. L., Fiaschetti, G., Castelletti, D., Masayuki, T., Nagasawa, K., Arcaro, A., Jelesarov, I., Shin-ya, K., and Grotzer, M. (2010) Disabling c-Myc in childhood medulloblastoma and atypical teratoid/rhabdoid tumor cells by the potent G-quadruplex interactive agent S2T1-6OTD, *Mol. Cancer Ther.* **9**, 167-179.
 73. Kuryavyi, V., Phan, A. T., and Patel, D. J. (2010) Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human *c-kit2* promoter, *Nucleic Acids Res.* **38**, 6757–6773.
 74. Hsu, S. T. D., Varnai, P., Bugaut, A., Reszka, A. P., Neidle, S., and Balasubramanian, S. (2009) A G-Rich sequence within the *c-kit* oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics, *J. Am. Chem. Soc.* **131**, 13399–13409.
 75. Lim, K. W., Lacroix, L., Yue, D. J. E., Lim, J. K. C., Lim, J. M. W., and Phan, A. T. (2010) Coexistence of two distinct G-quadruplex conformations in the hTERT promoter, *J. Am. Chem. Soc.* **132**, 12331–12342.
 76. Rankin, S., Reszka, A. P., Huppert, J., Zloh, M., Parkinson, G. N., Todd, A. K., Ladame, S., Balasubramanian, S., and Neidle, S. (2005) Putative DNA quadruplex formation within the human *c-kit* oncogene, *J. Am. Chem. Soc.* **127**, 10584–10589.
 77. Dexheimer, T. S., Sun, D., and Hurley, L. H. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter, *J. Am. Chem. Soc.* **128**, 5404-5415.
 78. Sun, D. Y., Guo, K. X., Rusche, J. J., and Hurley, L. H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents, *Nucleic Acids Res.* **33**, 6070–6080.
 79. De Armond, R., Wood, S., Sun, D. Y., Hurley, L. H., and Ebbinghaus, S. W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1 α promoter, *Biochemistry* **44**, 16341–16350.
 80. Palumbo, S. L., Memmott, R. M., Uribe, D. J., Krotova-Khan, Y., Hurley, L. H., and Ebbinghaus, S. W. (2008) A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity, *Nucleic Acids Res.* **36**, 1755-1769.

81. Cogoi, S., and Xodo, L. E. (2006) G-quadruplex formation within the promoter of the *KRAS* proto-oncogene and its effect on transcription, *Nucleic Acids Res.* **34**, 2536–2549.
82. Palumbo, S. L., Ebbinghaus, S. W., and Hurley, L. H. (2009) Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands, *J. Am. Chem. Soc.* **131**, 10878–10891.
83. Kumari, S., Bugaut, A., Huppert, J. L., and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation, *Nat. Chem. Biol.* **3**, 218–221.
84. Bugaut, A., and Balasubramanian, S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting, *Nucleic Acids Res.* **40**, 4727-4741.
85. Schaeffer, C., Bardoni, B., Mandel, J. L., Ehresmann, B., Ehresmann, C., and Moine, H. (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif, *EMBO J.* **20**, 4803-4813.
86. Arora, A., Dutkiewicz, M., Scaria, V., Hariharan, M., Maiti, S., and Kurreck, J. (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif, *RNA* **14**, 1290-1296.
87. Gomez, D., Guedin, A., Mergny, J. L., Salles, B., Riou, J. F., Teulade-Fichou, M. P., and Calsou, P. (2010) A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells, *Nucleic Acids Res.* **38**, 7187-7198.
88. Morris, M. J., and Basu, S. (2009) An unusually stable G-quadruplex within the 5'-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells, *Biochemistry* **48**, 5313-5319.
89. Lammich, S., Kamp, F., Wagner, J., Nuscher, B., Zilow, S., Ludwig, A. K., Willem, M., and Haass, C. (2011) Translational repression of the disintegrin and metalloprotease ADAM10 by a stable G-quadruplex secondary structure in its 5'-untranslated region, *J. Biol. Chem.* **286**, 45063-45072.
90. Bonnal, S., Schaeffer, C., Creancier, L., Clamens, S., Moine, H., Prats, A. C., and Vagner, S. (2003) A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons, *J. Biol. Chem.* **278**, 39330-39336.
91. Morris, M. J., Negishi, Y., Pazsint, C., Schonhoft, J. D., and Basu, S. (2010) An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES, *J. Am. Chem. Soc.* **132**, 17831-17839.
92. Halder, K., Largy, E., Benzler, M., Teulade-Fichou, M. P., and Hartig, J. S. (2011) Efficient suppression of gene expression by targeting 5'-UTR-based RNA quadruplexes with bisquinolinium compounds, *ChemBioChem* **12**, 1663-1668.

93. Katsuda, Y., Sato, S., Asano, L., Morimura, Y., Furuta, T., Sugiyama, H., Hagihara, M., and Uesugi, M. (2016) A Small Molecule That Represses Translation of G-Quadruplex-Containing mRNA, *J. Am. Chem. Soc.* 138, 9037-9040.
94. Rhodes, D., and Lipps, H. J. (2015) G-quadruplexes and their regulatory roles in biology, *Nucleic Acids Res.* 43, 8627-8637.
95. Cheung, I., Schertzer, M., Rose, A., and Lansdorp, P. M. (2002) Disruption of dog-1 in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA, *Nat. Genet.* 31, 405-409.
96. Bharti, S. K., Sommers, J. A., George, F., Kuper, J., Hamon, F., Shin-ya, K., Teulade-Fichou, M. P., Kisker, C., and Brosh, R. M., Jr. (2013) Specialization among iron-sulfur cluster helicases to resolve G-quadruplex DNA structures that threaten genomic stability, *J. Biol. Chem.* 288, 28217-28229.
97. Rossi, M. L., Ghosh, A. K., and Bohr, V. A. (2010) Roles of Werner syndrome protein in protection of genome integrity, *DNA Repair (Amst)* 9, 331-344.
98. Sun, H., Karow, J. K., Hickson, I. D., and Maizels, N. (1998) The Bloom's syndrome helicase unwinds G4 DNA, *J. Biol. Chem.* 273, 27587-27592.
99. Goto, M., Ishikawa, Y., Sugimoto, M., and Furuichi, Y. (2013) Werner syndrome: a changing pattern of clinical manifestations in Japan (1917~2008), *Bioscience trends* 7, 13-22.
100. German, J., Sanz, M. M., Ciocci, S., Ye, T. Z., and Ellis, N. A. (2007) Syndrome-causing mutations of the BLM gene in persons in the Bloom's Syndrome Registry, *Hum. Mutat.* 28, 743-753.
101. Wu, Y., Shin-ya, K., and Brosh, R. M., Jr. (2008) FANCDJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability, *Mol Cell Biol* 28, 4116-4128.
102. Paeschke, K., Bochman, M. L., Garcia, P. D., Cejka, P., Friedman, K. L., Kowalczykowski, S. C., and Zakian, V. A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs, *Nature* 497, 458-462.
103. Ribeyre, C., Lopes, J., Boule, J. B., Piazza, A., Guedin, A., Zakian, V. A., Mergny, J. L., and Nicolas, A. (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo, *PLoS Genet.* 5, e1000475.
104. Piazza, A., Boule, J. B., Lopes, J., Mingo, K., Largy, E., Teulade-Fichou, M. P., and Nicolas, A. (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*, *Nucleic Acids Res.* 38, 4337-4348.
105. Piazza, A., Adrian, M., Samazan, F., Heddi, B., Hamon, F., Serero, A., Lopes, J., Teulade-Fichou, M. P., Phan, A. T., and Nicolas, A. (2015) Short loop length and high thermal stability determine genomic

- instability induced by G-quadruplex-forming minisatellites, *EMBO J.* **34**, 1718-1734.
106. Sarkies, P., Reams, C., Simpson, L. J., and Sale, J. E. (2010) Epigenetic instability due to defective replication of structured DNA, *Mol. Cell.* **40**, 703-713.
 107. Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J. M., and Lemaitre, J. M. (2012) Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs, *Nat. Struct. Mol. Biol.* **19**, 837-844.
 108. Valton, A. L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintomé, C., Riou, J. F., and Prioleau, M. N. (2014) G4 motifs affect origin positioning and efficiency in two vertebrate replicators, *EMBO J.* **33**, 732–746.
 109. Hoshina, S., Yura, K., Teranishi, H., Kiyasu, N., Tominaga, A., Kadoma, H., Nakatsuka, A., Kunichika, T., Obuse, C., and Waga, S. (2013) Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA, *J. Biol. Chem.* **288**, 30161-30171.
 110. Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M. P., and Phan, A. T. (2014) Solution structure of a G-quadruplex bound to the bisquinolinium compound Phen-DC(3), *Angew. Chem. Int. Ed.* **53**, 999-1002.
 111. Dai, J., Carver, M., Hurley, L. H., and Yang, D. (2011) Solution structure of a 2:1 quindoline-c-MYC G-quadruplex: insights into G-quadruplex-interactive small molecule drug design, *J. Am. Chem. Soc.* **133**, 17673-17680.
 112. Haider, S. M., Parkinson, G. N., and Neidle, S. (2003) Structure of a G-quadruplex-ligand complex, *J. Mol. Biol.* **326**, 117-125.
 113. Chung, W. J., Heddi, B., Tera, M., Iida, K., Nagasawa, K., and Phan, A. T. (2013) Solution structure of an intramolecular (3 + 1) human telomeric G-quadruplex bound to a telomestatin derivative, *J. Am. Chem. Soc.* **135**, 13495-13501.
 114. Campbell, N. H., Patel, M., Tofa, A. B., Ghosh, R., Parkinson, G. N., and Neidle, S. (2009) Selectivity in ligand recognition of G-quadruplex loops, *Biochemistry* **48**, 1675-1680.
 115. Alzeer, J., Vummidi, B. R., Roth, P. J., and Luedtke, N. W. (2009) Guanidinium-modified phthalocyanines as high-affinity G-quadruplex fluorescent probes and transcriptional regulators, *Angew. Chem. Int. Ed.* **48**, 9362-9365.
 116. Wang, J., Lu, K., Xuan, S., Toh, Z., Zhang, D., and Shao, F. (2013) A Pt(II)-Dip complex stabilizes parallel c-myc G-quadruplex, *Chem. Commun.* **49**, 4758-4760.
 117. Read, M., Harrison, R. J., Romagnoli, B., Tanious, F. A., Gowan, S. H., Reszka, A. P., Wilson, W. D., Kelland, L. R., and Neidle, S. (2001) Structure-based design of selective and potent G quadruplex-

- mediated telomerase inhibitors, *Proc. Natl. Acad. Sci. USA* 98, 4844–4849.
118. Shin-ya, K., Wierzba, K., Matsuo, K., Ohtani, T., Yamada, Y., Furihata, K., Hayakawa, Y., and Seto, H. (2001) Telomestatin, a novel telomerase inhibitor from *Streptomyces anulatus*, *J. Am. Chem. Soc.* 123, 1262–1263.
 119. Ge, B. X., Huang, Y. C., Sen, D., and Yu, H. Z. (2010) A robust electronic switch made of immobilized duplex/quadruplex DNA, *Angew. Chem. Int. Ed.* 49, 9965–9967.
 120. Dutta, K., Fujimoto, T., Inoue, M., Miyoshi, D., and Sugimoto, N. (2010) Development of new functional nanostructures consisting of both DNA duplex and quadruplex, *Chem Commun* 46, 7772–7774.
 121. Jeng, S. C., Chan, H. H., Booy, E. P., McKenna, S. A., and Unrau, P. J. (2016) Fluorophore ligand binding and complex stabilization of the RNA Mango and RNA Spinach aptamers, *RNA* 22, 1884–1892.
 122. Trachman, R. J., 3rd, Demeshkina, N. A., Lau, M. W. L., Panchapakesan, S. S. S., Jeng, S. C. Y., Unrau, P. J., and Ferre-D'Amare, A. R. (2017) Structural basis for high-affinity fluorophore binding and activation by RNA Mango, *Nat. Chem. Biol.* 13, 807–813.
 123. Russo Krauss, I., Ramaswamy, S., Neidle, S., Haider, S., and Parkinson, G. N. (2016) Structural Insights into the Quadruplex-Duplex 3' Interface Formed from a Telomeric Repeat: A Potential Molecular Target, *J. Am. Chem. Soc.* 138, 1226–1233.
 124. Lim, K. W., and Phan, A. T. (2013) Structural basis of DNA quadruplex–duplex junction formation, *Angew. Chem. Int. Ed.* 52, 8566–8569.
 125. Lim, K. W., Khong, Z. J., and Phan, A. T. (2014) Thermal stability of DNA quadruplex–duplex hybrids, *Biochemistry* 53, 247–257.
 126. Lim, K. W., Jenjaroenpun, P., Low, Z. J., Khong, Z. J., Ng, Y. S., Kuznetsov, V. A., and Phan, A. T. (2015) Duplex stem-loop-containing quadruplex motifs in the human genome: a combined genomic and structural study, *Nucleic Acids Res.* 43, 5630–5646.
 127. Adrian, M., Heddi, B., and Phan, A. T. (2012) NMR spectroscopy of G-quadruplexes, *Methods* 57, 11–24.
 128. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package, *J. Magn. Reson.* 160, 65–73.
 129. Kypr, J., Kejnovska, I., Renciuk, D., and Vorlickova, M. (2009) Circular dichroism and conformational polymorphism of DNA, *Nucleic Acids Res.* 37, 1713–1725.
 130. Vorlickova, M., Kejnovska, I., Sagi, J., Renciuk, D., Bednarova, K., Motlova, J., and Kypr, J. (2012) Circular dichroism and guanine quadruplexes, *Methods* 57, 64–75.
 131. Gray, D. M., Wen, J. D., Gray, C. W., Repges, R., Repges, C., Raabe, G., and Fleischhauer, J. (2008) Measured and calculated CD spectra

- of G-quartets stacked with the same or opposite polarities, *Chirality* 20, 431–440.
132. Randazzo, A., Spada, G. P., and da Silva, M. W. (2013) Circular Dichroism of Quadruplex Structures, In *Quadruplex Nucleic Acids* (Chaires, J. B., and Graves, D., Eds.), pp 67-86, Springer Berlin Heidelberg, Berlin, Heidelberg.
 133. Lilley, D. M. J. (2000) Structures of helical junctions in nucleic acids, *Q. Rev. Biophys.* 33, 109–159.
 134. Holliday, R. (1964) A mechanism for gene conversion in fungi, *Genet. Res.* 5, 282–304.
 135. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983) The double-strand-break repair model for recombination, *Cell* 33, 25–35.
 136. Jaeger, L., and Chworos, A. (2006) The architectonics of programmable RNA and DNA nanostructures, *Curr. Opin. Struct. Biol.* 16, 531–543.
 137. Ha, S. C., Lowenhaupt, K., Rich, A., Kim, Y. G., and Kim, K. K. (2005) Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases, *Nature* 437, 1183–1186.
 138. Seeman, N. C. (2010) Nanomaterials Based on DNA, *Annu. Rev. Biochem.* 79, 65–87.
 139. Yatsunyk, L. A., Mendoza, O., and Mergny, J. L. (2014) "Nano-oddities": unusual nucleic acid assemblies for DNA-based nanostructures and nanodevices, *Acc Chem Res* 47, 1836-1844.
 140. Lech, C. J., and Phan, A. T. (2017) Ball with hair: modular functionalization of highly stable G-quadruplex DNA nano-scaffolds through N2-guanine modification, *Nucleic Acids Res.* 45, 6265-6274.
 141. Yu, Z. B., Gaerig, V., Cui, Y. X., Kang, H. J., Gokhale, V., Zhao, Y., Hurley, L. H., and Mao, H. B. (2012) Tertiary DNA structure in the single-stranded hTERT promoter fragment unfolds and refolds by parallel pathways via cooperative or sequential events, *J. Am. Chem. Soc.* 134, 5157–5164.
 142. Yatsunyk, L. A., Mendoza, O., and Mergny, J. L. (2014) "Nano-oddities": Unusual nucleic acid assemblies for DNA-based nanostructures and nanodevices, *Acc. Chem. Res.* 47, 1836–1844.
 143. Do, N. Q., and Phan, A. T. (2012) Monomer–dimer equilibrium for the 5'–5' stacking of propeller-type parallel-stranded G-quadruplexes: NMR structural study, *Chem-Eur J* 18, 14752–14759.
 144. Phan, A. T., and Patel, D. J. (2002) A site-specific low-enrichment ¹⁵N,¹³C isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids, *J. Am. Chem. Soc.* 124, 1160–1161.
 145. Yue, D. J. E., Lim, K. W., and Phan, A. T. (2011) Formation of (3+1) G-quadruplexes with a long loop by human telomeric DNA spanning five or more repeats, *J. Am. Chem. Soc.* 133, 11462–11465.
 146. Huang, X. N., Yu, P. L., LeProust, E., and Gao, X. L. (1997) An efficient and economic site-specific deuteration strategy for NMR

- studies of homologous oligonucleotide repeat sequences, *Nucleic Acids Res.* 25, 4758–4763.
147. Trajkovski, M., da Silva, M. W., and Plavec, J. (2012) Unique structural features of interconverting monomeric and dimeric G-quadruplexes adopted by a sequence from the intron of the N-myc gene, *J. Am. Chem. Soc.* 134, 4132–4141.
 148. Wang, X., and Seeman, N. C. (2007) Assembly and characterization of 8-arm and 12-arm DNA branched junctions, *J. Am. Chem. Soc.* 129, 8169–8176.
 149. Onel, B., Carver, M., Wu, G., Timonina, D., Kalarn, S., Larriva, M., and Yang, D. (2016) A New G-Quadruplex with Hairpin Loop Immediately Upstream of the Human BCL2 P1 Promoter Modulates Transcription, *J. Am. Chem. Soc.* 138, 2563–2570.
 150. Harley, C. B. (2008) Telomerase and cancer therapeutics, *Nat. Rev. Cancer* 8, 167–179.
 151. Arndt, G. M., and MacKenzie, K. L. (2016) New prospects for targeting telomerase beyond the telomere, *Nat. Rev. Cancer* 16, 508–524.
 152. Wu, X., Zhang, J., Yang, S., Kuang, Z., Tan, G., Yang, G., Wei, Q., and Guo, Z. (2017) Telomerase antagonist imetelstat increases radiation sensitivity in esophageal squamous cell carcinoma, *Oncotarget* 8, 13600–13619.
 153. Lagah, S., Tan, I. L., Radhakrishnan, P., Hirst, R. A., Ward, J. H., O'Callaghan, C., Smith, S. J., Stevens, M. F., Grundy, R. G., and Rahman, R. (2014) RHPS4 G-quadruplex ligand induces anti-proliferative effects in brain tumor cells, *Plos One* 9, e86187.
 154. Zhang, Z., Dai, J., Veliath, E., Jones, R. A., and Yang, D. (2010) Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K⁺ solution: insights into the interconversion of human telomeric G-quadruplex structures, *Nucleic Acids Res.* 38, 1009–1021.
 155. Galer, P., Wang, B., Sket, P., and Plavec, J. (2016) Reversible pH Switch of Two-Quartet G-Quadruplexes Formed by Human Telomere, *Angew. Chem. Int. Ed.* 55, 1993–1997.
 156. Hélène, C., and Toulme, J. J. (1990) Specific regulation of gene expression by antisense, sense and antigene nucleic acids, *Biochim. Biophys. Acta* 1049, 99–125.
 157. Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E., and Dervan, P. B. (1997) Regulation of gene expression by small molecules, *Nature* 387, 202–205.
 158. Neidle, S. (2001) DNA minor-groove recognition by small molecules, *Nat Prod Rep* 18, 291–309.
 159. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1996) Recognition of DNA by designed ligands at subnanomolar concentrations, *Nature* 382, 559–561.
 160. Neidle, S. (2015) A Personal History of Quadruplex-Small Molecule Targeting, *Chem. Rec.* 15, 691–710.

161. Lerman, L. S. (1963) The structure of the DNA-acridine complex, *Proc. Natl. Acad. Sci.* 49, 94-102.
162. Peek, M. E., Lipscomb, L. A., Bertrand, J. A., Gao, Q., Roques, B. P., Garbay-Jaureguiberry, C., and Williams, L. D. (1994) DNA distortion in bis-intercalated complexes, *Biochemistry* 33, 3794-3800.
163. Brulikova, L., Hlavac, J., and Hradil, P. (2012) DNA interstrand cross-linking agents and their chemotherapeutic potential, *Curr. Med. Chem.* 19, 364-385.
164. Moser, H. E., and Dervan, P. B. (1987) Sequence-specific cleavage of double helical DNA by triple helix formation, *Science* 238, 645-650.
165. Beal, P. A., and Dervan, P. B. (1991) Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation, *Science* 251, 1360-1363.
166. Gavathiotis, E., Heald, R. A., Stevens, M. F., and Searle, M. S. (2003) Drug recognition and stabilisation of the parallel-stranded DNA quadruplex d(TTAGGGT)₄ containing the human telomeric repeat, *J. Mol. Biol.* 334, 25-36.
167. Wartell, R. M., Larson, J. E., and Wells, R. D. (1974) Netropsin. A specific probe for A-T regions of duplex deoxyribonucleic acid, *J. Biol. Chem.* 249, 6719-6731.
168. De Cian, A., Delemos, E., Mergny, J. L., Teulade-Fichou, M. P., and Monchaud, D. (2007) Highly efficient G-quadruplex recognition by bisquinolinium compounds, *J. Am. Chem. Soc.* 129, 1856-1857.
169. Martino, L., Virno, A., Pagano, B., Virgilio, A., Di Micco, S., Galeone, A., Giancola, C., Bifulco, G., Mayol, L., and Randazzo, A. (2007) Structural and thermodynamic studies of the interaction of distamycin A with the parallel quadruplex structure [d(TGGGGT)]₄, *J. Am. Chem. Soc.* 129, 16048-16056.
170. Cosconati, S., Marinelli, L., Trotta, R., Virno, A., De Tito, S., Romagnoli, R., Pagano, B., Limongelli, V., Giancola, C., Baraldi, P. G., Mayol, L., Novellino, E., and Randazzo, A. (2010) Structural and conformational requisites in DNA quadruplex groove binding: another piece to the puzzle, *J. Am. Chem. Soc.* 132, 6425-6433.
171. Cocco, M. J., Hanakahi, L. A., Huber, M. D., and Maizels, N. (2003) Specific interactions of distamycin with G-quadruplex DNA, *Nucleic Acids Res.* 31, 2944-2951.
172. Mergny, J. L., Duval-Valentin, G., Nguyen, C. H., Perrouault, L., Faucon, B., Rougee, M., Montenay-Garestier, T., Bisagni, E., and Helene, C. (1992) Triple helix-specific ligands, *Science* 256, 1681-1684.
173. Zhao, P., Jin, S. F., Lu, J. Z., Lv, J. L., Wu, G. Q., Chen, P. P., Tan, C. L., and Chen, D. W. (2015) Novel porphyrin-daunomycin hybrids: synthesis and preferential binding to G-quadruplexes over i-motif, *Spectrochim. Acta A* 137, 227-235.
174. Yangyuoru, P. M., Di Antonio, M., Ghimire, C., Biffi, G., Balasubramanian, S., and Mao, H. (2015) Dual Binding of an Antibody

- and a Small Molecule Increases the Stability of TERRA G-Quadruplex, *Angew. Chem. Int. Ed.* **127**, 924-927.
175. Rodrik-Outmezguine, V. S., Okaniwa, M., Yao, Z., Novotny, C. J., McWhirter, C., Banaji, A., Won, H., Wong, W., Berger, M., de Stanchina, E., Barratt, D. G., Cosulich, S., Klinowska, T., Rosen, N., and Shokat, K. M. (2016) Overcoming mTOR resistance mutations with a new-generation mTOR inhibitor, *Nature* **534**, 272-276.
 176. Kontermann, R. E., and Brinkmann, U. (2015) Bispecific antibodies, *Drug Discov. Today* **20**, 838-847.
 177. Murray, C. W., and Rees, D. C. (2009) The rise of fragment-based drug discovery, *Nat. Chem.* **1**, 187-192.
 178. Scott, D. E., Coyne, A. G., Hudson, S. A., and Abell, C. (2012) Fragment-based approaches in drug discovery and chemical biology, *Biochemistry* **51**, 4990-5003.
 179. Maizels, N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells, *Nat. Struct. Mol. Biol.* **13**, 1055–1059.
 180. Wu, R. Y., Zheng, K. W., Zhang, J. Y., Hao, Y. H., and Tan, Z. (2015) Formation of DNA:RNA hybrid G-quadruplex in bacterial cells and its dominance over the intramolecular DNA G-quadruplex in mediating transcription termination, *Angew. Chem. Int. Ed.* **54**, 2447-2451.
 181. Weldon, C., Behm-Ansmant, I., Hurley, L. H., Burley, G. A., Branlant, C., Eperon, I. C., and Dominguez, C. (2017) Identification of G-quadruplexes in long functional RNAs using 7-deazaguanine RNA, *Nat. Chem. Biol.* **13**, 18-20.
 182. Plateau, P., and Gueron, M. (1982) Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences, *J. Am. Chem. Soc.* **104**, 7310-7311.
 183. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *J Comput Chem* **26**, 1668–1688.